JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Unitary Representations of the Generalized Poincaré Groups

YURI V. NOVOZHILOV
*Department of Theoretical Physics, University of Leningrad, U.S.S.R., and Centre for Advanced Study in Theoretical Physics, University of Delhi, Delhi, India**

AND

IGOR A. TERENTJEV
Department of Theoretical Physics, University of Leningrad, Leningrad, U.S.S.R.

(Received 7 October 1966)

The problem of the explicit construction of unitary representations is solved for the generalized inhomogeneous Poincaré groups $ISL(n, C)$ and their subgroups. As a key to the solution, a method is developed to find the generalized Wigner operator, which transforms a given n^2 momentum to the standard one. Results are also specified for the $U(m, m)$ subgroup of the group $ISL(2m, C)$.

INTRODUCTION

Some achievements of the SU_6 symmetry[1] indicate that there might exist a generalized group of elementary-particle symmetries which includes the Poincaré group and the group $SU(3)$ in a nontrivial way. However, up to now all attempts at relativization of the SU_6 symmetry were unsuccessful from the physical point of view. Nevertheless, it seems worth trying, at least from the mathematical point of view, to explicitly build unitary representations of a generalized Poincaré group.

In the present paper we consider the problem of explicit construction of unitary representations for the inhomogeneous groups $ISL(n, C)$ and $IU(m, m)$.

The homogeneous group $SL(n, C)$ is a group of linear unimodular transformations in the n-dimensional complex vector space. As is well known, the group $SL(2, C)$ is the universal covering group of the Lorentz group. If the latter is enlarged by adding four translations, which transform according to the (real) vector representation, one gets the Poincaré group. Similarly, by adding n^2 translations to the inhomogeneous group $SL(n, C)$ we obtain the inhomogeneous group $ISL(n, C)$. It is the semidirect product of $SL(n, C)$ and the Abelian group of translations T_{n^2}.

The group $U(m, m)$ is a subgroup of the unimodular group $SL(2m, C)$, namely, the transformations of $U(m, m)$ leave unchanged the bilinear form

$$\text{(I)} \qquad \sum_{k=1}^{m} \{|\xi_k|^2 - |\xi_{k+m}|^2\}.$$

The additional requirement (I) also remains valid for the inhomogeneous group $IU(m, m)$.

An element of the $ISL(n, C)$ may be written as (A, V), where V is a unimodular $n \times n$ matrix, $\det \mathsf{V} = 1$, and A denotes a Hermitian $n \times n$ matrix. The matrix A characterizes translations, being composed of n^2 components α_B of the displacement vector and basis Hermitian $n \times n$ matrices Λ_A:

$$\text{(II)} \qquad \mathsf{A} = \alpha_B \Lambda_B, \quad B = 0, 1, \cdots, n^2 - 1.$$

Hereafter, we adopt the usual convention of summation over repeated indices.

The multiplication law of the two group elements $(\mathsf{A}_1, \mathsf{V}_1)$ and $(\mathsf{A}_2, \mathsf{V}_2)$ is given by

$$\text{(III)} \quad (\mathsf{A}_1, \mathsf{V}_1)(\mathsf{A}_2, \mathsf{V}_2) = (\mathsf{A}_1 + \mathsf{V}_1^{\dagger -1} \mathsf{A}_2 \mathsf{V}_1^{-1}, \mathsf{V}_1 \mathsf{V}_2),$$

with the unit element $\mathbf{e} = (0, 1)$.

The existence of unitary representations $\mathsf{U}(\mathsf{A}, \mathsf{V})$ of the group $ISL(n, C)$ and their classification according to unitary representations of stationary groups (or little groups in Wigner's terminology[2]) have been thoroughly investigated and proved.[2,3] However, the problem of explicit construction of $\mathsf{U}(\mathsf{A}, \mathsf{V})$ has been solved for the Poincaré group only (Wigner[2]). It should be stressed that Wigner's investigation[2] is of very general character, so that in the present paper we take Wigner's results, generalized in an obvious way, as the starting point.

Unitary representations of the group $ISL(n, C)$ are defined by the set of functions of n^2 momenta $\mathbf{p}_B$ with the scalar product

$$\text{(IV)} \quad (\varphi, \varphi') = \int dp_0\, dp_1 \cdots dp_{n^2-1} \mu(\mathbf{p}) \varphi^*(\mathbf{p}) \varphi'(\mathbf{p}),$$

where $\mu(\mathbf{p})$ denotes an invariant measure.

An irreducible unitary representation $\mathsf{U}(\mathsf{A}, \mathsf{V})$ is given up to unitary equivalence by the following relation[2]:

$$\text{(V)} \quad U_{\sigma\sigma'}(\mathsf{A}, \mathsf{V}) \varphi_{\sigma'}(\mathbf{p}) = \exp\,[i \mathbf{Sp}(\mathsf{PA})] \cdot Q_{\sigma\sigma'}(\mathsf{V}') \varphi_{\sigma'}(\mathbf{p}').$$

* Address until 15 July 1967.

[1] B. Sakita, Phys. Rev. **136**, B1756 (1964); F. Gursey and L. Radicati, Phys. Rev. Letters **13**, 173 (1964).

[2] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[3] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962).

Here the $n \times n$ matrix V' in the argument of Q depends not only on momenta $\mathbf{p}_B$, but also on a unimodular transformation V of the homogeneous group $SL(n, C)$. On the one hand, $V'(\mathbf{p})$ is a transformation of the little group $S(\mathbf{p}^0)$, belonging to a standard momentum $\mathbf{p}^0$. It means that under the transformations V' the standard momentum $\mathbf{p}^0$ is left unchanged, or that the matrices $P^0 = \mathbf{p}^0_B \mathbf{\Lambda}_B$ and V' satisfy the equation $V'P^0V'^\dagger = P^0$. On the other hand, V' is related to a given transformation matrix V:

$$\text{(VI)} \qquad V'(\mathbf{p}) = \alpha^{-1}(\mathbf{p})V\alpha(\mathbf{p}'), \quad VP'V^\dagger = P.$$

Thus, the $n \times n$ matrix $\alpha(\mathbf{p})$ enables us to find an argument of the unitary matrix Q in (V). We refer to $\alpha(\mathbf{p})$ as the Wigner operator. The Wigner operator is defined (in a nonunique way) as a transformation, which brings the standard momentum P^0 into a given one P: $\alpha P^0 \alpha^\dagger = P$.

The matrix Q in (V) acts on the "internal" variables only; matrices Q form an irreducible representation of the little group $S(\mathbf{p}^0)$.

According to (V) an irreducible unitary representation U(A, V) is fully characterized by a type of the standard momentum $\mathbf{p}^0$ [i.e., by the type of the little group $S(\mathbf{p}^0)$] and specification of the irreducible representation Q of the $S(\mathbf{p}^0)$. We consider the case where the standard momentum matrix P^0 is proportional to the unit matrix: $P^0 = \kappa E$, $\kappa > 0$. In this case the little group in $ISL(n, C)$ is the well-known unitary group $SU(n)$. However, for an explicit construction of Q(V') in (V), it is necessary to know the Wigner operator α too, because $\alpha(\mathbf{p})$ connects a given unimodular matrix V with a unitary matrix V' of the little group $S(\mathbf{p}^0)$. Generators of the group $ISL(n, C)$ in the unitary representation can also be found only after $\alpha(\mathbf{p})$ is calculated.

Therefore, in essence the present paper deals with the structure of the momentum space of $ISL(n, C)$ and the method of the explicit construction of $\alpha(\mathbf{p})$.

In Sec. 1 the main relations for basis matrices Λ_A are summarized. Section 2 reviews the group relations involving momenta. The groups $ISL(2m, C)$, $IU(m, m)$, and the group with reflection $ISL(m, c)$ are considered here in sequence, determined by their interrelation

$$ISL(2m, C) \ni IU(m, m) \ni \overline{ISL}(m, C).$$

In Sec. 3 the method of the calculation of the Wigner operator is proposed.[4] For the group $ISL(n, C)$ the result is given by (3.12) and (3.19). In Sec. 4, the structure of the momentum space is investigated. Section 5 contains a brief discussion of generators in unitary representation.[5]

Generalized Poincaré groups $ISL(6, C)$ and $IU(6, 6)$, in connection with the problem of particle symmetry, have been discussed in a number of physical papers.[6]

1. THE BASIC MATRICES

We give in this section a number of relations involving basis matrices of the $SL(m, C)$ group.

Consider the full set of m^2 linear independent Hermitian $m \times m$ matrices λ_a, where the indices a, b take values $a, b = 0, 1, 2, \cdots, m^2 - 1$. We choose normalization of λ_a in the form

$$\lambda_0 = (1/2^{\frac{1}{2}}m)E, \quad \mathrm{Sp}\lambda_a = 2^{-\frac{1}{2}}\delta_{a0},$$
$$\mathrm{Sp}\lambda_a\lambda_b = (2m)^{-1}\delta_{ab}. \tag{1.1}$$

Along with the matrices λ_a we make use of $n \times n$ matrices $\mathbf{\Lambda}_A$, where $n = 2m$; here indices $A, B, \cdots$, take the values $0, 1, \cdots, n^2 - 1$. Matrices $\mathbf{\Lambda}_A$ are normalized according to the same condition (1.1) by substituting $m \to n$ and $a, b \to A, B$.

Properties of the group $ISL(m, C)$ may also be expressed by means of quantities

$$\mathbf{f}_{abc} = -2im\mathrm{Sp}(\lambda_a[\lambda_b, \lambda_c]),$$
$$\mathbf{d}_{abc} = 2m\mathrm{Sp}(\lambda_a\{\lambda_b, \lambda_c\}); \quad \mathbf{d}_{0ab} = 2^{\frac{1}{2}}m^{-1}\delta_{ab}. \tag{1.2}$$

The corresponding quantities constructed out of matrices $\mathbf{\Lambda}_A$ are denoted by $\mathbf{F}_{ABC}$ and $\mathbf{D}_{ABC}$.

Is is sometimes convenient to use a particular representation of $\mathbf{\Lambda}_A$ as a direct product of Pauli matrices $\sigma_\mu(\sigma_0 = 1)$ and matrices λ_a:

$$\mathbf{\Lambda}_A \equiv \mathbf{\Lambda}^\mu_a = \tfrac{1}{2}\sigma_\mu \otimes \lambda_a;$$
$$\mu = 0, 1, 2, 3; \quad a = 0, 1, \cdots, m^2 - 1. \tag{1.3}$$

Here the index A is replaced by a pair of indices (μ, a). The orthogonality relations for $\mathbf{\Lambda}_A$ can be written as

$$2n(\mathbf{\Lambda}_A)_{ik}(\mathbf{\Lambda}_A)_{jl} = \delta_{il}\delta_{kj} \tag{1.4}$$

with $i, j, k, l = 1, 2, \cdots, n$. As a consequence of (1.4) the following relation holds for arbitrary $n \times n$

[4] Yu. V. Novozhilov and I. A. Terentjev, Jadernaja Fys. 3, N6 (1966).

[5] Yu. V. Novozhilov and I. A. Terentjev, Vestn. Levingr. Univ. N16, 5 (1966).

[6] T. Fulton and J. Wess, Phys. Letters **14**, 57, 334 (1965); H. Bacry and J. Nuyts, Nuovo Cimento **87**, 1702 (1965); W. Ruhl, *ibid.* **37**, 301, 319 (1965); Yu. V. Novozhilov and I. A. Terentjev, Phys. Letters **15**, 86 (1965); V. Kadyshevsky, R. Muradyan, A. Tavkhelidze, and I. Tedorov, Phys. Letters **65**, 182 (1965); S. K. Bose and Yu. M. Shirokov, Phys. Rev. Letters **14**, 398 (1965); V. H. Nguen, Jadernaja Fys. **2**, 517 (1965); R. Delbourgo, M. Rashid, A. Salam, and J. Strathdee, "High-Energy Physics and Elementary Particles," Trieste, 1965 (unpublished); L. Michel and B. Sakita, Ann. Inst. Henri Poincaré **11** (1965); B. Sakita and K. C. Wali, Phys. Rev. **139**, B1355 (1965); J. S. Bell and H. Ruegg, Nuovo Cimento **39**, 116 (1965); W. Rühl, *ibid.*, **38**, 675 (1965); J. M. Charap, P. T. Matthews, and R. F. Streater, Proc. Roy. Soc. (London) **A290**, N1420 (1966).

matrices R and S:

$$2n\mathrm{Sp}(\mathbf{\Lambda}_A\mathsf{R})\mathrm{Sp}(\mathbf{\Lambda}_A\mathsf{S}) = \mathrm{Sp}(\mathsf{RS}). \tag{1.5}$$

We enclose indices A and a in parentheses in the case when values $A = 0$ and $a = 0$ are excluded: $(A) = 1, 2, \cdots, n^2 - 1$ and $(a) = 1, 2, \cdots, m^2 - 1$. In terms of the representation (1.3) for $\mathbf{\Lambda}_A$ it means that $(A) = (\mu, a)$ does not include the pair (0, 0).

2. MOMENTUM SPACE AND GROUP RELATIONS

A. Group $ISL(n, C)$

We write the corresponding Lie algebra in the form resembling that of the quantum-mechanical Poincaré group

$$\begin{aligned}
[\mathbf{M}_{(A)}, \mathbf{M}_{(B)}] &= \mathbf{F}_{(A)(B)(C)}\mathbf{M}_{(C)},\\
[\mathbf{M}_{(A)}, \mathbf{N}_{(B)}] &= \mathbf{F}_{(A)(B)(C)}\mathbf{N}_{(C)},\\
[\mathbf{N}_{(A)}, \mathbf{N}_{(B)}] &= -\mathbf{F}_{(A)(B)(C)}\mathbf{M}_{(C)},\\
[\mathbf{M}_{(A)}, \mathbf{P}_B] &= \mathbf{F}_{(A)B(C)}\mathbf{P}_{(C)},\\
[\mathbf{N}_{(A)}, \mathbf{P}_B] &= \mathbf{D}_{(A)BC}\mathbf{P}_C,\\
[\mathbf{P}_A, \mathbf{P}_B] &= 0.
\end{aligned} \tag{2.1}$$

Here $\mathbf{M}_{(A)}$ are the generators of the compact subgroup $SU(n)$. Operators $\mathbf{M}_{(A)}$ and $\mathbf{N}_{(A)}$ taken together display themselves as generators of the homogeneous subgroup $SL(n, C)$ of $ISL(n, C)$. Momenta $\mathbf{P}_B$ transform according to (n, n^*) representation of $SL(n, C)$. Thus, there are n^2 momenta $\mathbf{P}_A$ with eigenvalues p_A.

If we consider $SL(n, C)$ not as subgroup of $ISL(n, C)$ but as a separate group with generators $\bar{\mathbf{M}}_{(A)}$ and $\bar{\mathbf{N}}_{(A)}$ then in the fundamental representation of $SL(n, C)$ we would get

$$\bar{\mathbf{M}}_{(A)} = -i\mathbf{\Lambda}_{(A)}, \quad \bar{\mathbf{N}}_{(A)} = \mathbf{\Lambda}_{(A)}.$$

Then an element of $SL(n, C)$ algebra in the fundamental representation can be written as

$$z(\alpha, \beta) = -i\mathbf{\Lambda}_{(A)}\alpha_{(A)} + \mathbf{\Lambda}_{(A)}\beta_{(A)}, \quad \mathrm{Sp}z = 0, \tag{2.2}$$

with real parameters α and β.

Instead of the n^2-component momentum $\mathbf{p}_A$ one can introduce a Hermitian $n \times n$ matrix

$$\mathsf{P} = \mathbf{\Lambda}_A\mathbf{p}_A, \quad \mathsf{P}^\dagger = \mathsf{P}, \tag{2.3}$$

which has one-to-one correspondence with the momentum

$$\mathbf{p}_A = 2n\mathrm{Sp}(\mathbf{\Lambda}_A\mathsf{P}). \tag{2.4}$$

Under a transformation U from $SL(n, C)$ the momentum matrix transforms according to the rule

$$\mathsf{P}' = \mathsf{UPU}^\dagger, \quad \mathsf{U} = \exp z(\alpha, \beta). \tag{2.5}$$

U is a unimodular matrix; therefore the determinant of P is an invariant under transformation (2.5)

$$Z(\mathbf{p}) = \det \mathsf{P} = \det \mathsf{P}' = \text{invariant}. \tag{2.6}$$

We consider in this paper only those representations of $ISL(n, C)$ where the vector $\mathbf{p}_A$ can be obtained from the vector $\mathbf{p}^0_A = \sqrt{2}\, n\boldsymbol{\delta}_{A0}$ by means of a continuous transformation belonging to $SL(n, C)$. The vector $\mathbf{p}^0_A$ is an analog of the Poincaré momentum in the rest frame; the matrix P^0 corresponding to $\mathbf{p}^0_A$ is the unity matrix: $\mathsf{P}^0 = \mathsf{E}$. Then P is normalized in the sense that $\det \mathsf{P} = 1$.

The stationary, or little group, of the momentum $\mathbf{p}$ is the set of those transformations $\bar{\mathsf{V}}$ which leave P unchanged: $\bar{\mathsf{V}}_{(p)}\mathsf{P}\bar{\mathsf{V}}^\dagger_{(p)} = \mathsf{P}$. The little group $S(\mathbf{p}^0)$ of $\mathsf{P}^0 = \mathsf{E}$ consists of all unitary transformations belonging to $SL(n, C)$, or $S(\mathsf{P}^0) = SU(n)$.

It is well known that an arbitrary element U of $SL(n, C)$ may be written as a product of a Hermitian matrix

$$\mathsf{H} = \exp z(0, \beta')$$

and a unitary matrix

$$\mathsf{R} = \exp z(\alpha', 0):$$

$$\mathsf{U} = \exp z(\alpha, \beta) = \exp z(0, \beta') \cdot \exp z(\alpha', 0) = \mathsf{HR}. \tag{2.7}$$

The matrix R belongs to the subgroup $SU(n)$, while H describes pure Lorentz transformation in the case of the Poincaré group ($n = 2$). It is easy to see now that the momentum matrix P in our case is a square of the Hermitian transformation matrix:

$$\mathsf{P}(\beta) = \mathsf{UP}^0\mathsf{U}^\dagger = \mathsf{H}^2 \equiv \boldsymbol{\alpha}^2(\mathbf{p}). \tag{2.8}$$

Thus, in order to find P it is necessary to construct a Hermitian matrix $\boldsymbol{\alpha}(\mathbf{p})$ (Wigner operator) which brings "rest-frame" momentum P^0 into P.

Let us also introduce the n^2-component momentum $\mathbf{b}_A$ in the dual space. Properties of $\mathbf{b}_A$ follow from the transformation law of the matrix $\mathsf{B} = \mathbf{\Lambda}_A\mathbf{b}_A$

$$\mathsf{B}' = \mathsf{U}^{\dagger-1}\mathsf{BU}^{-1}, \quad \mathsf{U}^{\dagger-1} = \exp z(\alpha, -\beta). \tag{2.9}$$

In order to establish a connection between momenta of both kinds we impose the condition that the product PB of the relevant matrices is the unity matrix, or

$$\mathsf{B} = \mathsf{P}^{-1}. \tag{2.10}$$

The condition implies the identity of little groups in both dual spaces. According to (2.10) there exists a generalized frame where both matrices P^0 and B^0 are equal to the unity matrix. One can conclude also that

$$\boldsymbol{\alpha}(\mathbf{p}) = \boldsymbol{\alpha}^{-1}(\mathbf{b}).$$

Equation (2.10) is equivalent to the following set of relations for $\mathbf{p}_A$ and $\mathbf{b}_A$:

$$D_{(A)BC}\mathbf{p}_B\mathbf{b}_C = 0, \quad \mathbf{b}_A \cdot \mathbf{p}_A = 2n^2. \tag{2.11}$$

Solution of (2.11) gives $\mathbf{b}_A$ as a function of $\mathbf{p}_A$. It is given in Sec. 3.

B. Group $IU(m, m)$

Transformations of the $U(m, m)$ group are those unimodular transformations of $SL(2m, C)$ which are restricted by condition (I). Generators of $U(m, m)$ include a part of generators $\bar{\mathbf{M}}_{(A)}$ and $\bar{\mathbf{N}}_{(A)}$ only, i.e., $\bar{\mathbf{M}}^0_a$, $\bar{\mathbf{M}}^3_a$, $\bar{\mathbf{N}}^1_a$, $\bar{\mathbf{N}}^2_a$. An algebra in this case can be obtained in a standard way by means of subdivision of the compact algebra $\mathfrak{U}$ corresponding to the group $SU(n)$ on eigensubspaces of the inner automorphism τ of $\mathfrak{U}$.

We define τ by a set of conditions

$$\begin{aligned} &\tau\bar{\mathbf{M}}^0_{(a)}\tau = \bar{\mathbf{M}}^0_{(a)}, \quad \tau\bar{\mathbf{M}}^3_a\tau = \bar{\mathbf{M}}^3_a, \\ &\tau\bar{\mathbf{M}}^1_a\tau = -\bar{\mathbf{M}}^1_a, \quad \tau\bar{\mathbf{M}}^2_a\tau = -\bar{\mathbf{M}}^2_a, \\ &\tau^2 = 1. \end{aligned} \tag{2.12}$$

The $SU(n)$ algebra now is split up into two parts

$$\mathfrak{U} = \mathfrak{k}_0 + \mathfrak{p}_0. \tag{2.13}$$

Then we get for the algebra $\mathfrak{L}$ corresponding to the group $U(m, m)$

$$\mathfrak{L} = \mathfrak{k}_0 + i\mathfrak{p}_0. \tag{2.14}$$

The compact subalgebra of $\mathfrak{L}$ is nothing but an algebra $\mathfrak{k}_0$ with generators $\bar{\mathbf{M}}^0_{(a)}$ and $\bar{\mathbf{M}}^3_a$. Hence, the maximal compact subgroup in this case is isomorphic to $SU(m) \otimes SU(m) \otimes U(1)$.

In the fundamental representation the operation τ is given by

$$\tau = 2^{\frac{1}{2}n}\Lambda^3_0 = \sigma_3 \otimes E_m. \tag{2.15}$$

Thus, the momentum space of $U(m, m)$ contains all the vectors $\mathbf{p}_A(\beta^1_a, \beta^2_a, \beta^3_a, \beta^0_{(a)})$ of (2.8) with an additional condition $\beta^0_{(a)} = \beta^3_a = 0$. We denote vectors $\mathbf{p}_A$ in such a $U(m, m)$ space, as well as any function of them, by characters indicated with a tilde so that

$$\mathbf{p}_A = \tilde{\mathbf{p}}_A, \quad f(\mathbf{p}) = \tilde{f}(\mathbf{p}), \quad \text{if} \quad \beta^0_{(a)} = \beta^3_a = 0.$$

One immediately gets an element of algebra $\tilde{z}(\alpha, \beta)$ in the fundamental representation from (2.2) by putting $\alpha^1_a = \alpha^2_a = \beta^0_{(a)} = \beta^3_a = 0$. By virtue of (2.12) we have

$$\tau\tilde{z}(\alpha, \beta)\tau = \tilde{z}(\alpha, -\beta), \quad \tau\mathbf{\Lambda}_A\tau = \mathbf{e}_A\mathbf{\Lambda}_A, \tag{2.16}$$

where $e^0_a = e^3_a = -e^1_a = -e^2_a = 1$. There is no summation over A in (2.16); this implies that the presence of $\mathbf{e}_A$ in formula will mean also no summation over indices A.

One can conclude from (2.10) and (2.16) that the momentum matrix $\tilde{\mathsf{P}} = [\exp \tilde{z}(0, \beta)]^2$ satisfies the relation

$$\tau\tilde{\mathsf{P}}\tau = \tilde{\mathsf{P}}^{-1} = \tilde{\mathsf{B}}. \tag{2.17}$$

Expressed in terms of momenta $\mathbf{p}_A$ and $\mathbf{b}_A$ this relation means that in case of $U(m, m)$ group $\mathbf{b}_A$ is related to $\mathbf{p}_A$ as $\mathbf{b}_A = \mathbf{e}_A\mathbf{p}_A$.

We can introduce a metric in the momentum space $\mathbf{g}_{AB} = \mathbf{e}_A\delta_{AB}$ because, as a consequence of (2.17), the bilinear form $\mathbf{p}_A\mathbf{g}_{AB}\mathbf{p}'_B$ is an invariant under $U(m, m)$ transformations.

The transformation properties of $\tilde{\mathsf{P}}$

$$\tilde{\mathsf{P}}' = \tilde{\mathsf{U}}\tilde{\mathsf{P}}\tilde{\mathsf{U}}^\dagger, \quad \tilde{\mathsf{U}} = \exp \tilde{z}(-\alpha', \beta') \tag{2.18}$$

can be derived from (2.2) and (2.5). However, it is more convenient to consider the matrix $\mathsf{Q} = \tilde{\mathsf{P}}\tau$ with the simpler transformation law

$$\mathsf{Q}' = \exp(-\tilde{z})\mathsf{Q} \exp \tilde{z}. \tag{2.19}$$

In particular, any polynomial of Q transforms like Q and the trace of any power of Q is an invariant. Identities

$$\mathbf{Sp}\mathsf{Q}^{2r} = n, \quad \mathbf{Sp}\mathsf{Q}^{2r+1} = 0, \quad r = 0, 1, 2 \cdots \tag{2.20}$$

are easily verified.

C. Group with a Reflection $\overline{ISL}(m, C)$

Let us put $\alpha^3_{(a)} = \beta^2_a = \beta^1_0 = 0$ in $\tilde{z}(\alpha, \beta)$ and fix the parameter α^0_3 such that the condition

$$\exp(-i\Lambda^3_0\alpha^3_0) = i\tau, \quad \alpha^3_0 = -(\pi n/\sqrt{2}) \tag{2.21}$$

is satisfied. Then we get the group $\overline{ISL}(m, c)$ with reflection τ, the generators of the group being $\bar{\mathbf{M}}^0_{(a)}$ and $\bar{\mathbf{N}}^1_{(a)}$, $(a) = 1, 2, \cdots, m^2 - 1$. The momentum matrix in this case is given by

$$\mathsf{P} = \tfrac{1}{2}(\mathsf{I} + \sigma_1) \otimes \mathsf{p} + \tfrac{1}{2}(\mathsf{I} - \sigma_1) \otimes \mathsf{p}^{-1}, \tag{2.22}$$

where p denotes the momentum matrix of the group $SL(m, c)$ without reflection

$$\mathsf{p} = \tfrac{1}{2}\lambda_a p_a, \quad \mathsf{p}^{-1} = \tfrac{1}{2}\lambda_a b_a; \quad \det \mathsf{p} = 1.$$

Only momenta $p^0_a = \frac{1}{2}(p_a + b_a)$, $p^1_a = \frac{1}{2}(\mathbf{p}_a - \mathbf{b}_a)$ are not equal to zero.

Reflection τ is equivalent to the replacement of p by p^{-1} in (2.22) or to the replacement $\mathbf{p}_a \leftrightarrow \mathbf{b}_a$.

3. CONSTRUCTION OF THE WIGNER OPERATOR

The Wigner operator $\boldsymbol{\alpha}(\mathbf{p})$ as defined in Sec. 2

$$\boldsymbol{\alpha}^2(\mathbf{p}) = \mathsf{P}, \quad \boldsymbol{\alpha}^\dagger = \boldsymbol{\alpha},$$

is a Hermitian matrix which if squared gives us a

momentum matrix P. Thus, $\alpha(\mathbf{p})$ is a kind of a square root of P. It is quite an elementary task to take the square root of a matrix if it is a diagonal. Therefore, our method of solving the problem includes as an essential step the calculation of the diagonal elements of P in terms of known quantities, i.e., momenta $\mathbf{p}_A$.

A. Group *ISL(n, C)*

Suppose that the eigenvalue problem of the matrix P has already been solved,

$$\mathsf{P}\xi(k) = \pi(k)\xi(k); \quad (\xi(k), \xi(i)) = \delta_{ki}, \tag{3.1}$$

where $\pi(k)$; $i, k = 1, 2, \cdots, n$ is the eigenvalue of P and $\xi(k)$ are the corresponding eigenfunctions. In view of our choice of standard momentum P^0, all eigenvalues are positive, i.e., $\pi(k) > 0$. The matrix P may be written in the form

$$\mathsf{P} = \sum_{k=1}^{n} \pi(k)\boldsymbol{\chi}(k), \quad (\chi(k))_{ij} = \xi_i(k)\xi_j^*(k), \tag{3.2}$$

if we introduce projection operators $\boldsymbol{\chi}(k)$ with properties

$$\boldsymbol{\chi}(k)\boldsymbol{\chi}(i) = \delta_{ik}\boldsymbol{\chi}(k), \quad \sum_{k=1}^{n}\boldsymbol{\chi}(k) = \mathsf{E}, \tag{3.3}$$

$\pi(k)$ and $\boldsymbol{\chi}(k)$ being functions of momenta $\mathbf{p}_A$.

We also use an expression for $\boldsymbol{\chi}$ in terms of basis matrices $\boldsymbol{\Lambda}_A$

$$\boldsymbol{\chi}(k) = \boldsymbol{\Lambda}_A\mathbf{n}_A(k), \quad \mathbf{n}_A(k) = 2n(\boldsymbol{\xi}(k), \boldsymbol{\Lambda}_A\boldsymbol{\xi}(k)), \tag{3.4}$$

which follows from (3.2).

Consider a matrix $d\mathsf{P} = \boldsymbol{\Lambda}_A\, d\mathbf{p}_A$. It may be treated as a perturbation of P. But according to (3.4), the first-order correction to the eigenvalue $\pi(k)$ is equal to $\mathbf{n}_A(k) \cdot d\mathbf{p}_A/2n$. Hence, $\mathbf{n}_A(k)$ is proportional to the derivative of $\pi(k)$,

$$2n\, \partial\pi(k)/\partial\mathbf{p}_A = \mathbf{n}_A(k). \tag{3.5}$$

This formula is often used in our calculations.

It follows from (3.2) and (3.4) that momenta $\mathbf{p}_A$ are related to "unit vectors" $\mathbf{n}_A$ almost in the usual way:

$$\mathbf{p}_A = \sum_{k=1}^{n}\pi(k)\mathbf{n}_A(k). \tag{3.6}$$

The set of "orthonormality" relations for "unit vectors" may be easily derived from (1.1), (1.2), (3.3), and (3.4)

$$\sum_{k=1}^{n}\mathbf{n}_A(k)\mathbf{n}_B(k) = 2n\boldsymbol{\delta}_{AB}, \quad \sum_{k=1}^{n}\mathbf{n}_A(k) = \sqrt{2}\, n\boldsymbol{\delta}_{A0},$$

$$\mathbf{n}_A(k)\mathbf{n}_A(i) = 2n\delta_{ki}, \quad D_{ABC}\mathbf{n}_A(k)\mathbf{n}_B(i) = 2\boldsymbol{\delta}_{ki}\mathbf{n}_C(k). \tag{3.7}$$

Now we can find the momentum $\mathbf{b}_A$ of the dual space as a function of $\mathbf{p}_A$.[7] B is an inverse matrix of P; therefore

$$\mathbf{b}_A = \sum_{k=1}^{n}\mathbf{n}_A(k)\pi^{-1}(k) = 2nZ^{-1}(\mathbf{p})\, \partial Z(\mathbf{p})/\partial\mathbf{p}_A. \tag{3.8}$$

The second line in (3.8) is obtained by differentiating $Z(\mathbf{p}) = \det \mathsf{P}$ with respect to $\mathbf{p}_A$ and using (3.5). Formula (3.8) provides us with the solution of (2.10).

We now proceed to the derivation of α as a function of momenta $\mathbf{p}_A$. Let us first write α as a square root of the diagonal P:

$$\boldsymbol{\alpha}(\mathbf{p}) = \boldsymbol{\Lambda}_A\sum_{k=1}^{n}[\pi(k)]^{\frac{1}{2}}\mathbf{n}_A(k). \tag{3.9}$$

In order to get rid of $\mathbf{n}_A$ in the sum (3.9) we use combination of formulas (3.5) and (3.7):

$$\pi(k)\mathbf{n}_A(k) = n\, D_{ABC}\mathbf{p}_B\, \partial\pi(k)/\partial\mathbf{p}_C \equiv n\mathbf{N}_A^0\pi(k). \tag{3.10}$$

Consequently, if we apply $\mathbf{N}_A^0$ to $\beta(k) = \ln\pi(k)$, then

$$n\mathbf{N}_A^0\beta(k) = \mathbf{n}_A(k). \tag{3.11}$$

Now it is possible to rewrite (3.9), in a compact form

$$\boldsymbol{\alpha}(\mathbf{p}) = 2n\hat{\mathbf{N}}d_{\frac{1}{2}}(\mathbf{p}), \tag{3.12}$$

where

$$\hat{\mathbf{N}} = \boldsymbol{\Lambda}_A\mathbf{N}_A^0$$

and

$$d_s(\mathbf{p}) = \mathrm{Sp}\mathsf{P}^s = \sum_{k=1}^{n}[\pi(k)]^s. \tag{3.13}$$

We can conclude from (3.13) that calculation of all the matrix elements of $\boldsymbol{\alpha}(\mathbf{p})$ is now reduced to the calculation of a single function $d_{\frac{1}{2}}(\mathbf{p})$. Let us note that quantities $d_s(\mathbf{p})$ with an *integer* s should be considered as given. As a matter of fact, the d_s may be easily found as traces of P^s if momenta $\mathbf{p}_A$ are known.

In the same way one can obtain the inverse matrix

$$\boldsymbol{\alpha}^{-1}(\mathbf{p}) = 4n\boldsymbol{\Lambda}_A(\partial d_{\frac{1}{2}}(\mathbf{p})/\partial\mathbf{p}_A). \tag{3.14}$$

Let us calculate next $d_{\frac{1}{2}}$ as a function of d_s with integer indices s. Consider an auxiliary function

$$G(t) = \prod_{k=1}^{n}(1 + \pi(k)t) = \sum_{k=1}^{n}f_k(\mathbf{p})t^k, \quad f_0 = 1 \tag{3.15}$$

with coefficients f_k depending on momenta $\mathbf{p}$ only. In the case $k = n$ we get the determinant of P: $f_n = \det \mathsf{P}$. Since functions $f_s(\mathbf{p})$ are invariant under unitary transformations of the group $SU(n)$, they can depend only on momenta through the quantities $d_k(\mathbf{p})$ [see (3.13)].

An explicit expression for f_k is

$$f_k(\mathbf{p}) = A_k(d_1(\mathbf{p}), d_2(\mathbf{p}), \cdots, d_k(\mathbf{p})), \tag{3.16}$$

[7] Yu. V. Novozhilov, Phys. Letters **16**, 348 (1965).

where A_k is defined by

$$A_k(x_1, x_2, \cdots, x_k) = (-1)^k \sum_{(l)} \frac{(-x_1)^{l_1}(-x_2)^{l_2}\cdots(-x_k)^{l_k}}{l_1!\, l_2!\cdots l_k!\, 1^{l_1}2^{l_2}\cdots k^{l_k}}. \quad (3.17)$$

Summation in (3.17) is extended over all positive integers $l_1, l_2, \cdots, l_k$ satisfying the condition $l_1 + 2l_2 + \cdots + kl_k = k$. Equations (3.15)–(3.17) determine completely the polynomial $G(t)$. From the other side, points $t = -1/\pi(k)$ are roots of $G(t)$ or poles of

$$\frac{d}{dt} \ln G(t) = \sum_{k=1}^{n} \left(t + \frac{1}{\pi(k)}\right)^{-1}. \quad (3.18)$$

Hence, if we make use of an integral representation

$$\frac{\sin(\nu\pi)}{\pi} \int_0^\infty \frac{x^{\nu-1}}{b + x}\, dx = b^{\nu-1}, \quad 0 < \nu < 1,$$

then the function d_s in the case $0 < s < 1$ becomes

$$d_s(\mathbf{p}) = \frac{\sin(s\pi)}{\pi} \int_0^\infty t^{-s}\, d \ln G(t), \quad 0 < s < 1. \quad (3.19)$$

As a particular choice of (3.19) we get $d_{\frac{1}{2}}$.

It is now sufficient to substitute $d_{\frac{1}{2}}$ in (3.12) in order to have a solution for $\boldsymbol{\alpha}(\mathbf{p})$.

It should be noted that only first n functions $d_1, d_2, \cdots, d_n$ are independent. Any other function $d_r(\mathbf{p})$ with an integer $r > n$ can be reduced to a combination of basic functions $d_1, d_2, \cdots, d_n$. The connection between different $d_r(\mathbf{p})$ can be derived from the identity

$$d_r(\mathbf{p}) = \frac{(-1)^{r-1}}{(r-1)!} \frac{d^r}{dt^r} \ln G(t)\Big|_{t=0}. \quad (3.20)$$

Moreover, by virtue of (3.11) and (3.13) there exists another useful identity

$$\mathsf{P}^r = (n/r)\hat{\mathbf{N}} d_r(\mathbf{p}). \quad (3.21)$$

By means of (3.21) one can find matrix elements of the matrix P^r with integer r in terms of functions $d_1, d_2, \cdots, d_n$.

B. Group $IU(m, m)$

Since the group $U(m, m)$ is a subgroup of $SL(2m, C)$, the results for $\boldsymbol{\alpha}(\mathbf{p})$ obtained in the preceding section can be taken over without change, if we only consider $\boldsymbol{\alpha}(\mathbf{p})$ on the subspace of momenta $\tilde{\mathbf{p}}_A$. However, the simplifications cannot be made distinctive in such an approach. Therefore it is reasonable to undertake a special investigation.

Let us consider again momentum matrix P. Using (2.17) one can establish a relation between the first m eigenvalues $\tilde{\pi}(k)$ and other ones $\tilde{\pi}(k + m)$:

$$\tilde{\pi}(k + m) = \tilde{\pi}^{-1}(k), \quad k = 1, 2, \cdots, m.$$

Let us denote projectors on eigenvalues $\tilde{\pi}(k)$ and $\tilde{\pi}(k + m)$ by $\boldsymbol{\theta}(k)$ and $\boldsymbol{\theta}'(k)$ correspondingly and set up their even and odd combinations (with respect to reflection τ):

$$\boldsymbol{\chi}^{(\pm)}(k) = \boldsymbol{\theta}(k) \pm \boldsymbol{\theta}'(k)$$

so that $\tau\boldsymbol{\chi}^{(\pm)}\tau = \pm\boldsymbol{\chi}^{(\pm)}$. In the same way instead of "unit vector" $\mathbf{n}_A(k)$, we introduce two "unit vectors"

$$\mathbf{n}_A^{(\pm)}(k) = \tilde{\mathbf{n}}_A(k) \pm \tilde{\mathbf{n}}_A(k + m),$$

which manifest themselves also as coefficients in the $\boldsymbol{\Lambda}_A$ expansion of projectors:

$$\boldsymbol{\chi}^{(\pm)}(k) = \boldsymbol{\Lambda}_A \mathbf{n}_A^{(\pm)}(k).$$

The new "unit vectors" satisfy the relations

$$\tfrac{1}{2}(1 \pm \mathbf{e}_A)\mathbf{n}_A^{(\pm)}(k) = \mathbf{n}_A^{(\pm)}(k); \quad \tfrac{1}{2}(1 \mp \mathbf{e}_A)\mathbf{n}_A^{(\pm)}(k) = 0 \quad (3.22)$$

with $\mathbf{e}_A$ defined by (2.16). It then follows from (3.22) that

$$n_a^{1(+)} = n_a^{2(+)} = n_a^{0(-)} = n_a^{3(-)} = 0. \quad (3.22')$$

The momentum matrix (3.2) now becomes

$$\tilde{\mathsf{P}} = \sum_{k=1}^{n} [\boldsymbol{\chi}^{(+)}(k) \cosh \tilde{\beta}(k) + \boldsymbol{\chi}^{(-)}(k) \sinh \tilde{\beta}(k)], \quad \tilde{\beta} = \ln \tilde{\pi}, \quad (3.23)$$

whereas (3.13) has to be replaced by

$$d_s(\mathbf{p}) = \sum_{k=1}^{m} [\tilde{\pi}^s(k) + \tilde{\pi}^{-s}(k)] = 2\sum_{k=1}^{m} \cosh(s\tilde{\beta}(k)). \quad (3.24)$$

Our main expression for $\boldsymbol{\alpha}(\mathbf{p})$ is (3.12), which contains the differential operator $\hat{\mathbf{N}}$. Therefore, in order to make use of (3.12), one should know, in general, the function $\pi(\mathbf{p})$ apart from the subspace of $\tilde{\mathbf{p}}_A$. However, one may notice that an operator $\frac{1}{2}(1 - \mathbf{e}_A)\mathbf{N}_A^0$ is a generator of $IU(m, m)$ and, consequently, its application to functions of $\mathbf{p}_A$ cannot bring these functions out of subspace $\tilde{\mathbf{p}}_A$. Thus, it is possible to apply $(1 - \mathbf{e}_A)\mathbf{N}_A^0$ to functions of $\tilde{\pi}(k)$. For example, we have from (3.11)

$$n(1 - \mathbf{e}_A)\mathbf{N}_A^0 \tilde{\beta}(k) = \mathbf{n}^{(-)}(k).$$

Let us split up the Wigner operator $\tilde{\boldsymbol{\alpha}}(\mathbf{p})$ into odd and even parts:

$$\tilde{\boldsymbol{\alpha}}(\mathbf{p}) = \boldsymbol{\alpha}^{(+)}(\mathbf{p}) + \boldsymbol{\alpha}^{(-)}(\mathbf{p}), \quad \tau\boldsymbol{\alpha}^{(\pm)}\tau = \pm\boldsymbol{\alpha}^{(\pm)}. \quad (3.25)$$

The odd-parity part becomes now

$$\boldsymbol{\alpha}^{(-)}(\mathbf{p}) = n\boldsymbol{\Lambda}_A(1 - \mathbf{e}_A)\mathbf{N}_A^0 \tilde{d}_{\frac{1}{2}}(\mathbf{p}). \quad (3.26)$$

The even-parity part of $\tilde{\alpha}(\mathbf{p})$ can be calculated if $\tilde{d}_{\frac{1}{4}}(\mathbf{p})$ is known

$$\alpha^{(+)}(\mathbf{p}) = 2\gamma^2(\mathbf{p}) + 1, \tag{3.27}$$

where

$$\gamma(\mathbf{p}) = 2n\Lambda_A(1 - \mathbf{e}_A)\mathbf{N}_A^0\tilde{d}_{\frac{1}{4}}(\mathbf{p}).$$

The function $\tilde{G}(t)$ in this case takes the form

$$\tilde{G}(t) = \prod_{k=1}^{m}[1 + (\tilde{\pi}(k) + \tilde{\pi}^{-1}(k))t + t^2] = T(y)(t^2 + 1)^m, \tag{3.28}$$

where $T(y)$ is a polynomial of degree m,

$$T(y) = \sum_{k=1}^{m}\varphi_k(\mathbf{p})y^k, \quad y = t(t^2 + 1)^{-1}. \tag{3.29}$$

Coefficients $\varphi_k(\mathbf{p})$ of the polynomial can be represented by the already known function A_k [see (3.17)] but with different arguments:

$$\varphi_k(p) = A_k(\rho_1, \rho_2, \cdots, \rho_k), \qquad \rho_k = \sum_{r=1}^{m}[\tilde{\pi}(r) + \tilde{\pi}^{-1}(r)]^k. \tag{3.30}$$

New arguments ρ_k are functions of $\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_m$:

$$\rho_{2r} = \sum_{k=0}^{r-1}\binom{2r}{k}\tilde{d}_{2(r-k)} + \binom{2r}{r}, \qquad \rho_{2r-1} = \sum_{k=0}^{r-1}\binom{2r-1}{k}\tilde{d}_{2r-2k-1}. \tag{3.31}$$

Polynomial $T(y)$ has its zeros at $(-2\cosh\tilde{\beta}(k))^{-1}$. These points can be calculated explicitly for the case $n = 8$ or $m = 4$. In the general case one has to return to the integral formula (3.19).

As a result, $\tilde{d}_s(\mathbf{p})$ becomes, for $0 < s < 1$,

$$\tilde{d}_s(\mathbf{p}) = \frac{\sin(s\pi)}{\pi}\int_0^\infty t^{-s}\, d\ln\left\{\sum_{k=1}^{m}\varphi_k(\mathbf{p})t^k(t^2 + 1)^{m-k}\right\}, \tag{3.32}$$

where functions $\psi_k(\mathbf{p})$ are given by (3.30), (3.17), and (3.31).

We can conclude that (3.32) is in fact a solution of the problem, because according to (3.26), (3.27) one can obtain $\tilde{\alpha}(\mathbf{p})$ by simply differentiating (3.32).

C. Group $\overline{ISL}(m, C)$

The Wigner operator can be obtained in a straightforward manner from (2.18):

$$\alpha(\mathbf{p}) = \tfrac{1}{2}(1 + \sigma_1)\otimes\mathbf{a}(\mathbf{p}) + \tfrac{1}{2}(1 - \sigma_1)\otimes\mathbf{a}^{-1}(\mathbf{p}). \tag{3.33}$$

Here $\mathbf{a}(\mathbf{p})$ denotes the $\alpha(\mathbf{p})$ operator for the group $ISL(m, C)$ *without* reflection, which has been already considered in Sec. 3A.

4. STRUCTURE OF MOMENTUM SPACE

Properties of $\alpha(\mathbf{p})$ are closely connected with the structure of the momentum space of $ISL(n, C)$. For example, the complexity of $\alpha(\mathbf{p})$ is closely connected with the fact that in the case of $ISL(n, C)$ there are $(n - 1)$ independent $SU(n)$-invariant quantities which can be built out of momenta $\mathbf{p}_A$. Therefore, an investigation of the momentum space may be instructive.

Let us write the transformation $\alpha(\mathbf{p})$ as a product of two transformations α^0 and v

$$\mathsf{P}^0 \to \pi, \quad \pi = \alpha^0\mathsf{P}^0\alpha^0, \tag{4.1}$$

$$\pi \to \mathsf{P}, \quad \mathsf{P} = \mathsf{v}\pi\mathsf{v}^\dagger, \quad \det\mathsf{v} = 1, \tag{4.2}$$

where $\pi = \operatorname{diag}\{\pi(1), \pi(2), \cdots, \pi(n)\}$, and v is a unitary $n \times n$ matrix. The transformation (4.1) is analogous, in some sense, to the pure Lorentz transformation along z axis; it contains $(n - 1)$ parameters which in our case replace the relative velocity. If one puts $\alpha^0 = \exp\mathsf{z}^0(\varphi)$, then $\pi = \exp 2\mathsf{z}^0(\varphi)$, while the matrix z^0 is given in the Weyl basis by

$$\mathsf{z}^0(\varphi) = \sum_{\zeta=1}^{n-1}h_\zeta\varphi_\zeta, \quad \zeta = 0, (\zeta), \quad (\zeta) = 1, 2, \cdots, n - 1. \tag{4.3}$$

Hereafter, we employ a notation $\mathbf{h}_\zeta$ for the diagonal matrices Λ_A. We adopt a convention that $\Lambda_1, \Lambda_2, \cdots, \Lambda_{n-1}$ coincide with $\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{n-1}$, and $\mathbf{h}_0 = \Lambda_0$. The matrix elements

$$(h_\zeta)_{ik} = m_\zeta(k)\delta_{ik}, \quad i, k = 1, 2, \cdots, n,$$

are related to components $m_\zeta(k)$ of the weight $\mathbf{m}(k)$ in the fundamental representation of $SU(n)$.

The normalization conditions (1.1) correspond to

$$2n\sum_{k=1}^{n}m_\zeta(k)m_{\zeta'}(k) = \delta_{\zeta\zeta'}, \quad \sum_k m_{(\zeta)}(k) = 0, \qquad 2n\sum_{\zeta=0}^{n-1}m_\zeta(k)m_\zeta(k') = \delta_{kk'}, \quad m_0(k) = (\sqrt{2}\,n)^{-1}. \tag{4.4}$$

As is well known, differences $r_\zeta(ik) = m_\zeta(i) - m_\zeta(k)$ are roots of the group $SU(n)$.

Now, the transformation (4.1) brings momentum from the "rest frame" $\mathbf{p}_A^0 = \sqrt{2}\,n\delta_{A0}$ into $\pi_A = 2n\mathbf{Sp}(\Lambda_A\pi)$, which has only first n nonvanishing components

$$\pi_\zeta = 2n\sum_{k=1}^{n}m_\zeta(k)\pi(k) = 2n\sum_{k=1}^{n}m_\zeta(k)\exp[2(\mathbf{m}(k)\boldsymbol{\varphi})]. \tag{4.5}$$

Parameters φ_ζ are related to β of (3.23):

$$\beta(k) = 2(\mathbf{m}(k)\boldsymbol{\varphi}), \quad \boldsymbol{\varphi} = n\sum_{k=1}^{n}\mathbf{m}(k)\beta(k), \qquad \sum_k\beta(k) = 0. \tag{4.6}$$

Unitary transformation v is described in momentum space by (n^2-1)-dimensional orthogonal matrix $V_{(A)(B)}$. The latter leaves unchanged the zero component of momentum $p_0 = \pi_0$:

$$p_0 = p_0(\varphi) = \sqrt{2}\sum_{k=1}^{n} \exp\,[2(\mathbf{m}(k)\boldsymbol{\varphi})]. \tag{4.7}$$

In fact, $V_{(A)(B)}$ acts only on "spacelike" components $\mathbf{p}_{(A)}$ and is analogous to the spatial rotations of the Poincaré group. The space spanned by the vectors $\mathbf{p}_{(A)}$ is denoted by Ω. A subspace of Ω, spanned by vectors $\boldsymbol{\pi}_A$, is denoted by ω. The latter coincides with the weight space of the fundamental representation. Equation (4.5) defines a vector $\boldsymbol{\pi} \in \omega$ as a sum of vectors $2n\mathbf{m}(k)\pi(k)$ of length determined by $\pi(k)$, and with direction coinciding with that of the weight k.

Under the transformation (4.2) subspace ω moves orthogonally inside Ω, while weights undergo an orthogonal transformation of $SU(n)$ (we recall that weights transform according to the adjoint representation).

Comparing (4.5) with (3.7) one can easily convince oneself that

$$\mathbf{n}_{(A)}(k) = 2n\sum_{\zeta} V_{(A)(\zeta)}\mathbf{m}_{(\zeta)}(k), \quad \mathbf{n}_0(k) = 2n\mathbf{m}_0(k). \tag{4.8}$$

Thus the set of "unit vectors" $\mathbf{n}_{(A)}(k)$ is nothing but a set of weights in fundamental representation, which is transformed according to v of $SU(n)$.

Relations (3.7) then select admissible sets of weights (in fundamental representation) oriented in space Ω by means of $V_{(A)(B)}$. It should be noted that in fact the first three relations (3.7) may be obtained from (4.4) taking into account only orthogonality of $V_{(A)(B)}$, whereas the last relation (3.7) extracts out of orthogonal transformations belonging to $SU(n)$.

All the momenta

$$\mathbf{p}_{(A)}(\varphi) = V_{(A)(B)}\boldsymbol{\pi}_{(B)}(\varphi) \tag{4.9}$$

corresponding to different v form a set in the space Ω, which is a minimal $SU(n)$-invariant set of points equivalent to $\boldsymbol{\pi}_{(B)}$. The set includes $\boldsymbol{\pi}_{(B)}$ as a particular point and is entirely determined by $\boldsymbol{\pi}_{(B)}$ or parameters (φ). Every vector $\mathbf{p}_{(A)}(\varphi)$ has the same stationary subgroup S, which coincides with that of $\boldsymbol{\pi}_{(B)}(\varphi)$.

The stationary subgroup is the most essential characteristics of a vector $\mathbf{p}_{(A)}$. Therefore we subdivide the space Ω into subspaces Ω_α containing vectors with the same stationary subgroup S_α,

$$\Omega = \sum_\alpha \Omega(S_\alpha) \equiv \sum_\alpha \Omega_\alpha. \tag{4.10}$$

The corresponding subdivision of ω into subspaces ω_α is

$$\omega = \sum_\alpha \omega(S_\alpha) \equiv \sum_\alpha \omega_\alpha \tag{4.11}$$

such that Ω_α is a minimal $SU(n)$-invariant space containing ω_α.

We also consider the space Φ of parameters (φ), defined by (4.5). Because of the one-to-one correspondence between vectors $\boldsymbol{\pi}_\zeta \in \omega$ and vectors $\boldsymbol{\varphi} \in \Phi$, the subdivision (4.11) implies in turn the subdivision of space Φ:

$$\Phi = \sum \Phi(S_\alpha) \equiv \sum_\alpha \Phi_\alpha. \tag{4.12}$$

The type of the stationary group of a given vector depends entirely on the number and position of the coinciding eigenvalues $\pi(k)$ of P. For the description of the degeneration of P, let us introduce the set of integer numbers $x_1, x_2, \cdots, x_l$ in the following manner:

$$\pi(1) = \pi(2) = \cdots = \pi(x_1),$$
$$\pi(x_1+1) = \cdots = \pi(x_1+x_2), \cdots,$$
$$\pi\left(\sum_{k=1}^{l-1} x_k + 1\right) = \cdots = \pi(n),$$

so that

$$\sum_{k=1}^{l} x_k = n.$$

The enumeration of $\pi(k)$ is of no importance, since interchanging of $\pi(k)$ can be obtained by means of the Weyl transformations which are a particular kind of transformations v and do not change the type of the stationary subgroup. Thus, stationary subgroup S_k is properly characterized by a set $(x_1, x_2, \cdots, x_l)$. One can easily verify that

$$S(x_1, x_2, \cdots, x_l) = SU(x_1) \otimes SU(x_2) \otimes \cdots \otimes SU(x_l) \otimes [U(1)]^{\oplus(l-1)}.$$

Such a subgroup contains

$$\left(\sum_{k=1}^{l} x_k^2\right) - 1$$

parameters. The dimension of the minimal invariant subspace of $\Omega(x_1, \cdots, x_l)$ is denoted by

$$\eta(x_1, \cdots, x_l) = n^2 - \sum_{k=1}^{l} x_k^2.$$

Then, the momentum $\mathbf{p}_{(A)} \in \Omega(x_1, \cdots, x_l)$ is

$$\mathbf{p}_{(A)} = \pi(1)\sum_{k=1}^{x_1} \mathbf{n}_{(A)}(k) + \cdots + \pi\left(\sum_{i=1}^{l-1} x_i + 1\right) \times \sum_{k=x_{l-1}+1}^{x_l} \mathbf{n}_{(A)}(k), \tag{4.13}$$

where all $\pi(k)$ are different.

If we adopt a lexicographic ordering and make use of the freedom of enumeration of $\pi(k)$, we can say without loss of generality that $\mathbf{m}(1) \geq \mathbf{m}(2) \geq \cdots \geq \mathbf{m}(n)$. The subspace $\omega(x_1, \cdots, x_l)$ is then spanned by vectors

$$(2n)^{-1}\boldsymbol{\pi} = \pi(1)\sum_{k=1}^{x_1}\mathbf{m}(k) + \cdots + \pi\left(\sum_{i=1}^{l-1}x_i + 1\right) \times \sum_{k=x_{l-1}+1}^{x_l}\mathbf{m}(k). \quad (4.14)$$

Vector $\mathbf{r}(k) = \mathbf{m}(k) - \mathbf{m}(k+1)$ is a simple root of $SU(n)$. Denote by $\tilde{\omega}(x_1, x_2, \cdots, x_l)$ a subspace orthogonal to simple roots $\mathbf{r}(1), \cdots, \mathbf{r}(x_1 - 1)$, $\mathbf{r}(x_1 + 1), \cdots, \mathbf{r}(x_1 + x_2 - 1), \cdots,$

$$\mathbf{r}\left(\sum_{k=1}^{l-1}x_k + 1\right), \cdots, \mathbf{r}(n-1).$$

This subspace contains $\omega(x_1, \cdots, x_l)$:

$$\tilde{\omega}(x_1, \cdots, x_l) = \omega(x_1, \cdots, x_l) + \sum_{\alpha}\omega'_{\alpha}. \quad (4.15)$$

The sum in (4.15) is taken over all the stationary subgroups containing $S(x_1, x_2, \cdots, x_l)$ among their subgroups. One can enumerate a group subdividing the set of l integer numbers x_k into various subsets and replacing each subset of by just one number—the sum of x_k in the subset.

The parameter space $\Phi(x_1, \cdots, x_l)$ is formed by all vectors $\boldsymbol{\varphi}$ with the structure very similar to (4.14),

$$n^{-1}\boldsymbol{\varphi} = \beta(1)\sum\mathbf{m}(k) + \cdots + \beta\left(\sum^{l-1}x_k + 1\right)\sum^{x_l}\mathbf{m}(k). \quad (4.16)$$

As in the case of $\boldsymbol{\pi}$ space, one can also introduce a subspace $\tilde{\Phi}(x_1, \cdots, x_l)$ orthogonal to the sequence of simple roots discussed in connection with (4.15). Then the spaces Φ and $\tilde{\Phi}$ are interrelated by means of the formula similar to (4.15).

Let us now obtain the condition satisfied by p_A so that it belongs to the invariant minimal space $\tilde{\Omega}(x_1, \cdots, x_l)$, which includes $\tilde{\omega}(x_1, \cdots, x_l)$. It follows from the definition of $d_s(p)$ that in the case of $\pi(k) \in \tilde{\omega}(x_1, \cdots, x_l)$,

$$d_s(\mathbf{p}) = \sum_{k=1}^{l}x_k\left[\pi\left(\sum_{i=1}^{k-1}x_i + 1\right)\right]^s. \quad (4.17)$$

As a result, n quantities $d_s(\mathbf{p})$ are connected by $f = n - l + 1$ relations of a type $F_\gamma(d_1, \cdots, d_n) = 0$, $\gamma = 1, \cdots, f$, which are the conditions singling out the space $\tilde{\Omega}(x_1, \cdots, x_l)$. Note that the dimensionality of both $\tilde{\omega}(x_1, \cdots, x_l)$ and $\tilde{\varphi}(x_1, \cdots, x_l)$ is equal to $l - 1$, while that of $\tilde{\Omega}(x_1, \cdots, x_l)$ is equal to

$$n^2 - \sum_{k=1}^{l}x_k^2 + l - 1.$$

5. GENERATORS IN UNITARY REPRESENTATION

In order to complete the discussion of unitary representations of $ISL(n, C)$ let us consider briefly the form of generators.

An infinitesimal transformation $\mathsf{U}(0, \mathsf{V})$ by definition of generators is given by

$$\mathsf{U}(0, \mathsf{V}) \simeq \mathsf{I} + \mathbf{M}_{(A)}\alpha_{(A)} + \mathbf{N}_{(A)}\beta_{(A)}. \quad (5.1)$$

On the other hand, the change of $\varphi(\mathbf{p})$ arising from $\mathsf{U}(0, \mathsf{V})$ can be calculated from (V):

$$\mathsf{U}(0, \mathsf{V})\varphi(\mathbf{p}) = \mathsf{Q}(\alpha^{-1}(\mathbf{p})\mathsf{V}\boldsymbol{\alpha}(\mathbf{p}'))\varphi(\mathbf{p}') \quad (5.2)$$

with $\mathsf{V}\mathsf{P}'\mathsf{V}^\dagger = \mathsf{P}$. To this end, one has to replace V in (5.2) by

$$\mathsf{V} = \mathsf{I} - i\boldsymbol{\Lambda}_{(A)}\alpha_{(A)} + \boldsymbol{\Lambda}_{(A)}\beta_{(A)} \quad (5.3)$$

and to find the dependence of P' on infinitesimal parameters α_A, β_A and momenta $\mathbf{p}$. The expression for $\mathbf{M}_{(A)}$ and $\mathbf{N}_{(A)}$ follows then from the comparison of (5.1) with the infinitesimal form (5.2).

Let us first consider the scalar representation, when $\mathsf{Q} = \mathsf{I}$. In this case the generators are denoted by $M^0_{(A)}$ and $N^0_{(A)}$. A straightforward calculation gives

$$\mathbf{M}^0_{(A)} = -F_{ABC}\mathbf{p}_B\,\partial/\partial\mathbf{p}_C; \quad \mathbf{N}^0_{(A)} = D_{(A)BC}\mathbf{p}_B\,\partial/\partial\mathbf{p}_C. \quad (5.4)$$

The operators $\mathbf{M}^0_{(A)}$ and $\mathbf{N}^0_{(A)}$ correspond to the orbital angular momentum in the case of the Poincaré group. Note that we have already used $\mathbf{N}^0_{(A)}$ [see (3.10) and (3.11)].

In nonscalar representations the generators $\mathbf{M}_{(A)}$ and $\mathbf{N}_{(A)}$ take the form

$$\mathbf{M}_{(A)} = \mathbf{M}^0_{(A)} + \mathfrak{M}_{(A)}, \quad \mathbf{N}_{(A)} = \mathbf{N}^0_{(A)} + \mathfrak{N}_{(A)}. \quad (5.5)$$

Here operators $\mathfrak{M}_{(A)}$ and $\mathfrak{N}_{(A)}$ have arisen from the variation of the matrix

$$\mathsf{Q}(\boldsymbol{\alpha}^{-1}\mathsf{V}\boldsymbol{\alpha}') = \mathsf{I} + \mathfrak{M}_{(A)}\boldsymbol{\alpha}_{(A)} + \mathfrak{N}_{(A)}\boldsymbol{\beta}_{(A)}$$

under an infinitesimal V with

$$\boldsymbol{\alpha}' = \boldsymbol{\alpha} + 2n\mathrm{Sp}(\boldsymbol{\Lambda}_A\,d\mathsf{P})\,\partial\alpha/\partial\mathbf{p}_A.$$

In the case of the Poincaré group, operators $\mathfrak{M}_{(A)}$ and $\mathfrak{N}_{(A)}$ are connected with the spin; for this case they were derived by Shirokov.[8]

[8] Yu. M. Shirokov, Dokl. Akad. Nauk (SSSR) **94**, 857 (1954); C. Fronsdal, Phys. Rev. **113**, 1367 (1959).

The operators $\mathfrak{N}_{(A)}$ contains $\mathfrak{M}_{(A)}$ linearly

$$\mathfrak{N}_{(A)} = \mathsf{K}_{(A)(B)}\mathfrak{M}_{(B)},$$
$$\mathsf{K}_{(A)(B)} = -2n \operatorname{Im} \mathbf{Sp}\{\mathbf{\Lambda}_{(B)}\boldsymbol{\alpha}^{-1}(\mathbf{p})[\mathbf{\Lambda}_{(A)} + \mathbf{N}^0_{(A)}]\boldsymbol{\alpha}(\mathbf{p})\}. \quad (5.6)$$

If one inserts now $\boldsymbol{\alpha}(\mathbf{p})$ into (5.6), one gets an expression of $\mathsf{K}_{(A)(B)}$ in terms of variables $d_1, \cdots, d_n$.

Evidently the method of explicit construction of unitary representations developed for the case of the group $ISL(n, C)$ may be applied to all groups which are subgroups of $ISL(n, C)$. The class of such groups are rather large so that the method described above is of a considerable generality. As an example we have treated the group $IU(m, m)$ in Sec. 3B; a special case $m = 2$, i.e., the group $IU(2, 2)$, was investigated in detail earlier.[9] No additional difficulties arise also if one considers the group $IU(m, m')$ with $m + m' = n$.[9]

The problem of the construction of $\boldsymbol{\alpha}(\mathbf{p})$ for the groups $IU(m, m)$ and $\overline{ISL}(m, C)$ is equivalent to that of solution of the linear covariant equation of the type $Q\psi = \psi$. Therefore, it would be superfluous to discuss these equations separately.

[9] Yu. V. Novozhilov and I. A. Terentjev, in "International Summer School in Elementary Particles Theory," Jalta (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Note on Statistical Mechanics of Random Systems*

T. BURKE AND J. L. LEBOWITZ
Belfer Graduate School of Science, Yeshiva University, New York

(Received 23 October 1967)

We discuss some equilibrium properties of random systems, i.e., systems whose Hamiltonian depends on some random variables y with a distribution $P(y)$ which is independent of the dynamic state of the system. For a system of noninteracting particles which interact with randomly distributed scattering centers, the important quantity is the average density of states of a single particle per unit volume, $n(E)$. Feynman's path-integral formulation of quantum statistics is used to derive some properties of the average partition function for one particle $\langle Z_1 \rangle$, which is the Laplace transform of $n(E)$. In particular, it is shown that $\langle Z_1 \rangle$ is an analytic function of the density of scatterers ρ for a wide class of particle-scattering center potentials $V(\mathbf{r})$, including those with hard cores and those with negative parts. The analyticity in ρ of the equilibrium properties of these systems is very remarkable and is in contrast to the conjectured nonanalytic behavior of their transport (i.e., diffusion) coefficients. We find also upper and lower bounds on $\langle Z_1 \rangle$ for a particle acted upon by a random potential $V(\mathbf{r})$ obeying Gaussian statistics with $\langle V(\mathbf{r})V(\mathbf{r}') \rangle \sim \exp[-\alpha^2(\mathbf{r} - \mathbf{r}')^2]$. In the limit of "white noise," $\langle V(\mathbf{r})V(\mathbf{r}') \rangle \sim \delta(\mathbf{r} - \mathbf{r}')$, $\langle Z_1 \rangle$ is shown to diverge in two and three dimensions but remains finite in one dimension. This agrees with approximate results on the density of states. In appendices we prove the existence, in the thermodynamic limit, of the free-energy density of a system with random scatterers and also of the frequency distribution and, thus, the free-energy density for a random harmonic crystal.

1. INTRODUCTION

It is the purpose of this paper to investigate some equilibrium properties of certain random systems. By a random system we mean a system whose Hamiltonian depends on some parameters y which are, in a sense to be specified, random variables with a distribution $P(y)$ which is independent of the dynamical state of the system. The dynamical (canonical) variables of the system are denoted by x so that the Hamiltonian $H(x, y)$ is a function both of x and y with x varying in time according to some dynamical law.

Typical examples of model random systems discussed in the literature are (i) a lattice in which spins are located randomly on sites with a probability p independent of the other sites, (ii) a harmonic crystal in which the masses or spring constants have random values with some specified distribution, and (iii) a set of particles, usually electrons, acted on by an external potential due to the presence of centers of force (scatterers, impurities) at positions $\{\mathbf{R}_1, \cdots, \mathbf{R}_N\} \equiv \{R\} \equiv y$. In the last case, which will be our primary concern here, the Hamiltonian of a set of M particles with positions $\{\mathbf{r}_1, \cdots, \mathbf{r}_M\}$ and momenta $\{\mathbf{p}_1, \cdots, \mathbf{p}_M\}$ in a box of volume Ω is given by

$$H = \sum_{i=1}^{M} p_i^2/2m + U(\mathbf{r}_1, \cdots, \mathbf{r}_M) + \lambda \sum_{i=1}^{M}\sum_{n=1}^{N} V(\mathbf{r}_i - \mathbf{R}_n), \quad (1.1)$$

where U is the interparticle potential which is independent of the location (or presence) of the impurities

* Supported by the Air Force Office of Scientific Research under Grant No. 508-66.

The operators $\mathfrak{N}_{(A)}$ contains $\mathfrak{M}_{(A)}$ linearly

$$\mathfrak{N}_{(A)} = \mathsf{K}_{(A)(B)}\mathfrak{M}_{(B)},$$
$$\mathsf{K}_{(A)(B)} = -2n \operatorname{Im} \mathbf{Sp}\{\mathbf{\Lambda}_{(B)}\boldsymbol{\alpha}^{-1}(\mathbf{p})[\mathbf{\Lambda}_{(A)} + \mathbf{N}^0_{(A)}]\boldsymbol{\alpha}(\mathbf{p})\}. \quad (5.6)$$

If one inserts now $\boldsymbol{\alpha}(\mathbf{p})$ into (5.6), one gets an expression of $\mathsf{K}_{(A)(B)}$ in terms of variables $d_1, \cdots, d_n$.

Evidently the method of explicit construction of unitary representations developed for the case of the group $ISL(n, C)$ may be applied to all groups which are subgroups of $ISL(n, C)$. The class of such groups are rather large so that the method described above is of a considerable generality. As an example we have treated the group $IU(m, m)$ in Sec. 3B; a special case $m = 2$, i.e., the group $IU(2, 2)$, was investigated in detail earlier.[9] No additional difficulties arise also if one considers the group $IU(m, m')$ with $m + m' = n$.[9]

The problem of the construction of $\boldsymbol{\alpha}(\mathbf{p})$ for the groups $IU(m, m)$ and $\overline{ISL}(m, C)$ is equivalent to that of solution of the linear covariant equation of the type $Q\psi = \psi$. Therefore, it would be superfluous to discuss these equations separately.

[9] Yu. V. Novozhilov and I. A. Terentjev, in "International Summer School in Elementary Particles Theory," Jalta (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Note on Statistical Mechanics of Random Systems*

T. BURKE AND J. L. LEBOWITZ
Belfer Graduate School of Science, Yeshiva University, New York

(Received 23 October 1967)

We discuss some equilibrium properties of random systems, i.e., systems whose Hamiltonian depends on some random variables y with a distribution $P(y)$ which is independent of the dynamic state of the system. For a system of noninteracting particles which interact with randomly distributed scattering centers, the important quantity is the average density of states of a single particle per unit volume, $n(E)$. Feynman's path-integral formulation of quantum statistics is used to derive some properties of the average partition function for one particle $\langle Z_1 \rangle$, which is the Laplace transform of $n(E)$. In particular, it is shown that $\langle Z_1 \rangle$ is an analytic function of the density of scatterers ρ for a wide class of particle-scattering center potentials $V(\mathbf{r})$, including those with hard cores and those with negative parts. The analyticity in ρ of the equilibrium properties of these systems is very remarkable and is in contrast to the conjectured nonanalytic behavior of their transport (i.e., diffusion) coefficients. We find also upper and lower bounds on $\langle Z_1 \rangle$ for a particle acted upon by a random potential $V(\mathbf{r})$ obeying Gaussian statistics with $\langle V(\mathbf{r})V(\mathbf{r}') \rangle \sim \exp[-\alpha^2(\mathbf{r} - \mathbf{r}')^2]$. In the limit of "white noise," $\langle V(\mathbf{r})V(\mathbf{r}') \rangle \sim \delta(\mathbf{r} - \mathbf{r}')$, $\langle Z_1 \rangle$ is shown to diverge in two and three dimensions but remains finite in one dimension. This agrees with approximate results on the density of states. In appendices we prove the existence, in the thermodynamic limit, of the free-energy density of a system with random scatterers and also of the frequency distribution and, thus, the free-energy density for a random harmonic crystal.

1. INTRODUCTION

It is the purpose of this paper to investigate some equilibrium properties of certain random systems. By a random system we mean a system whose Hamiltonian depends on some parameters y which are, in a sense to be specified, random variables with a distribution $P(y)$ which is independent of the dynamical state of the system. The dynamical (canonical) variables of the system are denoted by x so that the Hamiltonian $H(x, y)$ is a function both of x and y with x varying in time according to some dynamical law.

Typical examples of model random systems discussed in the literature are (i) a lattice in which spins are located randomly on sites with a probability p independent of the other sites, (ii) a harmonic crystal in which the masses or spring constants have random values with some specified distribution, and (iii) a set of particles, usually electrons, acted on by an external potential due to the presence of centers of force (scatterers, impurities) at positions $\{\mathbf{R}_1, \cdots, \mathbf{R}_N\} \equiv \{R\} \equiv y$. In the last case, which will be our primary concern here, the Hamiltonian of a set of M particles with positions $\{\mathbf{r}_1, \cdots, \mathbf{r}_M\}$ and momenta $\{\mathbf{p}_1, \cdots, \mathbf{p}_M\}$ in a box of volume Ω is given by

$$H = \sum_{i=1}^{M} p_i^2/2m + U(\mathbf{r}_1, \cdots, \mathbf{r}_M) + \lambda \sum_{i=1}^{M}\sum_{n=1}^{N} V(\mathbf{r}_i - \mathbf{R}_n), \quad (1.1)$$

where U is the interparticle potential which is independent of the location (or presence) of the impurities

* Supported by the Air Force Office of Scientific Research under Grant No. 508-66.

and $\lambda V(\mathbf{r})$ is the potential energy of a particle in the presence of an impurity located at $\mathbf{r} = 0$.

In all these cases the physical situation, idealized in these models, corresponds to starting with the system at a high temperature and quenching it. The imperfections will then remain fixed in space and can be described by the parameters y, whereas the other degrees of freedom x come to equilibrium with a Hamiltonian which depends on y. (The actual physical situation is more complex, but this is a reasonable approximation to it.) For a given system, y has, of course, some specified set of values $\{\mathbf{R}_1, \cdots, \mathbf{R}_N\}$, and it is therefore necessary to state what we mean by saying that the $\mathbf{R}_i$ are distributed at random. To be more specific, how do we compute the properties, such as the internal energy or specific heat, of a macroscopic system for which the values of the $\mathbf{R}_i$ are unknown to us? The simplest attitude one can take is to make measurements on an ensemble of systems prepared in the same way.[1] (This is a superensemble of thermal ensembles for each specified $\{\mathbf{R}_1, \cdots, \mathbf{R}_N\}$.) Hence the average of any function $A(x, y)$ is given by

$$\langle A \rangle = \int dy P(y) \langle A \rangle_y, \tag{1.2}$$

where

$$\langle A \rangle_y = [\mathrm{Tr}_x (A(x, y) e^{-\beta H(x,y)})]/Z(y), \tag{1.3}$$

$$Z(y) = \mathrm{Tr}_x\, e^{-\beta H(x,y)}, \tag{1.4}$$

is the partition function for the system represented by a canonical ensemble with inverse temperature β. Here $P(y)$ is the distribution of random variables y which, for case (iii) above, is

$$P(y)\, dy = \Omega^{-N}\, d\mathbf{R}_1, \cdots, d\mathbf{R}_N; \quad \mathbf{R}_i \text{ in } \Omega. \tag{1.5}$$

It is also possible to consider the case where the number of scatterers in Ω is not fixed, but has an average value N. (This is actually used in Appendix B.) The thermodynamic properties of this ensemble would then be obtained from the free energy F defined as[1]

$$F = -\beta^{-1} \langle \ln Z(y) \rangle = -\beta^{-1} \int P(y) \ln Z(y)\, dy. \tag{1.6}$$

This prescription for obtaining the properties of a system will be satisfactory only when the dispersal in the values of $\ln Z(y)$ is small. This will be true for a macroscopic system when $\ln Z(y)$ is an additive quantity. The existence and some properties of the free-energy density in the thermodynamic limit $\lim_{\Omega \to \infty} (F/\Omega)$ for the spin system have been discussed previously.[2] A proof of the existence of this quantity, as well as the thermodynamic limit of the frequency distribution function for random harmonic crystals, is given in Appendix A.

In this paper we are primarily interested in case (iii). The existence of $\lim_{\Omega \to \infty} (F/\Omega)$ for this system is discussed in Appendix B. In the special situation when the particles are noninteracting ($U = 0$) and may be treated by Boltzmann statistics, the problem reduces to an investigation of one of the most frequently used models in the study of irreversible phenomena, that in which a single particle interacts with a set of scattering centers randomly distributed in space. General properties of perturbation expansions for the resolvent operator and the density matrix have been discussed in many papers.[3] In Sec. 2 we study properties of this system using the path-integral formulation of quantum statistics which was introduced by Feynman.[4] For M noninteracting particles obeying Boltzmann statistics, (1.4) becomes

$$Z_M = [Z_1(y)]^M / M! \tag{1.7}$$

where $Z_1(y)$ is the (quantum-mechanical) partition function for a single particle interacting in a volume Ω with N scatterers located at positions $\{\mathbf{R}_1, \cdots, \mathbf{R}_N\} \equiv y$. From (1.6)

$$\begin{aligned}(F/\Omega) &= -(\beta\Omega)^{-1} \langle \ln (Z_1^M/M!) \rangle \\ &= -(M/\beta\Omega)[1 - \ln (M/\Omega) + \langle \ln [Z_1(y; \Omega)/\Omega] \rangle],\end{aligned} \tag{1.8}$$

where we have used Stirling's formula for $\ln M!$.

We now argue somewhat heuristically that, since Z_1 is an extensive quantity in this case, we will have $\langle \ln (Z_1/\Omega) \rangle \sim \ln \langle Z_1/\Omega \rangle$ in the limit $\Omega \to \infty$, $N \to \infty$, $N/\Omega = \rho = \text{const}$. To see this more formally we divide up the volume Ω into J cubes, ω_i ($i = 1, \cdots, J$), each of volume ω; we consider the limits $J \to \infty$ followed by $\omega \to \infty$. For reasonable forms of the interaction potential $V(\mathbf{r})$ we should be able to neglect, for sufficiently large ω, the interaction between a particle in one cube ω_i with impurities outside ω_i as well as the precise boundary conditions. Thus

$$\Omega^{-1} Z_1(y; \Omega) \sim \Omega^{-1} \sum_{i=1}^{J} Z_1(y^{(i)}; \omega) \equiv J^{-1} \sum_{i=1}^{J} (z_i/\omega). \tag{1.9}$$

In (1.9) the z_i can be considered independent random variables so that, by the central limit theorem, in the

[1] R. Brout, Phys. Rev. **115**, 824 (1959); M. W. Klein and R. Brout, Phys. Rev. **132**, 2412 (1963).

[2] R. B. Griffiths and J. L. Lebowitz, J. Math. Phys. (to be published).

[3] See, e.g., papers cited in Refs. 6 and 7.

[4] R. P. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Co., New York, 1965). See also A. Siegert, Phys. Rev. **86**, 621 (1952).

limit of large J, $Z_1(y;\Omega)$ will have a normal distribution

$$P(Z_1) = \exp\left[-\frac{(Z_1 - \langle Z_1\rangle)^2}{2\langle(\Delta Z_1)^2\rangle}\right], \tag{1.10}$$

where

$$\langle(\Delta Z_1)^2\rangle = J\langle(\Delta z)^2\rangle$$

and

$$\begin{aligned}\langle \ln (Z_1/\Omega)\rangle &= \ln\langle Z_1/\Omega\rangle + \left\langle \ln\left[1 + \frac{Z_1 - \langle Z_1\rangle}{\langle Z_1\rangle}\right]\right\rangle \\ &= \ln\langle Z_1/\Omega\rangle - \frac{1}{2}\frac{\langle(\Delta z)^2\rangle}{J\langle z\rangle^2} + \cdots \\ &= \ln\langle Z_1/\Omega\rangle\end{aligned}$$

in the limit $J\to\infty$ when $P(Z_1)$ becomes infinitely sharply peaked about $\langle Z_1\rangle$. Thus from (1.8) in the limit $M\to\infty$, $\Omega\to\infty$, $M/\Omega = \text{const}$,

$$(F/\Omega) = -\beta^{-1}(M/\Omega)[1 - \ln(M/\Omega) + \ln(\langle Z_1\rangle/\Omega)]. \tag{1.11}$$

Alternatively, we can consider this system to be represented by a grand canonical ensemble with chemical potential μ, activity $\xi = e^{\beta\mu}$, and mean number of particles $\langle M\rangle$. The grand partition function

$$\Xi(y;\Omega) = \sum_{M=1}^{\infty} \xi^M Z_1^M/M! = e^{\xi Z_1}, \tag{1.12}$$

so that the grand potential

$$F - \mu\langle M\rangle = -\beta^{-1}\langle \ln \Xi\rangle = -\beta^{-1}\xi\langle Z_1\rangle \tag{1.13}$$

and

$$\langle M\rangle = \xi\left\langle \frac{\partial \ln \Xi(y;\Omega)}{\partial\xi}\right\rangle = \xi\langle Z_1\rangle.$$

For the free energy this gives

$$F = -\beta^{-1}\langle M\rangle[1 - \ln(\langle M\rangle/\langle Z_1\rangle)], \tag{1.14}$$

which coincides with (1.11) when $\langle M\rangle$ is associated with the M of the canonical ensemble.

The thermodynamics of this system is thus given entirely in terms of $\langle Z_1\rangle/\Omega$ in the limit when the system is large. In Sec. 2, we use the path-integral method of Feynman to investigate the existence of an expansion of $\langle Z_1\rangle/\Omega$ in powers of the density of scattering centers $\rho = N/\Omega$ in the limit $N, \Omega\to\infty$, and of an expansion in the coupling constant λ when certain conditions on the potential $V(\mathbf{r})$ between electron and scattering center are satisfied. This analyticity in λ was shown first by Doniach,[5] who used a different representation for the many-temperature Green's functions which arise in the expansion.

The importance of the quantity $\langle Z_1\rangle$ here arises from its linear relation to the average density of states

$$n(E) \equiv \langle n(E;y)\rangle = \Omega^{-1}\left\langle \sum_i \delta(E - E_i(y))\right\rangle, \tag{1.15}$$

where E_i are the energy eigenvalues of (1.1). We then have

$$\langle Z_1(\beta)\rangle/\Omega = \int_{-\infty}^{\infty} e^{-\beta E}\langle n(E;y)\rangle\, dE = \int_{-\infty}^{\infty} e^{-\beta E} n(E)\, dE. \tag{1.16}$$

Equation (1.16) is valid for both classical and quantum systems with their respective standard interpretations. For quantum systems in which the particles obey B.E. or F.D. statistics, the grand ensemble approach gives

$$\begin{aligned}\Omega^{-1}\langle \ln \Xi(y)\rangle &= \mp\int dE\, n(E) \ln(1 \mp e^{-\beta(E-\mu)}) \\ &\to e^{\beta\mu}\langle Z_1\rangle/\Omega,\end{aligned} \tag{1.17}$$

when the particle density is low. In principle we can compute the density of states from $\langle Z_1(\beta)\rangle$ by the inverse of (1.16):

$$n(E) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} d\beta\, e^{\beta E}\langle Z_1(\beta)\rangle/\Omega, \tag{1.18}$$

but this requires knowledge of $\langle Z_1(\beta)\rangle$ for complex β. The quantity $n(E)$ has been investigated recently[3] for various kinds of potential and we hope that our results will be of relevance.

In Sec. 3 we consider the properties of a particle in a potential random in space and obeying Gaussian statistics. This corresponds to the high-density limit of the random scatterers model discussed above.[6,7] In some nontrivial cases we obtain explicit bounds on the average partition function, which, in the limiting case of "white noise," are consistent with recent approximate calculations of the asymptotic value of the average density of states.

2. EVALUATION OF $\langle Z_1\rangle$

In terms of Feynman's path-integral formalism the partition function for a particle of mass m interacting through a potential $\lambda V(\mathbf{r})$, with N scattering centers at positions $\mathbf{R}_1, \cdots, \mathbf{R}_N$ in a volume Ω, is given by

$$\begin{aligned}Z_1(\beta;\{R\}) = \int d\mathbf{r}_0 \oint_{\mathbf{r}_0}^{\mathbf{r}_0} \delta\mathbf{r} \exp\Bigg[&-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt \\ &- \int_0^\beta dt \sum_{n=1}^{N} \lambda V(\mathbf{r}(t) - \mathbf{R}_n)\Bigg],\end{aligned} \tag{2.1}$$

[5] S. Doniach, "Greens Function Theory of Multiple Scattering. I. Convergence of the Perturbation Series for the Partition Function"; "II. Variational Estimate for the Conductivity under Strong Coupling Conditions" (unpublished).

[6] B. I. Halperin and M. Lax, Phys. Rev. **148**, 722 (1966); **153**, 802, (1967).

[7] J. Zittartz and J. S. Langer, Phys. Rev. **148**, 741 (1966).

where β is the inverse temperature, the symbol $\oint_{\mathbf{r}_0}^{\mathbf{r}_0} \delta\mathbf{r}$ denotes the appropriately normalized sum over all paths $\mathbf{r}(t)$ in Ω such that $\mathbf{r}(0) = \mathbf{r}(\beta) = \mathbf{r}_0$, and we have taken $\hbar = 1$. We define the configurational average of Z_1 by

$$\langle Z_1(\beta)\rangle = \int_\Omega \cdots \int_\Omega \prod_{n=1}^{N} d\mathbf{R}_n P(\mathbf{R}_1, \cdots, \mathbf{R}_N) Z_1(\beta; \{R\}). \tag{2.2}$$

The case of randomly distributed scatterers corresponds to $P = \Omega^{-N}$ so that[8]

$$\langle Z_1\rangle = \int_\Omega d\mathbf{r}_0 \oint_{\mathbf{r}_0}^{\mathbf{r}_0} \delta\mathbf{r} \exp\left(-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt\right) \times \prod_{n=1}^{N}\left\{\Omega^{-1}\int d\mathbf{R}_n \exp\left[-\lambda\int_0^\beta dt V(\mathbf{r}(t) - \mathbf{R}_n)\right]\right\}.$$

Each term in the product is identical, and, by adding and subtracting unity, we can write the product as an exponential in the limit N, $\Omega \to \infty$, $\rho = N/\Omega =$ const, so that

$$\langle Z_1(\beta)\rangle = \int d\mathbf{r}_0 \oint_{\mathbf{r}_0}^{\mathbf{r}_0} \delta\mathbf{r} \exp\left\{-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt + \rho\int d\mathbf{R}\left[\exp\left(-\lambda\int_0^\beta dt V(\mathbf{r}(t) - \mathbf{R})\right) - 1\right]\right\}.$$

If we introduce the following notation for the normalized average of a functional of the path

$$E\{F[\mathbf{r}(t)]\} = \frac{1}{\Omega}\int d\mathbf{r}_0 \left\{\frac{\oint_{\mathbf{r}_0}^{\mathbf{r}_0} \delta\mathbf{r} \exp\left(-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt\right) F[\mathbf{r}(t)]}{\oint_{\mathbf{r}_0}^{\mathbf{r}} \exp\left(-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt\right)\delta\mathbf{r}}\right\},$$

then we can write the partition function in the form

$$\langle Z_1(\beta)\rangle/Z_0 = E\left\{\exp\rho\int d\mathbf{R}\left[\exp\left(-\lambda\int_0^\beta dt V(\mathbf{r}(t) - \mathbf{R})\right) - 1\right]\right\}. \tag{2.3}$$

Here Z_0 is the partition function for a free particle so that

$$Z_0 = \int d\mathbf{r}\oint_{\mathbf{r}_0}^{\mathbf{r}_0}\delta\mathbf{r}\exp\left(-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt\right) = \Omega\left(\frac{m}{2\pi\beta}\right)^{\frac{3}{2}}. \tag{2.4}$$

An expression of the type (2.3) can be obtained also for a nonuniform distribution of scatterers of the type

$$P(\mathbf{R}_1, \cdots, \mathbf{R}_N) = \Omega^{-N}\prod_{n=1}^{N} f(\mathbf{R}_n),$$

where f is normalized by

$$\Omega^{-1}\int d\mathbf{R} f(\mathbf{R}) = 1.$$

However, this will only introduce an additional factor $f(\mathbf{R})$ into the $\mathbf{R}$ integration in (2.3) and will not significantly alter the following discussion. A more realistic distribution which takes into account correlations in the positions of scattering centers will not lead to a simple expression of the form (2.3).

A. Inequalities

Making use of the standard inequality

$$\bar{f}(x) \geq f(\bar{x}), \tag{2.5}$$

where the bar represents averaging with respect to a weight function normalized to unity and the function $f(x)$ is convex downward, i.e.,

$$d^2f/dx^2 \geq 0,$$

gives

$$\exp\left[-\lambda\int_0^\beta dt V(\mathbf{r}(t) - \mathbf{R})\right] \leq \beta^{-1}\int_0^\beta dt \exp\left[-\lambda\beta V(\mathbf{r}(t) - \mathbf{R})\right],$$

so that one obtains the following upper bound for $\langle Z_1\rangle$:

$$\langle Z_1(\beta)\rangle \leq \int d\mathbf{r}_0 \oint_{\mathbf{r}_0}^{\mathbf{r}_0}\delta\mathbf{r}\exp\left[-\int_0^\beta \tfrac{1}{2}m\dot{\mathbf{r}}^2\, dt + \rho\int d\mathbf{R}(e^{-\lambda\beta V(\mathbf{R})} - 1)\right].$$

Therefore,

$$\langle Z_1(\beta)\rangle/Z_0 \leq \exp\left[\rho\int d\mathbf{R}(e^{-\lambda\beta V(\mathbf{R})} - 1)\right] \equiv \langle Z_1^{cl}(\beta)\rangle/Z_0, \tag{2.6}$$

where $\langle Z_1^{cl}(\beta)\rangle$ is the configurational average of the classical partition function.

B. Analyticity in ρ

Using some results obtained by Ginibre[9] in his work on the analyticity properties of regular systems, it is readily shown that $\langle Z_1(\beta; \rho)\rangle/Z_0$ is an entire function of ρ (for real positive β), for all reasonable (noncoulomb) potentials $V(r)$. To be more specific, we assume that $V(r)$ can be written as a sum of two terms

$$V(r) = V_1(r) + V_2(r) \tag{2.7}$$

satisfying the conditions

$$V_1(r) \geq 0 \quad \text{for} \quad r \leq a, \quad V_1(r) = 0 \quad \text{for} \quad r > a, \tag{2.8}$$

[8] S. F. Edwards and Y. B. Gulyaev, Proc. Phys. Soc. (London) **83**, 495 (1964).

[9] J. Ginibre, J. Math. Phys. **6**, 1432 (1965); in particular, see Appendix 2.

and

$$\int d\mathbf{r}[e^{\beta|\lambda V_2(r)|} - 1] < \infty. \tag{2.9}$$

For the case where the particle-impurity has a hard core, a may be taken to be the range of this core. Writing (2.3) in the form

$$\langle Z_1(\beta; \rho)\rangle/Z_0 = E\{\exp[\rho Q([\mathbf{r}(t)], \beta)]\}, \tag{2.10}$$

we have

$$Q([\mathbf{r}(t)], \beta) = Q_1([\mathbf{r}(t)], \beta) + Q_2([\mathbf{r}(t)], \beta), \tag{2.11}$$

where

$$Q_1([\mathbf{r}(t)], \beta) = \int d\mathbf{R}\left[\exp\left(-\lambda\int_0^\beta dt V_1(\mathbf{r}(t) - \mathbf{R})\right) - 1\right] \tag{2.12}$$

and

$$Q_2([\mathbf{r}(t)], \beta) = \int d\mathbf{R}\left[\exp\left(-\lambda\int_0^\beta dt V_1(\mathbf{r}(t) - \mathbf{R})\right)\right] \times \left[\exp\left(-\lambda\int_0^\beta dt V_2(\mathbf{r}(t) - \mathbf{R})\right) - 1\right]. \tag{2.13}$$

Using (2.5) and (2.8),

$$\begin{aligned}
|Q_2([\mathbf{r}(t)], \beta)| &\leq \int d\mathbf{R}\left|\exp\left[-\lambda\int_0^\beta dt V_2(\mathbf{r}(t) - \mathbf{R})\right] - 1\right| \\
&\leq \int d\mathbf{R}\left[\exp\left(\left|\lambda\int_0^\beta dt V_2(\mathbf{r}(t) - \mathbf{R})\right|\right) - 1\right] \\
&\leq \int d\mathbf{R}\left[\exp\left(\int_0^\beta dt\,|\lambda V_2(\mathbf{r}(t) - \mathbf{R})|\right) - 1\right] \\
&\leq \int d\mathbf{R}\beta^{-1}\int_0^\beta dt\left[\exp(\beta|\lambda V_2(\mathbf{r}(t) - \mathbf{R})|) - 1\right].
\end{aligned}$$

Thus finally

$$|Q_2([\mathbf{r}(t)], \beta)| \leq \int d\mathbf{R}[e^{\beta|\lambda V_2(\mathbf{R})|} - 1], \tag{2.14}$$

which is bounded by (2.9). Also, using (2.8), we have

$$|Q_1([\mathbf{r}(t)], \beta)| \leq v_a([\mathbf{r}(t)], \beta), \tag{2.15}$$

the volume of the region containing all points within a distance a of the path $\mathbf{r}(t)$. The equality in (2.15) holds when V_1 is a hard-core potential. Ginibre was able to obtain very strong bounds on

$$E\{\exp[\rho v_a(\mathbf{r}(t), \beta)]\}$$

from which the analyticity of $\langle Z_1\rangle/Z_0$ now follows.

Having established the analyticity in ρ of $\langle Z_1\rangle/Z_0$, it is now possible to show that the average "pressure" of an ideal gas with quantum statistics is also analytic in ρ for $V(R) \geq 0$ and $e^{\beta\mu} < 1$. This is readily seen by expanding the integrand in (1.17) in powers of $e^{\beta\mu}$ for $\mu < 0$, since $n(E) = 0$ for $E < 0$, when $V(r) \geq 0$. The coefficient of $e^{l\beta\mu}$ is $(\pm 1)^{l+1}(l\Omega)^{-1}\langle Z_1(l\beta)\rangle$ and is bounded by

$$l^{-1}\left(\frac{m}{2\pi\beta l}\right)^{\nu/2}\exp\left[\rho\int d\mathbf{R}(e^{-\lambda l\beta V(\mathbf{R})} - 1)\right]$$

according to (2.6), where ν is the dimensionality of the space considered. The series will therefore converge to an analytic function of ρ for $e^{\beta\mu} < 1$. It is an interesting but unanswered question whether the Bose–Einstein condensation of an ideal gas, $\rho = 0$, will disappear at some finite ρ. This could happen, for example, if $n(E; \rho)$ was zero for $E < E_0$ and

$$n(E_0 +; \rho) \neq 0.$$

C. Analyticity in λ

By expansion of each term in the ρ series, when (2.11) is satisfied, in powers of the coupling constant λ and rearranging, one can obtain an expansion of $\langle Z_1(\beta)\rangle$ in powers of λ. This series was considered by Doniach,[5] who showed that $\langle Z_1(\beta)\rangle$ is an analytic function of λ for certain potentials. If we expand the potential in its Fourier components, we find

$$\begin{aligned}
Q([\mathbf{r}(t)], \beta) &= \int d\mathbf{R}\left[\exp\left(-\lambda\int_0^\beta dt V(\mathbf{r}(t) - \mathbf{R})\right) - 1\right] \\
&= \int d\mathbf{R}\sum_{m=1}^\infty \frac{(-\lambda)^m}{m!}\left[\int_0^\beta dt V(\mathbf{r}(t) - \mathbf{R})\right]^m \\
&= \sum_{m=1}^\infty \frac{(-\lambda)^m}{m!}\sum_{\mathbf{k}_1,\cdots,\mathbf{k}_m} V_{\mathbf{k}_1}\cdots V_{\mathbf{k}_m}\delta\left(\sum_{i=1}^m \mathbf{k}_i\right) \\
&\quad \times \int_0^\beta \cdots \int_0^\beta dt_1 \cdots dt_m \\
&\quad \times \exp[i\mathbf{k}_1 \cdot \mathbf{r}(t_1)] \cdots \exp[i\mathbf{k}_m \cdot \mathbf{r}(t_m)].
\end{aligned} \tag{2.16}$$

Each term in the λ expansion will involve many-temperature Green's functions of the form

$$\begin{aligned}
G_{\mathbf{k}_1\cdots\mathbf{k}_m}(t_1, \cdots, t_m) &= E\{e^{i\mathbf{k}_1\cdot\mathbf{r}(t_1)}\cdots e^{i\mathbf{k}_m\cdot\mathbf{r}(t_m)}\} \\
&= G^*_{\mathbf{k}_1\cdots\mathbf{k}_m}(t_1, \cdots, t_m).
\end{aligned}$$

Since

$$\exp\left[i\sum_{i=1}^m \mathbf{k}_i \cdot \mathbf{r}(t_i)\right] = \cos\left(\sum_i \mathbf{k}_i \cdot \mathbf{r}(t_i)\right) + i\sin\left(\sum_i \mathbf{k}_i \cdot \mathbf{r}(t_i)\right)$$

and for every path for which

$$\mathbf{r}(t_1) = \mathbf{r}_0 + \mathbf{x}_1, \quad \mathbf{r}(t_2) = \mathbf{r}_0 + \mathbf{x}_2, \cdots$$

there is a corresponding path with

$$\mathbf{r}(t_1) = \mathbf{r}_0 - \mathbf{x}_1, \quad \mathbf{r}(t_2) = \mathbf{r}_0 - \mathbf{x}_2, \cdots,$$

then the imaginary parts cancel out because $\mathbf{r}_0 \cdot \sum_i \mathbf{k}_i = 0$. Furthermore,

$$|G| \leq E\{|\exp[i\mathbf{k}_1 \cdot \mathbf{r}(t_1)] \cdots \exp[i\mathbf{k}_m \cdot \mathbf{r}(t_m)]|\} = 1. \tag{2.17}$$

Each term in the λ expansion will be majorized by the absolute value of a product of terms of the type on the right-hand side of Eq. (2.16) in which the exponentials have been replaced by unity. Then the λ expansion will exist for all values of λ if the following conditions on the potential are satisfied:

$$\text{(i)} \qquad \sum_{\mathbf{k}} |V_{\mathbf{k}}| \leq \bar{V},$$

$$\text{(ii)} \qquad |V_{\mathbf{k}}| \leq \left(\frac{\alpha^3}{\Omega}\right)\bar{V}, \tag{2.18}$$

where α is a length representing the range of the potential. These conditions were derived by Doniach,[5] who introduced a harmonic-oscillator potential which was finally set to zero and used properties of oscillator averages in order to prove the required properties of G. It is easy to show that, when the conditions (2.18) are satisfied, the ρ expansion also exists.

3. GAUSSIAN RANDOM POTENTIAL

For a particle in a potential $\varphi(\mathbf{r})$,

$$Z_1(\beta, [\varphi])/Z_0 = E\left\{\exp\left[-\int_0^\beta dt\varphi(\mathbf{r}(t))\right]\right\}. \tag{3.1}$$

By "Gaussian averaging" [7] we mean

$$\langle Z_1(\beta)\rangle = \frac{\int \delta\varphi Z_1(\beta, [\varphi]) \exp\left[-\frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}'\varphi(\mathbf{r})\varphi(\mathbf{r}')K(\mathbf{r}-\mathbf{r}')\right]}{\int \delta\varphi \exp\left[-\frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}'\varphi(\mathbf{r})\varphi(\mathbf{r}')K(\mathbf{r}-\mathbf{r}')\right]}.$$

Thus,

$$\langle Z_1(\beta)\rangle/Z_0 = E\left\{\exp\left[\int_0^\beta dt_1 \int_0^\beta dt_2 K^{-1}(\mathbf{r}(t_1) - \mathbf{r}(t_2))\right]\right\} \tag{3.2}$$

where

$$\int K(\mathbf{x}-\mathbf{y})K^{-1}(\mathbf{y}-\mathbf{z})\,d\mathbf{y} = \delta(\mathbf{x}-\mathbf{z}). \tag{3.3}$$

Comparison with (2.3) shows that (3.2) "corresponds" to the random scatterers case in the limit $\rho \to \infty$, $\lambda \to 0$ such that $\rho\lambda^2 = \text{const}$ with the identification

$$K^{-1}(\mathbf{r}) = \lambda^2\rho\int d\mathbf{R}V(\mathbf{R})V(\mathbf{R}+\mathbf{r}),$$

when

$$\int d\mathbf{R}V(\mathbf{R}) = 0. \tag{3.4}$$

It has also been pointed out by Halperin and Lax[6] that the Gaussian random potential is the high-density limit of the random scatterers model as a consequence of the central limit theorem.

The path integral (3.2) cannot be explicitly evaluated for any physically interesting potential correlation K^{-1}. If K^{-1} satisfies the inequalities

$$M_1 \leq K^{-1}(\mathbf{r}) \leq M_2,$$

then $\langle Z_1(\beta)\rangle/Z_0$ has corresponding bounds

$$e^{\beta^2 M_1} \leq \langle Z_1(\beta)\rangle/Z_0 \leq e^{\beta^2 M_2} \tag{3.5}$$

and one easily sees that an expansion in λ^2 exists. One can find a more useful lower bound using the inequality which follows from (2.5):

$$\langle Z_1(\beta)\rangle/Z_0 \geq \exp\left[E\left\{\int_0^\beta dt_1 \int_0^\beta dt_2 K^{-1}(\mathbf{r}(t_1) - \mathbf{r}(t_2))\right\}\right]. \tag{3.6}$$

When K^{-1} has a Gaussian form

$$K^{-1}(\mathbf{r}) = \gamma\left(\frac{\alpha^2}{\pi}\right)^{\nu/2} e^{-\alpha^2 r^2}, \tag{3.7}$$

where ν is the dimensionality and $\int d\mathbf{r}K^{-1}(\mathbf{r}) = \gamma$, the right-hand side of (3.6) can be evaluated to give in the one-dimensional case

$$\langle Z_1(\beta)\rangle/Z_0 \geq \exp\left[\gamma\beta\left(\frac{2m}{\pi}\beta\right)^{\frac{1}{2}} \sin^{-1}\left\{\frac{\alpha}{(\alpha^2 + 2m/\beta)^{\frac{1}{2}}}\right\}\right]. \tag{3.8}$$

In two dimensions we have

$$\langle Z_1(\beta)\rangle/Z_0 \geq \exp\left[\frac{-\gamma m\beta\alpha}{\pi(\alpha^2 + 2m/\beta)^{\frac{1}{2}}} \times \ln\left\{\frac{\alpha - (\alpha^2 + 2m/\beta)^{\frac{1}{2}}}{\alpha + (\alpha^2 + 2m/\beta)^{\frac{1}{2}}}\right\}\right]. \tag{3.9}$$

In three dimensions we have

$$\langle Z_1(\beta)\rangle/Z_0 \geq \exp\left[\frac{2\gamma\lambda^2\beta m\alpha^3}{\pi^{\frac{3}{2}}(\alpha^2 + 2m/\beta)}\right]. \tag{3.10}$$

From (3.5) the upper bound in the case of Gaussian correlation is given by

$$\langle Z_1(\beta)\rangle/Z_0 \leq \exp\left[\beta^2\gamma\left(\frac{\alpha^2}{\pi}\right)^{\nu/2}\right].$$

When $\alpha \to 0$, which can be considered as the limiting

case of a long-range potential correlation, the upper and lower bounds become identical and thus $\langle Z_1(\beta)\rangle$ is obtained exactly in this limit. In the limit $\alpha \to \infty$, K^{-1} becomes a δ function and we can make a connection with the work of Halperin and Lax[6] and Zittartz and Langer,[7] who calculated the low-energy behavior ($E \to -\infty$) of the average density of states $n(E)$ in the case of "white noise":

$$K^{-1}(\mathbf{r}) = \gamma\delta(\mathbf{r}). \tag{3.11}$$

In this limit we obtain in the one-dimensional case (3.8)

$$\langle Z_1(\beta)\rangle/Z_0 \geq \exp\,[\gamma\beta(\beta m\pi/2)^{\frac{1}{2}}], \tag{3.12}$$

but the two- and three-dimensional bounds (3.9) and (3.10) diverge as $\alpha \to \infty$ (the exponent diverging logarithmically and linearly, respectively). This is in agreement with the density of states results[6,7] which, as $E \to -\infty$, predict a behavior

$$n(E) \sim \exp\left[\frac{-\text{const}\,|E|^{2-\nu/2}}{\gamma}\right], \tag{3.13}$$

which, from (1.16), will lead to a finite value for $\langle Z_1(\beta)\rangle$ in the one-dimensional case and to an infinite value in three dimensions. The divergence in the two-dimensional problem may be associated with the fact that, as discussed by Halperin and Lax,[6] in the case of two- and three-dimensional "white noise," the second-order corrections to the variational energy diverge due to short wavelength potential fluctuations. This leads to an infinite constant in $\langle Z_1(\beta)\rangle$ unless those fluctuations are cut off below a certain wavelength, which is the case if K^{-1} is not a true δ function.

A further lower bound on (3.2) may be obtained by application of the methods used in Feynman's treatment of the polaron problem.[4] This leads to a complicated expression which, however, has the same behavior in the "white noise" case as the bounds (3.8)–(3.10). The expression (3.2) for $\langle Z_1(\beta)\rangle$ also satisfies the condition for applicability of the extension of Feynman's variational principle for the free energy to include dissipative processes, which was introduced by Doniach in his second paper.[5] This condition is that $\langle Z_1(\beta)\rangle$ is a convex-downward function of any parameter, say λ, multiplying K in (3.2)

$$\partial^2\langle Z_1(\beta)\rangle/\partial\lambda^2 \geq 0,$$

which, from (3.2), is true for any function K^{-1}.

ACKNOWLEDGMENTS

We wish to thank Dr. S. Doniach, whose work interested us in this subject, and Dr. R. B. Griffiths, Dr. A. Lenard, and Dr. E. Lieb for helpful comments. We are particularly indebted to Dr. Ginibre for pointing out to us the results he obtained in Ref. 9.

APPENDIX A

The free energy of a harmonic crystal can be written in the form[10]

$$F = \int_0^\infty d\omega F(\omega)g(\omega), \tag{A1}$$

where $F(\omega)$ is a smooth function of ω and $g(\omega)$ is the distribution function of normal mode frequencies of the crystal (eigenvalues of the dynamical matrix). If, for simplicity, we consider a crystal with one atom per unit cell, then we can write $g(\omega)$ in the form

$$g(\omega) = 2\omega G(\omega^2), \tag{A2}$$

where

$$G(\omega^2) = N^{-1}\frac{d}{d\omega^2}N(\omega^2) \tag{A3}$$

and

$$N(\omega^2) = \int_0^{\omega^2} dx \sum_{\mathbf{k}} \delta(x - \omega^2(\mathbf{k})) \tag{A4}$$

is the number of modes in the crystal with

$$(\text{frequencies})^2 \leq \omega^2.$$

In this Appendix we show the existence of

$$n_\Omega(\omega^2) = N_\Omega(\omega^2)/\Omega \tag{A5}$$

in the limit where the number of atoms N in the crystal and its volume Ω become infinite in such a way that N/Ω is constant. (A1)–(A4) then lead to the existence of the free energy density (F/Ω) in this limit. The proof also applies to a crystal with random masses or a random distribution of atoms on lattice sites.

For m identical crystals of cubic geometry each with N atoms in volume Ω, $N_\Omega(\omega^2)$ and the eigenfrequencies are the same for each crystal. For the system taken as a whole, the number of

$$(\text{frequencies})^2 \leq \omega^2,$$

$$N_{m\Omega}(\omega^2) = mN_\Omega(\omega^2) = \int_0^{\omega^2} dx \sum_{\mathbf{k}} \delta(x - \omega_m^2(\mathbf{k})),$$

when $\omega_m(\mathbf{k})$ are the eigenvalues of a dynamical matrix which can be decomposed into matrices referring to the individual crystals alone, each having $O(N)$ elements in the case of finite range interatomic forces.

If the m crystals are joined together in some way to form a new crystal I with mN atoms and volume $m\Omega$, then

$$N^I_{m\Omega}(\omega^2) = \int_0^{\omega^2} \sum_{\mathbf{k}} \delta(x - \omega_m^{I^2}(\mathbf{k})),$$

[10] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press Inc., New York, 1963).

in which $\omega_m^I(\mathbf{k})$ are eigenvalues of a matrix which differs from the above only by interaction terms between component crystals. These interaction terms correspond to $O(N^{\frac{2}{3}})$ elements since only surface atoms will contribute such terms to the dynamical matrix. By Ledermann's Theorem[10] the number of eigenvalues of this matrix $\omega_m^{I^2} \leq \omega^2$ will differ from the number for the previous matrix $\omega_m^2 \leq \omega^2$ by, at most, the number of rows and columns which are changed, i.e., by less than $\alpha_m N^{\frac{2}{3}}$ where α_m is some number which depends on m and on how the crystals are joined but is independent of N.

Hence

$$n_{m\Omega}^I \equiv N_{m\Omega}^I(\omega^2)/m\Omega = N_{m\Omega}(\omega^2)/m\Omega + \alpha_m N^{\frac{2}{3}}/m\Omega$$
$$= n_\Omega(\omega^2) + R_{m\Omega}(\omega^2),$$

where

$$R_{m\Omega}(\omega^2) \sim N^{\frac{2}{3}}/\Omega.$$

Following the method of Griffiths,[11] consider a simple cubic lattice of lattice constant 1. Let the cube Ω_s for $s = 2, 3, 4, \cdots$, be of volume $\Omega_s = 2^{3s}\Omega$ and contain $2^{3s}N$ particles so that it is composed of eight cubes Ω_{s-1}. The corresponding values of $n(\omega^2)$ satisfy an inequality of the type

$$n^{(s)}(\omega^2) \leq n^{(s-1)}(\omega^2) + |R^{(s)}(\omega^2)|,$$

where

$$|R^{(s)}(\omega^2)| \leq \frac{|\bar{\alpha}_8|\,(2^{3s}N)^{\frac{2}{3}}}{2^{3s}\Omega} = C2^{-s}\Omega^{-\frac{1}{3}}$$

in which $\bar{\alpha}_8$ is the maximum possible value of α_8. Therefore

$$n^{(s)}(\omega^2) \leq n^{(s-1)}(\omega^2) + C\Omega^{-\frac{1}{3}}2^{-s}.$$

Thus the quantity

$$n^{(s)}(\omega^2) + C\Omega^{-\frac{1}{3}}2^{-s}$$

is monotonically decreasing as $s \to \infty$. Since this is bounded below by zero, it will approach a limit $n(\omega^2)$ in the limit $s \to \infty$. By a method similar to that of Griffiths, one can show that an arbitrary sequence of cubes of increasing volume will yield the same limit $n(\omega^2)$ as the particular sequence Ω_s.

For a random crystal the above argument follows through where each crystal has some particular configuration θ and one takes the average

$$\langle n(\omega^2)\rangle = \sum_\theta P(\theta) n(\omega^2[\theta]),$$

in which $P(\theta)$ is the normalized probability of configuration θ. It is easily seen that

$$\langle n^{(s)}(\omega^2)\rangle \leq \langle n^{(s-1)}(\omega^2)\rangle + C\Omega^{-\frac{1}{3}}2^{-s}$$

and the argument follows as above.[2]

[11] R. B. Griffiths, J. Math. Phys. **6**, 1447 (1965).

APPENDIX B

In this Appendix we shall prove the existence of the thermodynamic limit of the free energy per unit volume defined in (1.6) for a system with a Hamiltonian given in (1.1).

The problem here is more complicated than in the case without scatterers, where the existence of the free-energy density in the thermodynamic limit was proved for very general interactions by Fisher[12] and Ruelle.[13] While we need not assume any restrictions on the interaction between the particles beyond those necessary for the existence of the thermodynamic limit in the absence of any scattering centers, we shall assume that the potential $V(\mathbf{r})$ between particle and scattering centers satisfies the "strong tempering" condition of Fisher[12]:

$$V(\mathbf{r}) \leq 0 \quad \text{when} \quad r \geq D, \tag{B1}$$

as well as the more usual condition that

$$\int [e^{-\beta V(\mathbf{r})} - 1]\, d\mathbf{r} < \infty. \tag{B2}$$

Furthermore, we shall assume that the particles of our system, of which there are M, are confined to a cubical box with sides of length L, $L^3 = \Omega$. The scattering centers, on the other hand, are located inside a larger cube of sides $L + 2D$, $(L + 2D)^3 = \Omega'$, centered on the original cube Ω. The probability density for having precisely l scatterers at positions $\mathbf{R}_1, \cdots, \mathbf{R}_l$ in some volume element $\omega \in \Omega'$ is assumed to be given by

$$P(y) = P_\omega(\mathbf{R}_1, \cdots, \mathbf{R}_l) = \frac{e^{-\rho\omega}\rho^l}{l!}, \tag{B3}$$

so that ρ is the average density of scatterers.

To prove the existence of

$$\lim_{\Omega\to\infty} -\frac{\beta\langle F\rangle}{\Omega} = \lim_{\Omega'\to\infty} \frac{\langle \ln Z(y, M, \Omega')\rangle}{\Omega'} \tag{B4}$$

for M/Ω fixed under the assumptions (B1) and (B2) and the special kind of boundary conditions we have used requires only a small modification of the methods of Fisher and Ruelle. We shall therefore give only a brief outline of the proof. There are two parts to the proof: (1) finding an upper bound on $\langle \ln Z(\Omega')\rangle/\Omega'$

[12] M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964).

[13] D. Ruelle, *Boulder Lectures 1963* (University of Colorado, Theoretical Physics Institute, 1963).

which is independent of Ω'; and (2) showing that, when eight cubes of volume Ω' are put together to form a larger cube of volume $\Omega'' = 2^3\Omega'$ with 2^3M particles in it, then

$$\langle \ln Z(y'', 2^3M, 2^3\Omega')\rangle/2^3\Omega' \geq \langle \ln Z(y, M, \Omega')\rangle/\Omega' + 0(\Omega'), \quad \text{(B5)}$$

where y'' is the configuration of scatterers in Ω''.

An upper bound on $Z(y, M, \Omega')$ is obtained from the condition that

$$U(\mathbf{r}_1, \cdots, \mathbf{r}_M) \geq -M\Phi, \quad \Phi \text{ const}, \quad \text{(B6)}$$

for all values of the $\mathbf{r}_i$. This condition is required for the existence of the thermodynamic limit of the free energy in the absence of scatterers. Hence

$$\frac{1}{\Omega'}\langle \ln Z(y, M, \Omega')\rangle \leq \frac{\beta M\Phi}{\Omega'} + \frac{1}{\Omega'}\left\langle \ln\left\{\frac{1}{M!}\left[\int_\Omega e^{-\beta\Sigma V(\mathbf{r}-\mathbf{R}_i)}\, d\mathbf{r}\right]^M\right\}\right\rangle \quad \text{(B7)}$$

$$\leq \frac{\beta M}{\Omega'}\Phi + 1 - \ln\left(\frac{M}{\Omega'}\right) + \frac{M}{\Omega'}\rho\int |e^{-\beta V(\mathbf{R})} - 1|\, d\mathbf{R} \leq C, \quad \text{(B8)}$$

where C is some constant independent of Ω and use was made of (2.5) in deriving the second inequality. Having obtained the bound in (B8), we can now proceed with our construction of the larger cubes *à la* Fisher and Ruelle.

The inequality (B5) is obtained by first noting that $\ln Z(y'', 2^3M, 2^3\Omega') \geq \ln Z'(y'', 2^3M, 2^3\Omega')$, where Z' is the partition function when the 2^3M particles are constrained to be inside, and evenly divided between, the original eight cubes Ω_i, $i = 1, \cdots, 8$, of volume Ω. A lower bound on Z' is now obtained using (B1) if we neglect the interactions between particles in Ω_i and scattering centers outside Ω_i'. The interaction between particles in different boxes Ω_i is now bounded by a term of $o(\Omega)$ which is independent of the configuration of scattering centers. Finally, we obtain

$$\ln Z(y'', 2^3M, 2^3\Omega') \geq \sum_{i=1}^{8} \ln Z(y_i, M, \Omega_i') + o(\Omega), \quad \text{(B9)}$$

where y_i is the configuration of scatterers in Ω_i. Due to the independence of the distribution of scatterers in the different boxes, averaging of (B9) now yields (B5).

JOURNAL OF MATHEMATICAL PHYSICS · VOLUME 9, NUMBER 10 OCTOBER 1968

Quantum Corrections to the Second Virial Coefficient at High Temperatures

ROBERT NYDEN HILL
Department of Physics, University of Delaware, Newark, Delaware

(Received 13 September 1967)

The Laplace transform of $\exp(-\beta H)$ is the Green's operator of the negative-energy Schrödinger equation $(H + W)^{-1}$. Conditions are stated under which a large $|W|$ asymptotic series for the Green's operator can be inverse-Laplace-transformed term-by-term to obtain a small β expansion for $\exp(-\beta H)$. This approach and the Watson transformation are used to calculate the first few terms of high-temperature asymptotic expansions for the exchange second virial coefficient for hard spheres and for the Lennard–Jones potential. The known results for the direct second virial coefficient for hard spheres are extended. The Wigner–Kirkwood expansion is calculated to order $\hbar^6$ and used to calculate the direct second virial coefficient for the Lennard–Jones potential through order $\hbar^6$.

I. INTRODUCTION AND SUMMARY

The problem of calculating quantum corrections to the second virial coefficient at high temperatures has been only partially solved by previous authors.[1] The exchange contribution has been particularly difficult, and it is only in the last year that Lieb,[2] by calculating rigorous upper and lower bounds, has obtained the leading term of an asymptotic expansion for the exchange contribution in the particular case of

[1] References to work prior to 1952 may be found in J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *The Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), pp. 407–424. The problem has been considered more recently by H. E. DeWitt [J. Math. Phys. **3**, 1003 (1962)] and by F. Mohling [Phys. Fluids **6**, 1097 (1963)], who give references to more recent numerical work.

[2] E. Lieb, J. Math. Phys. **8**, 43 (1967). An upper bound to the exchange second virial coefficient had been found previously by S. Larsen, J. Kilpatrick, E. Lieb, and H. Jordan, Phys. Rev. **140**, A129 (1965).

which is independent of Ω'; and (2) showing that, when eight cubes of volume Ω' are put together to form a larger cube of volume $\Omega'' = 2^3\Omega'$ with 2^3M particles in it, then

$$\langle \ln Z(y'', 2^3M, 2^3\Omega')\rangle/2^3\Omega' \geq \langle \ln Z(y, M, \Omega')\rangle/\Omega' + 0(\Omega'), \quad \text{(B5)}$$

where y'' is the configuration of scatterers in Ω''.

An upper bound on $Z(y, M, \Omega')$ is obtained from the condition that

$$U(\mathbf{r}_1, \cdots, \mathbf{r}_M) \geq -M\Phi, \quad \Phi \text{ const}, \quad \text{(B6)}$$

for all values of the $\mathbf{r}_i$. This condition is required for the existence of the thermodynamic limit of the free energy in the absence of scatterers. Hence

$$\frac{1}{\Omega'}\langle \ln Z(y, M, \Omega')\rangle \leq \frac{\beta M\Phi}{\Omega'} + \frac{1}{\Omega'}\left\langle \ln\left\{\frac{1}{M!}\left[\int_\Omega e^{-\beta\Sigma V(\mathbf{r}-\mathbf{R}_i)}\, d\mathbf{r}\right]^M\right\}\right\rangle \quad \text{(B7)}$$

$$\leq \frac{\beta M}{\Omega'}\Phi + 1 - \ln\left(\frac{M}{\Omega'}\right) + \frac{M}{\Omega'}\rho\int |e^{-\beta V(\mathbf{R})} - 1|\, d\mathbf{R} \leq C, \quad \text{(B8)}$$

where C is some constant independent of Ω and use was made of (2.5) in deriving the second inequality. Having obtained the bound in (B8), we can now proceed with our construction of the larger cubes *à la* Fisher and Ruelle.

The inequality (B5) is obtained by first noting that $\ln Z(y'', 2^3M, 2^3\Omega') \geq \ln Z'(y'', 2^3M, 2^3\Omega')$, where Z' is the partition function when the 2^3M particles are constrained to be inside, and evenly divided between, the original eight cubes Ω_i, $i = 1, \cdots, 8$, of volume Ω. A lower bound on Z' is now obtained using (B1) if we neglect the interactions between particles in Ω_i and scattering centers outside Ω_i'. The interaction between particles in different boxes Ω_i is now bounded by a term of $o(\Omega)$ which is independent of the configuration of scattering centers. Finally, we obtain

$$\ln Z(y'', 2^3M, 2^3\Omega') \geq \sum_{i=1}^{8} \ln Z(y_i, M, \Omega_i') + o(\Omega), \quad \text{(B9)}$$

where y_i is the configuration of scatterers in Ω_i. Due to the independence of the distribution of scatterers in the different boxes, averaging of (B9) now yields (B5).

JOURNAL OF MATHEMATICAL PHYSICS · VOLUME 9, NUMBER 10 OCTOBER 1968

Quantum Corrections to the Second Virial Coefficient at High Temperatures

ROBERT NYDEN HILL
Department of Physics, University of Delaware, Newark, Delaware

(Received 13 September 1967)

The Laplace transform of $\exp(-\beta H)$ is the Green's operator of the negative-energy Schrödinger equation $(H + W)^{-1}$. Conditions are stated under which a large $|W|$ asymptotic series for the Green's operator can be inverse-Laplace-transformed term-by-term to obtain a small β expansion for $\exp(-\beta H)$. This approach and the Watson transformation are used to calculate the first few terms of high-temperature asymptotic expansions for the exchange second virial coefficient for hard spheres and for the Lennard–Jones potential. The known results for the direct second virial coefficient for hard spheres are extended. The Wigner–Kirkwood expansion is calculated to order $\hbar^6$ and used to calculate the direct second virial coefficient for the Lennard–Jones potential through order $\hbar^6$.

I. INTRODUCTION AND SUMMARY

The problem of calculating quantum corrections to the second virial coefficient at high temperatures has been only partially solved by previous authors.[1] The exchange contribution has been particularly difficult, and it is only in the last year that Lieb,[2] by calculating rigorous upper and lower bounds, has obtained the leading term of an asymptotic expansion for the exchange contribution in the particular case of

[1] References to work prior to 1952 may be found in J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *The Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), pp. 407–424. The problem has been considered more recently by H. E. DeWitt [J. Math. Phys. **3**, 1003 (1962)] and by F. Mohling [Phys. Fluids **6**, 1097 (1963)], who give references to more recent numerical work.

[2] E. Lieb, J. Math. Phys. **8**, 43 (1967). An upper bound to the exchange second virial coefficient had been found previously by S. Larsen, J. Kilpatrick, E. Lieb, and H. Jordan, Phys. Rev. **140**, A129 (1965).

hard-sphere interactions.[3] The present paper employs Laplace transform methods to give, for the first time, a constructive procedure for the computation of a high-temperature asymptotic expansion of the exchange second virial coefficient for any potential $V(r)$ which is more strongly repulsive than r^{-2} as $r \to 0$. An application of these methods to the direct second virial coefficient simplifies the derivation of known results to the point where extending them becomes feasible.

Section II discusses a general approach to the computation of high-temperature expansions in statistical mechanics which uses the Laplace transform to relate the Green's function of the Schrödinger equation at large negative energies to the thermal Green's function at high temperatures. Section III formulates the second virial coefficient problem (the relation of the present formulation to the usual phase-shift formulation is demonstrated in Appendix C).

Section IV is devoted to the direct second virial coefficient for hard spheres. The Green's function of the negative-energy Schrödinger equation is decomposed into partial waves (sum over l). Because the computation is carried out at large negative energies, where the radial equation has no turning points, the Debye series for the modified Bessel functions provides a large negative-energy asymptotic series for the solutions of the radial equation which is uniformly valid in l. After separating off certain singular parts and treating them exactly, the sum over l is performed with the Euler–MacLaurin sum formula. The first six terms of the resulting high-temperature asymptotic expansion of the direct second virial coefficient for hard spheres are given in Eq. (53).

In Sec. V, the Laplace transform and the Watson transformation are used to give a general method [the general results are in Eqs. (57)–(60)] for the computation of the exchange second virial coefficient. Results are computed explicitly for hard spheres [Eqs. (67)–(71) and Table I] and for the Lennard–Jones potential [Eqs. (72), (79), and (81)–(83)].

In Sec. VI the terms of the Wigner–Kirkwood expansion are calculated through order $\hbar^6$ [Eq. (92)] by methods considerably less laborious than those used by previous authors. The integrals which appear are evaluated to give a high-temperature expansion of the direct second virial coefficient for the Lennard–Jones potential [Eqs. (93) and (94)].

[3] Numerical calculations on the hard-sphere gas have been performed by M. Boyd, S. Larsen, and J. Kilpatrick, J. Chem. Phys. **45**, 499 (1966). The first few terms of the high-temperature asymptotic series for the *direct* second virial coefficient of a hard-sphere gas have been calculated by R. A. Handelsman and J. B. Keller, Phys. Rev. **148**, 94 (1966); their results were extended by P. C. Hemmer and K. J. Mork, Phys. Rev. **158**, 114 (1967).

II. BASIC IDEA

The basic concern of quantum-statistical mechanics is the operator $e^{-\beta H}$, where $\beta = 1/kT$ and H is the Hamiltonian. The Laplace transform of this operator,

$$(H + W)^{-1} = \int_0^\infty e^{-\beta W} e^{-\beta H}\, d\beta,$$

is the Green's operator of the negative-energy Schrödinger equation. Our method is based on the fact that, if $\bar{f}(W)$ is the Laplace transform of $f(\beta)$ so that

$$\bar{f}(W) = \int_0^\infty e^{-\beta W} f(\beta)\, d\beta, \tag{1}$$

then *the behavior of f for small β determines the behavior of $\bar{f}$ for large W and, under some conditions, conversely.* We shall be interested in the possibility of inverse-Laplace-transforming term-by-term an asymptotic expansion of $(H + W)^{-1}$ for $W \to \infty$ to obtain an asymptotic expansion of $e^{-\beta H}$ for $\beta \to 0^+$. When such term-by-term inverse-Laplace transformation can be justified, high-temperature asymptotic series for thermodynamic quantities can be calculated by studying the Schrödinger equation at large negative energies.

To lend some precision to the discussion we state a definition and consider three theorems.

Definition: That $f(z)$ has the asymptotic expansion

$$f(z) \sim \sum_{n=0}^{\infty} a_n \varphi_n(z) \quad \text{for} \quad z \to z_0$$

in the domain D means that

(a) $$\lim_{z \to z_0 \text{ in } D} \varphi_{n+1}(z)/\varphi_n(z) = 0, \quad n = 0, 1, 2 \cdots \text{ and}$$

(b) $$\lim_{z \to z_0 \text{ in } D} \left[f(z) - \sum_{n=0}^{m} a_n \varphi_n(z) \right] \Big/ \varphi_m(z) = 0, \quad m = 0, 1, 2 \cdots.$$

The functions $\varphi_n(z)$ are commonly called gauge functions. The statement that the small β behavior of $f(\beta)$ determines the large W behavior of $\bar{f}(W)$ is contained in the following theorem.

Theorem 1: Assume that

$$f(\beta) \sim \sum_{n=0}^{\infty} a_n \varphi_n(\beta) \quad \text{for} \quad \beta \to 0^+,$$

and that there exists a real constant W_0 such that, for $\operatorname{Re} W > W_0$, $e^{-\beta W} f(\beta)$ and $e^{-\beta W} \varphi_n(\beta)$ are bounded and integrable on $(0, \infty)$ so that the Laplace transforms $\bar{f}(W)$ and $\bar{\varphi}_n(W)$ exist. Assume further that $\varphi_n(\beta) > 0$ for $\beta > 0$ and that $e^{\delta W} \bar{\varphi}_n(W) \to \infty$ as

$W \to \infty$ for each $\delta > 0$. Then

$$\tilde{f}(W) \sim \sum_{n=0}^{\infty} a_n \tilde{\varphi}_n(W) \quad \text{for} \quad W \to +\infty.$$

A somewhat more general version of Theorem 1 has been given by Erdelyi.[4] For completeness, we include a proof in Appendix A. In the special case where the $\varphi_n(\beta)$ are powers of β, Theorem 1 is known as Watson's lemma.[5] That a converse to Theorem 1 does not hold without additional conditions is clear from considering the function $f_1(\beta) \equiv \beta^{-\frac{1}{2}} \sin(1/\beta)$, which has as its Laplace transform

$$\tilde{f}_1(W) = \pi^{\frac{1}{2}} W^{-\frac{1}{2}} \exp[-(2W)^{\frac{1}{2}}] \sin[(2W)^{\frac{1}{2}}].$$

If we take $\varphi_n(\beta) = \beta^n$, then $\tilde{\varphi}_n(W) = n! W^{-n-1}$ and

$$\lim_{W \to +\infty} \tilde{f}_1(W)/\varphi_n(W) = 0$$

for all n, but

$$\lim_{\beta \to 0} f_1(\beta)/\varphi_n(\beta)$$

does not exist for any $n \geq -\frac{1}{2}$. Clearly the converse of Theorem 1 is true (and is demonstrable by contradiction) if an asymptotic expansion of $f(\beta)$ with respect to the gauge functions $\varphi_n(\beta)$ exists. The necessary existence theorem can be quite difficult to prove when only $\tilde{f}(W)$ is known explicitly. However, if the series for $\tilde{f}(W)$ is a convergent one, the following theorem[6] may be used to justify term-by-term inverse-Laplace transformations.

Theorem 2: Let

$$\tilde{f}(W) = \sum_{n=0}^{\infty} a_n \tilde{\varphi}_n(W)$$

be a convergent series for Re $W \geq W_0$. Assume that (a) all integrals

$$\tilde{\varphi}_n(W) = \int_0^{\infty} e^{-\beta W} \varphi_n(\beta)\, d\beta$$

exist in the half-plane Re $W > W_0$, (b) that the integrals

$$\psi_n(W) = \int_0^{\infty} e^{-\beta W} |\varphi_n(\beta)|\, d\beta$$

exist in the half-plane Re $W > W_0$, and (c) that the series $\sum_{n=0}^{\infty} \psi_n(W_0)$ converges. Then

$$f(\beta) = \sum_{n=0}^{\infty} a_n \varphi_n(\beta)$$

converges absolutely for almost all $\beta \geq 0$, and $\tilde{f}(W)$ is the Laplace transform of $f(\beta)$.

For convergent series, Theorem 2 guarantees the existence of an expansion of $f(\beta)$ with respect to the gauge functions $\varphi_n(\beta)$. If $\varphi_n(\beta) > 0$ for $\beta > 0$, the determination of small β behavior by large W behavior then follows. The case of asymptotic power series is covered by the following theorem which involves only $\tilde{f}(W)$.

Theorem 3: Let $\tilde{f}(W)$ be the Laplace transform of a function $f(\beta)$. Assume that $\tilde{f}(W)$ is analytic except on the real axis for $W < W_0$, and that $\tilde{f}(W)^* = \tilde{f}(W^*)$. Assume further that

$$\tilde{f}(W) \sim \sum_{n=0}^{\infty} a_n W^{-\alpha_n - 1}$$

uniformly in arg W for $|W| \to \infty$ in the right half-plane and uniformly in Re W for Im $W \to \infty$ in the left half-plane where $\alpha_{n+1} > \alpha_n > 0$ for all n. Then

$$f(\beta) \sim \sum_{n=0}^{\infty} a_n \beta^{\alpha_n} / \Gamma(\alpha_n - 1) \quad \text{for} \quad \beta \to 0^+.$$

Theorem 3 is proved in Appendix B. Generalizations of Theorem 3 to cover asymptotic series with respect to other gauge functions can be obtained by generalizing the proof of Appendix B.

Before turning to the specific problem of the second virial coefficient, we remark that for some cases the Laplace transform of some of the matrix elements of $e^{-\beta H}$ may not exist (this is the case in Sec. IV). When this happens, the present methods may still be useful if the singular part can be subtracted off and handled exactly.[7]

III. SECOND VIRIAL COEFFICIENT

We begin with the formulation of Boyd, Larsen, and Kilpatrick,[8] and write the second virial coefficient B in the form

$$B = B_{\text{direct}} + B_{\text{exch}}, \tag{2}$$

where

$$B_{\text{direct}} = \tfrac{1}{2} N \int d^3\mathbf{r}[1 - 2^{\frac{3}{2}} \lambda^3 G(\mathbf{r}, \mathbf{r}; \beta)] \tag{3}$$

and

$$B_{\text{exch}} = \mp \tfrac{1}{2} N (2S+1)^{-1} 2^{\frac{3}{2}} \lambda^3 \int d^3\mathbf{r} G(\mathbf{r}, -\mathbf{r}; \beta). \tag{4}$$

[4] A. Erdelyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), pp. 29–34.

[5] G. N. Watson, Proc. London Math. Soc. (2) **17**, 113 (1918); E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford University Press, London, 1935), pp. 218–219.

[6] For a proof of Theorem 2, see G. Doetsch, *Einführung in Theorie und Anwendung der Laplace-Transformation* (Birkhäuser Verlag, Basel und Stuttgart, 1958), p. 186.

[7] After the present manuscript had been submitted, the author became aware of a paper of J. Lavoine [Ann. Inst. Poincaré A (France) **4**, 49 (1966)] on the Abelian and Tauberian asymptotics of the Laplace transform. Theorem 3 is a special case of Lavoine's Theorem VI. Readers of the present paper interested in the use of the Laplace transform to replace a study of $e^{-\beta H}$ by a study of $(H + W)^{-1}$ should find his paper of considerable interest.

[8] M. Boyd, S. Larsen, and J. Kilpatrick, Ref. 3, Eqs. (12)–(14).

The minus (upper) sign in B_{exch} is associated with Bose statistics and the plus sign with Fermi statistics. Here S is the spin, $\lambda \equiv (2\pi\hbar^2\beta/m)^{\frac{1}{2}}$ is the thermal deBroglie wavelength, and

$$G(\mathbf{r}, \mathbf{r}'; \beta) \equiv \langle \mathbf{r} \mid \exp(-\beta H_{\text{rel}}) \mid \mathbf{r}' \rangle, \tag{5}$$

where

$$H_{\text{rel}} = -(\hbar^2/m)\nabla^2 + V(\mathbf{r}) \tag{6}$$

is the Hamiltonian for the relative motion of a pair of particles of mass m. The Laplace transform

$$\bar{G}(\mathbf{r}, \mathbf{r}'; W) = \int_0^\infty e^{-\beta W} G(\mathbf{r}, \mathbf{r}'; \beta)\, d\beta \tag{7}$$

is the Green's function of the negative-energy Schrödinger equation:

$$[-(\hbar^2/m)\nabla^2 + V(\mathbf{r}) + W]\bar{G}(\mathbf{r}, \mathbf{r}'; W) = \delta(\mathbf{r} - \mathbf{r}'). \tag{8}$$

Equation (8) follows either from the general considerations of Sec. II or from Laplace-transforming the Bloch equation satisfied by $G(\mathbf{r}, \mathbf{r}'; \beta)$. We now introduce

$$\begin{aligned} g(\mathbf{r}, \mathbf{r}'; \gamma) &\equiv (4\pi\hbar^2/m)\bar{G}(\mathbf{r}, \mathbf{r}'; W), \\ \gamma &\equiv \hbar^{-1}(mW)^{\frac{1}{2}}, \\ U(\mathbf{r}) &\equiv (m/\hbar^2)V(\mathbf{r}). \end{aligned} \tag{9}$$

Equation (8) then becomes

$$[\nabla^2 - U(\mathbf{r}) - \gamma^2]g(\mathbf{r}, \mathbf{r}'; \gamma) = -4\pi\delta(\mathbf{r} - \mathbf{r}'). \tag{10}$$

If the potential $V(\mathbf{r})$ and the boundary conditions are spherically symmetric, the solution of (10) can be written in the form

$$g(\mathbf{r}, \mathbf{r}'; \gamma) = \sum_{l=0}^{\infty} (2l + 1)P_l(\cos\Theta)g^{(l)}(r, r'; \gamma), \tag{11}$$

where Θ is the angle between $\mathbf{r}$ and $\mathbf{r}'$. In spherical coordinates,

$$\delta(\mathbf{r} - \mathbf{r}') = r^{-2}\delta(r - r')\csc\theta\delta(\theta - \theta')\delta(\varphi - \varphi'). \tag{12}$$

By inserting (11) and (12) in (10), multiplying by $P_l(\cos\Theta)\sin\theta$, and integrating over θ and φ, it follows that $g^{(l)}$ satisfies[9]

$$(\mathfrak{L} - \gamma^2)g^{(l)}(r, r'; \gamma) = -r^{-2}\delta(r - r'), \tag{13}$$

where

$$\mathfrak{L} \equiv \frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}\right) - \left[\frac{l(l+1)}{r^2} + U(r)\right]. \tag{14}$$

If u_1 is a solution of $(\mathfrak{L} - \gamma^2)u = 0$ which satisfies the inner boundary condition, and u_2 is a solution of $(\mathfrak{L} - \gamma^2)u = 0$ which satisfies the outer boundary condition, then the solution of (13) is[9]

$$g^{(l)}(r, r'; \gamma) = -\frac{1}{(r')^2\Delta[u_1(r'), u_2(r')]}\begin{cases} u_1(r)u_2(r'); & r \le r', \\ u_2(r)u_1(r'); & r \ge r', \end{cases} \tag{15}$$

where

$$\Delta(u_1, u_2) \equiv (u_1u_2' - u_2u_1') \tag{16}$$

is the Wronskian of the two solutions.

An alternate form of the solution to (13), which will be of use later, is[9]

$$g^{(l)}(r, r'; \gamma) = \sum_n \frac{\psi_n^*(r)\psi_n(r')}{\gamma^2 - \gamma_n^2}, \tag{17}$$

where the ψ_n are the normalized eigenfunctions of $\mathfrak{L}$: $\mathfrak{L}\psi_n = \gamma_n^2\psi_n$.

In the case of free particles, denoted by a subscript zero, the solution of $(\mathfrak{L}_0 - \gamma^2)u_0 = 0$ which satisfies the inner boundary condition (regularity at $r = 0$) is $r^{-\frac{1}{2}}I_{l+\frac{1}{2}}(\gamma r)$; the solution which satisfies the outer boundary condition (regularity at ∞) is $r^{-\frac{1}{2}}K_{l+\frac{1}{2}}(\gamma r)$. Hence, for $r < r'$,

$$g_0^{(l)}(r, r'; \gamma) = (rr')^{-\frac{1}{2}}I_{l+\frac{1}{2}}(\gamma r)K_{l+\frac{1}{2}}(\gamma r'). \tag{18}$$

If (18) is inserted in (11), we obtain

$$g_0(\mathbf{r}, \mathbf{r}'; \gamma) = \sum_{l=0}^{\infty}(2l + 1)P_l(\cos\Theta)(rr')^{-\frac{1}{2}}I_{l+\frac{1}{2}}(\gamma r)K_{l+\frac{1}{2}}(\gamma r'), \quad r < r'. \tag{19}$$

An alternative form is

$$g_0(\mathbf{r}, \mathbf{r}'; \gamma) = |\mathbf{r} - \mathbf{r}'|^{-1}\exp(-\gamma|\mathbf{r} - \mathbf{r}'|). \tag{20}$$

The relation of the present formulation to the usual phase-shift formulation given by Gropper and by Beth and Uhlenbeck[10] is shown in Appendix C. We turn now to specific interactions.

IV. B_{direct} FOR HARD SPHERES

For hard spheres of radius a, the potential $V(\mathbf{r})$ is replaced by the condition $g(\mathbf{r}, \mathbf{r}'; \gamma) = 0$, $r \le a$. If we decompose g as $g = g_0 + g_s$, we have $g_s = -g_0$ for $r \le a$, and for $r \ge a$,

$$g_s(\mathbf{r}, \mathbf{r}'; \gamma) = -(rr')^{-\frac{1}{2}}\sum_{l=0}^{\infty}(2l + 1)P_l(\cos\Theta) \times [I_{l+\frac{1}{2}}(\gamma a)/K_{l+\frac{1}{2}}(\gamma a)]K_{l+\frac{1}{2}}(\gamma r)K_{l+\frac{1}{2}}(\gamma r'). \tag{21}$$

For large γa, the sum in (21) does not begin to

[9] This method of obtaining Green's functions is discussed by P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., New York, 1953), Vol. I, pp. 825–833.

[10] L. Gropper, Phys. Rev. **51**, 1108 (1937); E. Beth and G. Uhlenbeck, Physica **4**, 915 (1937). The phase-shift formulation may also be found in K. Huang's textbook *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1963), pp. 303–311.

converge until $l \approx \gamma a$. We now show how to handle (21) for the case $\mathbf{r} = \mathbf{r}'$.

The computation of B_{direct} from (3) is easier if we can perform the integration over $\mathbf{r}$ before inverse-Laplace-transforming back. First we must separate off those pieces of $g(\mathbf{r}, \mathbf{r}; \gamma)$ for which this $\mathbf{r}$ integral is not finite. To begin with, we note that $g_0(\mathbf{r}, \mathbf{r}; \gamma) = \infty$. From Eq. (21),

$$g_s(\mathbf{r}, \mathbf{r}; \gamma) = -r^{-1} \sum_{l=0}^{\infty} (2l + 1)[I_{l+\frac{1}{2}}(\gamma a)/K_{l+\frac{1}{2}}(\gamma a)] \times [K_{l+\frac{1}{2}}(\gamma r)]^2. \quad (22)$$

We note that $g_s(\mathbf{r}, \mathbf{r}; \gamma)$ has a first-order pole at $r = a$: For fixed x, $|\nu| \to \infty$, we have the asymptotic estimate

$$I_\nu(x) \sim (x/2)^\nu [\Gamma(\nu + 1)]^{-1}. \quad (23)$$

Using $\Gamma(\nu)\Gamma(1 - \nu) = \pi \csc(\pi\nu)$ and

$$K_\nu(x) = \tfrac{1}{2}\pi \csc(\pi\nu)[I_{-\nu}(x) - I_\nu(x)], \quad (24)$$

we see that for fixed x, $|\nu| \to \infty$ with $\operatorname{Re} \nu > 0$,

$$K_\nu(x) \sim \tfrac{1}{2}\Gamma(\nu)(x/2)^{-\nu}. \quad (25)$$

From (23) and (25) we see that for $l \gg \gamma r \geq \gamma a$, the summand of (22) becomes $-r^{-1}(a/r)^{2l+1}$. Hence (22) converges for $r > a$, but diverges when $r = a$. Thus, for r sufficiently close to a, only the terms for which (23) and (25) are valid estimates matter, and we see that

$$g_s(\mathbf{r}, \mathbf{r}; \gamma) \xrightarrow[r \to a]{} -r^{-1} \sum_{l=0}^{\infty} (a/r)^{2l+1} = -ar^{-2}[1 - (a/r)^2]^{-1}. \quad (26)$$

The existence of a pole at $r = a$ is easily understood from the viewpoint of the method of images, which is applicable to the $|\mathbf{r} - \mathbf{r}'|^{-1}$ singularity of g_0 which produces the delta function on the right-hand side of (10) upon application of the Laplacian. Both g_0 and a piece of g_s which contains the pole at $r = a$ must be separated off before we can perform the integration over $\mathbf{r}$. Hence for $r \geq a$ we make the decomposition

$$g = g_0 + g_s^{(1)} + g_s^{(2)} \quad (27)$$

with corresponding decompositions of $\bar{G}$ and G. Here $g_s^{(1)}$ is a piece of g_s which contains the pole at $r = a$. We exploit our freedom to choose any $g_s^{(1)}$ which has the same residue at $r = a$ as g_s and pick a $g_s^{(1)}$ which gives the first quantum correction found by Uhlenbeck and Beth[11] exactly with nothing left over:

$$g_s^{(1)}(\mathbf{r}, \mathbf{r}; \gamma) \equiv -\frac{a^2 \exp[-2\gamma(r - a)]}{2r^2(r - a)}. \quad (28)$$

[11] G. Uhlenbeck and E. Beth, Physica 3, 729 (1936).

Inverse-Laplace-transforming back, we find

$$G_0(\mathbf{r}, \mathbf{r}'; \beta) = 2^{-\frac{3}{2}}\lambda^{-3} \exp[-\tfrac{1}{2}\pi\lambda^{-2}|\mathbf{r} - \mathbf{r}'|^2]. \quad (29)$$

Hence

$$G_0(\mathbf{r}, \mathbf{r}; \beta) = 2^{-\frac{3}{2}}\lambda^{-3}, \quad (30)$$

and, for $r \geq a$,

$$G_s^{(1)}(\mathbf{r}, \mathbf{r}; \beta) = -2^{-\frac{3}{2}}\lambda^{-3}a^2r^{-2} \exp[-2\pi\lambda^{-2}(r - a)^2]. \quad (31)$$

Using (31) in (3), remembering that $G = 0$ for $r \geq 0$, we get

$$B_{\text{direct}} = \tfrac{2}{3}N\pi a^3\left[1 + \frac{3}{2\sqrt{2}}\left(\frac{\lambda}{a}\right)\right] + B_{\text{direct}}^{(2)}, \quad (32)$$

where

$$B_{\text{direct}}^{(2)} = -2^{\frac{1}{2}}N\lambda^3 \int_a^\infty G_s^{(2)}(\mathbf{r}, \mathbf{r}; \beta) 4\pi r^2 \, dr. \quad (33)$$

We have now succeeded in separating out the pieces of G for which the integration over $\mathbf{r}$ in Eq. (3) cannot be done before inverse-Laplace-transforming back; the r integration in (33) *can* be performed on $\bar{G}_s^{(2)}(\mathbf{r}, \mathbf{r}; W)$.

To compute $g_s^{(2)} = g_s - g_s^{(1)}$ from (22) and (28), we expand (28) and define

$$g_s^{(1)}(\mathbf{r}, \mathbf{r}; \gamma; L) \equiv -\frac{a(a + r)}{2r^3} e^{-2\gamma(r-a)} \sum_{l=0}^{L} \left(\frac{a}{r}\right)^{2l+1}. \quad (34)$$

Similarly, we define

$$g_s(\mathbf{r}, \mathbf{r}; \gamma; L) \equiv -r^{-1} \sum_{l=0}^{L} (2l + 1)[I_{l+\frac{1}{2}}(\gamma a)/K_{l+\frac{1}{2}}(\gamma a)] \times [K_{l+\frac{1}{2}}(\gamma r)]^2. \quad (35)$$

Then,

$$\int_a^\infty g_s^{(2)}(\mathbf{r}, \mathbf{r}; \gamma) 4\pi r^2 \, dr = \lim_{L \to \infty} \left\{ \int_a^\infty g_s(\mathbf{r}, \mathbf{r}; \gamma; L) 4\pi r^2 \, dr - \int_a^\infty g_s^{(1)}(\mathbf{r}, \mathbf{r}; \gamma; L) 4\pi r^2 \, dr \right\}. \quad (36)$$

By using the finite-geometric series to perform the sum in Eq. (34), one can easily show that

$$\int_a^\infty g_s^{(1)}(\mathbf{r}, \mathbf{r}; \gamma; L) 4\pi r^2 \, dr = -2\pi a^2 \ln(L/\gamma a) + O\left(\frac{1}{L}\right). \quad (37)$$

By using the indefinite integral[12]

$$\int [K_\nu(x)]^2 x \, dx = \frac{x^2}{2}\left\{\left(1 + \frac{\nu^2}{x^2}\right)[K_\nu(x)]^2 - [K_\nu'(x)]^2\right\} \quad (38)$$

[12] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin, 1966), p. 88.

and the Wronskian relation[13]

$$I_\nu(x)K'_\nu(x) - I'_\nu(x)K_\nu(x) = -x^{-1}, \tag{39}$$

one can show that

$$\int_a^\infty g_s(\mathbf{r}, \mathbf{r}; \gamma; L)4\pi r^2\, dr = 2\pi a^2 \sum_{l=0}^{L} F_{l+\frac{1}{2}}(\gamma a), \tag{40}$$

where F_ν is defined by

$$F_\nu(x) \equiv 2\nu\left\{\left(1 + \frac{\nu^2}{x^2}\right)I_\nu(x)K_\nu(x) - I'_\nu(x)K'_\nu(x) + x^{-1}K'_\nu(x)[K_\nu(x)]^{-1}\right\}. \tag{41}$$

For x large, $F_\nu(x)$ is a slowly varying function of ν.[14] Thus we can perform the sum in (40) by using the Euler–Maclaurin sum formula[15]:

$$\sum_{j=m}^{j=n} f(j) = \int_m^n f(x)\, dx + \tfrac{1}{2}[f(m) - f(n)] + \sum_{k\geq 1}(-1)^k \frac{B_k}{(2k)!}[f(m)^{(2k-1)} - f(n)^{(2k-1)}], \tag{42}$$

where the numbers B_k are the Bernoulli numbers: $B_1 = 1/6$, $B_2 = 1/30$, $B_3 = 1/42$, etc. For x large, Debye's series[16] provide asymptotic expansions of $I_\nu(x)$, $K_\nu(x)$, $I'_\nu(x)$, and $K'_\nu(x)$, which are uniformly valid for $0 \leq \nu \leq \infty$ and can be used to compute an asymptotic expansion of $F_\nu(x)$ uniformly valid in ν for large x. The needed uniform approximation is much more easily obtained for the modified Bessel functions I_ν and K_ν than for the J_ν and Y_ν, which appear in the other authors' formulations of the problem, because the modified Bessel equation has no turning points. This absence of turning points from the relevant radial equation is one of the advantages of the present Laplace-transform method, which permits one to work with the Schrödinger equation at large *negative* energies.

Debye's series have the form

$$\begin{aligned}
I_\nu(x) &\sim (2\pi)^{-\frac{1}{2}}(\nu^2 + x^2)^{-\frac{1}{4}} \times e^\mu \sum_{k=0}^{\infty}[t^{-k}u_k(t)](\nu^2 + x^2)^{-k/2},\\
K_\nu(x) &\sim (\pi/2)^{\frac{1}{2}}(\nu^2 + x^2)^{-\frac{1}{4}} \times e^{-\mu} \sum_{k=0}^{\infty}[(-t)^{-k}u_k(t)](\nu^2 + x^2)^{-k/2},\\
I'_\nu(x) &\sim (2\pi)^{-\frac{1}{2}}x^{-1}(\nu^2 + x^2)^{\frac{1}{4}} \times e^\mu \sum_{k=0}^{\infty}[t^{-k}v_k(t)](\nu^2 + x^2)^{-k/2},\\
K'_\nu(x) &\sim -(\pi/2)^{\frac{1}{2}}x^{-1}(\nu^2 + x^2)^{\frac{1}{4}} \times e^{-\mu} \sum_{k=0}^{\infty}[(-t)^{-k}v_k(t)](\nu^2 + x^2)^{-k/2},
\end{aligned} \tag{43}$$

where

$$\mu = (\nu^2 + x^2)^{\frac{1}{2}} - \nu \sinh^{-1}(\nu/x)$$

and

$$t = \nu(\nu^2 + x^2)^{-\frac{1}{2}}.$$

The polynomial coefficients $u_k(t)$, $v_k(t)$ are determined recursively[17] from

$$\begin{aligned}
u_{k+1}(t) &= \tfrac{1}{2}t^2(1 - t^2)u'_k(t) + \frac{1}{8}\int_0^t (1 - 5t^2)u_k(t)\, dt,\\
v_k(t) &= u_k(t) + t(t^2 - 1)[\tfrac{1}{2}u_{k-1}(t) + tu'_{k-1}(t)],
\end{aligned} \tag{44}$$

where $u_0(t) = 1$. Using the series (43) in the definition (41) produces the asymptotic expansion

$$F_\nu(x) \sim \frac{\nu}{x^2}\sum_{k=0}^{\infty}\{t^{-k}[\xi_k(t) + \eta_k(t) - 2\zeta_k(t)] \times (x^2 + \nu^2)^{-(k-1)/2}\}. \tag{45}$$

The coefficients ξ_k arise from multiplying the series for I_ν and K_ν; the η_k come from the term $I'_\nu K'_\nu$. We have

$$\begin{aligned}
\xi_k(t) &= \sum_{l=0}^{k}(-1)^l u_{k-l}(t)u_l(t),\\
\eta_k(t) &= \sum_{l=0}^{k}(-1)^l v_{k-l}(t)v_l(t).
\end{aligned} \tag{46}$$

The coefficients ζ_k come from dividing the series for K'_ν and K_ν, and are most easily computed recursively from

$$\zeta_k(t) = (-1)^k v_k(t) - \sum_{l=0}^{k-1}(-1)^{k-l}u_{k-l}(t)\zeta_l(t). \tag{47}$$

[13] W. Magnus, F. Oberhettinger, and R. P. Soni, Ref. 12, p. 68.

[14] This constructive interference of the various partial waves on and near the caustic $\theta = 0$ gives rise to the phenomenon of "the glory" in the corresponding optical problem.

[15] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., 1940), p. 431; E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, Eng., 1952) 4th ed., pp. 127–128. The modification of the Watson transform used by V. A. Fock [Sov. Phys.—JETP **9**, 255 (1945); reprinted in *Electromagnetic Diffraction and Propagation Problems* (Pergamon Press, Inc., New York, 1965), [pp. 191–212] and by W. Franz [Z. Naturforsch. 9a, 705 (1954)] in the illuminated region of the corresponding optical problem is not useful here because it depends on the existence of a saddle point which arises from destructive interference between various partial waves away from the caustic $\theta = 0$.

[16] W. Magnus, F. Oberhettinger, and R. P. Soni, Ref. 12, p. 140; M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover Publications, Inc., New York 1965), p. 378. The Debye expansions are also uniformly valid in arg x^2 as $|x^2| \to \infty$ with Re $x^2 \geq 0$ and uniformly valid in Re x^2 as Im $x^2 \to \infty$ with Re $x^2 \leq 0$, so that the conditions for the use of Theorem 3 to justify term-by-term inverse-Laplace-transformation of (52) are satisfied.

[17] M. Abramowitz and I. A. Stegun (Ref. 15, p. 366) tabulate the first few u_k and v_k.

The first few coefficients are

$$\xi_{2k+1}(t) = 0, \quad \xi_0(t) = 1, \quad \xi_2(t) = \tfrac{1}{8}t^2(1 - 6t^2 + 5t^4),$$
$$\xi_4(t) = (1/128)t^4(27 - 580t^2 + 2170t^4 - 2772t^6 + 1155t^8);$$
$$\eta_{2k+1}(t) = 0, \quad \eta_0(t) = 1,$$
$$\eta_2(t) = \tfrac{1}{8}t^2(-3 + 10t^2 - 7t^4),$$
$$\eta_4(t) = \tfrac{1}{128}t^4(-45 + 812t^2 - 2790t^4 + 3388t^6 - 1365t^8);$$
$$\zeta_0(t) = 1, \quad \zeta_1(t) = \tfrac{1}{2}t(1 - t^2),$$
$$\zeta_2(t) = \tfrac{1}{8}t^2(-1 + 6t^2 - 5t^4),$$
$$\zeta_3(t) = \tfrac{1}{8}t^3(1 - 13t^2 + 27t^4 - 15t^6),$$
$$\zeta_4(t) = \tfrac{1}{128}t^4(-25 + 556t^2 - 2078t^4 + 2652t^6 - 1105t^8),$$
$$\zeta_5(t) = \tfrac{1}{32}t^5(13 - 439t^2 + 2550t^4 - 5514t^6 + 5085t^8 - 1695t^{10}).$$

Using these coefficients, (45) becomes

$$F_\nu(x) = -\frac{\nu}{(\nu^2 + x^2)} - \frac{\nu^3}{(\nu^2 + x^2)^2} + \frac{\nu}{(\nu^2 + x^2)^4}[-\nu^4 + \tfrac{5}{2}\nu^2x^2 - \tfrac{1}{4}x^4] + \frac{\nu}{(\nu^2 + x^2)^{\frac{11}{2}}}[-\nu^6 + \tfrac{17}{2}\nu^4x^2 - \tfrac{47}{8}\nu^2x^4 + \tfrac{1}{4}x^6] + \frac{\nu}{(\nu^2 + x^2)^7}[-\nu^8 + 23\nu^6x^2 - \tfrac{231}{4}\nu^4x^4 + \tfrac{187}{8}\nu^2x^6 - \tfrac{13}{16}x^8] + \nu O[(\nu^2 + x^2)^{-\frac{7}{2}}]. \quad (48)$$

For $\nu \ll x$, (48) can be re-expanded to get

$$F_\nu(x) = -\frac{\nu}{x^2} + \frac{\nu^3}{x^4} - \frac{\nu}{4x^4} + O\left(\frac{1}{x^5}\right). \quad (49)$$

Using the Euler–Maclaurin sum formula (42), writing

$$\int_{\frac{1}{2}}^{L} F_\nu(x)\,d\nu = \int_0^{L} F_\nu(x)\,d\nu - \int_0^{\frac{1}{2}} F_\nu(x)\,d\nu,$$

and using (49) for $0 \le \nu \le \frac{1}{2}$, we obtain

$$\sum_{l=0}^{L} F_{l+\frac{1}{2}}(x) = \int_0^{L} F_\nu(x)\,d\nu - \frac{1}{24x^2} - \frac{17}{960x^4} + O\left(\frac{1}{x^5}\right) + O\left(\frac{1}{L}\right). \quad (50)$$

The integral in (50) can be performed by using (48) and the formula

$$\int_0^\infty \frac{\nu^m\,d\nu}{(x^2 + \nu^2)^{n/2}} = \frac{\Gamma[(m+1)/2]\Gamma[(n-m-1)/2]}{2\Gamma(n/2)x^{n-m-1}}. \quad (51)$$

Insertion of the result in (40) yields

$$\int_a^\infty g_s(r, r; \gamma; L)4\pi r^2\,dr = 2\pi a^2\left\{-\ln(L/\gamma a) - \frac{2}{3\gamma a} - \frac{1}{24(\gamma a)^2} + \frac{2}{315(\gamma a)^3} - \frac{1}{480(\gamma a)^4} + O\left(\frac{1}{(\gamma a)^5}\right) + O\left(\frac{1}{L}\right)\right\}. \quad (52)$$

Using (37) and (52) in (36), letting $L \to \infty$, and inverse-Laplace-transforming back yields finally

$$B_{\text{direct}} = \tfrac{2}{3}N\pi a^3\left\{1 + \frac{3}{2\sqrt{2}}\left(\frac{\lambda}{a}\right) + \frac{1}{\pi}\left(\frac{\lambda}{a}\right)^2 + \frac{1}{16\,\pi\sqrt{2}}\left(\frac{\lambda}{a}\right)^3 - \frac{1}{105\pi^2}\left(\frac{\lambda}{a}\right)^4 + \frac{1}{640\pi^2\sqrt{2}}\left(\frac{\lambda}{a}\right)^5 + O\left(\left(\frac{\lambda}{a}\right)^6\right)\right\}. \quad (53)$$

The result (53) is in agreement with the numerical results of Boyd, Larson, and Kilpatrick.[3] All but the last coefficient $[1/(640\pi^2\sqrt{2}) = 0.0001119]$ have been obtained analytically by previous authors.[18] Clearly additional terms can be calculated by computing more of the polynomials $\xi_k(t)$, $\eta_k(t)$, and $\zeta_k(t)$ from (44), (46), and (47), although this becomes an increasingly tedious task with increasing k.

V. EXCHANGE SECOND VIRIAL COEFFICIENT

The computation of $g(\mathbf{r}, -\mathbf{r}; \gamma)$ for large γ is easily done with the aid of the Watson transformation.[19] We consider only hard spheres and potentials more strongly repulsive than r^{-2} as $r \to 0$. The original series (11) does not converge for $r = r'$ because of the singularity of $g(\mathbf{r}, \mathbf{r}'; \gamma)$ at $\mathbf{r} = \mathbf{r}'$; thus we keep $r < r'$ until the transformation is completed.

We begin by extending the domain of definition of $g^{(l)}(r, r'; \gamma)$ to complex l by means of the differential equation (13), wherein we replace l by $\nu - \frac{1}{2}$ where ν is an arbitrary complex number. We set $\Theta = \pi$ in the expansion (11) and exploit the fact that $-\pi \sec(\pi\nu)$ has first-order poles with residue $(-1)^l$ at $\nu = l + \frac{1}{2}$ to replace the sum in (11) by a contour

[18] G. Uhlenbeck and E. Beth, Ref. 10, obtained the (λ/a) term. The $(\lambda/a)^2$ term, apart from a missing factor 2, was obtained by F. Mohling, Ref. 1. The correct coefficient $1/\pi$, together with the $(\lambda/a)^3$ term, was obtained by Handelsman and Keller, Ref. 3; the $(\lambda/a)^4$ term was obtained by Hemmer and Mork, Ref. 3, using the method of R. A. Handelsman and J. B. Keller, viz., an expansion of the thermal Green's function (5) and its boundary conditions in powers of (λ/a).

[19] G. N. Watson, Proc. Roy. Soc. (London) **95**, 83 (1918). The application of the Watson method to quantum mechanics initiated by T. Regge is discussed by R. G. Newton in *The Complex j-Plane* (W. A. Benjamin, Inc., New York 1964).

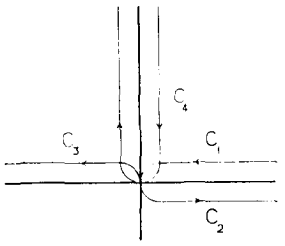


FIG. 1. Contours for the Watson transformation.

integral:

$$g(\mathbf{r}, \mathbf{r}'; \gamma)\Big|_{\Theta=\pi} = i\int_{C_1+C_2} \nu \sec(\pi\nu) g^{(\nu-\frac{1}{2})}(r, r'; \gamma)\, d\nu. \tag{54}$$

The contour $C_1 + C_2$, which encloses the positive real axis, is indicated in Fig. 1. Now $l(l+1) = \nu^2 - \frac{1}{4}$; hence the differential operator in (13) is even in ν. Furthermore, the boundary conditions for $r \to \infty$ and for $r \to 0$ do not involve ν. Hence the solution $g^{(\nu-\frac{1}{2})}$ is an even function of ν,[20] from which it follows that the contour C_2 can be replaced by its reflection C_3 traversed in the indicated direction.

For ν pure imaginary, the centrifugal potential $(\nu^2 - \frac{1}{4})r^{-2}$ becomes attractive, giving rise to a bound state[21] and a pole in the Green's function $g^{(\nu-\frac{1}{2})}(r, r'; \gamma)$. For sufficiently large real γ, these poles on the imaginary axis are the only singularities of $g^{(\nu-\frac{1}{2})}$ regarded as a function of the complex variable ν.[22] We now replace the contour $C_1 + C_3$ by the contour C_4 which surrounds these poles on the imaginary axis. We can now let $r \to r'$. We compute the residue of $g^{(\nu-\frac{1}{2})}$ by using (17); with γ fixed, the poles occur when $\gamma_n(\nu) = \gamma$. The residue is

$$-\tfrac{1}{2}\gamma_n^{-1}(\partial\nu/\partial\gamma_n)\psi_n^*(r)\psi_n(r').$$

We denote the values of ν at which these poles occur by $i\alpha_n(\gamma)$. Thus we obtain

$$g(\mathbf{r}, -\mathbf{r}; \gamma) = \pi\gamma^{-1}\sum_n \alpha_n \operatorname{sech}(\pi\alpha_n)(\partial\alpha_n/\partial\gamma)\,|\psi_n(r)|^2, \tag{55}$$

where the sum runs only over the positive α_n.

The complex Laplace inversion integral gives

$$G(\mathbf{r}, \mathbf{r}'; \beta) = \frac{1}{2\pi i}\int_{W_0-i\infty}^{W_0+i\infty} e^{\beta W}\bar{G}(\mathbf{r}, \mathbf{r}'; W)\, dW$$
$$= \frac{1}{4\pi^2 i}\int_{\gamma_0-i\infty}^{\gamma_0+i\infty} \exp\left(\frac{\lambda^2\gamma^2}{2\pi}\right) g(\mathbf{r}, \mathbf{r}'; \gamma)\gamma\, d\gamma, \tag{56}$$

where the second equality follows from using (9) and deforming the image in the γ plane of the original contour in the W plane to lie parallel to the imaginary axis. We now use (55) and (56) in (4). The $\mathbf{r}$ integration can be done trivially by virtue of the fact that the $\psi_n(r)$ are normalized radial functions:

$$\int_0^\infty |\psi_n(r)|^2 r^2\, dr = 1.$$

Thus we obtain

$$B_{\text{exch}} = \mp\frac{N}{2S+1}\sum_n A_n, \tag{57}$$

where

$$A_n = -2^{\frac{3}{2}}i\lambda^3\int_{\gamma_0-i\infty}^{\gamma_0+i\infty} g_n(\gamma)\exp[f_n(\gamma)]\, d\gamma \tag{58}$$

with

$$g_n(\gamma) = \alpha_n(\gamma)\partial\alpha_n(\gamma)/\partial\gamma \tag{59}$$

and

$$f_n(\gamma) = \frac{\lambda^2\gamma^2}{2\pi} - \ln[2\cosh\pi\alpha_n(\gamma)]. \tag{60}$$

At high temperatures, the integral in (58) will have a saddle point for large real γ which can be exploited to evaluate A_n by the method of steepest descent. Such an evaluation, when explicitly carried out, verifies that only the first term in (57) is important at high temperatures as is expected from the discussion of Sec. II.

A. Hard Spheres

For hard spheres of radius a, bound states occur when the function $K_\nu(\gamma r)$, which satisfies the boundary condition for $r \to \infty$, also satisfies the boundary condition of vanishing at the surface of the sphere. Hence the α_n in this case are determined by

$$K_{i\alpha_n}(\gamma_a) = 0. \tag{61}$$

[The fact that $g^{(\nu-\frac{1}{2})}$ has a pole when (61) is satisfied is also clear from (21).]

An asymptotic series for the roots $\alpha_n(\gamma)$ can be obtained from the Nicholson-type expansion of $K_{ip}(x)$, which has the form[23]

$$K_{ip}(x) = \pi e^{-\pi p/2}(\tfrac{1}{2}x)^{-\frac{1}{3}}\left[\sum_{l=0}^{m}(-1)^l p_l(\xi)(-\tfrac{1}{2}ix)^{-2l/3}\right]$$
$$\times Ai(\rho e^{-2\pi i/3}) + O[(-ix)^{-2(n+1)/3}],$$

where

$$\rho = \sum_{k=0}^{m}(-1)^k q_k(\xi)(-\tfrac{1}{2}ix)^{-2k/3},$$
$$\xi = i(x-p)(-\tfrac{1}{2}ix)^{-\frac{1}{3}}.$$

[20] This happens only for potentials more strongly repulsive than r^{-2} as $r \to 0$, since only for this case is the boundary condition as $r \to 0$ independent of ν. See R. G. Newton, Ref. 18, p. 22, and footnote 2 of Chap 4.

[21] For potentials more strongly repulsive than r^{-2} as $r \to 0$, see R. G. Newton, Ref. 18, p. 49.

[22] This statement, as well as the replacement of $C_1 + C_3$ by C_4, are justified in Appendix D.

[23] W. Magnus, F. Oberhettinger, and R. P. Soni, Ref. 11, p. 145. See also, C. B. Balogh, thesis Oregon State University, June 1965; C. B. Balogh, Bull. Am. Math. Soc. 72, 40 (1966); S. I. A. M. J. Appl. Math. 15, 1315 (1967).

The p_l and q_k are polynomials. We only need the q_k, the first few of which are[24]

$$q_0(\xi) = \xi, \quad q_1(\xi) = \tfrac{1}{60}\xi^2,$$
$$q_2(\xi) = \tfrac{2}{1575}\xi^3 + \tfrac{1}{140},$$
$$q_3(\xi) = \tfrac{41}{283500}\xi^4 + \tfrac{4}{1575}\xi,$$
$$q_4(\xi) = \tfrac{6553}{32744250 0}\xi^5 + \tfrac{1409}{1819125}\xi^2.$$

The roots α_n occur when $\rho e^{-2\pi i/3}$ is equal to a zero of the Airy function. Inverting the series for ρ yields for the nth root of $K_{ip}(\gamma a)$

$$\alpha_n(\gamma) = \gamma a + \beta_n(\gamma a)^{\frac{1}{3}} + \tfrac{1}{30}\beta_n^2(\gamma a)^{-\frac{1}{3}} + (\tfrac{1}{70} - \tfrac{1}{350}\beta_n^3)(\gamma a)^{-1} + (-\tfrac{29}{31150}\beta_n + \tfrac{281}{5670000}\beta_n^4)(\gamma a)^{-\frac{5}{3}} + O(\gamma^{-\frac{7}{3}}), \tag{62}$$

where $-2^{\frac{1}{3}}\beta_n$ is the nth root of the Airy function:

$$Ai(-2^{\frac{1}{3}}\beta_n) = 0. \tag{63}$$

For $(\lambda/a) \ll 1$, we evaluate the integral in (58) by steepest descent.[25] With an error of the same order as that already inherent in the method of steepest descent, we replace (60) by

$$f_n(\gamma) \approx (2\pi)^{-1}\lambda^2\gamma^2 - \pi\alpha_n(\gamma). \tag{64}$$

We choose γ_0 in (58) to be the saddle point; it is the root of $f_n'(\gamma_0) = 0$. Using (62) and (64)

$$\gamma_0 a = \left(\frac{\pi a}{\lambda}\right)^2 + \tfrac{1}{3}\beta_n\left(\frac{\pi a}{\lambda}\right)^{\frac{2}{3}} - \tfrac{23}{270}\beta_n^2\left(\frac{\pi a}{\lambda}\right)^{-\frac{2}{3}} + (-\tfrac{1}{70} + \tfrac{149}{3150}\beta_n^3)\left(\frac{\pi a}{\lambda}\right)^{-2} + (\tfrac{53}{1890}\beta_n - \tfrac{30346 7}{9185400}\beta_n^4)\left(\frac{\pi a}{\lambda}\right)^{-\frac{10}{3}} + O\left[\left(\frac{\pi a}{\lambda}\right)^{-\frac{14}{3}}\right]. \tag{65}$$

Expanding f_n and g_n in Taylor series about γ_0 in standard fashion and performing the integration in (58) yields

$$A_n = 4\lambda^3 \exp\left[f_n(\gamma_0)\right]\{g_n(\gamma_0)[f_n''(\gamma_0)/\pi]^{-\frac{1}{2}} - [1/(2\pi)]g_n''(\gamma_0)[f_n''(\gamma_0)/\pi]^{-\frac{3}{2}} + (1/\pi^2)[f_n''(\gamma_0)/\pi]^{-\frac{5}{2}} \times [\tfrac{1}{2}f_n'''(\gamma_0)g_n'(\gamma_0) + \tfrac{1}{8}f_n^{iv}(\gamma_0)g_n(\gamma_0)] + O(\gamma_0^{-\frac{7}{3}})\}. \tag{66}$$

Working out the expansions of the quantities in (66) yields finally

$$A_n = 4\pi^3 a^3 h_n \exp(f_n), \tag{67}$$

where

$$h_n = 1 + \tfrac{14}{9}\beta_n\left(\frac{\pi a}{\lambda}\right)^{-\frac{4}{3}} + \tfrac{130}{405}\beta_n^2\left(\frac{\pi a}{\lambda}\right)^{-\frac{8}{3}} + \frac{7}{81\pi}\beta_n\left(\frac{\pi a}{\lambda}\right)^{-\frac{10}{3}} - \tfrac{92}{3645}\beta_n^3\left(\frac{\pi a}{\lambda}\right)^{-4} + O\left[\left(\frac{a}{\lambda}\right)^{-\frac{14}{3}}\right] \tag{68}$$

and

$$f_n = -\pi\left\{\frac{1}{2}\left(\frac{\pi a}{\lambda}\right)^2 + \beta_n\left(\frac{\pi a}{\lambda}\right)^{\frac{2}{3}} + \tfrac{4}{45}\beta_n^2\left(\frac{\pi a}{\lambda}\right)^{-\frac{2}{3}} + (\tfrac{1}{70} - \tfrac{2688}{14175}\beta_n^3)\left(\frac{\pi a}{\lambda}\right)^{-2} + (-\tfrac{22}{1575}\beta_n + \tfrac{156864}{1913625}\beta_n^4)\left(\frac{\pi a}{\lambda}\right)^{-\frac{10}{3}} + O\left[\left(\frac{\pi a}{\lambda}\right)^{-\frac{14}{3}}\right]. \tag{69}$$

The roots of the Airy function are tabulated by Abramowitz and Stegun.[26] Using their numbers yields $\beta_1 = 1.85576$, $\beta_2 = 3.24460$. Using this β_1,

$$h_1 = 1 + 0.627394(\lambda/a)^{\frac{4}{3}} + 0.0522153(\lambda/a)^{\frac{8}{3}} + 0.00112413(\lambda/a)^{\frac{10}{3}} - 0.00165598(\lambda/a)^4 + O[(\lambda/a)^{\frac{14}{3}}], \tag{70}$$

$$f_1 = -15.5031(\lambda/a)^{-2} - 12.5056(\lambda/a)^{-\frac{2}{3}} - 0.448339(\lambda/a)^{\frac{2}{3}} + 0.0339141(\lambda/a)^2 - 0.00492278(\lambda/a)^{\frac{10}{3}} + O[(\lambda/a)^{\frac{14}{3}}]. \tag{71}$$

The results of comparing $A_1/(\frac{8}{3}\pi a^3)$ calculated from these series with the high-speed computer calculations of B^*_{exch} by Boyd, Larsen, and Kilpatrick[3] for a few values of (λ/a) are given in Table I. A_2 is negligible compared to A_1 for these values of (λ/a). Clearly the agreement worsens with increasing (λ/a). The fact that our methods guarantee only an asymptotic series for A_1 makes it unclear whether calculating additional terms would improve the agreement.

TABLE I. Calculated values of $A_1/(\frac{8}{3}\pi a^3)$ compared with corresponding values of B^*_{exch} (see Ref. 3).

λ/a	$A_1/(\frac{8}{3}\pi a^3)$	$-B^*_{\text{exch}}$
0.875	1.167×10^{-13}	1.16×10^{-13}
1.000	4.481×10^{-11}	4.4840×10^{-11}
1.250	7.29×10^{-8}	7.3070075×10^{-8}
1.500	5.66×10^{-6}	$5.68623166 \times 10^{-6}$
1.750	9.7×10^{-5}	$9.830673706 \times 10^{-5}$
2.000	7.2×10^{-4}	$7.332542607 \times 10^{-4}$
3.000	5.3×10^{-2}	$5.7408314497 \times 10^{-2}$

[24] q_3 and q_4 are from W. Schöbe, Acta Math. **92**, 267 (1954), in particular, p. 290.

[25] Morse and Feshbach, Ref. 8, pp. 437–443. More rigorous discussions based on Watson's lemma are given by A. Erdelyi, Ref. 4, pp. 34–41, and by E. T. Copson, Ref. 5, pp. 330–331.

[26] M. Abramowitz and I. A. Stegun, Ref. 15, p. 478.

B. Lennard-Jones 6–12 Potential

Next we consider the potential

$$V(r) = V_0[(r_0/r)^{12} - (r_0/r)^6]. \tag{72}$$

If we put $l = ip - \frac{1}{2}$, the relevant radial equation is

$$\frac{d^2}{dr^2}[ru(r)] - \left[\frac{mV(r)}{\hbar^2} - \frac{p^2+\frac{1}{4}}{r^2} + \gamma^2\right]ru(r) = 0. \tag{73}$$

As p increases with γ large and fixed, the potential well $\{[mV(r)/\hbar^2] - (p^2+\frac{1}{4})/r^2\}$ deepens until its bottom is below γ^2 and the first bound state occurs. Additional bound states occur as the well deepens further. For γ large, this happens for p large; for p large it is a good approximation to expand in Taylor series about the bottom of the well, which occurs at

$$r_{\min} = r_0\sigma^{\frac{1}{6}}[1 - (1/20)\sigma + (15/4000)\sigma^2 + O(\sigma^3)], \tag{74}$$

where

$$\sigma \equiv [6mV_0r_0^2\hbar^{-2}(p^2+\tfrac{1}{4})^{-1}]^{\frac{3}{5}}. \tag{75}$$

The result of expanding about $r_{\min}$ is

$$\begin{aligned}&\{[mV(r)/\hbar^2] - (p^2+\tfrac{1}{4})/r^2\}\\ &= 10(p^2+\tfrac{1}{4})r_0^{-2}\{-\tfrac{1}{12}\sigma^{-1}[1 + \tfrac{1}{5}\sigma + \tfrac{3}{100}\sigma^2 + O(\sigma^3)]\\ &\quad + \sigma^{-\frac{2}{3}}[1 + \tfrac{1}{2}\sigma + O(\sigma^2)]x^2 - \tfrac{1}{3}\sigma^{-\frac{5}{6}}[1 + O(\sigma)]x^3\\ &\quad + \tfrac{89}{4}\sigma^{-1}[1 + O(\sigma)]x^4 + O(x^5)\},\end{aligned} \tag{76}$$

where $x \equiv (r - r_{\min})/r_0$. If (76) is inserted in (73), one obtains an anharmonic-oscillator equation in which the anharmonic terms can be treated as small perturbations for p large. The condition that the wavefunction vanish at $r = 0$ can be replaced by the same condition at $-\infty$ with an exponentially small error; the eigenvalues are then[27]

$$\begin{aligned}(r_0\gamma_n)^2 &= \tfrac{5}{6}(p^2+\tfrac{1}{4})\sigma^{-\frac{1}{3}}[1 + \tfrac{1}{5}\sigma + \tfrac{3}{100}\sigma^2]\\ &\quad - 2[10(p^2+\tfrac{1}{4})]^{\frac{1}{2}}\sigma^{-\frac{1}{3}}[1 + \tfrac{1}{4}\sigma](n+\tfrac{1}{2})\\ &\quad + \sigma^{-\frac{1}{3}}[\tfrac{161}{18}(n^2+n) + \tfrac{97}{18}] + O(p^{-\frac{2}{3}}),\end{aligned} \tag{77}$$

where n is the quantum number indexing the levels. Insertion of (75) into (77) followed by inversion of the resulting series to obtain $p = \alpha_n$ as a function of γ yields

$$\begin{aligned}\alpha_n(\gamma) &= (\tfrac{6}{5})^{\frac{5}{12}}\tau^{\frac{1}{12}}(r_0\gamma)^{\frac{5}{2}} + 10^{\frac{1}{2}}(n+\tfrac{1}{2})\\ &\quad - \tfrac{1}{12}(\tfrac{6}{5})^{-\frac{1}{12}}\tau^{\frac{7}{12}}(r_0\gamma)^{-\frac{1}{2}}\\ &\quad - \tfrac{77}{12}(\tfrac{6}{5})^{-\frac{5}{12}}(n^2+n+\tfrac{1}{6})\tau^{-\frac{1}{12}}(r_0\gamma)^{-\frac{5}{2}}\\ &\quad + \tfrac{1}{4}3^{\frac{1}{2}}(n+\tfrac{1}{2})\tau^{\frac{1}{2}}(r_0\gamma)^{-1}\\ &\quad - \tfrac{1349}{1440}(\tfrac{6}{5})^{-\frac{7}{12}}\tau^{\frac{13}{12}}(r_0\gamma)^{-\frac{7}{2}} + O[(r_0\gamma)^{-\frac{5}{3}}],\end{aligned} \tag{78}$$

where

$$\tau \equiv 6mV_0r_0^2\hbar^{-2}. \tag{79}$$

We again replace (60) by (64) and evaluate (58) by steepest descent. The saddle point γ_0, which is the root of $f_n'(\gamma) = 0$, occurs at

$$\begin{aligned}r_0\gamma_0 &= (\tfrac{5}{6})^{\frac{1}{2}}\tau^{\frac{1}{14}}\left(\frac{\pi r_0}{\lambda}\right)^{\frac{12}{7}} + \tfrac{1}{70}(\tfrac{5}{6})^{\frac{1}{2}}\tau^{\frac{1}{2}}\\ &\quad + \tfrac{11}{2}(\tfrac{5}{6})^{\frac{1}{2}}(n^2+n+\tfrac{1}{6})\tau^{-\frac{3}{14}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{8}{7}}\\ &\quad - \tfrac{9}{35}3^{\frac{1}{2}}(n+\tfrac{1}{2})\tau^{\frac{5}{14}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{10}{7}}\\ &\quad + \tfrac{13}{1225}(\tfrac{5}{6})^{\frac{1}{2}}\tau^{\frac{13}{14}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{12}{7}} + O\left[\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{18}{7}}\right].\end{aligned} \tag{80}$$

The saddle-point calculation yields

$$A_n = 4(\tfrac{5}{7})^{\frac{1}{2}}\tau^{\frac{3}{14}}\lambda^3\left(\frac{\pi r_0}{\lambda}\right)^{\frac{15}{7}}h_n\exp(f_n), \tag{81}$$

where

$$\begin{aligned}h_n &= 1 + \left[(n+\tfrac{1}{2})10^{\frac{1}{2}} + \frac{29}{280\pi}\right]\tau^{-\frac{1}{7}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{10}{7}}\\ &\quad - \tfrac{9}{140}\tau^{\frac{3}{7}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{12}{7}} + O\left[\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{20}{7}}\right]\end{aligned} \tag{82}$$

and

$$\begin{aligned}f_n &= -\pi\left\{\tfrac{7}{12}\tau^{\frac{1}{7}}\left(\frac{\pi r_0}{\lambda}\right)^{\frac{10}{7}} + (n+\tfrac{1}{2})10^{\frac{1}{2}}\right.\\ &\quad - \tfrac{1}{12}\tau^{\frac{4}{7}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{2}{7}} - \tfrac{77}{12}(n^2+n+\tfrac{1}{6})\tau^{-\frac{1}{7}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{10}{7}}\\ &\quad + \tfrac{3}{2}10^{-\frac{1}{2}}(n+\tfrac{1}{2})\tau^{\frac{3}{7}}\left(\frac{\pi r_0}{\lambda}\right)^{-\frac{12}{7}} + \tfrac{1}{112}\tau\left(\frac{\pi r_0}{\lambda}\right)^{-2}\\ &\quad \left. + O\left[\left(\frac{\pi r_0}{\gamma}\right)^{-\frac{20}{7}}\right]\right\}.\end{aligned} \tag{83}$$

The present methods can be used for any potential which is more repulsive than r^{-2} as $r \to 0$. Clearly all such potentials will exhibit the exponential suppression of the exchange second virial coefficient at high temperatures, which has been computed explicitly here for hard spheres and for the Lennard-Jones potential.

VI. WIGNER-KIRKWOOD EXPANSION

For potentials with strongly repulsive cores but no sharp corners, such as the Lennard-Jones 6–12, the Wigner-Kirkwood expansion[28–30] in powers of $\hbar^2$ gives a suitable high-temperature expansion of the

[27] The results of applying Rayleigh-Schrödinger perturbation theory to the anharmonic oscillator are given by L. D. Landau and E. M. Lifchitz, *Quantum Mechanics—Non Relativistic Theory* (Addison-Wesley Publishing Company, Inc., Reading. Mass., 1958), p. 136.

[28] E. P. Wigner, Phys. Rev. **40**, 749 (1932).

[29] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).

[30] M. L. Goldberger and E. N. Adams, J. Chem. Phys. **20**, 240 (1952); A. J. F. Siegert, J. Chem. Phys. **20**, 572 (1952); F. Oppenheim and J. Ross, Phys. Rev. **107**, 28 (1957); A. M. Yaglom, Teoriya Veroyatnostei i ee Primeniya **1**, 161 (1956); H. E. DeWitt, Ref. 1.

direct second virial coefficient. The Laplace-transform methods of the present article provide a considerably less laborious means of computing the terms of this expansion. We first derive this expansion in its most general form, beginning with a Hamiltonian

$$H = -(\hbar^2/2M)\nabla^2 + V(\mathbf{r}), \tag{84}$$

where ∇^2 is the Laplacian in n dimensions and $\mathbf{r}$ is an n-dimensional vector. By setting $n = 3N$, we obtain the Hamiltonian for N identical-mass particles in three dimensions. We follow the original method of Kirkwood[29] to the extent that we use the coordinate-space representatives of momentum eigenfunctions for evaluating the partition function

$$\mathrm{Tr}\, e^{-\beta H} = \frac{\Omega}{h^n}\int d^n\mathbf{p}\,\langle\mathbf{p}|\, e^{-\beta H}\,|\mathbf{p}\rangle, \tag{85}$$

where

$$\langle\mathbf{p}|\, e^{-\beta H}\,|\mathbf{p}\rangle = \frac{1}{\Omega}\int d^n\mathbf{r} e^{-i\hbar^{-1}\mathbf{p}\cdot\mathbf{r}} e^{-\beta H} e^{i\hbar^{-1}\mathbf{p}\cdot\mathbf{r}}. \tag{86}$$

Here Ω is the coordinate-space volume. We now employ the Laplace transform and introduce

$$\begin{aligned}\langle\mathbf{p}|\,(H+W)^{-1}\,|\mathbf{p}\rangle &= \int_0^\infty e^{-\beta W}\langle\mathbf{p}|\, e^{-\beta H}\,|\mathbf{p}\rangle \\ &= \frac{1}{\Omega}\int d^n\mathbf{r} e^{-i\hbar^{-1}\mathbf{p}\cdot\mathbf{r}}(H+W)^{-1} e^{i\hbar^{-1}\mathbf{p}\cdot\mathbf{r}}.\end{aligned} \tag{87}$$

Now let (where $\mathbf{p}$ is a c number)

$$\begin{aligned}A &\equiv \mathbf{p}^2/(2M) + V(\mathbf{r}) + W, \\ B &\equiv [\hbar^2/(2M)]\nabla^2 + \mathbf{p}^2/(2M),\end{aligned} \tag{88}$$

so that $H + W = A - B$. Here $\mathbf{p}^2/(2M)$ has been added and subtracted to improve convergence by making $\langle\mathbf{p}|\, B\,|\mathbf{p}\rangle = 0$. We can now use the formal operator expansion

$$(H+W)^{-1} = A^{-1} + A^{-1}BA^{-1} + A^{-1}BA^{-1}BA^{-1} + \cdots \tag{89}$$

to obtain the expansion of $\langle\mathbf{p}|\,(H+W)^{-1}\,|\mathbf{p}\rangle$ in powers of $\hbar$. The labor of working out the terms can be further reduced by exploiting the hermiticity of $i\hbar\nabla$ to do half of the differentiations to the left; thus we obtain

$$\begin{aligned}&\langle\mathbf{p}|\,(H+W)^{-1}\,|\mathbf{p}\rangle \\ &= \frac{1}{\Omega}\int d^n\mathbf{r}\{A^{-1} + [\hbar^2/(2M)] \\ &\quad\times [2A^{-1}M^{-1}(\mathbf{p}\cdot\nabla A^{-1})^2 - (\nabla A^{-1})^2] \\ &\quad + [\hbar^2/(2M)]^2[A^{-1}(\nabla^2 A^{-1} - 2M^{-1}\mathbf{p}\cdot\nabla A^{-1}\mathbf{p}\cdot\nabla A^{-1})^2 \\ &\quad - 2M^{-1}(\nabla A^{-1}\mathbf{p}\cdot\nabla A^{-1})^2] \\ &\quad + [\hbar^2/(2M)]^3[2A^{-1}M^{-1}(\nabla^2 A^{-1}\mathbf{p}\cdot\nabla A^{-1} \\ &\quad + \mathbf{p}\cdot\nabla A^{-1}\nabla^2 A^{-1} + 2M^{-1}\mathbf{p}\cdot\nabla A^{-1}\mathbf{p}\cdot\nabla A^{-1}\mathbf{p}\cdot\nabla A^{-1})^2 \\ &\quad - (\nabla A^{-1}\nabla^2 A^{-1} - 2M^{-1}\nabla A^{-1}\mathbf{p}\cdot\nabla A^{-1}\mathbf{p}\cdot\nabla A^{-1})^2] \\ &\quad + \text{terms odd in } \hbar\mathbf{p} + O(\hbar^8)\}.\end{aligned} \tag{90}$$

Here the terms odd in $\hbar\mathbf{p}$ have been omitted since they drop out anyhow after performing the momentum integration. We now carry out the differentiations, inverse-Laplace-transform back, insert the result in (85), and perform the momentum integrations to obtain, where $U \equiv \beta V$:

$$\begin{aligned}\mathrm{Tr}\, e^{-\beta H} &= \left(\frac{2\pi M}{\beta h^2}\right)^{n/2} \\ &\quad\times\int d^n\mathbf{r} e^{-U(\mathbf{r})}\Big\{1 - \frac{1}{12}\left(\frac{\hbar^2\beta}{2M}\right)(\nabla U)^2 \\ &\quad + \frac{1}{7!}\left(\frac{\hbar^2\beta}{2M}\right)^2\Big[-\tfrac{1}{2}(\nabla U)^4 - 52(\nabla U)^2\nabla^2 U \\ &\quad + 70(\nabla^2 U)^2 + 36(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla U) \\ &\quad - 28\sum_{i,j}(\partial^2 U/\partial x_i\partial x_j)^2\Big] \\ &\quad + \frac{1}{9!}\left(\frac{\hbar^2\beta}{2M}\right)^3\Big[-\frac{61}{11}(\nabla U)^6 \\ &\quad - 120(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla U) \\ &\quad - 32(\nabla U)^2(\nabla^2 U)^2 - 72(\nabla\nabla^2 U)^2 \\ &\quad + \tfrac{816}{11}(\nabla U)^2(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla U) \\ &\quad + \tfrac{288}{11}(\nabla U)^4\nabla^2 U + 12(\nabla U)^2(\nabla U)\cdot(\nabla\nabla^2 U) \\ &\quad + 24(\nabla^2 U)(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla U) \\ &\quad - 96(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla\nabla^2 U) \\ &\quad + 96(\nabla^2 U)(\nabla U)\cdot(\nabla\nabla^2 U) \\ &\quad - 112(\nabla U)^2\sum_{i,j}(\partial^2 U/\partial x_i\partial x_j)^2 \\ &\quad - 114\sum_{i,j,k}(\partial^3 U/\partial x_i\partial x_j\partial x_k)^2 \\ &\quad - 96\sum_{i,j,k}(\partial U/\partial x_i)(\partial U/\partial x_j)(\partial U/\partial x_k) \\ &\quad\times(\partial^3 U/\partial x_i\partial x_j\partial x_k) \\ &\quad + 480\sum_{i,j,k}(\partial U/\partial x_i)(\partial^2 U/\partial x_j\partial x_k) \\ &\quad\times(\partial^3 U/\partial x_i\partial x_j\partial x_k) + O(\hbar^8)\Big]\Big\}.\end{aligned} \tag{91}$$

This expression can be further simplified by partial integration to obtain finally

$$\begin{aligned}\mathrm{Tr}\, e^{-\beta H} &= \left(\frac{2\pi M}{\beta h^2}\right)^{n/2} \\ &\quad\times\int d^n\mathbf{r} e^{-U(\mathbf{r})}\Big\{1 - \frac{1}{12}\left(\frac{\hbar^2\beta}{2M}\right)(\nabla U)^2 \\ &\quad + \frac{1}{2\cdot 6!}\left(\frac{\hbar^2\beta}{2M}\right)^2 \\ &\quad\times[(\nabla U)^4 - 8(\nabla U)^2\nabla^2 U + 12(\nabla^2 U)^2] \\ &\quad + \frac{1}{9!}\left(\frac{\hbar^2\beta}{2M}\right)^3[3(\nabla U)^6 - 40(\nabla U) \\ &\quad\cdot(\nabla\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla U)\end{aligned}$$

$$+ 12(\nabla U)^2(\nabla^2 U)^2 - 216(\nabla\nabla^2 U)^2$$
$$- 50(\nabla U)^4\nabla^2 U$$
$$+ 40(\nabla^2 U)(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla U)$$
$$+ 216(\nabla U)\cdot(\nabla\nabla U)\cdot(\nabla\nabla^2 U)$$
$$+ 264(\nabla^2 U)(\nabla U)\cdot(\nabla\nabla^2 U)] + O(\hbar^8)\Big\}. \quad (92)$$

The terms of order $\hbar^2$ and $\hbar^4$ have been obtained previously,[11,30] but to the author's knowledge, the $\hbar^6$ term has not appeared previously in the literature.[31]

The application of (92) to the computation of the direct second virial coefficient for the Lennard–Jones potential is straightforward and has been given previously[32] through the $\hbar^4$ term. Using the notation of Eq. (72), the result is

$$B_{\text{direct}} = \tfrac{2}{3}\pi N r_0^3 \sum_{n=0}^{\infty}\left(\frac{\hbar^2}{m r_0^2 V_0}\right)^n b_n, \quad (93)$$

where

$$b_0 = -\sum_{n=0}^{\infty}\frac{\Gamma(\frac{1}{2}n-\frac{1}{4})}{4n!}(\beta V_0)^{(2n+1)/4},$$
$$b_1 = \sum_{n=0}^{\infty}\frac{(36n-11)\Gamma(\frac{1}{2}n-\frac{1}{12})}{192\pi^2 n!}(\beta V_0)^{(6n+13)/12},$$
$$b_2 = -\sum_{n=0}^{\infty}\frac{(3024n^2+4728n+767)\Gamma(\frac{1}{2}n+\frac{1}{12})}{30720\pi^4 n!}(\beta V_0)^{(6n+23)/12},$$
$$b_3 = \sum_{n=0}^{\infty}\frac{(53568n^3+303216n^2+491076n+180615)\Gamma(\frac{1}{2}n+\frac{1}{4})}{1146880\pi^6 n!}(\beta V_0)^{(2n+11)/4}. \quad (94)$$

ACKNOWLEDGMENT

The author would like to thank Professor Fritz Oberhettinger for several helpful conversations.

APPENDIX A: PROOF OF THEOREM 1

Theorem 1 is a consequence of the following lemma.

Lemma 1: Let $\bar{g}(W)$ and $\bar{h}(W)$ be the Laplace transforms of functions $g(\beta)$ and $h(\beta)$ which are integrable on (O, R) for any $R < \infty$ and for which there exist constants A, W_0 such that $|g(\beta)| < Ae^{\beta W_0}$ and $|h(\beta)| < Ae^{\beta W_0}$ for $\beta \geq 0$. Assume further that $h(\beta) > 0$ for $\beta > 0$, and that $e^{\delta W}\bar{h}(W) \to \infty$ as $W \to \infty$ for each $\delta > 0$. Then $\lim_{\beta\to 0^+} g(\beta)/h(\beta) = 0$ implies $\lim_{W\to\infty} \bar{g}(W)/\bar{h}(W) = 0$.

Proof of Lemma 1: Decompose $\bar{g}(W)$ as $\bar{g}(W) = \bar{g}_1(W;\delta) + \bar{g}_2(W;\delta)$, where

$$\bar{g}_1(W;\delta) = \int_0^{\delta} e^{-\beta W} g(\beta)\, d\beta$$

and

$$\bar{g}_2(W;\delta) = \int_{\delta}^{\infty} e^{-\beta W} g(\beta)\, d\beta.$$

Now $\lim_{\beta\to 0^+} g(\beta)/h(\beta) = 0$ implies that for each $\epsilon > 0$ there exists a $\delta > 0$ such that $|g(\beta)| < \frac{1}{2}\epsilon h(\beta)$ for $0 \leq \beta \leq \delta$. Then $|\bar{g}_1(W;\delta)| < \frac{1}{2}\epsilon\bar{h}(W)$. Furthermore, for $W \geq W_1 > W_0$ and $\delta \leq \delta_0$ we have $|\bar{g}_2(W;\delta)| \leq ce^{-\delta W}$, where $c = Ae^{\delta_0 W_0}/(W_1 - W_0)$ is an upper bound independent of both W and δ. Then, for each $\delta > 0$, $e^{\delta W}\bar{h}(W) \to \infty$ as $W \to \infty$ implies the existence of a W_1 such that for $W > W_1$, $ce^{-\delta W} < \frac{1}{2}\epsilon\bar{h}(W)$. Hence for each $\epsilon > 0$ we can make $|\bar{g}(W)| < \epsilon\bar{h}(W)$ by taking W sufficiently large, and the result follows.

Proof of Theorem 1: Define

$$g_N(\beta) = f(\beta) - \sum_{n=0}^{N} a_n\varphi_n(\beta).$$

Theorem 1 then follows from letting $g = g_N$ and $h = \varphi_N$ in Lemma 1.

APPENDIX B: PROOF OF THEOREM 3

Theorem 3 is a consequence of the following lemma.

Lemma 2: Let $\bar{g}(W)$ be the Laplace transform of a function $g(\beta)$. Assume that $\bar{g}(W)$ is analytic except on the real axis for $W < W_0$, and that $\bar{g}(W)^* = \bar{g}(W^*)$. Assume further that for some $\alpha > 0$ $\lim_{|W|\to\infty} W^{\alpha+1}\bar{g}(W) = 0$ holds uniformly in $\arg W$ for $|W| \to \infty$ in the right half-plane and uniformly in $\operatorname{Re} W$ for $\operatorname{Im} W \to \infty$ in the left half-plane. Then

$$\lim_{\beta\to 0^+} \beta^{-\alpha} g(\beta) = 0.$$

[31] After the present manuscript had been submitted, the author became aware of a paper of Kihara, Midzuno, and Shizume [J. Phys. Soc. Japan 10, 249 (1955)], in which the Wigner–Kirkwood expansion was carried to order $\hbar^6$ for the special case of the direct second virial coefficient. The terms were evaluated for the Lennard–Jones potential; the present result agrees with theirs. The author is indebted to Dr. S. Y. Larsen and Dr. M. E. Boyd for calling his attention to their work.

[32] J. De Boer and A. Michels, Physica 5, 945 (1938).

Proof of Lemma 2: Consider the Laplace inversion integral

$$g(\beta) = \frac{1}{2\pi i}\int_{W_1-i\infty}^{W_1+i\infty} e^{\beta W}\bar{g}(W)\,dW. \tag{B1}$$

Introduce polar coordinates r, θ, so that $W = re^{i\theta}$, and deform the contour in (B1) to the curve c defined by

$$r = [(\alpha+1)/\beta](\theta/\sin\theta). \tag{B2}$$

Note that the phase of $e^{\beta W}W^{-\alpha-1}$ is zero on c. Writing $dW = dU + i\,dV$ and using $\bar{g}(W)^* = \bar{g}(W^*)$, we then have

$$g(\beta) = \frac{1}{2\pi}\int_c e^{\beta W}\bar{g}(W)\,dV. \tag{B3}$$

Now for each $\epsilon > 0$, there exists an L such that for $|W| > L$ in the right half-plane and for $|\mathrm{Im}\ W| > L$ in the left half-plane

$$|\bar{g}(W)| < \epsilon W^{-\alpha-1}. \tag{B4}$$

Hence for $\beta < (\alpha+1)/L$, (B4) holds on the entire contour c, and

$$|g(\beta)| \le \frac{\epsilon}{2\pi}\int_c |e^{\beta W}|\,|W^{-\alpha-1}|\,dV = \frac{\epsilon\beta^{\alpha}}{\Gamma(\alpha-1)}.$$

Hence $\lim_{\beta\to 0}\beta^{-\alpha}g(\beta) = 0$.

Proof of Theorem 3: Define

$$\bar{g}_N(W) = f(W) - \sum_{n=0}^{N} a_n W^{-\alpha_n-1}.$$

Theorem 3 now follows from applying lemma 2 with $\bar{g} = \bar{g}_N$ and $\alpha = \alpha_N$. Clearly the present method of proof can be extended to a more general set of gauge functions $\bar{\varphi}_n(W)$ if the contour in (B1) can be deformed into a contour on which the phase of $e^{\beta W}\bar{\varphi}_n(W)$ is stationary and on which the asymptotic expansion for large $|W|$ is uniformly valid.

APPENDIX C: PHASE SHIFT FORMULATION

In this Appendix we show how the present formulation can be transformed into the phase-shift formulation given by Gropper and by Beth and Uhlenbeck.[10] For free particles, where G is replaced by G_0, it is easily shown that $B^0_{\mathrm{direct}} = 0$ and $B_{\mathrm{exch}} = \mp 2^{-\frac{5}{2}}\lambda^3 N(2S+1)^{-1}$. In the interacting case, then, Eqs. (3) and (4) imply that

$$B_{\mathrm{direct}} = -2^{\frac{1}{2}}N\lambda^3\Delta_+, \tag{C1}$$

$$B_{\mathrm{exch}} = B^0_{\mathrm{exch}} \mp 2^{\frac{1}{2}}N\lambda^3(2S+1)^{-1}\Delta_-, \tag{C2}$$

where

$$\Delta_\pm \equiv \int G_s(\mathbf{r}, \pm\mathbf{r}; \beta)\,d^3\mathbf{r}. \tag{C3}$$

If we use the inversion integral (56) and the expansion (11) to express $G_s(\mathbf{r}, \pm\mathbf{r}; \beta)$ and perform the integration over $\mathbf{r}$ in (C3), we obtain

$$\Delta_\pm = \sum_{l=0}^{\infty}(\pm 1)^l(2l+1)B_l, \tag{C4}$$

where

$$B_l = \frac{1}{4\pi^2 i}\int_{\gamma_0-i\infty}^{\gamma_0+i\infty}\exp\left(\frac{\lambda^2\gamma^2}{2\pi}\right) \times\left[\int_0^\infty g_s^{(l)}(r, r; \gamma)4\pi r^2\,dr\right]\gamma\,d\gamma. \tag{C5}$$

In (C5), γ_0 is sufficiently large for the contour to lie to the right of any bound-state poles. We now shift the contour to lie ϵ to the right of the imaginary axis, picking up the residues at the bound-state poles with the aid of (17). We put $\gamma = \epsilon + ik$ above the real axis and $\gamma = \epsilon - ik$ below the real axis to get

$$B_l = \sum_n \exp\left[\frac{\lambda^2\gamma_n^2(l)}{2\pi}\right] + \frac{1}{\pi}\int_0^\infty \exp\left(-\frac{\lambda^2k^2}{2\pi}\right)\varphi_l(k)\,dk, \tag{C6}$$

where

$$\varphi_l(k) = ik\lim_{\epsilon\to 0}\left[\int_0^\infty g_s^{(l)}(r, r; \epsilon+ik)r^2\,dr - \int_0^\infty g_s^{(l)}(r, r; \epsilon-ik)r^2\,dr\right]. \tag{C7}$$

We can evaluate the integrals in (C7) by using the radial equation $(\mathfrak{L} - \gamma^2)u = 0$. Denote the solutions by $u_i(r; \gamma)$, where u_1 satisfies the inner boundary condition and u_2 satisfies the outer boundary condition. By integrating both sides of the relation

$$u_1(r; \gamma')\mathfrak{L}u_2(r; \gamma) - u_2(r; \gamma)\mathfrak{L}u_1(r; \gamma') = [\gamma^2 - (\gamma')^2]u_1(r; \gamma')u_2(r; \gamma),$$

one obtains the indefinite integral

$$\int u_1(r; \gamma')u_2(r; \gamma)r^2\,dr = \frac{r^2}{(\gamma')^2-\gamma^2}\left[u_2(r; \gamma)\frac{\partial u_1(r; \gamma')}{\partial r} - u_1(r; \gamma')\frac{\partial u_2(r; \gamma)}{\partial r}\right].$$

By letting $\gamma' \to \gamma$, and discarding an infinite integration constant, it follows that

$$\int u_1(r; \gamma)u_2(r; \gamma)r^2\,dr = \frac{r^2}{2\gamma}\left[u_2(r; \gamma)\frac{\partial^2 u_1(r; \gamma)}{\partial r\partial\gamma} - \frac{\partial u_1(r; \gamma)}{\partial\gamma}\frac{\partial u_2(r; \gamma)}{\partial r}\right]. \tag{C8}$$

We normalize u_1 by imposing an inner boundary condition which is independent of γ. The right-hand side of (C8) then vanishes at the lower limit. We

normalize u_2 by the demand that

$$u_2 \underset{r\to\infty}{\longrightarrow} r^{-\frac{1}{2}}K_{l+\frac{1}{2}}(\gamma r). \tag{C9}$$

For large r,

$$u_1 \underset{r\to\infty}{\longrightarrow} A_l(\gamma)r^{-\frac{1}{2}}I_{l+\frac{1}{2}}(\gamma r) + B_l(\gamma)r^{-\frac{1}{2}}K_{l+\frac{1}{2}}(\gamma r). \tag{C10}$$

Different choices of the γ-independent inner normalization for u_1 multiply A and B by a constant which is *independent of* γ. Evaluating the Wronskian for large r,

$$\Delta(u_1, u_2) = -r^{-2}A_l(\gamma). \tag{C11}$$

Using (15), (C8), (C9), (C10), and (C11), we obtain

$$\int_0^\infty g_s^{(l)}(r, r; \gamma)r^2\,dr = \frac{1}{2\gamma}\frac{A'(\gamma)}{A(\gamma)}, \tag{C12}$$

which is obviously unchanged by different choices of the γ-independent inner normalization. By using the asymptotic forms of the modified Bessel functions, we can replace (C10) by

$$u_1 \underset{r\to\infty}{\longrightarrow} (2\pi\gamma)^{-\frac{1}{2}}r^{-1}\{A_l(\gamma)e^{\gamma r} + [\pi B_l(\gamma) - (-1)^l A_l(\gamma)]e^{-\gamma r}\}. \tag{C13}$$

The phase shift η_l is defined by the statement that for the wave problem in which $\gamma = \pm ik$, the asymptotic form of u_1 is

$$u_1 \underset{r\to\infty}{\propto} \frac{\sin(kr - \frac{1}{2}l\pi + \eta)}{r}. \tag{C14}$$

If we compare (C13) for the cases $\gamma = \epsilon \pm ik$ with (C14), we find that

$$A_l(\epsilon + ik)/A_l(\epsilon - ik) = (-1)^l e^{2i\eta_l}. \tag{C15}$$

Taking the logarithmic derivative of (C15) and using (C7) and (C12) show that

$$\varphi_l(k) = d\eta_l/dk. \tag{C16}$$

We remark at this point that (C16) can also be established by considering the radial equation on (O, R) instead of (O, ∞), where R is much larger than the range of the potential and where $kR \gg l$; for finite R, the continuum is replaced by a discrete set of levels and the representation (17) of $g^{(l)}$ can be used to evaluate (C7).

Inserting (C16) into (C6), integrating by parts, and using $\lambda^2\gamma_n^2(l)/2\pi = -\beta e_n(l)$, where $e_n(l)$ is the bound-state energy, yield

$$B_l = \sum_n \exp[-\beta\epsilon_n(l)] + (\lambda^2/\pi^2)\int_0^\infty \exp\left(-\frac{\lambda^2k^2}{2\pi}\right)\eta_l(k)k\,dk, \tag{C17}$$

which is the result of the usual phase-shift formulation.

APPENDIX D: JUSTIFICATION OF THE WATSON TRANSFORMATION

The differential equation $(\mathfrak{L} - \gamma^2)u = 0$ is analytic in ν and the boundary conditions on u_1 and u_2 are ν independent[20]; hence u_1 and u_2 are analytic functions of ν. Hence $g^{(\nu-\frac{1}{2})}$ is analytic in ν except where bound states occur. Bound states can occur only if $(\mathfrak{L} - \gamma^2)u = 0$ has a turning point; hence the only singularities of $g^{(\nu-\frac{1}{2})}$ for real γ are bound-state poles in the imaginary axis.

Application of the WKB method for radial equations[33] to $(\mathfrak{L} - \gamma^2)u = 0$ [with $\mathfrak{L}$ defined by (14)] yields for the solutions u_1 and u_2 satisfying, respectively, the inner- and outer-boundary conditions

$$u_1 \cong r^{-1}[(\nu/r)^2 + U(r) + \gamma^2]^{-\frac{1}{4}}e^{f(r)}, \qquad u_2 \cong r^{-1}[(\nu/r)^2 + U(r) + \gamma^2]^{-\frac{1}{4}}e^{-f(r)}, \tag{D1}$$

where $\nu = l + \frac{1}{2}$ and

$$f(r) \equiv \gamma r - \int_r^\infty \{(\nu/r)^2 + U(r) + \gamma^2]^{\frac{1}{2}} - \gamma\}\,dr. \tag{D2}$$

In the case $U(r) = 0$, (D1) gives the first term of the Debye expansions. The approximation (D1) is valid if

$$\left|\frac{\partial}{\partial r}[U(r) + (\nu/r)^2]\right| \ll |\gamma^2 + U(r) + (\nu/r)^2|^{\frac{3}{2}}. \tag{D3}$$

The use of (D1) in (15) yields

$$g^{(\nu-\frac{1}{2})}(r, r'; \gamma) \cong \tfrac{1}{2}(r, r')^{-1}[(\nu/r)^2 + U(r) + \gamma^2]^{-\frac{1}{4}} \times [(\nu/r')^2 + U(r') + \gamma^2]^{-\frac{1}{4}}\begin{cases} e^{f(r)-f(r')}; & r \le r', \\ e^{f(r')-f(r)}; & r \ge r'. \end{cases} \tag{D4}$$

Clearly (D3) holds for large γ if ν is not pure imaginary; the estimate (D4) then justifies the replacement of the contour $C_1 + C_3$ by C_4 in (54).

[33] P. M. Morse and H. Feshbach, Ref. 8, pp. 1092–1105, especially p. 1101.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Crossing, Hermitian Analyticity, and the Connection between Spin and Statistics

HENRY P. STAPP
Lawrence Radiation Laboratory, University of California, Berkeley, California

(Received 17 November 1966)

The analytic S-matrix framework is further developed. First some results of earlier works are collected and the physical-region analyticity properties recently derived from macroscopic causality conditions are described. These entail that scattering functions are analytic at physical points except on positive-α Landau surfaces, and that there they are $i\epsilon$ limits of analytic functions from certain well-defined directions, except possibly at certain points where four or more positive-α surfaces intersect. A general $i\epsilon$ rule that also covers these exceptional points is then stated. It is then shown that the scattering function defined by analytic continuation is either symmetric or antisymmetric under interchange of variables describing identical particles and that the sign induced by the interchange is independent of the particular scattering function in which the variables appear. The physical-region analyticity properties of bubble-diagram functions are then derived from the general $i\epsilon$ rule. These functions are products of scattering functions and conjugate scattering functions integrated over physical internal-particle variables, as in the terms of unitarity equations. They are shown to be analytic in the physical region except on Landau surfaces and, more specifically, except on those Landau surfaces that correspond to Landau diagrams that are supported by the bubble diagram in question, with the further restriction that the Landau α's must be positive or negative for lines lying within positive or negative bubbles, respectively. Also, the basic rule for continuation around these singularities is derived. A new general derivation of the pole-factorization theorem is given, which is based on slightly weaker assumptions than earlier proofs. Particular attention is paid to the over-all sign. A general derivation of the crossing and Hermitian analyticity properties of scattering functions is then given. On the basis of the deduced general rules for constructing the paths of continuation that connect the crossed and Hermitian conjugate points, the various related points are found to be boundary values of a single physical sheet. In particular, a certain sequence of continuations is shown to take one back to the original point. From this fact it follows that abnormal statistics are incompatible with simultaneous unitarity in both the direct and crossed channels. The proof given here does not depend on the notion of interchange of variables other than those of identical particles. Earlier proofs depended on the unphysical notion of interchange of variables representing conjugate particles. Finally it is shown that the analytically continued M functions with normal-ordered variables are precisely the scattering functions: no extra signs are needed or permitted. Aside from the general i rule, the analyticity assumptions are these: (1) The discontinuity around a singularity of a bubble diagram B has no residue at a physical-particle mass value (in an appropriate variable) unless the singularity corresponds to a diagram that is supported by B and has the single-particle-exchange form that corresponds to a pole at that mass value. (2) The residue just described has the pole-factorization property. (3) Confluences of infinite numbers of singularity surfaces do not invalidate the results established by assuming that this number is locally finite. Assumption (1) entails that all relevant singularities of scattering functions lie on Landau surfaces. That is the basic assumption.

1. INTRODUCTION

An earlier S-matrix proof of the normal connection between spin and statistics given by this author[1] depended on an assumption that self-conjugate combinations of particle and antiparticle amplitudes were in principle observable. The assumption is objectionable because it has no experimental basis in the case of charged particle and in fact conflicts with a conjectured superselection rule.[2]

In that original paper, the beginning of a second proof, not depending on this special assumption, was also given.[3] This alternative proof depended on an apparent conflict between abnormal statistics and the crossing and Hermitian analyticity properties of scattering functions. The crossing property of (multiparticle) scattering functions is the property whereby the scattering function describing one reaction is connected by analytic continuation to the scattering function describing certain other reactions, called crossed reactions. The Hermitian analyticity property is the property whereby the scattering function representing a given process is analytically connected to the complex conjugate of the scattering function for the transposed process at certain real boundary points.

This second argument was not a full proof. In the first place, the statistics involved was the sign change under interchange of variables describing relative antiparticles, whereas the spin-statistics connection involves the sign change under interchange of two variables describing particles of the same type (identical particles). And in the second place, the required

[1] H. P. Stapp, Phys. Rev. **125**, 2139 (1962).
[2] G. C. Wick, A. S. Wightman, and E. Wigner, Phys. Rev. **88**, 101 (1952).
[3] Reference 1, Appendix I. There, crossing was assumed to hold without any extra phase factors, but the no-scattering parts were allowed to be different from unity.

properties of crossing and Hermitian analyticity were not derived, but simply assumed.

The first defect was partially remedied in a later paper,[4] where it was shown that a certain assumption on the phase factors occurring in the cluster decomposition expansion implied that the sign change under interchange of identical-particle variables was the same as the sign change under the interchange of conjugate-particle variables. This assumption on the phases was that they be such that the disconnected contributions to a unitarity equation be equivalent to a product of the unitarity equations in the various disconnected sectors considered by themselves. This assumption, although reasonable, is replaced in the present work by direct physical requirements.

The main object, however, of the present paper is to give proofs of the crossing and Hermitian analyticity properties. The work is a development of a line of approach initiated by Gunson[5] and explored by Olive[6,7] and is based on an exploitation of the pole-factorization property of scattering functions. This is the property whereby the residues of poles of scattering functions at certain points in the physical regions of multiparticle processes are given essentially as the products of scattering functions for certain other reactions involving fewer particles. More specifically, the work is an elaboration of an unpublished work of this author[8] in which the procedure, later adopted also by others,[7,9] was developed whereby the paths of continuation to crossed (or Hermitian conjugate) points are defined by distorting certain paths, originally lying in the physical region of the larger process and running between different parts of the pole manifold $S_\nu = \mu_p^2$, into paths lying completely within the pole manifold $S_\nu = \mu_p^2$. These latter paths specify the mass-shell continuations between the relevant points. The present work goes beyond earlier attempts[7,9] to exploit this idea in that it covers all possible reactions and all Landau singularities. The work of Ref. 7 treated only the simplest reactions and ignored all but the trivial normal-threshold singularities, and the work of Ref. 9 referred to a simple reaction with special mass ratios.

Having established the required crossing and Hermitian analyticity properties, we give a new version of the remainder of the proof of the normal connection between spin and statistics. This new version is more simple and direct than the one given earlier.[3,4] It avoids completely the introduction of the notion of a phase change induced by an interchange of conjugate variables (variables that refer to relative antiparticles). The need to introduce this notion was a disagreeable feature of the earlier proofs, for this phase change, unlike the sign change under the interchange of like variables (variables referring to identical particles), apparently has no direct physical significance. And in order to deal with the interchange of unlike variables, certain stipulations had to be introduced relating the phases in the cluster decomposition equation to special orders of variables. The present proof circumvents these difficulties.

This new proof of spin and statistics, which is given in Sec. 7, is largely independent of the details of the work preceding it. That section is therefore designed to be largely self-contained. It depends, in fact, only on a very gross feature of the proof of the pole-factorization theorem and on a rather trivial-sounding property of the paths connecting crossed reactions and Hermitian conjugate points.

This property is that the relationship of Hermitian conjugateness is maintained when the two related functions are continued to their respective crossed-reaction points; the abnormal connection between spin and statistics would demand that these two continuations lead to functions differing by a sign, which is just the trouble noted in Ref. 3. The proving of this analyticity property is the main object of the earlier sections.

The proof of spin and statistics given in the present paper resembles one recently proposed by Lu and Olive[10,11] in that both hinge on the analyticity property just mentioned. However, this property was simply assumed by Lu and Olive. Also, their proof, which seems more complicated than the present one, makes use of the notion of interchange of unlike variables and is based, like the one in Refs. 3 and 4, on a special stipulation relating the phases in the cluster-decomposition equation to particular orders of variables.

[4] H. P. Stapp, "The Decomposition of the S Matrix and the Connection Between Spin and Statistics," Lawrence Radiation Laboratory Report UCRL-10289, 1962 (unpublished).

[5] J. Gunson, Birmingham Preprint (1962) and J. Math. Phys. **6**, 827 (1965).

[6] D. Olive, "Towards an Axiomation of S Matrix Theory" (July 1963) (preliminary version of Ref. 7.).

[7] D. Olive, Phys. Rev. **135**, B745 (1964).

[8] H. P. Stapp, "Seminars in Problems in S-Matrix Theory," August 1963 (privately circulated). That Gunson's method might be used to justify an effective continuation in mass, which is the central idea of both this reference and of the present paper, was suggested earlier in the *Proceedings of the 1963 Midwest Conference on Theoretical Physics* (Notre Dame, May–June 1963). The main difficulty in achieving this is to take all possible Landau singularities into account.

[9] J. B. Boyling, Nuovo Cimento **33**, 1356 (1964).

[10] E. Y. C. Lu and D. I. Olive, Nuovo Cimento **45A**, 205 (1966).

[11] David Olive (private communication). The proof of spin and statistics given in Sec. 7 was constructed upon receiving this communication, in which Olive stated that he and Lu had obtained a proof based on his version of the crossing argument.

Because the present paper is long and much of it is devoted to establishing rather fine points, it will be useful to give an outline of the contents, with emphasis on the nature of the main problems being tackled.

Section 2 contains a resume of results needed from earlier works. The first eleven subsections give the basic formulas of the M-function formalism developed in Refs. 1 and 12. A principal reason for developing that formalism was to provide a framework for a satisfactory proof of the normal connection between spin and statistics. That is, in order to obtain this result as a statement with physical relevance, it seems necessary, in the present uncertain situation, to clear away abstract formal assumptions and to base the framework directly on assumed fundamental physical relationships. Then the conclusions cannot be altered by any adjustment of formal connections in a way that still preserves these basic physical connections.

A statement of the cluster-decomposition principle is given in Sec. 2L, and in Sec. 2M a "scattering function" is defined to be a connected part of an M function divided by the conservation δ function.

In Secs. 2N–2Q a description is given of the physical-region analyticity properties of scattering functions that have recently been derived within the S-matrix framework from a *macroscopic* causality condition.[13] What has been derived is the location of all possible physical-region singularities, and the rules for continuing around these singularities in such a way as to arrive at the physical function lying on the other side of the singularity manifold. The possible physical region singularities are, in fact, confined to positive-α Landau surfaces, and the paths of continuation about these singularities (the $i\epsilon$ rule) can be explicitly constructed from a knowledge of the corresponding Landau diagram.

The final six subsections of Sec. 2 are devoted to proving that the scattering functions must be either symmetric or antisymmetric under the interchange of like variables (variables representing identical particles). Moreover, the sign (change) under interchange of a particular type of variable is a universal quantity that is independent both of the particular scattering function in which the variables appear and of the particular location of the variables among the arguments of these functions.

The fact that parastatistics is precluded here is a direct consequence of our basic assumption[1] that the observables of the theory are squares of amplitudes. This is not true in parastatistics models. Thus the work of these subsections is not to be construed as a general disproof of parastatistics but rather as a proof that, within the framework adopted, in which observables are squares of amplitudes, the continuation of the scattering function through a region near physical points from an original region of definition to some region where like variables are interchanged must give back the original function, apart from a sign that depends only on the type of particle involved.

Having established that interchange of like variables leads to a sign (change) σ_p that depends only on the type of particle p, one may then ask what the value of this sign is. The normal connection between spin and statistics is the relationship

$$\sigma_p = (-1)^{2j_p},$$

where j_p is the spin of particle p.

Section 3 is devoted to the proof of some physical-region analyticity properties of functions represented by bubble diagrams. These functions are functions of the kind occurring in unitarity equations and are formed by integrating products of scattering functions and their conjugates over the physical phase space associated with certain internal particles. The singularities of any such function are shown to be confined to a certain corresponding subclass of Landau surfaces. The singular parts of the surfaces are not always the positive-α parts, however. The rules that determine which parts of the Landau surfaces are singular and the $i\epsilon$ rules for continuing around these surfaces are derived.

The fact that the singularities of these functions are confined to Landau surfaces is a result somewhat similar to one obtained by Polkinghorne.[14] The result of Polkinghorne does not refer specifically to the physical region, however, and the possibility of non-Landau (i.e., second-type) singularities arises. The $i\epsilon$ rules derived here for detouring around physical-region singularities generalize results about physical-region singularities recently obtained by Landshoff and Olive.[15]

Section 4 is devoted to a general proof of the pole-factorization theorem. The original S-matrix proof by Olive[7] was for a simple case and was based on an assumption (called a theorem) that has recently been shown by Branson[16] to be not valid in general.

[12] H. P. Stapp, "The Analytic S-Matrix Framework," in "The Trieste Lectures," *High Energy Physics and Elementary Particles* (IAEA, Vienna, 1965).

[13] C. Chandler and H. P. Stapp, URCL-17734.Certain infinitely differential singularities, not excluded by this work, are ruled out by an extension due to Iagolnitzer and Stapp (in preparation).

[14] J. C. Polkinghorne, Nuovo Cimento **25**, 901 (1962).

[15] P. V. Landshoff and D. I. Olive, J. Math. Phys. **7**, 1464 (1966). See note Ref. 12.

[16] David Branson, Nuovo Cimento **44A**, 1081 (1966).

Branson's counterexample also contradicts an assumption made in an earlier proof by this author.[17] That assumption was that almost all singularities lying on the "pole manifolds" $S_\nu = \mu_p^2$ are associated with "pole-type" (i.e., one-particle-exchange-type) Landau diagrams, these being Landau diagrams that can be reduced by contraction to connected Landau diagrams having just two vertices and just a single internal line connecting these two vertices. Branson has shown that other types of singular Landau surfaces can lie at $S_\nu = \mu_p^2$. Our earlier assumption is therefore weakened to the assertion that almost all *pole* (or worse) singularities lying on the manifold $S_\nu = \mu_p^2$ are associated with pole-type Landau diagrams. (That the assumption in Ref. 17 should be weakened in this way was already suggested there.)

This "pole assumption" is a fundamental assumption in the present work. It is believed that it can be verified by an examination of the nature of the possible Landau singularities lying at $S_\nu = \mu_p^2$, but this verification is not attempted here.

The derivation of the pole-factorization theorem given in Sec. 4 is different from the one given in Ref. 17. The form given here is useful because essentially the same technique can be used to derive the general normal-threshold discontinuity equation, as will be discussed in a later paper. Also, the present derivation is given in greater detail than the earlier one and covers particles with spin (a trivial extension in the M-function formalism). More importantly, the phase factors occurring in the cluster decomposition are taken into account. These will play an important role in the discussion of spin and statistics.

Section 5 contains a proof of the Hermitian analyticity property. The "pole assumption" is again fundamental, and now it is extended to points lying outside the physical region. The essential idea of the proof is to consider a larger process from which the scattering function of interest can be extracted as a factor in the residue of a product of poles.[5–9] The unitarity equation for the larger process at a point corresponding to null energy–momentum vectors for the reaction of interest is effectively continued to the pole position by exploiting the fact that most contributions to the larger process do not contribute to the residue.

The arguments in Sec. 5 deal individually with individual Landau surfaces. There is a tacit assumption that results that hold for the Landau surfaces individually will hold for them collectively. In particular, it is assumed that no natural boundaries formed from confluences of infinite numbers of Landau surfaces invalidate results that are valid when the surfaces are considered individually.

In Sec. 6 the reader is first referred to the proof of crossing given in Ref. 17. That proof is then extended in such a way as to obtain a compatibility condition on the paths of continuation connecting crossed and Hermitian conjugate points. This compatibility condition, which says essentially that the Hermitian conjugate points for crossed reactions are connected by the complex conjugate of the crossing path, plays a key role in the proof of spin and statistics. It is also shown that this compatibility requirement carries over to the case in which the paths of continuation jump across various cuts, rather than detouring around them, provided the pole-factorization property carries over to the relevant discontinuity functions, as it would do if these functions were given by a Cutkosky rule.

The proof of the normal connection between spin and statistics is given in Sec. 7. It is quite simple. It is noted that the residue functions in the pole-factorization property were obtained from a corresponding pole term in the unitarity equation for the process in which the pole appears and that the phase factor in the residue formula is, consequently, precisely the phase factor of this contribution to unitarity. In special cases, this contribution to unitarity is just one of the absolute-value - squared contributions to a "forward scattering" process, apart from some signs coming from interchanges of certain identical particles. Thus the phase factors in these pole contributions are determined by the statistics of certain particles. The residue functions associated with crossed reactions, which are connected by analytic continuation, are compared and shown to have a sign incompatibility in case any scattered particle has abnormal statistics.

In Sec. 8 the phase factor in the crossing relation is shown to be unity for the functions $M_c(K)$: these particular functions, without any added phase factors, give, when continued along the paths of continuation connecting crossed regions, the scattering functions for the various crossed process.

To obtain this result, a special stipulation relating phases in the decomposition equation to order of variables is invoked. This stipulation is, in effect, used in the proof of spin and statistics given by Lu and Olive. Since this stipulation is of a formal rather than physical nature, it renders that proof, like the proof of Ref. 4, not completely satisfactory from the pure S-matrix viewpoint.

[17] H. P. Stapp, "Antiparticles in S-Matrix Theory," in "The Trieste Lectures," *High Energy Physics and Elementary Particles* IAEA, Vienna, 1965).

This stipulation, although objectionable as an element of a proof of spin and statistics, is quite natural and is adopted in the final specification of the formalism. It eliminates an indeterminate factor in the crossing relations and also places conditions on the phases induced by interchange of variables associated with different particles. In particular, it implies, as is shown in Sec. 8, that the interchange of adjacent conjugate variables induces the same sign change as the interchange of the corresponding like variables.

Note added in proof: It has now been proven in an excellent work by Froissart and Taylor (Princeton University Preprint) that this special stipulation can be derived from physical requirements. Also, some results similar to the structure theorems of Sec. 3 have now been derived by Landshoff, Olive, and Polkinghorne (Cambridge University Preprint).

2. GENERAL FORMALISM

Certain basic definitions and fundamental results from earlier works will be summarized in this section.

A. The S Functions

Scattering processes are described by functions $S(K'; K'')$, where K' and K'' are sets of variables of the form

$$K' \equiv \{p_i', m_i', t_i'\}$$

and

$$K'' \equiv \{p_i'', m_i'', t_i''\},$$

describing the final and initial particles, respectively. The index t_i is a positive integer that specifies the type of particle—electron, proton, positron, etc. The p_i is the physical energy–momentum vector of particle i and is constrained by the mass condition

$$p_i^2 \equiv (p_i^o)^2 - (\mathbf{p}_i)^2 = \mu^2(t_i) \equiv \mu_i^2 > 0. \quad (2.1)$$

The masses μ_i are assumed positive in this work. The m_i are positive and negative integers or half-integers, called spin indices, ranging in unit steps from $j(t_i) \equiv j_i$ to $-j_i$:

$$j(t_i) \equiv j_i \geq m_i \geq -j_i. \quad (2.2)$$

B. Unitarity

The S functions satisfy the unitarity equations

$$\sum_K S(K'; K)S^*(K''; K) = \delta(K'; K'') \quad (2.3)$$

and

$$\sum_K S(K; K'')S^*(K; K') = \delta(K'; K''), \quad (2.4)$$

where the asterisk denotes complex conjugate and the summation over K means a summation over the discrete spin and particle-type labels and an integration with positive weight factor over the mass-shell momentum–energy vectors p_i. The weight factor $\rho(K)$ can be taken to be any positive function; different definitions of $\rho(K)$ lead to different S, but observables, which involve always an integration over p_i, are independent of $\rho(K)$. The δ function is defined by the properties

$$\begin{aligned}\sum_K \delta(K'; K)\delta(K''; K) &= \delta(K'; K'') \\ &= \delta(K''; K') \\ &= \delta^*(K''; K') \\ &= 0 \quad \text{for} \quad K'' \neq K'. \end{aligned} \quad (2.5)$$

C. M Functions

Covariance properties are most conveniently expressed in terms of functions

$$M(K'; K'') \equiv L(V')S(K'; K'')L(V''), \quad (2.6)$$

where

$$L(V) = \prod_i L_i(v_i). \quad (2.7)$$

The $L_i(v_i)$ is a matrix in the spin space of particle i that transforms spin functions from values coordinated to a rest frame Σ_i of particle i to values coordinated to the general coordinate frame Σ. For a particle of spin $\frac{1}{2}$,

$$\begin{aligned} L_i^j(v_i) &\equiv L_i^{\frac{1}{2}}(v_i) \\ &\equiv (\sigma_i v_i)^{\frac{1}{2}} \\ &\equiv (\sigma_{i_\mu} v_i^\mu)^{\frac{1}{2}} \\ &\equiv (v_i^o + \boldsymbol{\sigma} \cdot \mathbf{v}_i)^{\frac{1}{2}} \\ &\equiv \cosh\frac{\alpha_i}{2} + \frac{\left(\boldsymbol{\sigma} \cdot \mathbf{v}_i \sinh\frac{\alpha_i}{2}\right)}{|\mathbf{v}_i|} \\ &\equiv (v_i^o + 1 + \boldsymbol{\sigma} \cdot \mathbf{v})(2v_i^o + 2)^{-\frac{1}{2}}, \end{aligned} \quad (2.8)$$

where v_i is the covariant velocity vector

$$v_i = \frac{p_i}{|(p_i \cdot p_i)^{\frac{1}{2}}|} \quad (2.9)$$

and

$$\sinh \alpha_i \equiv |\mathbf{v}_i|. \quad (2.10)$$

The matrix $L_i^j(v_i)$ for particles of spin $j > \frac{1}{2}$ is obtained by extracting, by means of Clebsch–Gordan coefficients, the spin-j part of a tensor product of $2j$ factors $L^{\frac{1}{2}}(v_i)$. (See Eq. C14 of Ref. 12.)

The rule for contraction of the spin indices in Eq. (2.6) is not always the matrix rule of contraction of adjacent indices. The exact rule is given in Sec. 2H below.

D. Covariance Property

From the assumed relativistic invariance of probability correlations one derives the covariance

property

$$M(K'; K'') = \Lambda_s^{-1} M(\Lambda K'; \Lambda K''), \tag{2.11}$$

where

$$\Lambda K = \{\Lambda p_i, m_i, t_i\}. \tag{2.12}$$

Here Λ is any element of the real, orthochronous, proper, homogeneous Lorentz group $L_+^\uparrow$, and Λ_s is a corresponding p_i-independent spin-space transformation (see Ref. 12). To obtain Eq. (2.11), the weight factor $\rho(k)$ has been taken to have covariant form. In particular, we take

$$\sum_K = \Sigma' \int dK = \Sigma' \prod_i \int \frac{d^4 p_i}{(2\pi)^4} \theta(p_i^0) 2\pi \delta(p_i^2 - \mu_i^2), \tag{2.13}$$

where Σ' is the sum over discrete indices.

E. Physical Irrelevancy of Order of Variables

The experimental result labeled by the set $K = \{p_i, m_i, t_i\}$ is assumed to be completely specified by the values of the arguments p_i, m_i, and t_i. In particular, no additional information having to do with the ordering of the variables is needed to determine the experimental result. This assumption, in conjunction with our quantum postulate of Ref. 17, means that in the integration (2.13) one should include only once the contribution from each value of K, considered as an *unordered* set of variables.

F. Fundamental Analyticity Property of M Functions

We introduce the following definitions.

Definition 2.1: A function $F(p_i)$, defined only over a subset W of the space of complex numbers p_i, will be said to be an analytic function of the p_i at point P of W if and only if, for every mapping $p_i(z_j)$ from an open set in a space of complex numbers z_i into W, the functions $F'(z_j) \equiv F(p_i(z_j))$ are analytic functions of the z_j, in the usual sense, at all points $\{\bar{z}_j\}$ satisfying $\{p_i(\bar{z}_j)\} = P$ for which the functions $p_i(z_i)$ are analytic at $\{\bar{z}_j\}$.

Definition 2.2: A point K is a set $\{p_i, t_i\}$. It is distinguished from an argument $K = \{p_i, m_i, t_i\}$.

Definition 2.3: "*$F(K)$ is analytic at a point K*" will mean that the functions $F(K)$ corresponding to the various values of the spin indices m_i are all analytic functions of the momentum–energy vectors p_i at point K of W, where W is the set of points K over which $F(K)$ is defined.

A consequence[18–20] of the covariance property (2.11) is that, if $M(K)$ is analytic at a point $\bar{K}$, then $M(K)$ can be extended to a function that is analytic at all points K of the set generated from $\bar{K}$ by application of any element of the proper homogeneous complex Lorentz group $L_+^\uparrow(C)$, which is the subgroup of $L(C)$ continuously connected to the identity. The property of $M(K)$ being analytic at a point K is therefore Lorentz-invariant. This is not the case for the S functions, since the functions $L(V)$ have singularities whose positions depend on the coordinate frame, as is seen from (2.8). The M functions have, in this sense, simpler analyticity properties and are the more convenient functions to use in a relativistic theory based on analyticity.

G. Expressions for Observables

It is advantageous to express observables directly in terms of M functions, rather than passing to S functions, for in this way manifest covariance is maintained. Let s_i be the spacelike four-vector satisfying

$$s_i \cdot s_i = -1 \tag{2.14a}$$

and

$$s_i \cdot p_i = 0 \tag{2.14b}$$

that specifies the axis relative to which the spin-quantum number m_i is measured. Equation (2.14b) says that s_i is purely spacelike in any rest frame of particle i. With M functions one uses, in place of the usual spin-projection operators $P_i(m_i, s_i)$, rather the covariant-spin operators $\tilde{P}_i(m_i, -s_i, v_i)$. For spin-$\frac{1}{2}$ particles

$$\begin{aligned}\tilde{P}_i^{\alpha\beta}(m_i, -s_i, v_i) &= (\tfrac{1}{2} v_i - m_i s_i) \cdot \tilde{\sigma}_i^{\alpha\beta} \\ &= (\tfrac{1}{2} v_i^\mu - m_i s_i^\mu) \tilde{\sigma}_{i\mu}^{\alpha\beta},\end{aligned} \tag{2.15}$$

where

$$\tilde{\sigma}_\mu = (1, -\boldsymbol{\sigma}) \tag{2.16}$$

and $\boldsymbol{\sigma}$ is the usual Pauli spin-matrix vector. Notice that the $\tilde{P}_i(m_i, -s_i, v_i)$ in (2.15) reduces the usual $j = \frac{1}{2}$ projection operator in a rest frame of particle i. For $j > \frac{1}{2}$ the $\tilde{P}_i(m_i, -s_i, v_i)$ is obtained by extracting, by means of Clebsch–Gordan coefficients, the spin-j part of the symmetrized tensor product of $2j$ spin-$\frac{1}{2}$ spin operators $\tilde{P}_k(m_k, -s_i, v_i)$ subject to

$$\Sigma m_k = m_i.$$

[18] A proof of this fact is given in Ref. 8 and also in Ref. 19. A proof explicitly making use of the definition (2.1) of analyticity is given in Ref. 20.

[19] Peter Minkowsky, David N. Williams, and Rudolf Seiler, in *Proceedings of the Symposium of the Lorentz Group, Seventh Annual Summer Institute for Theoretical Physics, 1964* (University of Colorado, Boulder, Colorado), Lemma A.

[20] H. P. Stapp, *1963 Madras Lectures on Analytic S-Matrix Theory* (Matscience, Madras, India, 1963), p. 72.

The $2j$ upper-dotted (undotted) indices of the $\tilde{P}_k^{\alpha\beta}$ are combined to give the $(2j+1)$-valued upper-dotted (undotted) index of $\tilde{P}_i^{\alpha\beta}$. (More details are given in Ref. 12.)

Each spin index m_i of $M(K';K'')$ and $M^*(K';K'')$ is defined to be lower-dotted or lower-undotted, according to whether it is contracted in the calculation of observables with an upper-dotted or upper-undotted index of $\tilde{P}_i^{\alpha\beta}$. Define, accordingly, a quantity λ_i:

$$\lambda_i = +1 \tag{2.17a}$$

if the spin index m_i of $M(K';K'')$ is undotted;

$$\lambda_i = -1 \tag{2.17b}$$

if the spin index m_i of $M(K';K'')$ is dotted.

Define ϵ_i by

$$\epsilon_i = +1 \tag{2.18a}$$

if particle i is final;

$$\epsilon_i = -1 \tag{2.18b}$$

if particle i is initial.

Then m_i is the projection of physical spin-angular momentum on the "physical" spin direction

$$s_i^{\text{phys}} = \epsilon_i \lambda_i s_i, \tag{2.19}$$

where s_i is the "mathematical" spin vector s_i appearing in $\tilde{P}_i(m_i, -s_i, v_i)$. The result (2.19) follows from the covariance property (2.11) and the requirement that spin-angular momentum plus orbital-angular momentum be a conserved quantity (see Ref. 12).

In certain other formalisms, the initial and final particles are associated with kets and bras, respectively, and one always gets $\epsilon_i\lambda_i = +1$. This special condition does not occur naturally in the development of the M-function formalism from basic physical postulates, and it is advantageous not to introduce it. For in the development of the theory, we shall be led to analytically continue our functions to regions where ϵ_i is reversed. Under analytic continuation, the transformation property (2.11), hence the index type and hence λ_i, necessarily remains unaltered. To resolve this conflict with a condition $\epsilon_i\lambda_i = +1$, new functions would have to be introduced into the theory. This unneccessarily complicates the formalism and leads to possible phase ambiguities. It therefore is better never to introduce the artificial condition $\epsilon_i\lambda_i = +1$. Then a single function will describe both the direct and cross reactions. However, the connection between the physical spin vector s_i^{phys} and the mathematical spin vector s_i will be reversed under continuation to crossed channels. This relationship between the physical and mathematical spin vectors is completely analogous to the one that will be obtained for the momentum–energy vectors.

H. Contraction Rule in the Definition of $M(K';K'')$

In the general development of the theory, the M functions are originally defined by their connection to observables through contractions with the covariant spin operators, and Eq. (2.6) emerges as a consequence. The index m_1 of $S(K';K'')$ turns out to be contracted with the adjacent index of $L(V')$ or $L(V'')$ if $\lambda_i\epsilon_i = +1$, and the nonadjacent index otherwise. The λ_i can be specified at will by specifying the index of $\tilde{P}^{\alpha\beta}$ with which the index m_i of $M(K';K'')$ is contracted in the calculation of observables.

I. Unitarity for M Functions

One can specify the conventions for λ_i so that the $\epsilon_i\lambda_i$ for each individual particle, whether occurring initially or finally, is a fixed sign depending only on the particle type. (This specification relates initial particle to final particle—not to final antiparticle; the crossing concept is not involved.) If the λ_i are specified in this way, then unitarity takes the form

$$\sum_K M(K';K)\tilde{G}(V)M^*(K'';K) - G(V')\delta(K';K'') = 0, \tag{2.20}$$

where

$$\tilde{G}(V) = \prod_i \tilde{G}_i(v_i) = [G(V)]^{-1} \tag{2.21}$$

and

$$\tilde{G}_1(v_i) = [L_i(v_i)]^{-2}. \tag{2.22}$$

For spin-$\frac{1}{2}$ particles,

$$\tilde{G}_i(v_i) = v_i \cdot \tilde{\sigma} = v_i^o - \mathbf{v}\cdot\boldsymbol{\sigma}, \tag{2.23}$$

while for spin $j > \frac{1}{2}$ the $\tilde{G}_i(v_i)$ is obtained by extracting by means of Clebsch–Gordan coefficients the spin-j part from a tensor product of $2j$ spin-$\frac{1}{2}$ matrices (2.23). Thus $\tilde{G}_i(v_i)$ is of degree $2j_i$ in the vectors v_i and

$$\tilde{G}_i(-v_i) = (-1)^{2j_i}\tilde{G}_i(v_i). \tag{2.24}$$

J. Momentum–Energy Conservation

The M functions are nonzero only at points satisfying

$$\Sigma p_i' = \Sigma p_i''. \tag{2.25}$$

This conservation-law constraint is equivalent to the statement of translational invariance if space and time are introduced by Fourier transformation. In order to give finite effects, an M function must have a conservation-law δ function $(2\pi)^4\delta(\Sigma p_i' - \Sigma p_i'')$ as a factor.

K. $M(K)$ Functions

We define $M(K)$ (without the semicolon) by

$$M(K) = M(K', -\tilde{K}'') = M(K';K''), \tag{2.26}$$

where

$$(-\tilde{K}'') \equiv \{-p_i, m_i, -t_i\}. \tag{2.27}$$

For later convenience, the order of writing the variables of $(-\tilde{K}'')$ is reversed relative to K''. (See Sec. 2R below and Sec. 8.) The momentum–energy arguments of $M(K)$ will be called the "mathematical" momentum–energy vectors k_i, where

$$k_i = \epsilon_i p_i, \tag{2.28}$$

and ϵ_i is $+1$ or -1 according to whether particle i is final or initial. In terms of the k_i the momentum–energy conservation law δ function becomes

$$(2\pi)^4\delta^4(\Sigma k_i) \equiv (2\pi)^4\delta^4(\Sigma p_i' - \Sigma p_i''). \tag{2.29}$$

L. Cluster Decomposition

$M(K)$ is assumed to satisfy the cluster property[4]

$$M(K) = \sum_p M_p(K), \tag{2.30a}$$

where

$$M_p(K) = \alpha_p \prod_s M_1(K_{ps}). \tag{2.30b}$$

Here K_{ps} is the sth subset of the pth partition of K. The first (and only) subset of the first partition of K is K itself,

$$K_{11} \equiv K, \tag{2.30c}$$

and the function $M_1(K)$ is asserted to have no conservation-law δ function aside from the over-all one given by (2.29). The α_p are phase factors depending on the ordering of variables of K and of the K_{ps} but not on the values of the momentum–energy arguments k_i. The phase of $M_1(K)$ is defined by

$$\alpha_1 = 1. \tag{2.30d}$$

The other phase factors α_p must evidently depend on the orders of the variables. They are asserted to be restricted by the following two conditions.

E2: Let K_{ps} be an ordered set of variables consisting of the variables of the sth subset of the pth partition of the ordered set K. Let $\tilde{K}$ be some ordered subset of the variables of K. Suppose there are two partitions $p = a$ and $p = b$ of $\tilde{K}$, and also two partitions $p = a$ and $p = b$ of the set K, such that these two partitions of K coincide with the corresponding partitions of $\tilde{K}$ over the set $\tilde{K}$ and coincide with each other over the remaining variables. That is, for some arrangement of the indices s,

$$\tilde{K}_{as} = K_{as} \quad \text{for} \quad s \leq \tilde{s}_a, \tag{2.30e}$$

$$\tilde{K}_{bs} = K_{bs} \quad \text{for} \quad s \leq \tilde{s}_b, \tag{2.30f}$$

$$K_{as} = K_{bs'} \quad \text{for} \quad s - \tilde{s}_a = s' - \tilde{s}_b > 0, \tag{2.30g}$$

where $\tilde{s}_p$ is the number of terms of partition p of $\tilde{K}$. Then the α_p and $\tilde{\alpha}_p$ in the cluster decompositions of K and $\tilde{K}$ satisfy

$$\tilde{\alpha}_a/\tilde{\alpha}_b = \alpha_a/\alpha_b. \tag{2.30h}$$

E3: Let a and b denote two initial sets of particles and let c and d denote two final sets. Suppose each of these four sets is divided into n subsets. Suppose the first $n-1$ subsets of set a are identical to the first $n-1$ subsets of set b and the first $n-1$ subsets of set c are identical to the first $n-1$ subsets of set d. Let ac, ad, bc, and bd be values of p that denote the partitions of the four sets of variables $a+c$, $a+d$, $b+c$, and $b+d$, respectively, into n subsets, with the first initial subset of a or b grouped with the first final subset of c or d, etc. Then the four α_p satisfy

$$\frac{\alpha_{ac}}{\alpha_{ad}} = \frac{\alpha_{bc}}{\alpha_{bd}}. \tag{2.30i}$$

Postulates *E2* and *E3* are extensions of *E1* of Ref. 12. Postulate *E2* asserts that the phase difference between two different cluster terms of a given scattering amplitude that differ only over a certain subset of variables is independent of the remaining set of variables R over which they are identical. These phase differences are observable quantities, according to Postulate *B2* of Ref. 12. If they were not independent of R, then observable phenomena would depend on effects associated with disconnected bubbles in a manner contrary to the physical decomposition principle; phenomena would depend on "unconnected" phenomena, where "unconnected" means unconnected by energy–momentum transfer.

Postulate *E3* asserts that the ratio $\alpha_{ac}/\alpha_{ad}:\alpha_{bc}/\alpha_{bd}$ takes the same value (unity) that it would take if just the nth subsets alone were present. This ratio is an observable quantity (provided the various M functions are all nonzero—otherwise one of the phases can be defined at will). This observable corresponds to an interference effect in a transition from a combination of a and b to a combination of c and d. The postulate asserts that this observable quantity is independent of "unconnected" phenomena, as required by the physical decomposition principle.

It is easy to verify that postulates *E2* and *E3* imply that, in a unitarity equation, the sum of contributions having a given connectedness structure (i.e., having a given set of unintegrated conservation-law δ functions) combine to give a product of the connected parts of the unitarity equations for the appropriate subsectors. (The connected part of the unitarity equation is the part having only one unintegrated conservation-law δ function. The terminology comes from the diagrammatic representation discussed in Sec. 3.)

M. Scattering Function $M_c(K)$

The function

$$M_c(K) \equiv \frac{M_1(K)}{(2\pi)^4\delta^4(\Sigma k_i)} \tag{2.31}$$

is called a *scattering function.*

N. Landau Diagrams

A Landau diagram D is constructed from a set

$$\{L_j, V_n, \epsilon_{jn}\}$$

consisting of several directed line segments L_j and two or more vertices V_n. Each V_n contains endpoints of three or more of the L_j, but only one endpoint of any single L_j. The structure of D is defined by the set of numbers ϵ_{jn} defined by

$$\begin{aligned} &\epsilon_{jn} = +1 \quad \text{if} \quad L_j^+ \subset V_n, \\ &\epsilon_{jn} = -1 \quad \text{if} \quad L_j^- \subset V_n, \\ &\epsilon_{jn} = 0 \quad \text{otherwise,} \end{aligned} \tag{2.32}$$

where L_j^+ and L_j^- are the leading and trailing endpoints of L_j, respectively. With each L_j is associated a type of particle t_j, whose mass is μ_j. If particles of type t_j carry a_j units of an additively conserved quantum number a, then the conditions

$$\sum_j a_j\epsilon_{jn} = 0 \quad (\text{all } n) \tag{2.33}$$

are required of D.

The lines L_j are characterized as being incoming, outgoing, or internal according to the following rules:

L_j is outgoing if $\epsilon_{jn} \leq 0$ for all n, (2.34a)

L_j is incoming if $\epsilon_{jn} \geq 0$ for all n, (2.34b)

L_j is internal if neither of the above holds. (2.34c)

The incoming and outgoing lines are called external lines.

A Landau diagram $D(K)$ is a Landau diagram whose incoming and outgoing lines can be placed in one-to-one correspondences with the initial and final particles, respectively, associated with the set $K = (K', -\tilde{K}'')$. A $D_c(K)$ is a connected Landau diagram $D(K)$.

O. Landau Surfaces $\mathcal{M}[D]$

Consider an association

$$L_j \leftrightarrow (\alpha_j, p_j) \tag{2.35}$$

between lines of a Landau diagram D and pairs consisting of a nonzero number α_j and a (positive energy) energy-momentum vector p_j. The Landau surface $\mathcal{M}[D]$ is the set consisting of the p_j associated with the external lines of all associations (2.35) satisfying the conditions

$$p_j^2 = \mu_j^2 \quad (\text{all } j), \tag{2.36a}$$

$$\sum_j p_j\epsilon_{jn} = 0 \quad (\text{all } n), \tag{2.36b}$$

and

$$\sum_j \alpha_j p_j n_{jf} = 0 \quad (\text{all } f), \tag{2.36c}$$

where n_{jf} is the number of times the directed Feynman closed loop f passes along L_j in the positive sense, minus the number of times it passes along L_j in the negative sense.

Equations (2.36a) and (2.36b) are the mass-constraint and conservation-law equations, respectively, while the equations (2.36c) are called the loop equations.

The loop equations imply that the diagram D can be converted into an energy–momentum space diagram $\bar{D}$ of the same topological structure as D by replacing each L_j of D by the vector $\Delta_j = \alpha_j p_j$.

Definition 2.4: $\bar{D}$ is the energy–momentum diagram associated with $\mathcal{M}[D]$.

This equivalence between the existence of an energy–momentum diagram $\bar{D}$ and the validity of the Landau equations (2.36) is the basis of much that follows. It should be thoroughly understood.

Definition 2.5: $\mathcal{M}^+[D]$ is the part of $\mathcal{M}[D]$ that can be realized with all $\alpha_j > 0$.

Definition 2.6:

$$\mathcal{M}(K) \equiv \bigcup_{D(K)} \mathcal{M}[D(K)], \tag{2.37a}$$

$$\mathcal{M}^+(K) \equiv \bigcup_{D(K)} \mathcal{M}^+[D(K)], \tag{2.37b}$$

$$\mathcal{M}_c(K) \equiv \bigcup_{D_c(K)} \mathcal{M}[D_c(K)], \tag{2.37c}$$

$$\mathcal{M}_c^+(K) \equiv \bigcup_{D_c(K)} \mathcal{M}^+[D_c(K)], \tag{2.37d}$$

$$\bar{\mathcal{M}}_c^+(K) \equiv \text{closure of } \mathcal{M}_c^+(K), \tag{2.37e}$$

$$\mathcal{P}(K) \equiv \{K\colon K \text{ is a physical point}\}. \tag{2.37f}$$

Definition 2.7: A *physical point* is a point K such that $M(K)$ represents a physical scattering process in the manner described in Sec. 2G. Physical points are necessarily real points.

The Landau surface $\mathcal{M}[D]$ depends on the external vectors p_i only through the combinations

$$q_n = -\sum_{\text{ex } i} p_i\epsilon_{in} = \sum_{\text{ex } i} k_i\,|\epsilon_{in}|, \tag{2.38}$$

where the sum runs only over the i corresponding to

the external lines of D. The surface $\mathcal{M}^+[D]$ is, explicitly, the set of external p_i consistent with the set of q_n lying on the surface defined by

$$q_n = -\sum_{m\neq n}\sum_i \frac{\omega_n - \omega_m}{|\omega_n - \omega_m|}\epsilon_{in}\epsilon_{im}\mu_i, \qquad (2.39)$$

where the ω_n and ω_m are energy–momentum vector parameters that range over all real values such that the vectors $(\omega_n - \omega_m)\epsilon_{in}|\epsilon_{im}|$ are positive timelike for $\epsilon_{in}\epsilon_{im} \neq 0$. The denominator function is the Lorentz length

$$|\omega_n - \omega_m| \equiv [(\omega_n - \omega_m)(\omega_n - \omega_m)]^{\frac{1}{2}}, \qquad (2.40)$$

which is necessarily positive, since $\omega_n - \omega_m$ is timelike. The parameters ω_n can be interpreted as vectors from an arbitrary origin to the vertices V_n of the energy–momentum diagram $\bar{D}$ associated with $\mathcal{M}^+[D]$.[21]

Since Eq. (2.39) is invariant under translations and dilations, every point of $\mathcal{M}^+[D]$ is achieved by a five-fold continuum of sets $\{\omega_n\}$. Sets $\{\omega'_n\}$, not exhibiting these degeneracies, are therefore introduced:

Definition 2.8: A set $\{\omega'_n\}$ is a set $\{\omega_n\}$ satisfying $\Sigma\omega'_n = 0$ and

$$\sum_{n>m}\sum_i |\omega'_n - \omega'_m||\epsilon_{in}||\epsilon_{im}| = 1.$$

Definition 2.9: A *simple point* of $\bar{\mathcal{M}}_c^+(K)$ is a point $\bar{K}$ of $\bar{\mathcal{M}}_c^+(K)$ such that all points of $\bar{\mathcal{M}}_c^+(K)$ in some neighborhood of $\bar{K}$ are points of just a single surface $\bar{\mathcal{M}}_c^+[D_c(K)]$, and such that the inverse functions $\omega'_n(K)$ are single-valued, continuous functions of $K \in \bar{\mathcal{M}}_c^+(K)$ in some neighborhood of $\bar{K}$.

Surfaces $\bar{\mathcal{M}}^+[D_c(K)]$, corresponding to diagrams $D_c(K)$ that have parts that touch the rest of the diagram at only two points (vertices), but that contain other vertices, are not to be considered. They always coincide with surfaces of simpler diagrams (or are null).

P. Landau Condition for Physical Region Singularities

Definition 2.10: The *Landau condition for physical region singularities* is the condition that $M_c(K)$ be analytic at points K of $\mathcal{T}(K) - \bar{\mathcal{M}}_c^+(K)$.

[21] That the Landau surfaces can be expressed in this parametric form has been noticed also by Logunov, Todorov, and Chernikov. See *1962 International Conference on High Energy Physics at CERN*, p. 695. A. A. Logunov, I. T. Todorov, N. A. Chernikov, see also Nucl. Phys. **50**, 273 (1964). My derivation consists of expressing the q_n of Eq. (2.38) first in terms of the internal momenta incident on V_n, using (2.36b), and then expressing the internal p_j in terms of the $\Delta_j = \alpha_j p_j = \Delta_j\omega$ of diagram $\bar{D}$.

This condition was derived in Ref. 13 from a macroscopic causality condition formulated within the mass-shell S-matrix framework.

Q. The $i\epsilon$ Rules

The second chief result of Ref. 13 is the "basic $i\epsilon$ rule" defined as follows.

Definition 2.11: The *basic $i\epsilon$ rule* is the assertion that for any simple point $\bar{K}$ of $\bar{\mathcal{M}}_c^+(K) \cap \mathcal{T}(K)$ there is a neighborhood $N(\bar{K})$ of $\bar{K}$ and a function $M_c^N(K)$, defined and analytic at points of $[N(\bar{K}) \cap \operatorname{Im}\sigma(K;\bar{K}) \geq 0] - \bar{\mathcal{M}}_c^+(K)$ and coinciding with $M_c(K)$ in $[\mathcal{T}(K) \cap N(\bar{K})] - \bar{\mathcal{M}}_c^+(K)$. The function $\sigma(K;\bar{K})$ is

$$\sigma(K;\bar{K}) \equiv \Sigma q_n(K)\omega'_n(\bar{K}), \qquad (2.41)$$

where $q_n(K)$ and $\omega'_n(\bar{K})$ are the quantities defined in Sec. 2O. Furthermore, the contributions from small neighborhoods of points of $N(\bar{K}) \cap \mathcal{T}(K) \cap \bar{\mathcal{M}}_c^+(K)$ to a summation over physical points can be represented by an integration of $M_c^N(K)$ over a contour that passes around these points by detours into the domain of definition of $M_c^N(K)$. The $M_c^N(K)$ is an analytic extension of $M_c(K)$, and the superscript N is usually omitted. [Actually, $\operatorname{Im}\sigma(K;\bar{K})$ should, according to the result obtained in Ref. 13, be replaced by its minimum as the ω'_n in Eq. (2.41) range over arbitrarily small neighborhoods of the points $\omega'_n(\bar{K})$. This slight complication does not materially affect our arguments, and it will be ignored.]

The points of $\bar{\mathcal{M}}_c^+(K) \cap \mathcal{T}(K)$ that are not simple points fall into various classes:

Definition 2.12: An *almost-simple point* of $\bar{\mathcal{M}}_c^+(K)$ is a point $\bar{K}$ of $\bar{\mathcal{M}}_c^+(K)$ such that in some neighborhood $N(\bar{K})$ of $\bar{K}$ there is a function $\sigma'(K;K')$, defined and continuous in both K and K', when both K and K' are in $N(\bar{K})$, such that, for every K in $N(\bar{K})$ and K' in $N(\bar{K}) \cap \bar{\mathcal{M}}_c^+(K) \cap \mathcal{T}(K)$,

$$\sigma'(K;K') = \sigma(K;K'),$$

where $\sigma(K;K')$ is defined in (2.41).

At an almost-simple point $\bar{K}$ of $\bar{\mathcal{M}}_c^+(K) \cap \mathcal{T}(K)$ there is evidently no conflict between the $i\epsilon$ rules associated with different surfaces $\mathcal{M}^+[D_c(K)]$; the distortions around the various singularities are mutually compatible.

Definition 2.13: A *simply multiplicative point of* $\bar{\mathcal{M}}_c^+(K)$ is a physical point lying on several $\bar{\mathcal{M}}^+[D_c(K)]$,

the $D_c(K)$ of each of which is obtained from one single larger $D_c(K)$ by contracting to points all but one of various "independent parts" contained in it. An independent part of a $D_c(K)$ is a part having an independent dilation parameter in the energy–momentum diagram $\bar{D}$ associated with $\mathcal{M}^+[D_c]$. The various independent parts of a $D_c(K)$ touch each other only at single points, and the Feynman loops can all be confined to individual independent parts; i.e., no loop need pass through several independent parts.

Because the dilation parameters of independent parts are independent it follows, after some algebra, that the corresponding distortions can be made in independent combinations of the δp_i. The $i\epsilon$ rules for all of the surfaces $\bar{\mathcal{M}}^+[D_c(K)]$ passing through a simply multiplicative point can therefore be simultaneously satisfied. Thus there is no difficulty extending the basic $i\epsilon$ rule to simply multiplicative points of $\bar{\mathcal{M}}_c^+(K) \cap \mathfrak{I}(K)$.

Definition 2.14: The *extended Landau surface* $\mathcal{N}^+[D]$ is defined just like $\mathcal{M}[D]$, except that some, but not all, of the α's are allowed to be zero.

Definition 2.15: The *positive-α extended Landau surface* $\mathcal{N}^+[D]$ is the part of $\mathcal{N}^+[D]$ that can be realized with all $\alpha_j \geqq 0$.

Definition 2.16: A *semisimple* point of $\bar{\mathcal{M}}^+(K)$ is a point $\bar{K}$ of $\bar{\mathcal{M}}^+(K)$ such that each point of $\bar{\mathcal{M}}^+(K)$ in some neighborhood of $\bar{K}$ belongs to the Landau surface $\mathcal{N}^+[D]$ for some single fixed $D \equiv D(\bar{K})$.

Definition 2.17: $\mathcal{M}_0^+[D]$ is the part of $\mathcal{N}^+[D]$ that can be realized *only* with all $\alpha_i > 0$.

Definition 2.18: $\mathfrak{I}_0(K)$ is the part of $\mathfrak{I}(K)$ for which no two initial (and no two final) energy–momentum vectors are parallel.

The third chief result of Ref. 13 is the derivation from macrocausality of the "extended $i\epsilon$ rule," defined as follows.

Definition 2.19: The *extended $i\epsilon$ rule* is the assertion that for any semisimple point $\bar{K}$ of $\bar{\mathcal{M}}^+(K)$ in $\mathfrak{I}_0(K)$ there is a neighborhood $N(\bar{K})$ and a function of $M_c^N(K)$ defined and analytic at points of the *nonempty* set

$$[N(\bar{K}) \cap \prod_i \operatorname{Im} \sigma_i(K; \bar{K}) \geq 0] - \bar{\mathcal{M}}_c^+(K),$$

such that $M_c^N(K)$ coincides with $M_c(K)$ in

$$[\mathfrak{I}(K) \cap N(\bar{K})] - \bar{\mathcal{M}}^+(K).$$

The functions $\sigma_i(K; \bar{K})$ are the functions (2.41) for all those surfaces $\mathcal{M}^+[D_c^i(\bar{K})]$ such that $\mathcal{M}_0^+[D_c^i(\bar{K})]$ contains $\bar{K}$. These surfaces are codimension-1 analytic submanifolds (in the mass shell) at $\bar{K}$. [The sets $\operatorname{Im} \sigma_i(K; \bar{K}) \geqq 0$ should, strictly speaking, be interpreted as the slightly smaller sets described in Definition 2.11.]

The codimensionality of a set is the difference between the dimensionality of the set and the dimensionality of the imbedding space. It is shown in Ref. 13 that, except possibly at points of a set of codimension ≥ 3, the function $M_c(K)$ decomposes locally into a sum of functions such that each has only those singularities that lie on a single $\mathcal{N}^+[D]$. Moreover, the analytic structure of each term is what would be entailed by the extended $i\epsilon$ rule if the singularities were semisimple: the extended $i\epsilon$ rule applies to each term.

One can undoubtedly use the low dimensionality of the exceptional terms as the basis for a special treatment, and thus effectively circumvent them. But rather than pursuing these fine points here, we shall simply assume that the additivity property holds everywhere. In particular, we shall take as our primary $i\epsilon$ assumption the "general $i\epsilon$ rule."

Definition 2.20: The *general $i\epsilon$ rule* is the assertion that, in some neighborhood $N(\bar{K})$ of any point $\bar{K}$ of $\mathfrak{I}(K) \cap \bar{\mathcal{M}}_c^+(K)$, the function $M_c(K)$ breaks up into a finite number of terms to each of which the extended $i\epsilon$ rule applies. In particular, for each term the extended $i\epsilon$ rules specifies a region of continuation connecting points of $[\mathfrak{I}(K) \cap N(\bar{K})] - \bar{\mathcal{M}}_c^+(K)$, and summations over physical points of $N(\bar{K})$ are represented by integrals along contours distorted slightly into this region. Moreover, the decompositions in nearby neighborhoods $N(\bar{K})$ are "compatible" in the sense that the distortions of contours can be extended globally by patching together distortions allowed in nearby neighborhoods.

The general $i\epsilon$ rule stated above is very plausible if one considers all singularities to arise from the unitary equations. For if several singularity surfaces passing through a given point are associated with unrelated diagrams, then they will generally come from unrelated portions of the unitary integrals, and hence will be additive.

Definition 2.21: An *essentially real path* is a path that remains at real points except for arbitrarily small distortions around points of $\bar{\mathcal{M}}_c^+(K)$ made in accordance with the general $i\epsilon$ rules. The physical functions $M_c(K)$ at points of $\mathfrak{I}(K) - \bar{\mathcal{M}}_c^+(K)$ are analytically

connected by essentially real paths, according to the general $i\epsilon$ rules, and summations over physical points are represented by contour integrals over essentially real paths.

Remark 2.1: The results proved in Ref. 13 do not include the physical region "boundary points," where two or more initial or final energy–momenta are collinear. However, the concept of continuation around these singularities does not arise, for such singularity surfaces do not separate the neighboring physical points into disjoint regions: The neighboring nonsingular physical points can be connected by real paths that do not cross the singularity surface.

R. Persistence and the Interchange of Like Variables

A *variable* $\bar{V}_i$ is the triplet

$$\bar{V}_i = (k_i, m_i, t_i).$$

Thus

$$K = \{\bar{V}_i\}.$$

Two variables $\bar{V}_i$ and $\bar{V}_j$ are called *like variables* if and only if

$$t_i = t_j, \tag{2.42a}$$

$$m_i = m_j, \tag{2.42b}$$

and

$$k_i^0 k_j^0 > 0. \tag{2.42c}$$

[Equation (2.42c) is in fact implied by (2.42a), since, according to Eq. (2.27), the sign of t_i is the same as the sign of k_i^0. However, Eq. (2.42c) is included for emphasis.] Like variables refer to particles differing only in their energy–momentum vectors k_i.

The assumption was made (Sec. 2E) that a complete set of experimental results is labeled by the various possible sets K considered as *unordered* sets of variables. However, the variables of the analytic function $M_c(K)$ must originally be placed in some specific order. Let the set of points K for which $M_c(K)$ originally represents the physical function be called $\mathfrak{F}$.

Suppose $K_1 \in \mathfrak{F}$ has two like variables and suppose these occupy positions i and j. Let $\mathcal{E}_{ij}$ be the operator that exchanges the variables that occupy positions i and j. The point K_2 is defined by

$$K_2 = \mathcal{E}_{ij} K_1. \tag{2.43}$$

One may now inquire whether analytic continuation of $M_c(K)$ along essentially real paths from $K_1 \in \mathfrak{F}$ to K_2 is possible, and, if so, what significance the so-defined function $M_c(K_2)$ has, if any. The object of the remainder of this section is to show that the physical significance of the function $M_c(K)$ must persist when continued along essentially real paths to outside the original region of definition $\mathfrak{F}$, and that $M_c(K_2)$ has, consequently, the same physical significance as $M_c(K_1)$. It will further be shown that $M_c(K_2)$ must be equal to $M_c(K_1)$ up to a possible sign, and that this sign must be the same for all M functions in which these two like variables appear and, moreover, must be independent of the positions of these like variables within the sets K. The sign is therefore a universal quantity depending only on the type of variables interchanged. Once this is proved, the remainder of the spin-statistics problem is to establish the connection between this universal sign and the spin of the particles corresponding to the interchanged like variables.

The problem of proving the universality of this sign under interchange does not generally arise in field theory, because there one generally assumes that the interchange of like operators gives at most a change of sign, and that this sign under interchange is independent of states upon which the operators are acting. The reader willing to accept the corresponding proposition that the sign under interchange of like variables is a universal quantity, depending only on the particle type, may proceed to Sec. 3.

Our natural idea of the connection between physical functions and analytic functions is that, if a certain physical function is represented by a function analytic in some region, then this correspondence should "persist" as the variables move through a region where the physical function is defined and the mathematical function remains analytic; there should be no break in the correspondence so long as the mathematical function remains analytic at real points.

This persistence property follows, in fact, from the considerations of Ref. 13. There the M functions were considered initially to be distributions defined over test functions of compact support in momentum space. In the case of identical particles, these test functions can be initially restricted to those having supports containing no pairs of distinct points related by an interchange of like variables. This restriction is imposed to avoid possible ambiguities associated with indistinguishability.

For a given process (specified by $\{m_i\}$ and $\{t_i\}$), a distribution is defined over this restricted space of test functions. This distribution is defined by the set of physical transition amplitudes between initial and final systems represented by the allowed set of test functions. According to Postulate *B2* of Ref. 12, these physical transition amplitudes are well defined up to a possible over-all phase.

It follows from the work of Ref. 13 that, over the support of any allowed test function, the distribution can be represented by a function analytic except at points of $\bar{\mathcal{M}}_c^+(K)$. Furthermore, the functions defined

in the intersection of two support regions must agree up to an over-all phase, since they both represent the same physical process and hence must give the same relative amplitudes for various test functions defined over the overlap regions. By patching these functions together, one obtains a single function defined over the union of the allowed support regions. This function is single valued, since, by dimensional considerations, the regularity region is simply connected. By construction it represents the single specified physical process even when continued outside of some original region $\mathcal{R}$. That is, the physical significance *persists* under analytic continuation, so long as the real path of continuation reaches no point of $\bar{\mathcal{M}}_c^+$, where the analyticity property fails.

By the very same argument, the result extends past the points of $\bar{\mathcal{M}}_c^+(K)$, provided continuation is made along the essentially real paths; one simply patches together the functions over the various support regions, in each of which the result follows from the work of Ref. 13. Thus the physical significance of $M_c(K)$ cannot suddenly change; when continued along essentially real paths, $M_c(K)$ continues to represent the correspondingly continued physical function.

The physical continuation from K_1 to K_2 has the effect of exchanging the detectors of the two like particles. That this has no effect on the experimental observables is just the content of the assumption of Sec. 2E; the experimental results were there assumed to be specified by the sets $\{k_i, m_i, t_i\}$ considered as unordered sets of variables, and, in particular, no additional information having to do with order of variables is supposed to be needed to identify the experimental result. Such information would be required if the experimental results depended on which piece of apparatus detected which particle.

One concludes from the above arguments that if K_1 is such that the over-all conservation law is the only one satisfied at K_1, so that $M(K_1)$ is proportional to $M_c(K_1)$, then the experimental correlations are unaltered by the replacement of $M_c(K_1)$ by the $M_c(K_2)$ obtained by continuation along any essentially real path from K_1. In particular, at such a point K, we have

$$M_c(K_1) = \alpha M_c(\delta_{ij}K_1), \tag{2.44}$$

where $M_c(\delta_{ij}K_1)$ is defined by analytic connection from $K_1 \in \mathcal{R}$ along any essentially real path, and α is a phase factor depending on the arguments other than spin indices displayed in

$$\alpha = \alpha_{ij}(K). \tag{2.45}$$

The fact that the α are independent of the spin indices follows from the completeness of the set of spin matrices $\tilde{P}^{\alpha\beta}$ in spin space: interference effects between amplitudes labeled by different spin-quantum numbers are observable (see Ref. 12).

S. The Sign Change Under Interchange of Like Variables

By virtue of postulate *B2* of Ref. 12, linear combinations of amplitudes labeled by K, differing only in the values of k_i, are observable; these are just the usual interference phenomena. This implies, as a generalization of Eq. (2.44), which we now rewrite as

$$M_c(K'; K'') = \alpha_{ij}(K'; K'')M_c(\delta_{ij}K'; K''), \tag{2.46}$$

that

$$M_c(K'; K'') + M_c(K'''; K'') = \alpha_{ij}(K', K'''; K'') \times [M_c(\delta_{ij}K'; K'') + M_c(\delta_{ij}K'''; K'')], \tag{2.47}$$

where K' and K''' are sets differing only by values of the k_i. Substitution of (2.46) into (2.47) gives

$$\alpha_{ij}^{-1}(K', K'''; K'')[M_c(K'; K'') + M_c(K'''; K'')] = \alpha_{ij}^{-1}(K'; K'')M_c(K'; K'') + \alpha_{ij}^{-1}(K'''; K'')M_c(K'''; K''). \tag{2.48}$$

If $M_c(K'''; K'')$ is zero but $M_c(K'; K'')$ is not, then $\alpha^{-1}(K', K'''; K'') = \alpha^{-1}(K'; K'')$. If both functions are nonzero, then there are two possible solutions of (2.48). The first is

$$\alpha_{ij}(K', K'''; K'') = \alpha_{ij}(K'; K'') = \alpha_{ij}(K'''; K''). \tag{2.49}$$

This implies that $\alpha_{ij}(K'; K'')$ is independent of K'. On the other hand, we know that $\alpha_{ij}(K'; K'')\alpha_{ij}(\delta_{ij}K'; K'') = 1$, since a double interchange is the identity. The nondependence on K' then implies

$$\alpha_{ij}(K'; K'') = \alpha_{ij}(K'') = \pm 1. \tag{2.50}$$

The alternative solution to Eq. (2.48) gives

$$\frac{M_c(\delta_{ij}K'; K'')}{M_c^*(K'; K'')} = \frac{M_c(\delta_{ij}K'''; K'')}{M_c^*(K'''; K'')}. \tag{2.51}$$

This says that $M_c(\delta_{ij}K)$ equals $M_c^*(K)$ up to a phase factor $\alpha_{ij}(K'; K'')$ that is independent of K'. This K'-independent phase factor must again be $+1$ or -1, as before.

If $M_c(K)$ had singularities at real points, then (2.51) would contradict the $i\epsilon$ rules. Thus solution (2.49) must hold for M functions having singularities at real points. But the unitarity equations demand there be singularities at least at normal thresholds. Thus only the case (2.49) is possible, and we have

$$M_c(K'; K'') = \alpha_{ij}(K'')M_c(\delta_{ij}K'; K''), \tag{2.52}$$

where $\alpha_{ij}(K'')$ is either $+1$ or -1.

T. Equality of Sign Changes for Interchange of Like Variables

If the set K' of $M(K';K'')=M(K)$ contains several like variables located at positions $i, j, k, \cdots$, then there will be corresponding signs $\alpha_{ij}(K'')$, $\alpha_{ik}(K'')$, $\alpha_{jk}(K'')$, etc. These signs must all be equal. To see this, let the exchange $\mathcal{E}_{jk}$ be applied to both sides of Eq. (2.52). This gives, suppressing the K'' dependence of the α's,

$$\alpha_{jk}M(\mathcal{E}_{jk}K)=\alpha_{jk}\alpha_{ij}M(\mathcal{E}_{jk}\mathcal{E}_{ij}K), \tag{2.53}$$

which, with the replacement of K by $\mathcal{E}_{jk}K$ and cancellation of α_{jk}, becomes

$$\begin{aligned} M(\mathcal{E}_{jk}\mathcal{E}_{jk}K) &\equiv M(K) \\ &= \alpha_{ij}M(\mathcal{E}_{jk}\mathcal{E}_{ij}\mathcal{E}_{jk}K) \\ &= \alpha_{ij}M(\mathcal{E}_{ik}K), \end{aligned} \tag{2.54}$$

since, as may be readily confirmed,

$$\mathcal{E}_{jk}\mathcal{E}_{ij}\mathcal{E}_{jk}=\mathcal{E}_{ik}. \tag{2.55}$$

But Eq. (2.52), with j replaced by k, together with (2.54), gives

$$\alpha_{ik}=\alpha_{ij}. \tag{2.56}$$

This implies equations like

$$\alpha_{ik}=\alpha_{ij}=\alpha_{gj}. \tag{2.57}$$

Hence, all the $\alpha(K'')$ referring to exchanges of this particular kind of like variables are equal.

U. Order-Independence of Sign Changes Under Interchange of Like Variables

The sign $\alpha_{ij}(K'')$ is independent of the order of any like variables occurring in K'', for the relation (2.52) can be continued along essentially real paths to the point where the like variables of K'' are exchanged. That is, $\alpha_{ij}(K'')=\alpha_{ij}(\mathcal{E}_{kg}K'')$.

V. Persistence of Unitarity Equation

Initially arbitrary phases can be specified so that the no-scattering part is unity:

$$S_0(K';K'')=\delta(K';K''). \tag{2.58a}$$

This convention is uniformly adopted in each of the original regions arising in the proof of the persistence property. Thus the unitarity equation takes the form

$$\begin{aligned} &M(K';K'')+M^*(K'';K') \\ &\quad=-\sum_k M(K';K)\tilde{G}(V)M^*(K'';K) \\ &\quad=-\sum_k M(K;K'')\tilde{G}(V)M^*(K;K') \end{aligned} \tag{2.58b}$$

at all real points connected to a physical point by a real path. That is, the form (2.58) of unitarity "persists." Of course, individual terms continue in different ways around various singularities, but the entire equation, nonetheless, remains true.

W. Sign Change for Interchange of Related Like Variables

The sign change $\alpha_{ij}(K'')$ for the interchange of two like variables of K' in $M_c(K';K'')$ is the same as the sign change $\bar{\alpha}_{ij}(K')$ for the interchange of the two corresponding like variables of K'', in the special case where K' and K'' are originally equal. One sees this by applying both interchanges to the unitarity equations (2.58). The right-hand sides become the right-hand sides of unitarity at the new point. The two terms on the left, which are complex conjugates in the case $K'=K''$, become the terms on the left of this equation only if $\alpha_{ij}(K')=\bar{\alpha}_{ij}(K')$.

X. Universality of Sign Change Under Interchange of Like Variables

By virtue of the result of the above section the sign change $\alpha_{ij}(K'')$ is in fact independent of K''. If one interchanges like variables of K', but not K'', for the case in which K' originally equals K'', then the left-hand side of (2.58) is multiplied by $\alpha_{ij}(K'')=\bar{\alpha}_{ij}(K')$. Since the right-hand side is a sum of positive numbers, each of these must undergo this same sign change in order that the equation remain valid. That is, $\alpha_{ij}(K'')=\bar{\alpha}_{ij}(K')=\alpha_{ij}(K)=\bar{\alpha}_{ij}(K)$ for all K such that $M(K';K)$ is physical. The sign change is therefore a universal number depending on the type of variables interchanged but not on the position that these variables occupy in $M_c(K)$ or on the particular $M_c(K)$ in which the variables occur.

Remark 2.2: No interchange of variables between initial and final sets has been discussed. In terms of the variables $K\equiv\{K',-\tilde{K}''\}$ we consider only interchanges of variables having the same type of variables, including sign. However, the interchange of two like variables of type t_i induces the same sign change as the exchange of two like variables of type $-t_i$. This is a rephrasing of the result of Subsection 2W.

Remark 2.3: The sign change holding for the $M_c(K)$ must evidently hold for the $M(K)$ as well, by virtue of considerations of interference effects between different cluster terms.

3. STRUCTURE THEOREMS

Some properties of the functions $B(K)$ represented by bubble diagrams are derived in this section.

A *bubble diagram* B is a collection of directed line segments L_i and signed circles called bubbles. The L_i are directed leftward and each one either issues from the left side of some bubble, or terminates on the right side of some bubble, or does both. In this last

case the line L_i is called an internal line of B. In the other two cases the line is called a final or an initial line of B, respectively. The bubbles of B are partially ordered by the requirement that each internal line terminate on a bubble standing left of the bubble whence it issues.

Each line L_i of a bubble diagram B represents a variable (p_i, m_i, t_i) and each bubble b represents a function $F_b(K_b'; K_b'')$, where K_b' is the set of variables represented by the lines issuing from the left of b and K_b'' is the set of variables represented by the lines terminating on the right of b. The function $M^B(K)$ represented by B is a function of the variables represented by the external (noninternal) lines of B and is defined by

$$M^B(K) = \sum_{\text{int}} \prod_{b \in B} F_b(K_b'; K_b'') \prod_i \tilde{G}_i(v_i), \qquad (3.1)$$

where the summation is over all physical values of the variables represented by the internal lines of B, and the product over i runs over the indices i of all the internal lines L_i of B. The $\tilde{G}_i(v_i)$ are the spin-space factors (2.22) associated with the internal lines L_i, and their spin indices are covariantly contracted on corresponding indices of the F_b. The function $F_b(K_b'; K_b'')$ is either $M_1(K_b'; K_b'')$ or $M_1^*(K_b''^*; K_b'^*)$, according to whether the sign of b is plus or minus.

Remark 3.1: The summation over physical points is represented by integrations over contours that are distorted about singularities of the F_b in accordance with the $i\epsilon$ rules described in the earlier sections. Our first task will be to determine when the distortions prescribed by the various relevant $i\epsilon$ rules are mutually compatible.

Remark 3.2: The decomposition principle is, apart from the phase factors α_p, graphically exhibited by representing $M(K'; K'')$ as a sum of bubble diagrams. Each term in the sum consists of a column of plus bubbles such that every line represented by K' issues from the left of some bubble and every line represented by K'' terminates on the right of some bubble. The summation is over all different ways that the external lines can be connected to a column of bubbles. The contributions from certain of these terms will vanish due to the conservation law and mass constraints. Unitarity in the one-particle system requires that the "trivial" two-line bubble associated with an unscattered line be the "unit" operator $G_i(v_i)\delta(k_i'; k_i'')$, aside from a phase factor that can be defined to be unity. (This definition fixes relative initial and final phases.)

Definition 3.1: With respect to $M^B(K)$, the *physical points* $\mathfrak{P}(K)$ will mean the original points of definition of $M^B(K)$. At these points all the occurring $M_1(K_b'; K_b'')$ are evaluated at physical points or at points infinitesimally removed from them in the manner prescribed by the $i\epsilon$ rules. Analytic continuations from these original (physical) points will be discussed later.

Definition 3.2: A $D' \subset B$ is a Landau diagram D' that can be constructed by replacing each bubble b of the bubble diagram B by either a connected Landau diagram D_c^b or by a point vertex V^b. The D_c^b is required to be a $D_c^b(K)$ such that $\mathcal{M}^+[D_c^b(K)]$ is a Landau surface corresponding to b.

Definition 3.3: A contraction $D \supset D'$ of a Landau diagram D' is a Landau diagram D that can be obtained by shrinking to points certain internal line segments L_i of D' and then removing all the line segments that terminate at their own origin points. D' is considered a trivial contraction of itself.

Definition 3.4: A $D \supset \subset B$ is a Landau diagram D that is a contraction of some Landau diagram $D' \subset B$. If D is $D \supset \subset B$, then B is said to *support* D, and conversely.

Definition 3.5:

$$\mathcal{M}^B(K) \equiv \bigcup_{D(K) \supset \subset B} \mathcal{M}[D(K)], \qquad (3.2a)$$

$$\bar{\mathcal{M}}^B(K) \equiv \text{closure of } \mathcal{M}^B(K), \qquad (3.2b)$$

$$\mathcal{M}_c^B(K) \equiv \bigcup_{D_c(K) \supset \subset B} \mathcal{M}[D_c(K)], \qquad (3.2c)$$

$$\bar{\mathcal{M}}_c^B(K) \equiv \text{closure of } \mathcal{M}_c^B(K), \qquad (3.2d)$$

$$\mathfrak{P}(K) = \{K: K \text{ is a physical point}\}. \qquad (3.2e)$$

Remark 3.2: Every $M^B(K)$ contains a factor $(2\pi)^4\delta^4(\Sigma k_i)$.

Definition 3.6: For a connected B

$$M_c^B(K) \equiv \frac{M^B(K)}{(2\pi)^4\delta^4(\Sigma k_i)}. \qquad (3.3)$$

Theorem 3.1: (First Structure Theorem)
If the functions $M_c(K)$ are analytic at points K of $\mathfrak{P}(K) - \bar{\mathcal{M}}_c^+(K)$, and if the general $i\epsilon$ rules are valid, then the function $M_c^B(K)$ represented by a connected diagram B is analytic at points K of $\mathfrak{P}(K) - \bar{\mathcal{M}}_c^B(K)$.

Proof: Define $\mathfrak{R}(B, K)$ to be the set of physical points represented by the internal lines of B when the variables represented by the external lines of B are fixed at the physical point K. The general $i\epsilon$ rules then

imply that the summation over physical points occurring in the definition of $M_c^B(K)$ can be represented as an integral over a contour that coincides with $\mathcal{R}(B, K)$ except for infinitesimal distortions into the appropriate upper or lower half σ_b planes near the points of the sets $\bar{\mathcal{M}}_c^+(K_b)$ corresponding to the functions $M_c(K_b'; K_b'')$ or $M_c^*(K_b''^*; K_b'^*)$, respectively. The general $i\epsilon$ rules also assure that the various distortions associated with various surfaces $\mathcal{N}^+[D_c^b]$, corresponding to a single bubble b, do not conflict with one another. However, the distortions associated with surfaces $\mathcal{N}^+[D_c^b]$ corresponding to different bubbles b must also be compatible if the $i\epsilon$ rules are to assure a representation of $M_c^B(K)$ in which the contours can be made to avoid all singularities. We therefore examine the compatibility requirements on the distortions associated with a set of surfaces $\mathcal{N}^+[D_c^b]$, one for each bubble b of B.

The integration region is constrained by conservation-law and mass constraints. The conservation-law requirements are automatically satisfied if the integration variables are taken to be Feynman loop momenta h_f. Variations of these parameters h_f are subject, however, to the various mass constraints $\delta p_i^2 = 0$, where p_i is the momentum–energy vector associated with the internal line L_i of B. The variations δp_i^2 are given by

$$\delta p_i^2 = 2 \sum_f p_i n_{if} \delta h_f, \tag{3.4}$$

where n_{if} is the number of times loop f passes along line L_i in the positive sense minus the number of times f passes along L_i in the negative sense.

To calculate the variations $\delta\sigma_b$ one may write, using Eqs. (2.41) and (2.38) and suppressing the prime on ω_n',

$$\begin{aligned}\sigma_b &= \sum_n q_n^b \omega_n^b \\ &= -\sum_n \sum_{\text{ex } i} p_i^b \epsilon_{in} \omega_n^b \\ &= \sum_n \sum_{\text{int } j} p_j^b \epsilon_{jn} \omega_n^b \\ &= \sum_{\text{int } j} p_j^b \sum_n \epsilon_{jn} \omega_n^b \\ &= \sum_{\text{int } j} p_j^b \Delta_j^b, \end{aligned} \tag{3.5}$$

where p_j^b is the energy–momentum vector of the jth *internal* line L_j^b of D_c^b, and Δ_j^b is the difference of the two end points of the line L_j^b of the energy–momentum diagram $\bar{D}_c^b$ associated with $\mathcal{M}^+[D_c^b]$. The variation $\delta\sigma_b(K; \bar{K})$ for K in the neighborhood of a fixed point $\bar{K}$ of $\mathcal{M}^+[D_c^b]$ is then

$$\begin{aligned}\delta\sigma_b &= \sum_{\text{int } j} \delta p_j^b \Delta_j^b(\bar{K}) \\ &= \sum_{\text{int } j} n_{jf}^b \delta h_f \Delta_j^b, \end{aligned} \tag{3.6}$$

where n_{jf}^b is the algebraic number of times loop f passes along line L_j^b of the diagram D_c obtained by replacing the bubbles b of B by the connected Landau diagrams D_c^b. The particular path within D_c^b taken by the loop f is irrelevant to $\delta\sigma_b$, since the sum of $n_{jf}^b \Delta_j^b$ around any closed loop of D_c^b is zero: this sum is just the sum of vectors around a closed loop of the energy–momentum diagram $\bar{D}_c^b$.

By the theory of linear equations, the variations $\delta\sigma_b$ and δp_i^2 can be specified in any desired manner by an appropriate choice of the δh_f, unless there is a set of α's, not all zero, such that

$$\sum_i \alpha_i p_i n_{if} + \sum_{b,j} \alpha_b \Delta_j^b n_{jf}^b = 0 \tag{3.7}$$

for all f. Now the vectors Δ_j^b can be expressed as

$$\Delta_j^b = \alpha_j^b p_j^b, \tag{3.8}$$

since both sides represent the vector of $\bar{D}_c^b$ corresponding to the line L_j^b of D_c^b (see Def. 2.4). Thus Eq. (3.7) can be written

$$\sum_j \alpha_j' p_j n_{jf} = 0, \tag{3.7'}$$

where the sum now runs over all internal lines L_j of the $D_c \subset B$. The α_i' is α_i for internal lines L_i of B and α_j' is $\alpha_b \alpha_j^b$ for internal lines L_j^b of D_c^b.

The Eqs. (3.7') are just the Landau loop equations for D_c. Since the conservation-law and mass conditions are satisfied by the construction, the various $i\epsilon$ distortions associated with the various $\mathcal{N}^+[D_c^b]$ are mutually compatible at a point $\bar{K}$ of $\mathcal{R}(B, K)$ lying on the intersection of these surfaces $\mathcal{N}^+[D_c^b]$ unless the Landau equations are satisfied at $\bar{K}$. That is, the required distortions are mutually compatible at every point $\bar{K}$ of $\mathcal{R}(B, K)$ for every combination of surfaces $\mathcal{N}^+[D_c^b]$, one for each b of B, for all points K of $\mathcal{T}(K) - \bar{\mathcal{M}}_c^B(K)$.

The wording of this argument is for the case that there is only one σ_b for each b, which is the case for a point of $\mathcal{R}(B, K)$ that corresponds to a simple point of each $\bar{\mathcal{M}}_c^+(K_b)$. But for semisimple points of $\bar{\mathcal{M}}_c^+(K_b)$ there are several σ_b, and the distortion must be simultaneously into the upper half-planes of each. Thus the sum over b in (3.7) should include several terms for a single b, for any point of $\mathcal{R}(B, K)$ that corresponds to a semisimple point of $\bar{\mathcal{M}}_c^+(K_b)$. These several terms will have different values of the α_j^b in $\Delta_j^b = \alpha_j^b p_j^b$. These terms are necessarily linearly independent. [See Theorem 8 of Ref. 13] Thus again (3.7) is soluble only if (3.7') is soluble. The α_j' are now somewhat more complicated functions of the original α's, but the argument is not essentially different from the one given above.

It is also required, for analyticity, that the $i\epsilon$ distortion to imaginary values be a continuous function of the real point $\bar{K}$ of $\mathcal{R}(B, K)$. If $\bar{K}$ is a simple point, or a semisimple point, of all the $\bar{\mathcal{M}}^+[K_b]$, then each $\sigma_b(K; \bar{K})$ can be extended to a function $\sigma_b'(K; K')$ continuous in both variables in a neighborhood of $\bar{K}$. The space of the allowed δh_f can be solved for in terms of the $\delta\sigma_b$. Since the equations are nonsingular, this space of the allowed δh_f will be a continuous function of the $\delta\sigma_b$, and hence also of the point $\bar{K}$ in $\mathcal{R}(B, K)$, for simple and semisimple points.

The considerations for general singular points can be reduced to those for simple and semisimple points by means of the additivity property asserted by the general $i\epsilon$ rules.

It follows from the above argument that the contours can be distorted so as to remain in regions of analyticity for all K in $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$. To prove the theorem one needs, however, also to establish the analytic character of the surface $\mathcal{R}(B, K)$. To examine this question consider the transformation from the N_f variables h_f to the set of N_m variables p_i^2 and $N_f - N_m$ other variables x_j, where the N_m variables p_i^2 are the squares of the momentum–energy carried by the internal lines of B. For K in $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$ and $\bar{K}$ in $\mathcal{R}(B, K)$ the loop equations are not satisfied. Thus the variations δp_i^2 and $\delta\sigma_b$, considered as functions of the δh_f, are linearly independent. Thus the δp_i^2 are themselves linearly independent. Hence it is possible to choose, for any K in $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$ and any $\bar{K}$ in $\mathcal{R}(B, K)$, a set of $N_f - N_m$ variables x_j, linear in the h_f, so that $\partial(p_i^2, x_j)/\partial h_f$ is nonzero in a neighborhood of $\bar{K}$. This implies,[22] for K in $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$, that every point of $\mathcal{R}(B, K)$ is an "interior point" of $\mathcal{R}(B, K)$. In fact, $\mathcal{R}(B, K)$ is an analytic manifold, which means that each point of $\mathcal{R}(B, K)$ is contained in an open set of points of $\mathcal{R}(B, K)$ that is the image of an open set in the space of the $N_f - N_m$ real variables x_j, under an analytic mapping $k_i = k_i(x_j)$.

Since $\mathcal{R}(B, K)$ is defined as the common zeros of a finite set of analytic functions, it is necessarily a closed set. But a closed set consisting of interior points can have no boundary points. Thus, for K in $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$, the set $\mathcal{R}(B, K)$ is a closed $(N_f - N_m)$-dimensional surface without edges (i.e., a cycle). This surface is confined to a bounded region in k_i space, and is easily shown to be of finite measure. Moreover, the functions $k_i(x_j)$ are analytic (in fact linear) in K. Thus, for K in $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$, $\mathcal{R}(B, K)$ is a real analytic manifold of finite measure depending analytically on the variables of K. Moreover, as shown earlier, the contour can be distorted so that the integrand is analytic at all points K' on the contour. It therefore follows from Theorem A2 of Appendix A that the integral $M_c^B(K)$ is analytic at points K of $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$. Since all the relevant quantities are well defined and depend analytically on the relevant variables, it is, of course, highly plausible that the integral $M_c^B(K)$ should be analytic, though the proof is not completely trivial.

Definition 3.7: A *simple point of* $\bar{\mathcal{M}}_c^B(K)$ is a point $\bar{K}$ of $\bar{\mathcal{M}}_c^B(K)$ such that in some neighborhood $N(\bar{K})$ of $\bar{K}$ all points of $\bar{\mathcal{M}}_c^B(K)$ belong to the $\mathcal{M}[D_c]$ of only one $D_c \supset \subset B$, and such that the α_j' in (3.7′), when subjected to the constraint $\sum_j |\alpha_j' p_j| = 1$, are uniquely defined continuous functions of the K in $N(\bar{K}) \cap \bar{\mathcal{M}}_c^B(K)$, apart from a sign ambiguity.

Definition 3.8: $\mathcal{M}_c^{B+}(K)$ is the subset of $\mathcal{M}_c^B(K)$ that can be achieved by restricting the α_j' in Eq. (3.7′) to be positive or negative for lines L_j of D_c contained in D_c^b's corresponding to plus or minus bubbles b of B, respectively. The remaining lines of D_c, which are just the lines occurring in B itself, can be either positive or negative.

Definition 3.5′: $\bar{\mathcal{M}}_c^{B+}(K) =$ closure of $\mathcal{M}_c^{B+}(K)$.

Definition 3.9: $\tilde{\mathcal{M}}_c^B(K) = \{K: K$ is a point of $\bar{\mathcal{M}}_c^B(K)$ that is not a simple point of $\bar{\mathcal{M}}_c^B(K)\}$.

Theorem 3.2: (Second Structure Theorem)
In Theorem 3.1, the set $\mathcal{I}(K) - \bar{\mathcal{M}}_c^B(K)$ can be replaced by the set $\mathcal{I}(K) - \bar{\mathcal{M}}_c^{B+}(K) - \tilde{\mathcal{M}}_c^B(K)$.

Proof: For points on $\mathcal{M}_c^B(K)$, it is not possible to arbitrarily specify all the variations $\delta\sigma_b$ and δp_i^2. But it may nonetheless be possible to find variations that keep $\delta p_i^2 = 0$ and $\eta_b \operatorname{Im} \delta\sigma_b > 0$, where η_b is the sign of bubble b of B. This is a sufficient condition for regularity, since it allows one to keep the contour in the region of analyticity.

The variations are subject to the condition (3.7). If $\bar{K}$ is a simple point of $\bar{\mathcal{M}}_c^B(K)$, then there is only one such condition (3.7), since each such condition gives either another $\mathcal{M}[D]$ or another set of α's. When there is only one condition (3.7), all but one of the variations $\{\delta p_i^2, \delta\sigma_b\}$ can be specified, and this remaining one depends continuously on the specified ones. Suppose, for some pair of b, the sign of the ratio of the α_b of the unique (3.7) differs from the ratio of the corresponding η_b. Consider a variation in which the $\delta\sigma_b$ for one of these b's is the dependent

[22] S. Bochner and W. Martin, *Several Complex Variables* (Princeton Univ. Press, Princeton, N.J., 1948), p. 39.

variation and the $\delta\sigma_b$ of the other one of these b's is large compared to the remaining independent ones, which can be considered relatively infinitesimal If the δp_i^2 are taken to be zero, then Eq. (3.7), together with continuity, assures that $\eta_b \operatorname{Im} \delta\sigma_b > 0$ is satisfied for the one dependent variation $\delta\sigma_b$ if it is satisfied for all the independent ones. Thus, the function $M_c^B(K)$ is analytic at simple points of $\mathfrak{T}(K) \cap \bar{\mathcal{M}}_c^B(K)$ that are not on $\bar{\mathcal{M}}_c^{B+}(K)$, which is what the theorem says. The signs of the α_b in Eq. (3.7) carry directly over to the corresponding signs in (3.7′) because the α_j^b in (3.8) are all positive. Theorem 3.2 goes beyond Theorem 3.1 only if the single linear dependence relation (3.7) involves at least one α_b contribution. Thus the δp_i^2 contributions will still be linearly independent and $\mathfrak{R}(B, K)$ will be an analytic manifold, just as in Theorem 3.1.

Definition 3.10: Let D_c be the Landau diagram corresponding to a simple point $\bar{K}$ of $\bar{\mathcal{M}}_c^{B+}(K)$. The corresponding energy–momentum diagram $\bar{D}_c$ is the diagram obtained by replacing each L_j of D_c by the energy–momentum vectors $\Delta_j = \alpha_j' p_j$ of Eq. (3.7′), with $\eta_b \alpha_b \geq 0$ (η_b is still the sign of bubble b). The $\sigma(K; \bar{K})$ for $\bar{D}_c$ is now defined exactly as in Eq. (2.41).

Definition 3.11: The *basic $i\epsilon$ rule for the functions* $M_c^B(K)$ is the same as the "basic $i\epsilon$ rule" defined in Def. 2.11, except that $M_c^B(K)$ replaces $M_c(K)$ and $\bar{\mathcal{M}}_c^{B+}(K)$ replaces $\bar{\mathcal{M}}_c^{+}(K)$. That is, this rule asserts that $M_c^B(K)$ is analytic at points of the upper half $\sigma(K; \bar{K})$ plane near a simple point $\bar{K}$ of $\bar{\mathcal{M}}_c^{B+}(K)$. However, there is one proviso: at least one line of the diagram D_c must correspond to an internal line of some bubble of B; the basic $i\epsilon$ rule for the function $M_c^B(K)$ asserts (by definition) nothing about the case in which every line of D_c is a line of B.

Theorem 3.3: (Third Structure Theorem)
If the assumptions of Theorem 3.1 are satisfied, then the basic $i\epsilon$ rule for the function $M_c^B(K)$ is valid.

Proof: The arguments leading to Eqs. (3.6) and (3.7) give, similarly,

$$\begin{aligned}
\delta\sigma &= \sum_{\text{int } j} n_{jf} \delta h_f \Delta_j \\
&= \sum_{\text{int } j} n_{jf} \delta h_f \alpha_j' p_j \\
&= \sum_{\text{int } i} \alpha_i p_i n_{if} \delta h_f + \sum_{b,j} \alpha_b \Delta_j^b n_{jf}^b \delta h_f \\
&= \sum_{\text{int } i} \alpha_i p_i n_{if} \delta h_f + \sum_b \alpha_b \frac{\partial \sigma_b}{\partial h_f} \delta h_f \\
&= \sum_{\text{int } i} \alpha_i \delta p_i^2 + \sum_b \alpha_b \delta\sigma_b ,
\end{aligned} \tag{3.9}$$

where now the h_f include also the momentum–energy vectors carried along some paths similar to Feynman loops, but unclosed, that enter D_c at certain external vertices and leave at others. These vectors provide for the variations of the external variables and will be called the external parameters h_f. The actual paths they take along the lines of D_c are not relevant to our argument.

The basic $i\epsilon$ rule to be proved refers only to simple points $\bar{K}$ of $\mathfrak{T}(K) \cap \bar{\mathcal{M}}_c^B(K)$. At these points, there is only one equation (3.7′) and consequently all the δp_i^2 and $\delta\sigma_b$ except one can be fixed arbitrarily by appropriate choice of the internal δh_f, as mentioned in Theorem 2. Therefore if we shift $\sigma(K; \bar{K})$ into the upper half-plane by variations of the external δh_f, then the internal δh_f can be adjusted so that all δp_i^2 and all but one of the $\delta\sigma_b$ vanish. This last $\delta\sigma_b$, when multiplied by α_b, must therefore be shifted into the upper half-plane, since σ is. But then by a slight adjustment of the internal δh_f the remaining $\alpha_b \sigma_b$ can also be shifted into their upper half-planes, keeping the $\delta p_i^2 = 0$. This achieves the required result of moving all $\eta_b \sigma_b$ into their upper half-planes, which are the regularity regions, while keeping all the $\delta p_i^2 = 0$.

This argument depends on the assumption that D_c contains some line that is an internal line of the Landau diagram D_c^b corresponding to some bubble b of B, since otherwise the contributions $\delta\sigma_b$ in Eq. (3.9) all vanish. If D_c has only the lines L_i that are the lines occurring already in B itself, then continuation past the singularity is not possible in general. On the other hand, if D_c contains any line that comes from the interior of any bubble b of B, then Theorem 3.3 gives the rule for continuation past this singularity surface unless $\delta\sigma$ is necessarily zero. This circumstance can occur only if the external vertices of $\bar{D}_c$ all coincide or all lie on a single line that is parallel to every external line incident upon all but one of the external vertices. (The external vertices are vertices upon which external lines end.) In this situation the mass constraints on the external lines force all variations $\delta\sigma$ to vanish, and hence no rule for continuation past the singularity is provided by Theorem 3.3.

4. THE POLE-FACTORIZATION THEOREM

This section is devoted to a proof of the pole-factorization theorem. Some definitions are first introduced.

Definition 4.1: A pole diagram D_p is a connected Landau diagram having precisely two vertices V_v and $V_{v'}$ and precisely one internal line L_p. Each vertex therefore contains exactly one of the two

endpoints of L_p:

$$\epsilon_{pv} = +1, \tag{4.1a}$$

$$\epsilon_{pv'} = -1. \tag{4.1b}$$

Definition 4.2: The sets v and v' *defined* by a pole diagram D_p are the sets of external lines connected to the vertices V_v and $V_{v'}$, respectively. The $\bar{v}$ and $\bar{v}'$ represent the corresponding index sets

$$\bar{v} = \{i: \epsilon_{iv} \neq 0, i \neq p\}, \tag{4.2a}$$

$$\bar{v}' = \{i: \epsilon_{iv'} \neq 0, i \neq p\}. \tag{4.2b}$$

Definition 4.3:

$$q_v = -\sum_{\text{ex}} p_i \epsilon_{iv} = \sum_{i \in \bar{v}} k_i, \tag{4.3a}$$

$$q_{v'} = -\sum_{\text{ex}} p_i \epsilon_{iv'} = \sum_{i \in \bar{v}'} k_i = -q_v, \tag{4.3b}$$

$$S_v = q_v^2 = S_{v'}. \tag{4.3c}$$

Remark 4.1: For a pole diagram D_p, one has $\mathcal{M}^+[D_p] = \mathcal{M}[D_p] = \{K: S_v(K) = \mu_p^2\}$, where μ_p is the mass of the particle associated with line L_p of D_p. For $M_c(K)$, the condition of analyticity near $\bar{K}$ of $\mathcal{M}[D_p] \cap \mathcal{T}(K)$ in $\text{Im}\, \sigma(K; \bar{K}) > 0$ implies analyticity near $\bar{K}$ in $\text{Im}\, S_v(K) > 0$, as is shown by some simple algebra. For the function M_c^* represented by a minus bubble, this region of analyticity is switched to $-\text{Im}\, S_v(K) > 0$.

Definition 4.4: A *quasisimple point* $\bar{K} \in \mathcal{M}[D_p] \cap \mathcal{T}(K)$ of a function of the form

$$F(K) = \sum_{B \in \mathcal{B}} M_c^B(K) \tag{4.4}$$

is a point $\bar{K} \in \mathcal{M}[D_p] \cap \mathcal{T}(K)$ such that $F(K)$, considered as a distribution[23] over a real neighborhood of $\bar{K}$, admits a decomposition into analytic functions[24]

$$\begin{aligned} F(K) &= f(S_v(K), W(K)) \\ &= \lim_{\epsilon \to 0} [f_+(S(K) + i\epsilon, W(K)) \\ &\quad - f_-(S(K) - i\epsilon, W(K))], \end{aligned} \tag{4.5}$$

where f_+ and f_- are analytic functions of their arguments in the region corresponding to real K in a neighborhood of $\bar{K}$ and $0 \leq \epsilon < \eta > 0$, except possibly at points where $S(K) = \mu_p^2$ and $\epsilon = 0$. The set $W(K)$ is some set of arguments such that $[S_v(K), W(K)]$ gives a one-to-one analytic mapping of a neighborhood of $\bar{K}$ [in the domain of definition of $F(K)$] into a bounded open set in (S_v, W).

[23] Functionals $f[\varphi]$, more general than distributions, could be used here.

[24] H. J. Bremmerman, *Distributions, Complex Variables and Fourier Transformations* (Addision-Wesley Publishing Co., Inc., Reading, Mass., 1965), p. 48.

Definition 4.5: A function $F(K)$ of the form (4.4) is said to have no pole (or worse) singularity at a quasisimple point $\bar{K} \in \mathcal{M}[D_p] \cap \mathcal{T}(K)$ only if

$$\lim_{\epsilon \to 0} \epsilon f_\pm(S_v(\bar{K}) \pm i\epsilon, W(\bar{K})) = 0. \tag{4.6}$$

Definition 4.6: The *pole assumption* P_1 for a simple-point $\bar{K}$ of $\mathcal{M}^+(K)$ lying on $\mathcal{M}[D_p] \cap \mathcal{T}(K)$ is the assumption that all pole singularities of functions $F(K)$ of the form (4.4) are associated with pole diagrams in the limited sense that if $\bar{K}$ is a quasisimple point $\bar{K} \in \mathcal{M}[D_p] \cap \mathcal{T}(K)$ of $F(K)$, then $F(K)$ has no pole (or worse) singularity at $\bar{K}$ unless some $B \in \mathcal{B}$ supports D_p. Furthermore, the $i\epsilon$ rule for the part of $F(K)$ contributing to the residue (4.6) at such a point $\bar{K}$ is the same as the $i\epsilon$ rule for the various $D_p \supset \subset B \in \mathcal{B}$, provided these $i\epsilon$ rules are all the same (i.e., all have the same sign in $\pm \text{Im}\, S_v > 0$).

Definition 4.7: The *stability condition on physical-particle masses* is the condition that the mass of any (physical) particle is less than the sum of the masses of any set of particles into which is allowed by selection rules to decay. Thus any (nontrivial) bubble b that represents a nonvanishing $M_1(K_b)$ must have at least two initial lines and at least two final lines. And correspondingly, each vertex V_n of any Landau diagram must contain the leading endpoints of at least two lines and the trailing endpoints of at least two lines: formal Landau diagrams, not satisfying this condition, are spurious and can be ignored.

Theorem 4.1 (Pole-Factorization Theorem):
Assumptions:

(1) Unitarity [Eq. (2.20)],
(2) Cluster decomposition [Eq. (2.30)],
(3) Landau conditions for physical region singularities [Def. 2.10],
(4) Basic $i\epsilon$ rules [Def. 2.11],
(5) Stability conditions for physical-particle masses [Def. 4.7].

Consequences: Let $\bar{K}$ be a simple point of $\mathcal{M}^+(K)$ lying on $\mathcal{M}[D_p] \cap \mathcal{T}(K)$ such that the pole assumption P_1 is valid at $\bar{K}$. Then $M_c(K)$ has a pole singularity at $\bar{K}$, whose residue

$$r(\bar{K}) = \lim_{\substack{K \to \bar{K} \\ \text{Re}\, S_v(K) = \mu_p^2 \\ \text{Im}\, S_v(K) > 0}} [S_v(K) - \mu_p^2] M_c(K) \tag{4.7}$$

is

$$r(K) = i\alpha M_c(K_v) \tilde{G}_p(v_p) M_c(K_{v'}). \tag{4.8}$$

The sets K_v and $K_{v'}$ are the sets of variables associated with the lines, both internal and external, incident on vertices V_v and $V_{v'}$, respectively, of $D_p(K)$. The indices associated with particle p are covariantly

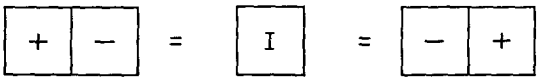


FIG. 1. Unitarity in box notation. The external lines are suppressed and a summation over all possible sets of internal lines is understood. The unit operator I is a product of factors $G(v_i)\delta(k_i'; k_i'')$.

contracted with the corresponding indices of $\tilde{G}_p(v_p)$, which is the spin matrix (2.22). The factor α is

$$\alpha = \frac{\alpha_a}{\alpha_b \sigma_p}, \tag{4.9}$$

where σ_p is the sign induced by interchange of two like variables p, and α_a and α_b are the phase factors occurring in the decomposition

$$\begin{aligned} M(K_{vv'}) = {} & \alpha_a M_1(K_v) M_1(K_{v'}) \\ & + \alpha_b M_1(K) G_p(v_p) \delta(k_p'; k_p'') \\ & + \sum_{p \neq a,b} \alpha_p \prod_s M_1(K_{ps}). \end{aligned} \tag{4.10}$$

Here $K_{vv'}$ is the set of variables consisting of all those in either K_v or $K_{v'}$. This set is just the set K plus two variables: one for an initial particle p and one for a final particle p. That α is independent of the order of variables in $K_{vv'}$ is assured by Postulate *E2*.

Proof: The functions $M(K', K'')$ and $M^*(K''^*; K'^*)$ will be represented by plus and minus boxes, respectively, with the sets of lines issuing from the left and terminating on the right of these boxes being the lines representing the sets K' and K'', respectively. Then the cluster property is the assertion that the plus (minus) box is equal, apart from the phases α_p, to a sum of bubble diagrams, each consisting of a column of plus (minus) bubbles, with the sum being over all ways of connecting the given external lines to a column of bubbles. In this notation unitarity takes the form shown in Fig. 1.

Multiplication of $M(K)$ by unitarity gives the equation represented by Fig. 2.

The terms on the left of Fig. 2 that support D_p will be classified with the aid of the following two lemmas, which depend upon the idea of "key bubble":

Definition 4.8: A *key bubble* (relative to D_p) of a bubble diagram B is a bubble b of B such that every path in B from a line in the set v defined by D_p to a line in the set v' defined by D_p passes through b.

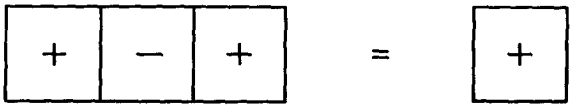


FIG. 2. Result of multiplying $M(K)$ by unitarity.

Lemma 4.1: If B supports D_p, then B has a key bubble (relative to D_p).

Proof: "B supports D_p" means there is a $D \subset B$ having an internal line $\bar{L}$ such that the contraction to points of all other internal lines of D gives D_p, with $\bar{L}$ of D becoming L_p of D_p. Every path in D from v to v' must pass along $\bar{L}$, for if there were one not passing along $\bar{L}$, then the contraction would give a path in D_p from v to v' not passing along L_p, which is not possible. Since every path from v to v' passes along $\bar{L}$, each of these paths passes also through any $\bar{b}$ such that $\bar{L}$ is an internal or external line of the D_c^b replacing $\bar{b}$ of B. At least one such $\bar{b}$ must exist, and any such $\bar{b}$ is a key bubble.

This argument proves, in addition to the lemma, the result asserted in the following corollary.

Corollary 4.1: Any $\bar{L}$ of $D \subset B$ that becomes the L_p of $D_p \supset D \subset B$ upon contraction of the other lines of D is an internal or external line of the D_c^b replacing some key bubble of B.

Convention 4.1: In this section all trivial (two-line) bubbles will be considered absent: the unscattered particles of $M(K)$ will be represented by single lines containing no bubbles.

Convention 4.2: Bubble diagrams that correspond to functions $B(K)$ that vanish because of combinations of mass constraints, conservation laws, and stability conditions will be considered not to exist.

Definition 4.9: A *direct path* connecting two bubbles is a path that touches these two bubbles at, but only at, its two endpoints.

Lemma 4.2: A B that supports D_p has precisely one or two key bubbles (relative to D_p, always). In the first case, no L of any $D \subset B$ can become the L_p of $D_p \supset D \subset B$ unless L is an internal line of the D_c^b that replaces the one key bubble. In the second case, the two key bubbles are connected by a line $\bar{L}$ of B. Moreover, the removal of $\bar{L}$ from B disconnects the part of the diagram connected to v from the part connected to v'. This line $\bar{L}$ of B becomes L_p of $D_p \supset \subset B$ upon contraction of the rest of the diagram. No other L of any $D \subset B$ can become the L_p of a $D_p \supset D \subset B$, in this case of two key bubbles.

Proof: By Lemma 4.1, there is at least one key bubble. If there is precisely one key bubble, then any

L of $D \subset B$ that becomes the L_p of $D_p \supset D \subset B$ upon contraction is an internal or external line of the D_c^b replacing this key bubble, by Corollary 4.1. If L were an external line of D_c^b, then it would have to be an internal line of B itself. But then the bubble of B lying on the other end of L would also be a key bubble, contradicting the supposition that there is only one key bubble. Thus L must be an internal line of the D_c^b that replaces the one key bubble, in this case of just one key bubble.

If there is more than one key bubble, then pick two. These two are connected by some path in B, since B must be connected in order to support the connected D_p. This path can be taken to be a direct path, by removing closed loops. If this direct path touches some other bubble b', then any path from b' to any external line L_e of B must pass through one of the two key bubbles. Otherwise L_e could be connected to any specified external line of B by a path passing through at most one key bubble: one could pass via b' directly to the last of these two bubbles lying on some original path to that specified external line. But then all external lines of B would belong to the same set ν or ν' to which that L_e belongs, since a path from ν to ν' must pass through all key bubbles, by definition. But, by virtue of our definitions, all external lines of B cannot belong to a single one of the two sets ν or ν', and hence any path from any b' to any L_e must pass through one of the two key bubbles.

This implies, in turn, that every b' lying on any direct path connecting the two key bubbles must stand to the right of one of these two key bubbles and must stand to the left of the other of these two key bubbles; otherwise the rightmost of the b''s could have no initial lines or the leftmost of the b''s could have no final lines, which is not possible because of the conservation-law requirement. Thus the two key bubbles must be ordered, with one standing to the right of the other, and this is (trivially) true also if there is no b', since the connecting path is then simply a single line segment L_j, which is directed. If b_r is a key bubble that stands right of a key bubble b_l, then all the lines of b_r lying on direct paths connecting b_r to b_l must be final lines of b_r (which issue from the left of b_r), since otherwise either b_l or some b' on some path from b_r to b_l would have to stand right of b_r, contrary to assumption or to the above result.

If b is a key bubble of B, its removal must give a diagram $B - b$ in which the parts connected to ν and ν', respectively, are relatively disjoint. The external lines of b belonging to these two parts will be called b^ν and $b^{\nu'}$, respectively. They are disjoint, and all external lines of b must belong to their union, since every one of these lines is connected in $B - b$ to some external line of B, by virtue of the fact that each bubble of B has both initial and final lines.

All external lines of b_r lying on direct paths to some other (fixed) key bubble b_l must belong to a single one of the two sets b_r^ν or $b_r^{\nu'}$; otherwise ν and ν' would be connected by a path that passes through b_l but not through b_r, which is impossible, since b_r is a key bubble. Moreover, all the external lines of b_r not lying on any direct path to the (fixed) key bubble b_l must belong to a single one of the two sets b_r^ν or $b_r^{\nu'}$; otherwise ν and ν' could be connected by a path passing through b_r but not through b_l, which is impossible since b_l is a key bubble. Thus the set of lines of b_r lying on direct paths to b_l constitute one of the two sets by b_r^ν or $b_r^{\nu'}$. This set consists of only final lines of b_r if b_r stands right of b_l. On the other hand, we have:

Proposition 4.1: An internal line L of a diagram D_c^b replacing a key bubble b of B can become an L_p of $D_p \supset \subset B$ only if b^ν and $b^{\nu'}$ both contain both initial and final lines of B.

Proof: The removal of L must disconnect b^ν from $b^{\nu'}$ in D_c^b, since otherwise ν and ν' would not become disconnected by the removal of L, as is required if L is to become an L_p of $D_p \supset \subset B$. If b^ν or $b^{\nu'}$ consisted of only initial lines or only final lines, then the energy–momentum carried by L would have to be the energy–momentum carried by this set of initial or final lines. This conflicts with the stability requirements unless b^ν consists of a single line. But this possibility is precluded by the requirement that vertices of Landau diagrams must contain endpoints of three or more lines, together with the mass, conservation, and stability conditions and the positive-α condition imposed on the D_c^b by Definition 3.2.

Combining Proposition 4.1 with the result stated just before it, we conclude that an L of $D \subset B$ that becomes L_p of D_p cannot be an internal line of the D_c^b replacing the key bubble b_r or, by exactly similar arguments, b_l. Thus any L of $D \subset B$ that becomes L_p or D_p must be an external line of some key bubble, hence an internal line of B, in this case in which there is more than one key bubble.

If there is more than one key bubble, then the L of D that becomes L_p or D_p can only be an internal L of B, as just shown. The two bubbles on either end of this internal line of B are both key bubbles. Since the line L connecting them becomes L_p of D_p, its removal must leave the parts of the diagram connected

to ν and ν' relatively disjoint. This proves the lemma for the case of precisely two key bubbles, since the existence of another L that becomes an L_p of $D_p \subset B$ would imply the existence of other key bubbles. It remains to be shown that there can be no other key bubbles.

Let b_r and b_l be the right-hand and left-hand key bubbles on the two ends of some L of B that becomes L_p of D_p, in the case of more than one key bubble. Suppose there is another key bubble. If this other key bubble stands right of b_r, then the lines of b_r lying on the direct paths to this other key bubble all lie on the right side of b_r and constitute one of the two sets b_r^ν or $b_r^{\nu'}$. The single line L, which is the only line of b_r lying on the direct path to b_l, also constitutes one of the two sets b_r^ν or $b_r^{\nu'}$, and in fact the other one of these two sets, since it lies on the left of b_r. Thus L is the only final line of b_r. This contradicts stability. Thus, this other key bubble cannot stand right of b_r. Neither can it stand left of b_l. Nor can it stand to the right of b_l and to the left of b_r, since this would imply the existence of a direct path between b_r and b_l that parallels L, and hence precludes the possibility that L becomes L_p of D_p as required. Thus there can be at most two key bubbles. This concludes the proof of Lemma 4.2.

Lemma 4.2 allows a classification of the terms on the left of Fig. 2 that can support D_p. First there are terms having only one key bubble. This single bubble can belong to any one of the three columns. Then there are the various terms having two key bubbles connected by a line of B, the removal of which separates B into two disjoint parts, one connected to the set ν and the other connected to the set ν'. And this line must become the L_p of D_p upon contraction of the rest of the diagram. The various terms supporting D_p are indicated in Fig. 3.

The protruding products of little plus and minus boxes in the first and third terms are just the identity, by virtue of unitarity, and can be dropped.

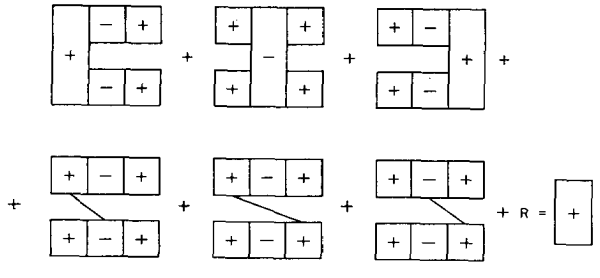


FIG. 3. Decomposition of left side of Fig. 2 into the six possible types of terms that support D_p plus a remainder term R that does not support D_p. The line terminating at the top or bottom of a box is supposed to end on some (nontrivial) bubble within that box.

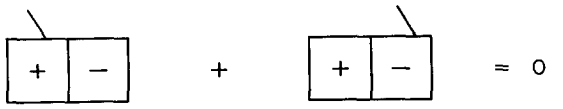


FIG. 4. Modified form of unitarity equation. The singled-out line emerging from the left side of the unitarity equation can come either from some nontrivial plus-bubble, from some nontrivial minus-bubble, or from the incident lines of the right. Terms of this third kind cancel the unity on the right of unitarity, leaving the equation represented in Fig. 4.

For the next terms we make use of the identity shown in Fig. 4.

The equation represented by Fig. 4 allows the line leaving the lower minus box of term four of Fig. 3 to be shifted to the lower right plus box. Then unitarity can be used to cancel the protruding products of plus and minus boxes. An equation similar to that represented in Fig. 4 allows the fifth and sixth terms to be cancelled. Extraction of the connected part then gives Fig. 5.

The result claimed in the theorem now follows essentially from the fact that first and third terms in Fig. 5 are analytic in the upper and lower half S_v planes, respectively, as far as the pole contribution is concerned. The detailed argument is as follows: Near a simple point $\bar{K}$ of $\mathcal{M}[D_p] \cap \mathcal{S}(K)$ the first term of Fig. 5 is (after the conservation-law δ function is factored out) the limit of a function analytic at points K near $\bar{K}$ in the upper half S_v plane, according to the basic $i\epsilon$ rule. The second term, which has a factor $2\pi\delta(S_v - \mu^2)$ coming from the phase-space factor (2.13), can be decomposed into a sum of two functions, one analytic at points K near $\bar{K}$ in the upper half S_v plane and the other analytic at points K near $\bar{K}$ in the lower half S_v plane. These two functions both have pole singularities at $S_v = \mu_p^2$, but otherwise are analytic at points K near $\bar{K}$, since the two M-function factors can have no singularities at a simple point $\bar{K}$.

Since the sum of the first two terms of Fig. 5 can be decomposed into functions, analytic in the upper and lower half S_v planes, the same must be true of the sum of the second two terms. These decompositions into upper and lower half-plane parts are unique up to a function analytic at $\bar{K}$, by virtue of Theorem C3 of Appendix C. Thus the residue at $\bar{K}$ of the sum over all four terms in Fig. 5 of either the upper or the lower half-plane parts must separately be zero. However, the

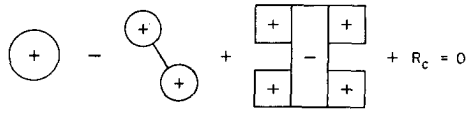


FIG. 5. Result of applying unitarity in lower-order sectors to Fig. 3. The subscript c denotes connected part. The phases α_p, here assumed to be unity, are discussed in the text.

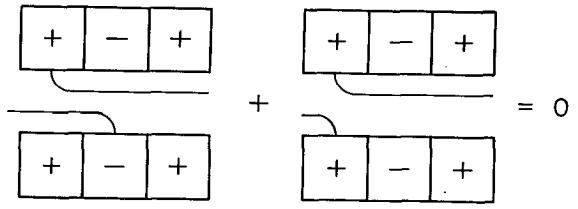


FIG. 6. Diagram indicating unitarity in the lower sector. Note that the left-hand column of the first term has an extra unscattered line that is not present in the fourth term in Fig. 3. The other two columns also contain extra unscattered lines. These lines induce phase changes.

residue at $\bar{K}$ of the sum of the last two terms is zero for the upper half-plane parts by virtue of the pole assumption P_1, which says that this sum is the limit of a function analytic in the lower half-plane, so far as the pole contribution is concerned. Thus the residues of the upper half-plane parts of the first two terms of Fig. 5 must cancel. This gives just the desired result (4.8), apart from the effect of the phase factors α_p.

To complete the proof, the case $\alpha_p \neq 1$ must be considered. Then one must be careful about the switching of lines on the boxes of Fig. 3 by means of Fig. 4. The equation to be used is shown in Fig. 6.

In order to bring the fourth term in Fig. 3 into a form where Fig. 6 can be applied, it must be multiplied by the phase factors α_p from the decomposition law that multiply a contribution to the left-hand term in Fig. 6 and divided by the phase factors that multiply this contribution in Fig. 3. By virtue of Postulates *E2* and *E3*, this ratio of phases is a single phase that is independent of the particular contribution considered. It is in fact just σ_p times the α_b of Eq. (4.10), as is shown in Appendix B. After multiplication by this phase, one can apply Fig. 6, which gives the product of the two plus-bubbles appearing with the phase that they have in the second term of Fig. 6. This phase is just the α_a of Eq. (4.10). Dividing now by the unwanted phase $\alpha_b\sigma_p$, one obtains the required Eq. (4.9).

Concluding Remarks

(1) If the point $\bar{K}$ were a simple point of both $\bar{\mathcal{M}}_c^+(K)$ and $\bar{\mathcal{M}}_c^{B+}(K)$ for all B occurring in the third term of Fig. 5, then the second part of the pole assumption P_1 would not be necessary; one could use Theorem 4.3 instead.

(2) It will be assumed in what follows that $\mathcal{M}^+[D_p] \cap \mathcal{S}(K)$ has a dense set of simple points so that the pole-factorization property is valid for almost all points of $\mathcal{M}^+[D_p] \cap \mathcal{S}(K)$. The validity of this assumption can be confirmed by the methods of Ref. 13.

5. HERMITIAN ANALYTICITY

Hermitian analyticity is the property of scattering functions whereby $M_c(K'; K'')$ and $-M_c^*(K''^*; K'^*)$ are different boundary values of a single analytic function. The central idea of the present proof is to justify, within a strictly mass-shell framework, an effective continuation in external masses.

Instead of the original M_c, one considers the M_c of a "larger" process, the external lines of which are those of a diagram constructed by connecting to each line L_i of the bubble representing the original M_c an "outer bubble" b_i, which is connected to other bubbles only along L_i. The M_c of the larger process will have poles at $S_i = \mu_i^2$, corresponding to these lines L_i, and will contain the original M_c as a factor of the residue of the product of these poles.

The unitarity equation for the larger process will be considered at a point P_0 where all $S_i = 0$. Certain continuations will then be made to points P where all $S_i = \mu_i^2$, and the residue of the product of the poles will be examined.

The unitarity equations at P_0 will consist of a sum of terms each represented by a bubble diagram. According to our basic pole assumption P_1, a function represented by a bubble diagram B can have a pole singularity at $S_i = \mu_i^2$ only if this diagram will support a corresponding pole diagram. In order to have a pole in each of the S_i the diagram must support each of the corresponding pole diagrams. Thus, according to Lemma 1 of Theorem 4, the function represented by the connected bubble diagram B can have pole singularities in each of the set of channel energies S_i only if for each i, individually, the removal of some single bubble b_i of B completely disconnects the external lines of one of the two complementary sets associated with S_i from those of the other of these two sets.

Definition 5.1: A line of a bubble diagram is said to be *directly connected* to another line if and only if these two lines end on a common bubble. A set of lines, consisting of one or more lines, is said to be *directly connected* to another set of lines if and only if some line of one set is directly connected to some line of the other set. A bubble is said to be *directly connected* to a line that ends on it or to a set containing a line that ends on it.

Definition 5.2: The bubbles of a bubble diagram B that represents a contribution to unitarity can be classed as *initial* or *final* according to whether they contain endpoints of initial or final lines of B, respectively.

Convention 5.1: In this section, trivial two-line bubbles will be considered inserted into the unscattered lines of each factor of a unitarity diagram, so that no bubble touches both initial and final lines of the diagram. Accordingly, each bubble is either an initial bubble or a final bubble, but not both. And each line is either initial, final, or internal.

Lemma 5.1: Let B be a connected bubble diagram representing a term in a unitarity equation. Suppose the initial lines $\mathfrak{I}$ of B are divided into $n > 1$ disjoint sets $\mathfrak{I}_i$, and the final lines $\mathfrak{F}$ of B are divided into n disjoint sets $\mathfrak{F}_i$. And suppose B is such that for each i, individually, there is a bubble b_i of B such that the removal of b_i from B completely disconnects the set $\mathfrak{F}_i \cup \mathfrak{I}_i$ from the complementary set $\mathfrak{F} \cup \mathfrak{I} - \mathfrak{F}_i \cup \mathfrak{I}_i$. Then the internal lines of B can be divided into n disjoint sets I_i, plus a remainder set I^R, such that for each i either

(a) I_i is directly connected to every line of $\mathfrak{F}_i$ but to no line of $\mathfrak{F} - \mathfrak{F}_i$, or

(b) I_i is directly connected to every line of $\mathfrak{I}_i$ but to no line of $\mathfrak{I} - \mathfrak{I}_i$.

The set $\mathfrak{E}_i$ is defined to be $\mathfrak{F}_i$ or $\mathfrak{I}_i$ in cases (a) and (b), respectively, and I_i contains every internal line directly connected to $\mathfrak{E}_i$. ($A - B$ is the set of elements belonging to A but not to B.)

Proof: If for some i the set $\mathfrak{I}_i$ is directly connected to the set $\mathfrak{I} - \mathfrak{I}_i$ (necessarily by an initial bubble) and also the set $\mathfrak{F}_i$ is directly connected to the set $\mathfrak{F} - \mathfrak{F}_i$ (necessarily by a final bubble), then the removal of no single bubble can completely disconnect $\mathfrak{F}_i \cup \mathfrak{I}_i$ from $\mathfrak{F} \cup \mathfrak{I} - \mathfrak{F}_i \cup \mathfrak{I}_i$. Thus, the stipulations of the lemma assure, for each i, either that $\mathfrak{I}_i$ is not directly connected to $\mathfrak{I} - \mathfrak{I}_i$ or that $\mathfrak{F}_i$ is not directly connected to $\mathfrak{F} - \mathfrak{F}_i$. If for any i only one of these two conditions is satisfied, then we define $\mathfrak{E}_i$ to be the set $\mathfrak{I}_i$ or $\mathfrak{F}_i$ that is not directly connected to $\mathfrak{I} - \mathfrak{I}_i$ or $\mathfrak{F} - \mathfrak{F}_i$, respectively, and define I_i to be the set of all internal lines of B directly connected to lines of $\mathfrak{E}_i$. On the other hand, if both conditions are satisfied, for some i, then we have two sets $\mathfrak{E}_i$ and $\mathfrak{E}'_i$, one $\mathfrak{I}_i$ and one $\mathfrak{F}_i$, and two corresponding sets I_i and I'_i.

In this latter case it is impossible that $I'_i - I_i$ and $I_i - I'_i$ both be nonempty. If $I'_i - I_i$ is nonempty, then b_i must lie at one end or the other of this set in order that the removal of b_i completely disconnect $\mathfrak{E}'_i$ from $\mathfrak{F} \cup \mathfrak{I} - \mathfrak{E}'_i \cup \mathfrak{E}_i$. And if $I_i - I'_i$ is nonempty, then b_i must lie on one end or the other of $I_i - I'_i$. Both these sets being nonempty would therefore require both that b_i either touch $\mathfrak{F}_i$ or $\mathfrak{I}$ but not $\mathfrak{I}_i$, and that b_i either touch $\mathfrak{I}_i$ or $\mathfrak{F}$ but not $\mathfrak{F}_i$. This is impossible, and hence one of the two sets $I'_i - I_i$ or $I_i - I'_i$ is empty. We adjust the definitions so that $I_i - I'_i$ is empty, which defines the I_i for this case. The definition is unique because $I_i - I'_i$ and $I'_i - I_i$ cannot both be empty, as this would make $I_i = I'_i$, and the diagram would not be connected.

Having defined the I_i, we must now prove them disjoint. Two I_i corresponding to two initial sets $\mathfrak{E}_i = \mathfrak{I}_i$ must evidently be disjoint. For if they contained a common line, then the initial bubble connected to this line would directly connect these two sets $\mathfrak{E}_i = \mathfrak{I}_i$. But the defining characteristic of these sets $\mathfrak{E}_i = \mathfrak{I}_i$ is that no initial bubble connect a line of $\mathfrak{I}_i$ to a line of $\mathfrak{I} - \mathfrak{I}_i$. Similarly, two I_i, corresponding to two final sets $\mathfrak{E}_i = \mathfrak{F}_i$, are disjoint.

Finally, the I_i corresponding to an initial $\mathfrak{E}_i = \mathfrak{I}_i$ must be disjoint from the I_j corresponding to a final $\mathfrak{E}_j = \mathfrak{F}_j$. For suppose $I_i \cap I_j$ were nonempty. The bubble b_i would then have to lie on one end or the other of the set $I_i \cap I_j$, in order that its removal disconnect $\mathfrak{I}_i$ from $\mathfrak{F}_j$. Suppose, first, that b_i were a final bubble directly connected to $\mathfrak{F}_j = \mathfrak{E}_j$, and hence not directly connected to $\mathfrak{F}_i$. In order that its removal completely disconnect the lines of $\mathfrak{F}_i$ from those of $\mathfrak{F} - \mathfrak{F}_i$ it must be true that $\mathfrak{F}_i$ is not directly connected within B to $\mathfrak{F} - \mathfrak{F}_i$. For any bubble directly connecting them could not be b_i, since b_i is not directly connected to $\mathfrak{F}_i$. But if the lines of $\mathfrak{F}_i$ are not originally directly connected to those of $\mathfrak{F} - \mathfrak{F}_i$, then we must have $\mathfrak{F}_i = \mathfrak{E}'_i$, since $\mathfrak{E}_i = \mathfrak{I}_i$. There is, then, a set I'_i, and the set $I'_i - I_i$ is nonempty, by construction. Thus b_i must lie on one end or the other of $I'_i - I_i$, as mentioned before. This means that b_i must either be a final bubble directly connected to $\mathfrak{E}'_i = \mathfrak{F}_i$ or it must be an initial bubble not directly connected to $\mathfrak{E}_i = \mathfrak{I}_i$. This contradicts the assumption that b_i was a final bubble directly connected to $\mathfrak{F}_j = \mathfrak{E}_j$, hence directly connected to no lines of $\mathfrak{F} - \mathfrak{F}_j$.

Suppose, alternatively, that b_i were an initial bubble directly connected to $\mathfrak{E}_i = \mathfrak{I}_i$. Again we must have $\mathfrak{F}_i = \mathfrak{E}'_i$ and $I'_i - I_i$ nonempty. Thus again b_i would have to be either a final bubble directly connected to $\mathfrak{E}'_i = \mathfrak{F}_i$ or an initial bubble not directly connected to $\mathfrak{E}_i = \mathfrak{I}_i$. This is again contradictory. Thus $I_i \cap I_j$ must be empty, which proves the lemma.

It follows from the definition of I_i that the energy carried by the internal set I_i is equal to the energy carried by the external set $\mathfrak{E}_i$. Moreover, the energy carried by the set of all internal particles is the sum of the energies carried by all of the final sets $\mathfrak{F}_i$ (or by all of the initial sets $\mathfrak{I}_i$). Thus the energy carried

by the set I^R of Lemma 5.1 is $E^R = \Sigma\, q_i^0$, where q_i^0 is the energy part of the vector q_i that is the energy–momentum carried by $\mathcal{F}_i$ minus the energy–momentum carried by $\mathfrak{I}_i$, and the sum is over those i for which $\mathcal{E}_i = \mathfrak{I}_i$. The point P_0 is taken to be a point where $q_i = 0$ for all i. Thus E^R is zero at P_0. This means I^R is empty at P_0.

Lemma 5.2: A connected diagram B such that the I^R of Lemma 2.1 is empty has either $\mathcal{E}_i = \mathfrak{I}_i$ for all i or $\mathcal{E}_i = \mathcal{F}_i$ for all i.

Proof: No bubble can be directly connected to a line of an I_i corresponding to an $\mathcal{E}_i = \mathfrak{I}_i$ and also to a line of an I_j corresponding to an $\mathcal{E}_j = \mathcal{F}_j$. For any initial bubble directly connected to a line of I_i cannot be directly connected to any internal line not in I_i, by virtue of the definition of I_i. Similarly, any final bubble directly connected to a line of I_j cannot be directly connected to any internal line not in I_j. Thus, since $I_i \cap I_j$ is empty, neither an initial nor a final bubble can be directly connected both to a line of I_i and to a line of I_j.

Let C_i be the set of bubbles of B directly connected to any line of any I_i corresponding to an $\mathcal{E}_i = \mathfrak{I}_i$. And let C_j be the set of bubbles of B directly connected to any line of any I_j corresponding to an $\mathcal{E}_j = \mathcal{F}_j$. According to the above result, the sets C_i and C_j are disjoint. If I^R is empty, then every bubble of B must be in either C_i or C_j. And moreover no line of B can connect a bubble of C_i to a bubble of C_j. Since B is connected, it follows that either C_i or C_j must be empty.

Lemma 5.3: If the I^R of Lemma 5.1 is empty, then all the b_i of B in Lemma 5.1 must be one and the same bubble $b = b'$.

Proof: It was shown at the end of Lemma 5.1 that b_i cannot be directly connected to $\mathcal{E}_i$. The same argument show that b_i must be directly connected to the other one of the two sets $\mathcal{F}_i$ or $\mathfrak{I}_i$.

Consider, in view of Lemma 5.2, the case in which all $\mathcal{E}_i = \mathcal{F}_i$. Then each b_i is an initial bubble directly connected to $\mathfrak{I}_i$. If b_i is directly connected also to either $\mathfrak{I}_j$ or I_j, then $b_i = b_j$, for the removal of no other bubble could then disconnect either $\mathfrak{I}_j$ or $\mathcal{F}_j$ from $\mathfrak{I}_i$. But b_i must be directly connected to some $\mathfrak{I}_j$ or I_j with $j \neq i$ for B to be connected: if b_i were directly connected to no $\mathfrak{I}_j$ or I_j with $j \neq i$, then the part of B connected to $\mathcal{F}_i \cup \mathfrak{I}_i$ could not be connected to the rest of B, since I^R is empty.

By this argument, b_i must in fact be directly connected to either I_j or $\mathfrak{I}_j$ for every j. For if it were directly connected to one of these two sets only for a proper subset J of the j's, then the replacement of b_i to given back B could not reconnect the part of B connected to the $\mathcal{F}_j \cup \mathfrak{I}_j$ for j in J to the rest of B. Thus, either I_j or $\mathfrak{I}_j$ is directly connected to b_i for every j and one has $b_i = b_j$ for every j. The case in which all $\mathcal{E}_i = \mathfrak{I}_i$ is essentially the same. This completes the proof.

According to the above lemmas and discussion, the unitarity equation (for the larger process) at P_0 has only two kinds of terms that will contribute, when continued to P, to the residue of the product of all the poles. The first is the kind in which all $\mathcal{E}_i = \mathcal{F}_i$ and the second is the kind in which all $\mathcal{E}_i = \mathfrak{I}_i$. The sums of the terms of these two kinds are represented by the two diagrams of Fig. 7.

The sums of terms of the first and second kinds will be denoted by A^+ and A^-, respectively. The functions A^+ and A^- will be continued from P_0 to points P^+ and P^-, respectively, by detouring around singularities of the terms of these functions in accordance with the basic $i\epsilon$ rules for the various functions $M_c^B(K)$ constituting these terms. Near the points $P^\pm$, the pole terms, indicated in Fig. 8, become dominant.

Figure 8 can be considered to represent the residue of the product of the poles at $P^\pm$; then the plus and minus lines represent the factors $\pm i\tilde{G}_i(v_i)$. In particular the residues have the forms

$$r^\pm = M_c^\pm(K) \prod_i [\pm i\tilde{G}_i(v_i) F_i^\pm(K_i)], \qquad (5.1)$$

where the $F_i^\pm(K_i)$ are the functions represented by the

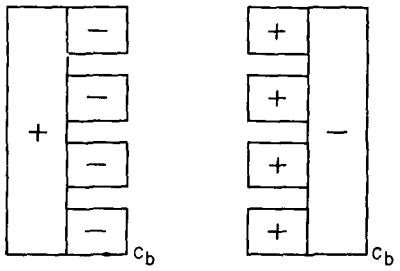


FIG. 7. Diagrams representing the sums of terms of the first and second kind. In this figure n, the number of outer bubbles, is 4. The small boxes represent the sums of sets of bubbles of the indicated sign connected to, and only to, the indicated set $\mathcal{E}_i$. The large rectangle consists of the sum of all sets of bubbles of the indicated sign such that the overall diagram is connected, and such that the removal of some single bubble b' of this rectangle disconnects each set $\mathcal{F}_i \cup \mathfrak{I}_i$ from every other one. Only connected diagrams are included because we consider here only the connected part of the unitarity equation, which is itself a valid equation, since the disconnected parts themselves give valid equations, by virtue of postulates *E2* and *E3*.

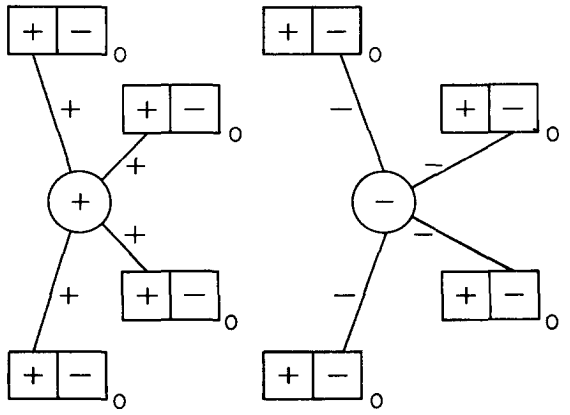


FIG. 8. Diagrams representing the dominant pole contributions near $P^{\pm}$. The multipole contributions come from insertion of the pole contributions into the key bubble b' of one of the large rectangles. The subscripts 0 denote connected parts, and also the possible limitations in the set of diagrams arising from the fact that the terms occurring in $A^{\pm}$ are those present at P_0.

outer boxes of the right- and left-hand diagrams of Fig. 8, and $M_c^{\pm}(K)$ represent the original $M_c(K'; K'')$ and its conjugate $M_c^*(K''^*; K'^*)$.

The unitarity equation at P_0 is $A^+ + A^- + A^0 = 0$, where A^0 is the sum of the terms of the unitarity equation appearing in neither A^+ nor A^-. One can think that A^+ is continued from P^+ to P_0; at P_0 one adds the function A^0 and obtains, according to unitarity, the result $-A^-$, and the function $-A^-$ is then continued to P^-. One has in this way a path of continuation leading from P^+ to P_0 to P^-, but there is a "jump" across a certain cut at P_0. The discontinuity across this cut at P_0 is A^0.

Following the method of Ref. 8, we now attempt to shift (distort) this path of continuation $P^+P_0P^-$ to a new path $P^+P_0'P^-$ that lies completely in the manifold $S_i = \mu_i^2$ for all i. This will be done by constructing a set of manifolds $S_i = a_i^2$ and gradually increasing the a_i^2 from zero to μ_i^2. The original path $P^+P_0P^-$ touches $S_i = 0$ (all i) at P_0. As one increases the a_i^2, the part of the path in $S_i \leq a_i^2$ is shifted into the manifold $S_i = a_i^2$, while the part of the original path $P^+P_0P^-$ in $S_i \geq a_i^2$ is left as originally. The original path $P^+P_0P^-$ then defines the end points of the part of the path lying at constant a_i.

In this distortion of the path, there is the jump originally at P_0 to consider. If the singularities of the function A_0 are confined, locally, to a finite number of singularity manifolds, then A_0 can be continued to a (continually shifting) point on the shifting path. Then one simply adds at this point the jump defined by the continued function A_0. When one arrives finally at the situation where all $a_i^2 = \mu_i^2$ and considers the residue of the product of poles, this jump will not contribute, since A^0, by construction, is the set of terms not having all the requisite poles. The pole assumption P_1 is now considered extended to non-physical points, as will be discussed momentarily.[25]

In the process of distortion, various singularities may be met by the shifting path. One can perhaps distort the path away from them. Alternatively, one can jump across the cut trailing such a singularity s by adding the discontinuity across the cut, just as one did at P_0. If the singularities of this discontinuity function are confined to a finite number of singularity manifolds, then the function can be continued to a (continually shifting) point P_s on the shifting path. Then in the analytic continuation from P^+ to P^- along the shifting path one simply adds at the point P_s this discontinuity function. The contribution from such a discontinuity will not affect the residue at $a_i^2 = \mu_i^2$, unless the discontinuity has all the required poles.

To discuss which singularities have cuts having discontinuities having the requisite poles some definitions will be introduced.

Definition 5.3: $D' \subset D$ is a Landau diagram D' that can be constructed by replacing some of the vertices V_n of D by connected Landau diagrams D_c^n.

Remark 5.1: It is easily confirmed that if D' is a $D' \subset D$, then D is a contraction $D \supset D'$ of D' and conversely (see Def. 3.3).

Definition 5.4: $D_1 \supset\subset D$ is a Landau diagram D_1 that is a contraction $D_1 \supset D'$ of some Landau diagram $D' \subset D$. D *supports* D_1 means D_1 is a $D_1 \supset\subset D$.

Lemma 5.4: A Landau diagram D supports a pole diagram D_p only if D has a key vertex V_p such that every path in D connecting a line in the set of external lines ν defined by D_p to a line in the conjugate set ν' passes through V_p. Such a vertex V_p is called a key vertex of D (relative to D_p).

Proof: The proof is essentially identical to the proof of Lemma 4.1.

[25] The path of continuation from P^+ to P_0 to P^- is really a collection of paths, one for each term in $A^{\pm}$. The shifting of these paths into the mass shell is to be done in such a way that they all pass through a common point $P_0(a_i)$, at which A_0 is added. This requirement may entail that some of the paths have to jump across certain cuts, as is discussed in the following paragraph. So long as only a finite number of singularity manifolds are relevant, the required distortions and continuations are possible. Natural boundaries are considered to be infinite collections of singularity manifolds and are assumed to give no contributions to the residue if none of the constituent manifolds do. The set of points where the manifold property fails is very thin and the paths can be moved by them without difficulty.

Definition 5.5: The pole assumption P_2 is the assumption that the discontinuity across the cut connected to a singularity surface $\mathcal{M}$ of a function $M_c^B(K)$ represented by a connected bubble diagram B has no pole [or worse] singularity at $\mathcal{M}[D_p]$ unless the surface $\mathcal{M}$ is a surface $\mathcal{M}[D]$ such that $D \subset B$ supports the pole diagram D_p [or possibly some other pole diagram D_p' with $\mathcal{M}[D_p] = \mathcal{M}[D_p']$]. In this sense, all pole singularities at $\mathcal{M}[D_p]$ are associated with the pole diagram D_p.

Remark 5.2: Pole assumption P_2 is similar to pole assumption P_1. It is more general in that it is not restricted to physical points $\mathfrak{P}(K)$. [The functions $M_c^B(K)$ and the sets $\mathcal{M}[D]$ are, of course, now analytically extended to include points not in $\mathfrak{P}(K)$.] Aside from this difference, the assumptions P_1 and P_2 would be very similar if the discontinuity across a surface $\mathcal{M}[D]$ were given by a Cutkosky rule. This would make the discontinuity function essentially a bubble diagram function $M_c^{B'}(K)$ represented by a bubble diagram B' obtained by replacing each vertex of $D \subset B$ by an appropriate bubble. Then pole assumption P_1, generalized to nonphysical points, would say that the discontinuity function has no pole (or worse) singularity unless $D \subset B$ supports D_p. By no pole (or worse) singularity we mean, here, no singularity that affects the residue of the pole.

The quantity of interest to us is the residue of the product over i of the poles at $S_i = \mu_i^2$. The only singularities having discontinuities contributing to this residue are, according to pole assumption P_2, those associated with diagrams D supporting each of the corresponding pole diagrams D_p^i. According to Lemma 5.4, a Landau diagram D can support a pole diagram D_p^i only if it contains a corresponding key vertex V_p^i. Thus the earlier arguments now show that A_0 will not contribute to the residue at either physical or nonphysical points.

The contributions to A^+ and A^- are represented by bubble diagrams B each having only one key bubble b'. If $D \subset B$ supports D_b^i, then the key vertex V_p^i of D must be a vertex of the diagram $D_c^{b'}$ replacing b' in the construction of D from B; the removal of a vertex from the D_c^b of a nonkey bubble cannot effect the required separation, since the removal of the entire bubble does not. Thus, for a B representing a term in A^+ or A^-, which we write as $B \in A^\pm$, any key vertex of any $D \subset B \in A^\pm$ must be a vertex of the diagram $D_c^{b'}$ replacing b' in the construction of D from B. It then follows that any $D \subset B \in A^\pm$ containing all the required key vertices V_p^i can be constructed by replacing the bubbles in one of the terms indicated in Fig. 8 by Landau diagrams D_c^b and then contracting certain lines. At points $S_i < \mu_i^2$ (all i), all the lines L_i, explicitly shown in Fig. 8, must be contracted to points, since the corresponding mass constraints are not satisfied at these points. The required key vertices V_p^i are then just these contracted lines L_i.

The Landau diagrams D associated with the singularities having all the required poles have, according to the above arguments, a very special structure. They consist of $n+1$ "independent" parts connected only at key vertices. For each positive $i \leq n$ there is one "outer"-independent part having the external lines $\mathcal{F}_i \cup \mathfrak{I}_i$ and precisely one of the n key vertices. For $i = 0$ there is the one inner part, which has no external lines but which has all n key vertices. These $n+1$ parts are independent in the sense used in Sec. 1: they have independent dilation parameters and the Feynman loops can be confined to individual independent parts. Because of this, the Landau surfaces are just the Landau surfaces for these independent parts. That is, the singularity surface $\mathcal{M}[D]$ is just a sum of singularity surfaces $\mathcal{M}[D_i]$, where each surface $\mathcal{M}[D_i]$ is a surface in the variables associated with just one of the $n+1$ independent parts. As a consequence, the path of continuation at fixed $S_i = a_i^2$ can be considered to be a product of paths $P_i(a_i)$, one in the variables associated with each of the $n+1$ independent parts. For the outer parts there are mass constraints on each of the external lines, and there is one additional "mass" constraint $S_i = a_i^2$ associated with the vertex V_p^i. For the inner part there is a "mass" constraint $S_i = a_i^2$ associated with each vertex V_p^i. As the a_i increase, the motion of the singularity surfaces $\mathcal{M}[D_i]$ can be viewed as the motion of the Landau singularity surfaces for the individual parts under a continuation in the "masses" a_i. These "masses" a_i are, of course, not physical-particle masses, but rather variables of the larger process.

Consider now $P^+(a_i)P_0'(a_i)P^-(a_i)$, the part of the shifting path of continuation lying in the surfaces $S_i = a_i^2$. At $a_i = 0$ the two end points $P^+(a_i)$ and $P^-(a_i)$ of this part of the path coincide with the point P_0. Then, as the a_i increase, the points $P^\pm(a_i)$ start moving along paths determined by the original paths P_0P^+ and P_0P^-. These original paths lie in the physical region of the larger process and their detours around the physical region singularities are specified by the basic $i\epsilon$ rules for functions $M_c^B(K)$.

Apart from these $i\epsilon$ detours, the two paths P_0P^+ and P_0P^- will be taken to be identical. Then

$P^+(a_i)P_0'(a_i)P^-(a_i)$ becomes a closed loop, except for the small $i\epsilon$ gap between the two endpoints $P^+(a_i)$ and $P^-(a_i)$. Thus, the only singularities that can get inside this loop are either physical-region singularities that have, for some value of the a_i, entered through this gap, or singularities that have emerged from the cuts trailing physical-region singularities that have entered through the gap.

One follows the motion of these singularities by a continuation in the "masses" a_i. Assuming, temporarily, that the paths can be kept away from the various singularities whose discontinuities contribute to the residue, one arrives finally at $a_i \simeq \mu_i$ (all i) and considers the residue of the product over all i of the poles at $S_i = \mu_i^2$. At the points P^+ and P^- one has the residues r^+ and r^- given in Eq. (5.1) and indicated in Fig. 8. By construction, continuation along the path $P^+P_0'P^-$ takes r^+ to $-r^-$.

As just discussed, $P^+P_0'P^-$ is effectively a product of paths P_i, with one path P_i in the variables associated with each factor in Eq. (5.1). Each factor in r^+ is, accordingly, continued along the corresponding path P_i. Under this continuation, r^+ goes to r^h,

$$r^h = M_c^h(K) \prod_i i\tilde{G}(v_i)F_i^h(K_i), \tag{5.2}$$

which we know to be $-r^-$.

Consider the paths P_i corresponding to the outer sets $\mathfrak{F}_i \cup \mathfrak{I}_i$. The path P_0P can, as we shall verify below, be chosen such that it crosses no singularity in these variables. In this case the outer paths P_i can be shrunk to points, which means that

$$F_i^h(K_i) = F_i^+(K_i). \tag{5.3}$$

Moreover, in this case, in which P_0P crosses no singularities in the outer variables, the terms contributing to the outer factors in Fig. 8 are the same at P_0 and P. Hence the subscript zero can be replaced by the subscript c, denoting connected part. Then unitarity (Fig. 2) gives

$$F_i^+(K_i) + F_i^-(K_i) = 0. \tag{5.4}$$

These equations convert the equation $r^h = -r^-$ into

$$M_c^h(K) = -M_c^-(K), \tag{5.5}$$

which is just Hermitian analyticity. The path of continuation h, which is constructed by the procedure described above, is called the path of Hermitian analyticity.

The phases α of the pole-factorization theorem are incorporated into the functions $F^{\pm}(K)$. That the phases of the various contributions to (5.4) are then such as to ensure its validity, by virtue of unitarity, follows from the fact that the various contributions to unitarity associated with different connected structures, hence different conservation-law δ functions, must satisfy separately the unitarity equations, by virtue of postulates *E2* and *E3*. The argument is similar to the one given in conjunction with these phases in the proof of the pole-factorization theorem, and need not be given again.

Equation (5.3) is valid, provided P_0P is chosen so as to cross no singularities in variables associated with the outer processes. This can be achieved, for instance, by taking the outer processes to be simple two-particle scattering processes, and holding fixed, in the continuation from P_0 to P, the total energies E_i of these outer processes. The "masses" a_i are varied by varying the momenta of the particles of the outer processes. The only possible Landau singularities in the variables associated with these outer two-particle processes are normal-threshold singularities at constant E_i. This follows from a simple enumeration of possible physical region Landau diagrams for a two-particle process. These singularities at constant E_i will not be crossed because the E_i are held fixed (at values not at a normal threshold).

When the path $P^+P_0P^-$, which lies at physical points of the larger process, is shifted to $P^+P_0'P^-$, which lies at $S_i = \mu_i^2$, the values of (at least some of) the momentum vectors q_i associated with lines L_i must become complex (at some points on the new path). This is because the part of the path associated with the variables of the inner process is forced to pass through the region below the physical threshold of the inner process, and such regions cannot be realized with real q_i. These complex values of the q_i can and will be obtained by performing appropriate complex Lorentz transformations on the corresponding outer parts; that is, the complexification of q_i will be obtained by a complex Lorentz transformation on all the vectors of the ith outer part. Since the singularity structure is not altered by a (real or complex) Lorentz transformation, the fact that q_i is complex will not reflect itself in the part of the path associated with the variables of the outer process; one can consider part of the path associated with the outer parts to lie at real values of the energy–momentum vectors, as far as the singularities in these parts themselves allow this.

In the above discussion it was assumed that the paths of continuation can be distorted so as to stay away from all singularity surfaces that contribute to the multipole residue. That this is possible follows in most cases from dimensional considerations: A one-dimensional curve is generally too "thin" to get trapped by a finite set of singularity surfaces. For instance, we know that contours of integration of

real dimension n in a space of complex dimension n can get trapped at a point to give a singularity of an integral. But then for $n > 1$, a one-dimensional curve will in general not be trapped: it can slide away from the usual pinch configuration in $n - 1$ directions.

One can confirm this in a simple example: Consider the surfaces $z_1 = 0$, $z_1 + \alpha = z_2^2$ and the curve $x_1 = -\epsilon$, $x_2 = 0$, $y_1 = \tau$, $y_2 = \tau$, where $\alpha > \epsilon > 0$. The curve intersects the real plane at a point lying in the region $\mathcal{R}$ of the real plane bounded by the restriction to real points of the two surfaces. As α approaches zero, this region $\mathcal{R}$ shrinks to a point. But the curve can be moved away from the pinch, by shifting it, for example, to the curve $x_1 = -\epsilon + g(\tau - 2)$, $x_2 = 0$, $y_1 = \tau$, $y_2 = \tau - 2$, where $g(\tau)$ is zero for $|\tau| \geq 1$ and greater than ϵ near $\tau = 0$.

This dimensional argument does not cover all cases, however. For example, a curve might get trapped between two surfaces that reduce to a single surface at a pinch configuration, for then the situation is essentially one-dimensional. Though such possibilities can probably be ruled out, we do not pursue this tack, for in any case the curve might get pulled into an unphysical sheet by some singularity surface. In order to avoid this, the path of continuation will be taken to jump across the cuts trailing certain singularities, rather than detouring around them. In particular, if a singularity moves across the path of continuation then one can define the discontinuity function in the situation before the singularity leaves the physical sheet or is pinched against another singularity. The path of continuation can then be taken to jump across the cut, by adding the discontinuity function at the cut. If these cuts are part of the boundary of the physical sheet, then the path of continuation will remain always on the physical sheet. The definition of the physical sheet given in Refs. 1 and 12 was in terms of essentially this same procedure of continuation in external "masses," and was such that the physical point and the Hermitian conjugate point are necessarily on the boundary of the physical sheet.

6. A CONNECTION BETWEEN PATHS OF CROSSING AND HERMITIAN ANALYTICITY

A derivation of the crossing property of scattering functions is given in Ref. 17, and it will not be repeated here. This section gives an extension of that argument that leads to an important connection between the paths connecting crossed reaction and Hermitian conjugate points. This connection will play a key role in the proof of the normal connection between spin and statistics given in the next section.

The basic idea in the S-matrix derivation of crossing properties is similar to the one used in the above derivation of Hermitian analyticity: One considers a "larger process" having pole singularities with residues containing the scattering functions of interest as factors. In the study of crossing, the larger process is selected so that its physical region, which is a connected set, intersects the "pole manifold" $S_\nu = \mu_p^2$ in two different disjoint regions, with these two regions corresponding to the two different signs of the energy part of the vector k_p whose square is S_ν. Let K and $\bar{K}$ be points of $S_\nu = \mu_p^2$ lying in these two different regions. The corresponding residues are

$$r(K) = i\alpha M_c(K_\nu)\tilde{G}_p(v_p)M_c(K_{\nu'}) \tag{6.1}$$

and

$$r(\bar{K}) = i\bar{\alpha}M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_{\nu'}), \tag{6.2}$$

respectively, by virtue of formula (4.8).

The sets K_ν and $\bar{K}_\nu$ both refer to the same subset ν of the external particles of the larger reaction, and the sets $K_{\nu'}$ and $\bar{K}_{\nu'}$ both refer to the same complementary subset ν' of the particles of the larger reaction. The additional particle referred to by both K_ν and $K_{\nu'}$, and associated with the pole at K, is denoted by p. The additional particle referred to by both $\bar{K}_\nu$ and $\bar{K}_{\nu'}$, and associated with the pole at $\bar{K}$, is denoted by $\bar{p}$. Since the poles at K and $\bar{K}$ lie on disjoint parts of the manifold $S_\nu = \mu_p^2$, the two particles p and $\bar{p}$ need not be identical, although their masses are equal. Indeed, the energy conservation-law requirement demands that the particles p and $\bar{p}$ have the opposite initial-final status and be therefore particles carrying opposite units of all additive quantum numbers. The particles p and $\bar{p}$ are called conjugate particles or relative antiparticles. Use is made in this argument of the converse pole-factorization theorem, which asserts, under the same assumptions, that if there is a pole at $S_\nu = \mu_p^2$ in the physical region, then there must be a corresponding physical particle contributing to unitarity summations; if there were no such particle, then the δ-function contribution needed for the pole would be absent.

Let C^o be a path from K to $\bar{K}$ that runs through the physical region of the larger process, passing around singularities in accordance with physical-region $i\epsilon$ rules. Following the procedure of Ref. 8, we distort (if possible) this path C^o into a path C between K and $\bar{K}$ lying in the mass shell $S_\nu = \mu_p^2$. The continuation of $r(K)$ from K to $\bar{K}$ along path C is designated by $r(K^c)$. By virtue of the definition of C we have

$$r(K^c) = r(\bar{K}), \tag{6.3}$$

or equivalently,

$$i\alpha M_c(K_\nu^c)\tilde{G}_p(v_p^c)M_c(K_{\nu'}^c) = i\bar{\alpha}M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_\nu). \tag{6.4}$$

In the distortion of the original path C^o into the path C one must, as in the case of Hermitian analyticity, avoid singularities having discontinuities with nonzero residue at $S_\nu = \mu_p^2$. The necessary distortions are examined by using an effective continuation in the mass of the pole particle. That is, the various paths of continuation intermediate between C^o and C are divided into three segments with the middle segment at constant $S_\nu = a_p^2$. The path C^o is carried to C by increasing a_p from zero to μ_p. The distortions required of the middle segment are those needed to avoid those Landau singularity surfaces of the larger process that become Landau singularity surfaces of one of the two subreactions associated with the residue when a_p reaches μ_p. The discontinuities across the cuts associated with the remaining singularities will not contribute to the residue, by virtue of pole assumption P_2.

The two endpoints of the middle segment lie on the portions of the original path C^o leading from the zero point to K and $\bar{K}$, respectively. As a_p increases, these end points detour around any encountered singularities in the manner specified by the physical region $i\epsilon$ rules.

In the procedure just sketched, the pole-factorization property was applied to just one particle of some original reaction of interest; the other particles were taken to be particles of the larger reaction. The crossing paths associated with these other particles can be constructed by applying this same procedure to each of these particles separately. However, to standardize the construction and obtain a connection to paths of Hermitian analyticity we shall apply the pole-factorization property simultaneously to all of the particles of the original reaction of interest. That is, the larger process will be chosen to be one having a pole singularity for each particle of this original reaction, so that the scattering function for this reaction occurs as a factor in the residue of the product of all these poles, much as in the case of the derivation of Hermitian analyticity.

Just as in the preceding case of Hermitian analyticity, the continuation in the a_p now starts from a point where all the a_p, and also all their associated k_p, are zero. There are paths in the physical region of the larger process leading from this starting point to the regions associated with each of the n different crossed reactions associated with original reaction. Each of these n paths passes around any encountered singularities in accordance with the physical region $i\epsilon$ rules. If C_i^o is a path in the physical region of the larger process leading from the starting point zero to the point K_i associated with the ith one of the crossed reaction, then $C_{ij}^o = C_i^o - C_j^o$ is a path in the physical region of the larger process leading from K_j to K_i. (Sums of paths are read from right to left.) The result of distorting C_{ij}^o into the mass shell $a_p = \mu_p$ (all p) in such a way as to avoid singularities that contribute to the residue of the product of the poles is denoted by C_{ij}. Since

$$C_{ij}^o + C_{jk}^o = C_{ik}^o, \tag{6.5}$$

and since the distortions are such as to avoid the relevant singularity surfaces, we have also

$$C_{ij} + C_{jk} \simeq C_{ij}, \tag{6.6}$$

where the $\simeq$ sign in Eq. (6.6) means equivalence with respect to continuation of the residue $r(K)$ of the product of poles. Equation (6.6) expresses the compatibility of the various crossing paths C that connect the various crossed reaction regions.

As discussed in the preceding section, and also in Ref. 17, the pole assumption P_2 implies that the singularity surfaces that have cuts having discontinuities contributing to the residue are just the Landau surfaces corresponding to the individual scattering functions of the residue, but with the external mass μ_p shifted to a_p. (Only the Landau surfaces are extended off the mass shell, not the M functions.) Because of this special character of the relevant singularity surfaces the parts of the paths of continuation at constant a_p can be considered to be products of paths, with one factor for each process referred to by the residue. The distortion of each individual path is then followed by following the motion of the Landau surfaces corresponding to the appropriate process, as the a_p increase from zero to μ_p. Only those Landau surfaces need be avoided that are not forced to be nonsingular by the positive-α requirement on the Landau singularities entering the physical region of the larger process.

The parts of the paths C_{ij}, C_{ij}^o, and C_i^o referring to the inner process will be represented by the corresponding lower-case quantities. Then the part of (6.6) referring to the inner process reads

$$c_{ij} + c_{jk} \simeq c_{ik}, \tag{6.6'}$$

where $\simeq$ means equivalence with respect to analytic continuation of any function whose singularities are confined to those Landau surfaces of the inner process that are restrictions to $a_p = \mu_p$ of the surfaces avoided in the distortions of the c_{ij} into the mass shell. Equation (6.6′) is certainly valid if in these distortions one retains the original structure of the paths wherein a single central point is connected to each of the various crossed reaction points. If it is possible to distort all the paths c_{ij}^o into the mass shell, so that the

relevant singularities are avoided, then this structure can certainly be retained. The case in which it is not possible to distort the paths so as to avoid all the relevant singularities will be discussed later.

For each path c_i^o there is a complex conjugate-path $\bar{c}_i^o$ that coincides with c_i^o except that it detours around the physical region singularities in accordance with the $i\epsilon$ rules associated with the conjugate function. According to the previous section, it is the path $\bar{c}_i^o - c_i^o$ that, distorted into the mass shell, gives the path of Hermitian analyticity h_i that takes the scattering function $M(K_i)$ into $-M^+(K_i)$. The rules for the distortion of the path $\bar{c}_i^o - c_i^o$ into the mass shell to give h_i are the same as the rules for the distortion of the path $c_j^o - c_i^o$ into the mass shell to give c_{ji}: One must avoid the points of a larger process that lie on the formal extension off the mass shell of the Landau surfaces of the (inner) reaction of interest. However, one need not avoid those Landau surfaces that are required to be nonsingular by the positive-α requirement on singularities that enter the physical region of the larger process.

Since the paths c_i^o and $\bar{c}_i^o$ leading to the n crossed reactions and their Hermitian conjugate points all start from a single point and the rule for distortion of these paths into the mass shell is a uniform one, the same set of Landau surfaces being avoided in all cases (see below), the compatibility requirement (6.6′) carries over also to paths connecting Hermitian conjugate points. In particular, we obtain relations such as

$$\bar{h}_i + c_{ij} \simeq \bar{c}_{ij} + h_j, \tag{6.7}$$

where $\bar{c}_{ij}$ is the result of distorting $\bar{c}_{ij}^o = \bar{c}_i^o - \bar{c}_j^o$ into the mass shell. Equation (6.7) says (reading from right to left) that the path from K_j to K_i to its conjugate point $\bar{K}_i$ is equivalent to the path from K_j to its conjugate point $\bar{K}_j$ to $\bar{K}_i$. Here "equivalent" means equivalent with respect to analytic continuation of a function having singularities only on Landau surfaces that are restrictions to mass shell $a_\nu = \mu_\nu$ of Landau surfaces associated with the inner process that are not required to be nonsingular for the larger processes by the positive-α requirements. The scattering function of interest must have its singularities confined to these surfaces, since it is a factor of the residue of the larger process, and this factor contains all the dependence on the variables associated with these singularities.

Equation (6.7) is certainly valid if in the distortion into the mass shell one maintains the structure wherein all the n crossed-reaction points and their conjugate points are connected to a single central point. Alternatively, the n crossed-reaction points can be connected to one central point, and the n conjugate points can be connected by conjugate paths to a conjugate central point that is connected by a single path to the unconjugate central point. More generally, Eq. (6.7) is certainly valid so long as no closed loops are introduced into the set of paths connecting the various points. Cases where closed loops are present require some additional discussion, which is given in Appendix D. However, there is no real need to introduce closed loops.

Because the Landau structure is invariant under Hermitian conjugation, the paths $\bar{c}_{ij}$ can be taken to be the complex conjugates of the paths $\tilde{c}_{ij}$, where $\tilde{c}$ is the path of crossing for the transposed function. The relationship of Hermitian conjugateness is maintained if the two related functions are continued along conjugate paths. Thus, from the Hermitian analyticity relationship

$$M_c(K^h) = -M_c^\dagger(K) = -M_c^T(K)^*, \tag{6.8}$$

where $M_c(K^h)$ is the result of continuing $M_c(K)$ along the path h to the conjugate point, and the superscript T represents transpose, one obtains

$$M_c(K^{h\bar{c}}) = -M^T(K^{\tilde{c}})^*, \tag{6.9}$$

where $M_c(k^{h\bar{c}})$ is the result of continuing $M_c(k)$ first along h and then along $\bar{c}$, and $M_c^T(K^{\tilde{c}})^*$ is the result of continuing $M_c^T(K)$ along its path of crossing $\tilde{c}$, and then complex-conjugating. The $\tilde{c}$ and $\bar{c}$ are conjugate paths. Applying (6.7) to the left-hand side of (6.9), we obtain our principal result

$$M_c(K^{c\bar{h}}) = -M_c^T(K^{\tilde{c}})^*, \tag{6.10}$$

which says that the result of continuing $M_c(K)$ first along c and then along the path of Hermitian analyticity h associated with the crossed point $\bar{K}$ gives minus the complex conjugate of the function $M_c^T(K^{\tilde{c}})$.

Equation (6.10) would follow directly from the Hermitian analyticity relation at the crossed point

$$M_c(\bar{K}^{\bar{h}}) = -M_c^T(\bar{K})^* \tag{6.11}$$

if we were in possession of the crossing relationships $M_c(K^c) = M_c(\bar{K})$ and $M_c^T(K^{\tilde{c}}) = M_c^T(\bar{K})$. However, we have so far obtained only the weaker condition (6.4). Because (6.4) has a product of two M functions, there is an ambiguity in the relative phase and normalization of $M_c(K^c)$ and $M_c(\bar{K})$, as was stressed by Olive.[7] And there are also the extra phase factors α and $\bar{\alpha}$ to be considered. These latter depend on the statistics of the particles and will be discussed in the next section.

In the discussion of crossing and Hermitian analyticity given so far it was assumed that the various physical-region paths can actually be distorted into

the mass shell without cutting across any of the singularity surfaces having cuts with discontinuities contributing to the residue. It is conceivable, however, that these singularities might pinch together in such a way as to make impossible the distortion into the mass shell of some of these paths.

Rather than distorting the paths around the various singularities, we can elect rather to jump across the associated cuts, by adding the corresponding discontinuity functions. This was in fact the procedure adopted for the various cuts whose discontinuities do not contribute to the residue. For cuts around which it is not always possible to detour, the discontinuity across the cut is defined for values of a_p for which the two sides are still connected, and this function is then continued to $a_p = \mu_p$.

The discontinuity functions associated with cuts around which it is possible to detour, within the mass shell, share with the original function the important pole-factorization property, since this property can be continued around these cuts. In particular, if a singularity under consideration occurs in the variables associated with the inner reaction, then one can detour around this cut without changing the functions in the residue formula that are associated with the outer processes. One makes use here of the relativistic invariance property, which allows the momentum–energy transferred to the outer reactions to be altered without changing their invariants. Hence the functions associated with these outer reactions will remain unchanged under continuation in the inner variables, except for the alteration of certain polynomials associated with the expansion of spin states. These polynomials return to their original values when the continuation is brought back to the other side of the cut and hence the outer factors return to their original values. The jump in the residue function across the cut is therefore represented by adding a certain discontinuity function to the factor associated with the inner process, upon whose variables the singularity was assumed to depend, the outer factors remaining unchanged. Moreover the discontinuity function for the factor associated with the inner reaction is independent of the particular larger process being considered. These properties of the discontinuity function will be called the *pole-factorization property of discontinuity functions*. It is the property whereby the discontinuity function of the residue across a cut in the variables associated with a given one of the functions occurring in the residue is obtained by adding a discontinuity to that particular one of these functions, this discontinuity being independent of the particular larger process under consideration.

The pole-factorization property of discontinuity functions is, as we have just said, automatically satisfied for cuts around which one can detour without leaving the mass shell. It is also satisfied for cuts with discontinuities given by a Cutkosky formula, for then the pole-factorization property of the discontinuity function is a consequence of the pole-factorization property of the individual functions occurring in the Cutkosky formula.

If the pole-factorization property of discontinuity functions is satisfied for all the cuts across which the paths of continuation jump, then the discussion of crossing and Hermitian analyticity given above is essentially unaltered. For then there are certain cuts across which the paths must jump, but the corresponding discontinuity functions are universal quantities that do not depend upon the particular larger process from which it is derived. Thus, the discontinuities that must be added as a path jumps across the various cuts will be independent of the particular end points being connected by this path, and the compatibility conditions (6.6′) and (6.7) still hold. Furthermore, the property whereby the Hermitian conjugateness relation is maintained when the related functions are continued along conjugate paths is also undisturbed by the cuts. For in the defining of the discontinuities on the two conjugate paths one can use for the larger processes two Hermitian conjugate reactions. Then the Hermitian conjugateness property will be valid for the discontinuity functions calculated at $a_p < \mu_p$ and will be carried into the mass shell by continuation in a_p to μ_p. Thus these discontinuities will not destroy the Hermitian conjugateness property and one still obtains (6.9) and hence (6.10).

7. CONNECTION BETWEEN SPIN AND STATISTICS

The residue of the pole at a point K on the manifold $S_\nu = \mu_p^2$ in the physical region of the (larger) process described by the scattering function $M_c(K)$ is given according to Eq. (4.8) as

$$r(k) = i\alpha M_c(K_\nu)\tilde{G}_p(v_p)M_c(K_{\nu'}), \tag{7.1}$$

where $\tilde{G}_p(v_p)$ is a metric tensor satisfying Eq. (2.24),

$$\tilde{G}(-v_p) = (-1)^{2j_p}\tilde{G}(v_p), \tag{7.2}$$

and α is a phase factor given by Eq. (4.9). This phase factor is important to our considerations but the formula (4.9) will not be needed.

At a physical point $\bar{K}$ lying on the crossed-region part of the manifold $S_\nu = \mu_p^2$, the residue of the scattering function for this same larger process is

$$r(\bar{K}) = i\bar{\alpha}M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_{\nu'}). \tag{7.3}$$

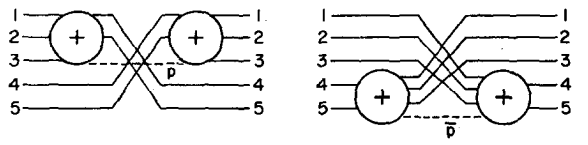


FIG. 9. Representation of direct and crossed residue formulas in a special case where the initial and final particles of the larger process are the same set of particles. The factors $i\alpha$ and $i\bar{\alpha}$ must be added. The order of incident lines, reading from top to bottom, will indicate the order of variables of the functions $M_c(K';K'')$ and $M(K';K'')$.

The point $\bar{K}$ has the same set of variables as K, but the values of the momentum-vector parts have been shifted. The sets of variables K_v and $K_{v'}$ each contain, in addition to certain of the variables of K, a variable associated with the particle p associated with the pole at K. Similarly, the sets of variables $\bar{K}_v$ and $\bar{K}_{v'}$ each contain, in addition to certain of the variables of $\bar{K}$, a variable associated with the particle $\bar{p}$ associated with the pole at $\bar{K}$. Particle $\bar{p}$ is called the antiparticle of particle p.

The result of continuing from K along the mass-shell path of continuation c is represented by placing the superscript c on the set of arguments K. The path c is constructed so that $r(K^c) = r(\bar{K})$. This gives

$$i\alpha M_c(K_v^c)\tilde{G}_p(v_p^c)M_c(K_{v'}^c) = i\bar{\alpha}M_c(\bar{K}_v)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_{v'}). \tag{7.4}$$

Since the point $\bar{K}$ was assumed to lie on the crossed-region part of the manifold $S_v = \mu_p^2$, we have $v_p^c = -\bar{v}_p$, which, with the help of Eq. (7.2), gives the result

$$(-1)^{2j_p}\alpha M_c(K_v^c)\tilde{G}_p(\bar{v}_p)M_c(K_{v'}^c) = \bar{\alpha}M_c(\bar{K}_v)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_{v'}). \tag{7.5}$$

This factor $(-1)^{2j_p}$ will be the origin of the normal connection between spin and statistics. It is also the origin of the connection between the intrinsic parities of particles and their conjugate antiparticles, as was shown earlier.[26]

The relationship between α and $\bar{\alpha}$ is determined in certain cases by the statistics of the particles of the larger process.[27,28] Consider in particular the residue formulas in the special cases indicated in Fig. 9.

The phase factors α and $\bar{\alpha}$ associated with these residues can be determined from statistics by a direct examination of the derivation of the residue formula. In the key step of the derivation the contributions

[26] H. P. Stapp, Phys. Rev. **128**, 1863 (1962).

[27] It is an earlier version of the following argument that, communicated to J. R. Taylor and converted by him to Hilbert space form, served as a key element in the proof of Lu and Olive.

[28] J. R. Taylor, J. Math. Phys. **7**, 181 (1966).

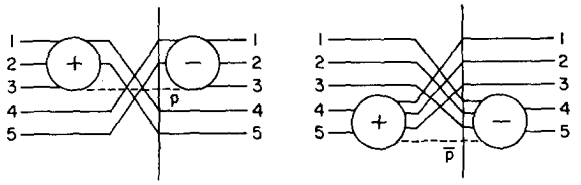


FIG. 10. The contributions to the unitarity equations of the larger process that are converted by means of the unitarity equations shown in Fig. 11 into the residue formulas shown in Fig. 9.

to a unitarity equation represented in Fig. 10 were converted by means of the unitarity equations represented in Fig. 11 to the residues represented in Fig. 9.

The two terms on the left of the equations in Fig. 11 are just the complex-conjugate transposes (Hermitian conjugates) of each other, according to the conventions adopted in Eqs. (2.30d) and (2.58). It is with these phase conventions that the Hermitian analyticity properties were proved.

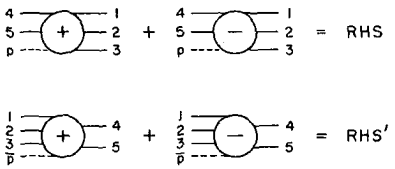


FIG. 11. Unitarity equations used to convert the pole contributions to the unitarity equations of the larger process shown in Fig. 10 into the residue formulas shown in Fig. 9. That the right-hand sides of these equations correctly cancel against terms of the unitarity equation not shown in Fig. 10 is shown in the proof of the pole-factorization theorem.

Because there are no extra phases on the left side of the equations represented in Fig. 11, the phases α and $\bar{\alpha}$ of the residue functions shown in Fig. 9 must be precisely the phases of the corresponding contribution to unitarity shown in Fig. 10. This is a key point.

The phases of the contributions to unitarity shown in Fig. 10 are determined by the statistics of the external particles of these diagrams. In particular, interchange of the pairs of identical particles 4 and 5 in the first factor of the first diagram of Fig. 10, and the pairs of identical particles 1, 2, and 3 in the first factor of the second diagram of Fig. 10, leads to Fig. 12.

The contributions to unitarity shown in Fig. 12 are just absolute-value-squared contributions, and

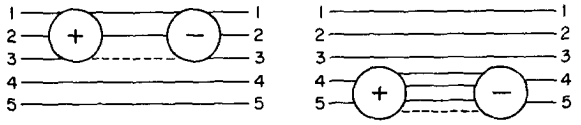


FIG. 12. Result of interchanging some pairs of identical particles in Fig. 10.

these must have no phase. More specifically, if the minus-bubbles in Fig. 12 represent precisely the complex conjugates of the corresponding plus-bubbles in this figure, then the phase factors of the terms in Fig. 12 are unity. This stipulation that the minus-bubbles be the complex conjugates of the plus-bubbles we record as

$$M_c^\dagger(K_{\nu'}) = M_c^*(K_\nu) \tag{7.6a}$$

and

$$M_c^\dagger(\bar{K}_{\nu'}) = M_c^*(\bar{K}_\nu). \tag{7.6b}$$

In the passage from Fig. 10 to Fig. 11, no reordering of the variables of the connected parts is performed; these functions $M_c(K'; K'')$ are kept fixed. Only the orderings of the variables of the larger process on the left are altered. This interchange induces an over-all sign in accordance with the statistics of the particles interchanged.

From the fact that the phase factors in Fig. 12 are unity, the phase factors in Fig. 11, hence Fig. 9, are immediately determined from the statistics of the five external particles. If the number of these particles obeying abnormal statistics $[\sigma_i = -(-1)^{2j_i}]$ is odd, then the quotient $a/\bar{a}$ is

$$-\prod_{i=1}^{5}(-1)^{2j_i} = -(-1)^{2j_p}, \tag{7.7}$$

where the equality (7.7) follows from the fact that the sum of the spins of the particles participating in a nonvanishing reaction is even, by virtue of the covariance condition (2.11). But if $a/\bar{a}$ is $-(-1)^{2j_p}$, then (7.5) becomes

$$-M_c(K_\nu^c)\tilde{G}_p(\bar{v}_p)M_c(K_{\nu'}^c) = M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_{\nu'}). \tag{7.8}$$

Continuation of the right-hand factors in (7.8) along the path of Hermitian analyticity $\bar{h}$ associated with the crossed point $\bar{K}_{\nu'}$ converts Eq. (7.8) to

$$M_c(K_\nu^c)\tilde{G}_p(\bar{v}_p)M_c^*(K_\nu^c) = -M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c^*(\bar{K}_\nu), \tag{7.9}$$

where use has been made of Eqs. (6.10), (6.11), and (7.6).

The functions M and M^* are complex-conjugate functions and the factor G is the square of a Hermitian matrix. Thus the right side of Eq. (7.9) is nonpositive and the left side is nonnegative. Therefore both sides must be zero, which implies immediately that

$$M_c(\bar{K}_\nu) = 0, \tag{7.10}$$

for this case in which an odd number of the external particles of the set $\bar{K}_\nu$, exclusive of $\bar{p}$, are abnormal. This result was derived for the particular case of an M function with three initial particles and three final particles, but the argument holds also for the cases of more than three. That is, any M_c function referring to a process with three or more initial particles and three or more final particles is zero if any subset of all but one of these particles has an odd number of particles with abnormal statistics. This immediately implies that all such M functions referring to a set of particles containing *any* abnormal particles must vanish, except possibly for M functions referring to an odd number of particles all of which are abnormal. This last possibility is ruled out by unitarity, since the non-vanishing process would contribute a term to the sum of positive terms giving the real part of the M function for a corresponding forward-scattering process, which must, however, vanish because it involves an even number of abnormal particles.

The possibility that abnormal particles occur in a reaction involving only two initial or two final particles, but in no reactions involving three or more initial and final particles, conflicts with the pole-factorization property plus unitarity. (Unitarity guarantees that the transpose process is nonzero.) Thus we conclude that the scattering function $M_c(K)$ vanishes if any of the particles referred to by K obey abnormal statistics: Only particles obeying normal statistics can react.

8. PHASE FACTOR IN THE CROSSING RELATIONSHIP

Having established that all particles obey normal statistics, we obtain, instead of (7.8) and (7.9), the relationships

$$M_c(K_\nu^c)\tilde{G}_p(\bar{v}_p)M_c(K_{\nu'}^c) = M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c(\bar{K}_{\nu'}) \tag{8.1}$$

and

$$M_c(K_\nu^c)\tilde{G}_p(\bar{v}_p)M_c^*(K_\nu^c) = M_c(\bar{K}_\nu)\tilde{G}_p(\bar{v}_p)M_c^*(\bar{K}_\nu). \tag{8.2}$$

From (8.2) it follows that

$$|M_c(K_\nu^c)| = |M_c(\bar{K}_\nu)|. \tag{8.3}$$

That is, the continuation of $M_c(K_\nu)$ is equal to $M_c(\bar{K}_\nu)$ up to a possible phase factor.

Equation (8.3) can be written as

$$M_c(K_\nu^c) = M_c(\bar{K}_\nu)\alpha(\bar{K}_\nu), \tag{8.4}$$

where $\alpha(\bar{K}_\nu)$ is a spin-independent phase factor. [The spin independence of $\alpha(\bar{K}_\nu)$ follows immediately from the covariance conditions (2.11).[28]] Insertion of (8.4) into (7.5) gives

$$\alpha(\bar{K}_\nu)\alpha(\bar{K}_{\nu'}) = Q(\bar{K}_\nu, \bar{K}_{\nu'}), \tag{8.5}$$

where $Q = (-1)^{2j_p}\bar{\alpha}/\alpha$.

All our equations are invariant under a transformation of the form

$$M_c(K) \to [\exp(\Sigma \pm i\varphi_p)]M_c(K), \tag{8.6}$$

where φ_p is a real number depending only on the particle type p. The sum in (8.6) is over the particles referred to by K, and the $\pm$ sign is plus for final particles and minus for initial. The numbers φ_p can be chosen so that for each particle p there is one particular M function that satisfies, instead of Eq. (8.4), the more stringent condition

$$M_c(K_\nu^c) = M_c(\bar{K}_\nu). \tag{8.7}$$

That is, the phases φ_p can be chosen so that $\alpha(\bar{K}_\nu) = 1$, which implies that for the particular $\bar{K}_{\nu'}$ associated with $\bar{K}_\nu$ in (8.1) one has also $\alpha(\bar{K}_{\nu'}) = 1$. If the phases α and $\bar{\alpha}$ in (7.5) were such that Q were always unity, then the above adjustment of phases to give $\alpha(\bar{K}_\nu) = 1$ would make $\alpha(\bar{K}_{\nu'}) = 1$ for all $\bar{K}_{\nu'}$, and the crossing relationship (8.4) would have no extra phase.

In order to discuss the value of Q, certain stipulations regarding the order of variables must apparently be made. In field-theoretic models one has a cluster decomposition law that yields

$$M(K_1', K_2', \cdots, K_n'; K_1'', K_2'', \cdots, K_n'') = M_c(K_1'; K_1'')M_c(K_2'; K_2'') \cdots M_c(K_n'; K_n'') + \cdots. \tag{8.8}$$

That is, if the variables in M are ordered according to a particular cluster term, in the manner shown, then the α_p for this particular cluster term is unity.

That this equation should continue to hold in a pure S-Matrix theory can be argued as follows.[29] Let all but one of the sets K_i' and K_j'' be held fixed and let this one remaining set be denoted by A. Then Eq. (8.8) will be written in the abbreviated form

$$M(A) = M_c(A)\Pi + \cdots, \tag{8.9}$$

where Π stands for the product of the remaining factors on the right. [The $M(A)$ on the left is, of course, a quite different function from the $M_c(A)$ on the right.]

From our general cluster-decomposition property we have, instead of Eq. (8.9), the more general equation

$$M(A) = \alpha_A M_c(A)\Pi + \cdots, \tag{8.10}$$

where α_A is the phase factor α_p of (2.30). Let B be a set of variables labeling an amplitude in the same superselection class as the result labeled by A. We have then also

$$M(B) = \alpha_B M_c(B)\Pi + \cdots, \tag{8.11}$$

where all other variables are still fixed as before.

Consider now a superposition $C = aA + bB$. That is, C labels the amplitude such that

$$M_c(C) = aM_c(A) + bM_c(B). \tag{8.12}$$

According to the general cluster property, one should have, in analogy to (8.10) and (8.11), also

$$M(C) = \alpha_C M_c(C)\Pi + \cdots. \tag{8.13}$$

This is actually a slight extension of our postulate *E1* of Ref. 12, which, as stated, referred only to amplitudes labeled by sets K, not to their superpositions. But exactly the same physical principle should apply to superpositions. This extension of Postulate *E1*, which we call *E1'*, gives Eq. (8.13).

We need also a stipulation that

$$M(C) = aM(A) + bM(B). \tag{8.14}$$

This requirement would be rather natural if we were dealing with a Hilbert-space formalism in which the S matrix were regarded as a unitarity mapping of free-particle states onto free-particle states, with these free-particle states regarded as tensor products of individual single-particle states, and in which the ordering of variables specifies the ordering of these states. Of course, since we know that the tensor products of these states in different orders are not all equal, it is not absolutely clear that Eq. (8.14) must be satisfied, since the addition of extra states might affect different states differently. Proofs of spin and statistics that depend on such extra stipulations are not completely satisfactory, since it is conceivable that a theory with abnormal statistics might be possible if one were to abandon the extra stipulations. This might be done in such a way as to leave the physical relationships of superposition, Lorentz invariance, etc., unaltered.

However, having proved the normal connection between spin and statistics without recourse to such stipulations, our objective now is to complete the specification of the basic formalism of a proposed S-Matrix theory. The stipulation (8.14) is therefore now adopted.

The stipulation (8.14) immediately gives the result

$$\alpha_A = \alpha_B, \tag{8.15}$$

as one sees by taking special values a and b satisfying

$$a{:}b = -M_c(B){:}M_c(A). \tag{8.16}$$

Then the relevant term on the left side of (8.13) vanishes, which implies, by virtue of Eqs. (8.10), (8.11), and (8.14), the result (8.15). The phase factor α_A therefore depends only on the superselection class of A, or, more generally, on the superselection classes of the various sets K_i''. In that case, however, one can

[29] J. R. Taylor, Phys. Rev. **42**, 1236 (1966).

take the sets K_i' to be equal to the sets K_i'', without altering α_A. Since the relative phases of the connected parts $M_c(K_i''; K_i'')$ and their corresponding no-scattering parts are fixed to be unity by virtue of *EI*, and since the phase of the no-scattering contribution to $M(K_1'', K_2'', \cdots K_\mu''; K_1'', K_2'', \cdots K_\mu'')$ is unity by virtue of our original conventions on the no-scattering parts (which we were free to choose), we find that the factor $\alpha_A = \alpha_B$ is also unity, and thus obtain (8.8).

From the fact that the phase factor α_p is unity for the decomposition of the type shown in (8.8) one can conclude that the factors α and $\tilde{\alpha}$ in Q are independent of the external variables of the process containing the pole. In particular, if one writes the $M(K_{\nu\nu'})$ on the left of (4.10) as

$$M(K_{\nu\nu'}) = M(K_{\nu'}', K_\nu', K_p'; K_{\nu'}'', K_\nu'', K_p''), \quad (8.17)$$

where K_ν'' and K_ν' contain the initial and final variables of K_ν, and similarly for $K_{\nu'}$, and K_p' and K_p'' are the variables associated with the exchanged particles, then, according to Eq. (8.8), one can say that α_b in (4.10) is unity.

The phase α_a in (4.10) is the inverse of the phase change induced by moving K_p' through K_ν' into the position where (8.8) is again applicable. But this phase change is independent of the variables associated with the external particles associated with $K_{\nu'}$. Here we are using the fact that the phase change induced by the interchange of any two adjacent variables is independent of the remaining variables of the M function. This is a consequence of (8.8) and unitarity. For unitarity ensures that the phase change induced by a reordering of the various final variables is independent of the particular initial variables (To see this, consider the contributions to forward scattering, which is a sum of absolute values squared. Thus all contributions must suffer the same phase change under a reordering of the final external variables.) But then (8.8) insures that the interchange of two adjacent final variables must induce a phase change that is independent of all of the other variables, since one can consider a decomposition in which these two final variables are the only two final variables of one of the individual factors on the right-hand side of Eq. (8.8). This factor can appear as a contribution to various reactions.

For the analogous calculation of $\bar{\alpha}$ one uses Eq. (8.17) with $K_{\bar{p}}$ in place of K_p.

Again $\bar{\alpha}_a = 1$, and $\bar{\alpha}_b$ is independent of the external variables of $K_{\nu'}$; now one must commute $K_{\bar{p}}''$ through K_ν'' to obtain the form where (8.8) is applicable. But then Q is independent of the external lines of $K_{\nu'}$, and one obtains from the special crossing relation

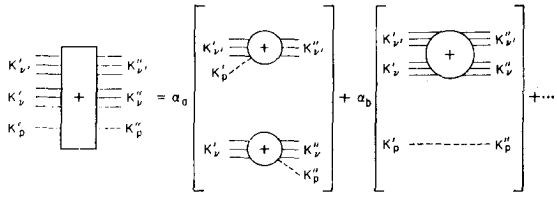


FIG. 13a. Decomposition of (8.17) according to (4.10).

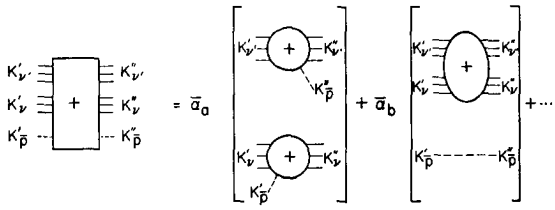


FIG. 13b. Decomposition of analogous to the one in Fig. 13a, but with $\bar{p}$ in place of p.

(8.7) (with the variables ordered as in Figs. 13a and 13b) the general crossing relation

$$M_c(K_{\nu'}'^c, K_p^c; K_{\nu'}''^c) = M_c(\bar{K}_{\nu'}; \bar{K}_{\nu'}'', \bar{K}_p). \quad (8.18)$$

That is, if the phases are adjusted so that (8.18) is valid for one particular ν', then it will be valid for every ν'.

In terms of the functions $M_c(K)$ defined in (2.26), the crossing relation (8.18) becomes

$$M_c(K_{\nu'}^c) = M_c(\bar{K}_{\nu'}), \quad (8.19)$$

which is just the statement that $(-p_p, m_p, -t_p)$ is equivalent to $(p_{\bar{p}}, m_p, t_{\bar{p}})$, with the understanding that the continuation from the original region of positive $(-p_p)$ to the new region of positive p_p has been made along the path of continuation c. That is, for the $M_c(K)$ functions the division between initial and final can be drawn arbitrarily: one has a single universal M function for all the crossed reactions.

The important fact that the sign change under interchange of adjacent conjugate variables is the same as the sign change under the interchange of the corresponding like variables follows immediately from Eq. (8.8) and the fact that the phase change under interchange of adjacent variables is independent of the other variables of M. One can consider Eq. (8.8) for two different orderings of the sets of variables of the M function on the left. Then to one of these sets of variables one adds a conjugate pair and finds that this pair must commute with all variables. In fact, any set of variables of zero quantum numbers must commute with any other variable, if (8.8) is to be consistent with unitarity.

APPENDIX A: A FUNDAMENTAL THEOREM ON ANALYTICITY OF INTEGRALS

Theorem A1: Let $F(K)$ be defined by

$$F(K) = \int f(\bar{K}; K) \prod_{i=1}^{m} \delta(g_i(\bar{K}; K))\, d\bar{K}, \quad \text{(A1)}$$

where the g_i are single-valued real analytic functions of the sets of real variables $K = (k_1, \cdots, k_N)$ and $\bar{K} = (\bar{k}_1, \cdots, \bar{k}_N)$. Suppose, for K in a set $\mathcal{S}$, that $\mathcal{R}(K) \equiv \{\bar{K}: g_j(\bar{K}; K) = 0 \text{ all } j\}$ is a bounded set over which $\partial g_j/\partial \bar{k}_i$ is of maximal rank $m \leq n$, and that $f(\bar{K}; K)$ is analytic at points $(\bar{K}; K)$ of $\{(\bar{K}; K): K \in \mathcal{S}, \bar{K} \in \mathcal{R}(K)\}$. Then $F(K)$ is analytic at points K of $\mathcal{S}$. The analyticity of $f(\bar{K}; K)$ and $F(K)$ is in the sense of Def. 2.1 of Sec. 2.

Proof: Let K be a fixed point in $\mathcal{S}$. Because the rank of $\partial g_j/\partial \bar{k}_i$ is maximal, the set $\mathcal{R}(K)$ is a real analytic submanifold.[30,31] That is, for any $\bar{K}$ in $\mathcal{R}(K)$ there is a real function $\bar{K}'_{\bar{K}}(X)$, defined and analytic on $\bar{U}_{\bar{K}}(K)$, the closure of a bounded open set $U_{\bar{K}}(K)$, in the space of points labeled by the set of local coordinates $X = \{x_1, \cdots, x_{n-m}\}$, such that $\bar{K}'_{\bar{K}}(X)$ maps $U_{\bar{K}}(K)$ onto an open neighborhood $\mathcal{R}_{\bar{K}}(K)$ of $\bar{K}$ in the space $\mathcal{R}(K)$. One can, in fact, evidently take the x_j to be linear functions of the $\bar{k}_i$ in such a way that $\partial(x_j, g_j)/\partial \bar{k}_i \equiv \partial(X, G)/\partial \bar{K}$ is nonzero at $G = 0$ (all $g_j = 0$) for X in a sufficiently small $\bar{U}_{\bar{K}}(K)$. This ensures[30] that the inverse function $\bar{K}'_{\bar{K}}(X, G)$ will be unique and analytic in both arguments at $G = 0$ for X in $\bar{U}_{\bar{K}}(K)$. The function $\bar{K}'_{\bar{K}}(X)$ is then $\bar{K}'_{\bar{K}}(X, 0)$.

The function $\bar{K}'_{\bar{K}}(X, G)$ depends also on K and will sometimes be written as $\bar{K}'_{\bar{K}}(X, G; K)$. It is, in fact, analytic in K, by virtue of the fact that it is analytic in G and that G is analytic in K. For one can write

$$d\bar{K}' = \frac{\partial \bar{K}'}{\partial X}\frac{\partial X}{\partial K}\, d\bar{K} + \frac{\partial \bar{K}'}{\partial G}\frac{\partial G}{\partial K}\, d\bar{K} + \frac{\partial \bar{K}'}{\partial G}\frac{\partial G}{\partial K}\, dK + \frac{\partial \bar{K}'}{\partial K}\, dK = d\bar{K}, \quad \text{(A2)}$$

which gives

$$\frac{\partial \bar{K}'}{\partial K} = -\frac{\partial \bar{K}'}{\partial G}\frac{\partial G}{\partial K}, \quad \text{(A3)}$$

as the well-defined derivative.

Because $\mathcal{R}(K)$ is bounded, it is also compact in the induced topology, in which the neighborhoods in $\mathcal{R}(K)$ are defined as the intersection of $\mathcal{R}(K)$ with neighborhoods in the imbedding K space. This result is well known.[32]

The basic neighborhoods in $\mathcal{R}(K)$ will be taken small enough so that each one is contained with its closure in one of the $\mathcal{R}_{\bar{K}}(K)$. This is possible because of the analytic character of $\bar{K}'_{\bar{K}}(X, G)$ and its inverse. In particular, given a point $\bar{K}$ of $\mathcal{R}(K)$, one can find a sufficiently small neighborhood $\Delta G_{\bar{K}}(K)$ of $G = 0$ such that $\bar{K}'_{\bar{K}}(X, G)$ is analytic with an analytic inverse over $U_{\bar{K}}(K) \otimes \Delta G_{\bar{K}}(K)$.[30] Thus, by taking the basic neighborhoods in $\bar{K}$ space small enough so each is contained with its closure in the image of one of the $U_{\bar{K}}(K) \otimes \Delta G_{\bar{K}}(K)$, we ensure that the closure of the restriction to $\mathcal{R}(K)$ of each of these neighborhoods is in one of the $\mathcal{R}_{\bar{K}}(K)$.

The basic neighborhoods in the space of real points $\bar{K}$ can be defined as the open sets bounded by surfaces at rational constant values of the $\bar{k}_i$. This provides also a set of neighborhoods in $\mathcal{R}(K)$. For the basic neighborhoods in $\mathcal{R}(K)$ a subset of these will be chosen. In particular, since for points on $\mathcal{R}(K)$ the g_j are analytic functions of the $\bar{k}_i$ with nonzero the gradients, the basic neighborhoods in $\mathcal{R}(K)$ can be taken small enough so that the gradients of the surfaces $g_j = 0$ are almost constant over any basic neighborhood in $\mathcal{R}(K)$.[33] Then the set of basic neighborhoods in $\mathcal{R}(K)$ is further restricted by the requirement that none of these neighborhoods be bounded by a surface corresponding to a certain constant $\bar{k}_i$ if the gradient to this constant $\bar{k}_i$ surface is "nearly parallel" to any linear combination of the gradients to the surfaces $g_j = 0$, at any point $\bar{K}$ of the neighborhood in question.

[The point here is first that one can certainly find $n - m$ constant $\bar{k}_i$ coordinate surfaces whose gradients are not "near" the subspace spanned by the gradients at $\bar{K}$ to the surfaces $g_j = 0$. For let $\{V_i\}$ be the orthonormal set of normalized gradients to the coordinate surfaces lying at constant $\bar{k}_i$, and let $\{W_i\}$ be an orthonormal set of vectors such that the first m of them span the space $\mathcal{W}(\bar{K})$ spanned by the m gradients ∇g_j at point $\bar{K}$. Suppose $m + 1$ of the V_i lie "near" the space $\mathcal{W}(\bar{K})$, in the sense that, with a suitable ordering of the V_i, the quantity

$$\sum_{i=1}^{m+1} \sum_{j=m+1}^{n} (V_i \cdot W_j)^2 \equiv \delta$$

[30] Robert C. Gunning and Hugo Rossi, *Analytic Functions of Several Complex Variables* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1965) pp. 16–18.

[31] B. A. Fuks, *Introduction to the Theory of Analytic Functions of Several Complex Variables* (American Mathematical Society, Providence, R.I., 1963), p. 6 and p. 203.

[32] N. Dunford and J. Schwartz, *Linear Operators* (Interscience Publishers Inc., New York, 1958), Vol. I, p. 17.

[33] That this can be achieved is assured by an application of Theorem 8, p. 8 of Ref. 30. "Almost constant" can mean, for instance, less $(1000n)^{-1}$ per cent variation.

is smaller than unity. Then the sum over the complementary set in i is

$$\sum_{i=m+2}^{n}\sum_{j=m+1}^{n}(V_i\cdot W_j)^2 = n-m-\delta$$
$$\leq \sum_{i=m+2}^{n}\sum_{j=1}^{n}(V_i\cdot W_j)^2 = n-m-1.$$

This gives $\delta \geq 1$, contradicting the assumption that $m+1$ of the V_i lie near the space $\mathfrak{W}(\bar{K})$. Thus at most m of the V_i can lie near $\mathfrak{W}(\bar{K})$, and one can find a set of at least $n-m$ vectors V_i, none of which is nearly parallel to any linear combination of the gradients to the surfaces $g_j = 0$. (By "nearly parallel" we can mean, specifically, that the lines make an angle of less than n^{-1} degrees.) One can choose any one of these vectors as one of the x_i. Working up by induction in m, one sees that one can complete the set of x_i by choosing from the set of $\bar{k}_i$ those whose gradients are not nearly parallel to the vectors of $\mathfrak{W}(\bar{K})$. Thus one can find arbitrarily small neighborhoods of $\bar{K}$ in $\mathfrak{R}(\bar{K})$ that are bounded only by manifolds corresponding to $\bar{k}_i$ whose gradients are not nearly parallel to any vectors of $\mathfrak{W}(\bar{K})$. And since $\mathfrak{W}(\bar{K})$ depends continuously on $\bar{K}$, the condition can be maintained for all $\bar{K}$ in sufficiently small neighborhoods.]

By virtue of the compactness of $\mathfrak{R}(K)$ there is a finite covering of $\mathfrak{R}(K)$ by neighborhoods of the type specified above. Since intersections of finite numbers of these neighborhoods are also neighborhoods of this same type, one can find a finite set $\{\mathfrak{R}_\alpha(K)\}$ of these neighborhoods whose sum is just $\mathfrak{R}(K)$, apart from the set of measure zero coming from the boundary points of the various $\mathfrak{R}_\alpha(K)$.

By virtue of the conditions imposed on the basic neighborhoods of $\mathfrak{R}(K)$, each $\mathfrak{R}_\alpha(K)$ is contained in one of the $\mathfrak{R}_{\bar{K}}(K)$. The inverse image of $\mathfrak{R}_\alpha(K)$ under the corresponding $\bar{K}'_{\bar{K}}(X)$ will be called $U_\alpha(K)$. Let $\bar{K}_\alpha(X;K)$ be the $\bar{K}'_{\bar{K}}(X,0;K) = \bar{K}'_{\bar{K}}(X)$ that maps $U_\alpha(K)$ onto $\mathfrak{R}_\alpha(K)$. Then one can write

$$F(K) = \sum_\alpha \int_{U_\alpha(K)} f_\alpha(X;K)J_\alpha(X;K)\,dX$$
$$\equiv \sum_\alpha F_\alpha(K), \tag{A4}$$

where

$$f_\alpha(X;K) = f(\bar{K}_\alpha(X;K);K) \tag{A5}$$

and

$$J_\alpha(X;K) = \frac{\partial \bar{K}_\alpha(X;K)}{\partial(X,G)}. \tag{A6}$$

Some straightforward formal manipulations give, for the derivative of $F_\alpha(K)$ with respect to K, the expression

$$\frac{\partial F_\alpha(K)}{\partial K} = \int_{U_\alpha(K)}\left(\frac{\partial f_\alpha J_\alpha}{\partial K}\right)dX$$
$$-\sum_\beta \int_{\partial U_\alpha(K)} f_\alpha J_\alpha \delta(h_{\beta\alpha})\frac{\partial h_{\beta\alpha}}{\partial K}\,dX, \tag{A7}$$

where

$$h_{\beta\alpha}(X;K) = \bar{k}_{j\beta\alpha}(X;K) - c_{\beta\alpha} = 0 \tag{A8}$$

are the equations for the surfaces in X bounding $U_\alpha(K)$. The function $\bar{k}_{j\beta\alpha}(X;K)$ is the component of $\bar{K}_\alpha(X;K)$ associated with the boundary surface $\beta\alpha$ of $U_\alpha(K)$.

The derivative with respect to K is to be interpreted, always, as the derivative on any variable upon which the k_i of K depend analytically. In particular, if the function is defined only over a restricted set, then the derivative is with respect to any appropriate local coordinate, in the sense of Def. 2.1. With this understanding, the derivative $\partial f_\alpha/\partial K$ exists over the closure of $\bar{U}_\alpha(K)$ by virtue of our original assumption. It is therefore also uniformly bounded over $U_\alpha(K)$. The derivative of J_α exists in the usual sense, hence also in the sense of Def. 2.1. It is therefore also uniformly bounded over $U_\alpha(K)$. The set $U_\alpha(K)$ is bounded and hence the first term on the right in (A7) is finite and, in the sense that it is independent of the phase of the variation dK, well defined.

The second term on the right-hand side of Eq. (A7) is also finite and well defined. The function $h_{\beta\alpha}$ is analytic in K in the usual sense, hence also in the sense of Def. 2.1. For any particular term β one can transform to a set of variables in which the $\bar{k}_i$ corresponding to the surface $\beta\alpha$ is one of the x_j. Then the $\delta(h_{\beta\alpha})$ just eliminates this one of the dx_j in dX. The remaining integral is finite because $U_\alpha(K)$, and hence $\partial U_\alpha(K)$, is of bounded extent. By virtue of the method of construction, the number of sides β is finite and hence so is the second term in (A7). The formal derivative (A7), is therefore finite and well defined.

Since the formal derivative $\partial F/\partial K$ given by Eq. (A7) is well defined, the function of $F(K)$ is analytic in K to the extent that the formal expression actually represents the limit of $\Delta F/\Delta K$. For real ΔK, this is true. However, for complex K, the meaning of $F(K)$ is not yet defined.

For complex K near a real $K_0 \in \mathcal{S}$, one can take $F(K)$ to be defined by Eq. (A4). The many-variable version of the Cauchy theorem[34] then permits the

[34] See Ref. 31, p. 264. It is important to note that if the region of integration $V = U_\alpha(K)$ is bounded in part by a portion of a manifold defined by $z_{j\gamma} = c_\gamma$, where $j\gamma \equiv j_\gamma$, then the values of the remaining z_j on the interior of this portion of the boundary can be varied in a smooth way through a region of analyticity of the integrand without altering the integral. This is because $dZ = dz_1 \wedge dz_2 \cdots \wedge dz_{n-m}$ is zero if any dz_j is zero, and hence the extra piece of contour gives no contribution.

contour for $F_\alpha(K_0 + \Delta K)$ to be taken to consist of a central part $U_\alpha(K_0)$ plus a boundary strip running between $\partial U_\alpha(K_0)$ and $\partial U_\alpha(K_0 + \Delta K)$. Using this form for the contour, one obtains Eq. (A7) as the limit of $\Delta F/\Delta K$ for all complex ΔK. Thus the function $F(K)$ defined by Eq. (A4) is analytic at points K_0 in $\mathcal{S}$.

For real K, the various possible ways of choosing the local coordinates and the $U_\alpha(K)$ all lead, via Eq. (A4), to the same function $F(K)$, by virtue of the factors J_α. Since the extension to complex K via any one of these choices gives an analytic function, the extension must be independent of the particular choice used in Eq. (A4).

Definition: A *local coordinate patch* in $\mathcal{R}(K)$ will mean the image in $\mathcal{R}(K)$, under an analytic one-to-one mapping $\bar{K}(X)$, of a bounded open set in the space of points $X = \{x_1, \cdots, x_{n-m}\}$. The set X is the set of local coordinates corresponding to the local coordinate patch. We further specify that the x_j be a subset of the set of $\bar{k}_j$. That this is possible follows from arguments given in Theorem A1.

Theorem A2: Let $\mathcal{R}(K)$ be the set of real $\bar{K}$ described in Theorem A1, and let $\{\bar{K}: G(\bar{K}; K) = 0\}$ be the set of all complex $\bar{K}$ satisfying the same conditions, $g_j(\bar{K}; K) = 0$ for all j. Suppose $R'(K) \subset \{\bar{K}: G(\bar{K}; K) = 0\}$ is the image of $\mathcal{R}(K)$ under a mapping $\bar{K} \to \bar{K}' = \bar{K} + iT'(\bar{K}; K)$, where $T'(\bar{K}; K)$ is real and continuous over $\mathcal{R}(K) \otimes \mathcal{S}$ with continuous (hence uniformly bounded) first derivatives with respect to the local coordinates corresponding to some (hence every) finite covering of $\mathcal{R}(K)$ by local coordinate patches. Suppose $\mathcal{R}'(K)$ is close to $\mathcal{R}(K)$ in the sense that the image $U'_\alpha(K)$ of $U_\alpha(K)$ is within the region where $\bar{K}_\alpha(X, G)$ is analytic with analytic inverse, the relevant minors of $\partial g_j/\partial \bar{k}_i$ still being nonzero. Suppose $f(\bar{K}; K)$ is analytic at points $(\bar{K}; K)$ of

$$\{(\bar{K}; K): K \in \mathcal{S}, \bar{K} \in R'(K)\}.$$

Then $F(K)$, defined by Eq. (A4), but with $U'_\alpha(K)$ in place of $U_\alpha(K)$, is analytic at K in $\mathcal{S}$, provided $\mathcal{R}'(K)$ is sufficiently close to $\mathcal{R}(K)$. [The final condition of closeness means that for some $\varepsilon > 0$ we have, using the metric in $\bar{K}$ space, $|T(\bar{K}; K)| < \varepsilon$ for all $\bar{K}$ in $\mathcal{R}(K)$, for any fixed point K of $\mathcal{S}$. Although this condition is used in the following proof, it probably is not necessary for the validity of the theorem. The explicit definition of $F(K)$ is given by Eq. (A9).]

Outline of Proof. The boundaries of the sets $U_\alpha(K)$ map into surfaces H_γ in $\mathcal{R}(K)$, the surface H_γ being, by construction, the intersection of $\mathcal{R}(K)$ with a portion of the manifold $\bar{k}_{i\gamma} = c_\gamma$. Here c_γ is a (real) rational number and $i\gamma$ means i_γ. The image of H_γ in $\mathcal{R}'(K)$ is H'_γ, on which $\bar{k}_{i\gamma}$ is given by $\bar{k}_{i\gamma} = c_\gamma + it'_{i\gamma}(\bar{K}/H_\gamma; K)$. The symbol $\bar{K}/H_\gamma$ means that $\bar{K}$ is considered restricted to H_γ.

Each of the real functions $t'_{i\gamma}(\bar{K}/H_\gamma; k)$ can, by virtue of the Weierstrass approximation theorem,[35] be approximated over H_γ, for fixed K, to arbitrary pointwise precision by a real polynomial $t''_{i\gamma}(\bar{K}; K)$ in the variables $\bar{k}_i$ of $\bar{K}$. Indeed all the $t'_{i\gamma}(\bar{K}/H_\gamma; K)$ having the same index i can be approximated by one single polynomial $t''_i(\bar{K}; K)$. We assume this is done so that the index $i\gamma$ on $t''_{i\gamma}(\bar{K}; K)$ can be interpreted as an i_γ. Sometimes, as in Eq. (A7), $\beta\alpha$ is used in place of γ to identify a boundary surface.

The $U_\alpha(K)$ can be selected so that for each point $\bar{K}$ of $\mathcal{R}(K)$ there is, in $\mathcal{R}(K)$, a coordinate patch $N(\bar{K})$ containing $\bar{K}$, such that the $\bar{k}_{i\gamma}$ associated with each H_γ that intersects $N(\bar{K})$ is a member of the set of local coordinates X corresponding to $N(\bar{K})$. If the original $U_\alpha(K)$ do not satisfy this condition, then the c_γ can be slightly shifted so that the condition is satisfied. [Suppose, for example, that a surface $h_\gamma(X) \equiv \bar{k}_{i\gamma}(X) - c_\gamma = 0$ intersects the intersection I_a of a set of coordinate surfaces $x_j = 0$, where j runs over the set $\{1, \cdots, a \leq n - m\}$. And suppose $\partial h_\gamma/\partial x_j = 0$, for $j = a + 1, \cdots, n - m$, at some point of $I_a \cap \{h_\gamma = 0\}$. This is a typical case where the gradient of h_γ is not independent of the gradients of some subset of the x_j and hence h_γ cannot be taken as one of the x_j. There may be a connected set of points in I_a for which this condition on the gradient remains satisfied, but all points of this set must lie at $\bar{k}_{i\gamma} = c_\gamma$. Thus a slight shift of c_γ will move this entire set of points on $\{h_\gamma = 0\}$ for which the gradient condition is satisfied out of the set I_a. A finite number of applications of this argument will give the required result.]

By the Heine–Borel covering theorem, the compact $\mathcal{R}(K)$ can be covered by a finite number of coordinate patches of the type specified above. These real neighborhoods $N(\bar{K})$ in $\mathcal{R}(K)$ can be extended to complex neighborhoods $N^*(\bar{K})$ in $\{\bar{K}: G(\bar{K}; K) = 0\}$ such that the mapping $\bar{K}(X)$ associated with $N(\bar{K})$ remains analytic and single-valued over $N^*(\bar{K})$. We shall require that, for some finite covering of $\mathcal{R}(K)$ by these local coordinate patches $N(\bar{K})$, the image in $\mathcal{R}'(K)$ of each $N(\bar{K})$ lies in $N^*(\bar{K})$. This requirement certainly can be satisfied if $R'(K)$ lies sufficiently close to $\mathcal{R}(K)$. This condition is far from necessary, however.

[35] See, e.g., R. Courant, *Methods of Mathematical Physics* (New York University, Institute for Mathematics and Mechanics, New York, 1950), p. 47.

The complexification of the neighborhoods $N(\bar{K})$ leads to a complexification $x_j \to z_j$ of the corresponding local coordinates. Because the z_j are independent variables over the corresponding $N^*(\bar{K})$, the requirement just imposed ensures that the values of the $\bar{k}'_{i\gamma}$ on H'_γ are independent variables. In particular, the $\bar{k}'_{i\gamma}$ associated with the various H'_γ can be simultaneously shifted by sufficiently small amounts without moving off the surface $\{\bar{K}: G(\bar{K}; K) = 0\}$.

Because the $\bar{k}'_{i\gamma}$ on H'_γ are independent variables, in this sense, a surface $\mathcal{R}''(K) \subset \{\bar{K}: G(\bar{K}; K) = 0\}$ can be defined by the mapping $\bar{K} \to \bar{K}'' = \bar{K} + iT''(\bar{K}; K)$, where the i_γ component of $T''(\bar{K}/H_\gamma; K)$ is a polynomial $t''_{i\gamma}(\bar{K}/H_\gamma; K)$ of the type discussed earlier, and where all components of $T''(\bar{K}; K)$ are, for fixed K, continuous in $\bar{K}$ over $\mathcal{R}(K)$ with continuous (hence uniformly bounded) first derivatives in the local coordinates of any fixed finite covering of $\mathcal{R}(K)$.

The surface $\mathcal{R}''(K)$ can be made to lie arbitrarily close to $\mathcal{R}'(K)$. Thus, by virtue of the many-variable Cauchy theorem,[34] the contour can be taken to run over $\mathcal{R}'(K)$ instead of $\mathcal{R}'(K)$, without changing the value of the integral.

The construction described above is carried out for the original real $K = K_0$ in S. For nearby real K, the boundaries of the real $U_\alpha(K)$ are taken to be defined by the same equations $h_\gamma = 0$ that are used at $K = K_0$. The boundaries of the images $U''_\alpha(K)$ of $U_\alpha(K)$, under $\bar{K} \to \bar{K}''$, are defined by taking the $t''_{i\gamma}(\bar{K}; K)$ to be independent of K. This can be done because they are independent variables, in the sense discussed above.

The function $F(K)$ is defined by Eq. (A4), but with the X in $\bar{K}_\alpha(X; K)$ replaced by $Z''_\alpha(X; K)$, which is the function that maps $U_\alpha(K)$ onto $U''_\alpha(K)$. In particular, we have

$$F_\alpha(K) = \int_{U''_\alpha(K)} f_\alpha(Z''; K) \frac{\partial \bar{K}_\alpha(Z''; K)}{\partial Z} dZ''$$

$$= \int_{U_\alpha(K)} f_\alpha(Z''_\alpha(X; K); K) \times \frac{\partial \bar{K}_\alpha(Z''_\alpha(X; K); K)}{\partial Z} \frac{\partial Z''}{\partial X} dX, \quad \text{(A9)}$$

where

$$f_\alpha(Z''; K) \equiv f(\bar{K}_\alpha(Z''; K); K).$$

As the real K varies from its original value K_0, certain of the boundaries of the $U_\alpha(K)$ may move. The integral $F_\alpha(K)$ can be considered to be composed of a central part lying over the fixed $U_\alpha(K_0)$ plus a boundary part that is the strip connecting $\partial U''_\alpha(K_0)$ to $\partial U''_\alpha(K)$. By virtue of Cauchy's theorem,[34] applied to the first form in Eq. (A9), the exact shape of the interior of the contour $U''_\alpha(K)$ is not important; it can be slightly shifted without changing $F_\alpha(K)$.

Because of this freedom in the choice of contour, the location of central part of the contour, lying over $U(K_0)$, can be prescribed by taking the function $Z''_\alpha(X; K)$ to be independent of K. This function is then analytic in K and, consequently, so is the part of the integrand in Eq. (A9) lying over $U(K_0)$. Thus, this central contribution to $F_\alpha(K)$, which corresponds to the first term in Eq. (A7), is analytic in K.

For the calculation of the contribution to $F_\alpha(K)$ coming from the boundary strip near $H_{\beta\alpha} = H_\gamma$ we choose a set of local coordinates X_γ which has an element $x_{j\gamma}$ that is $\bar{k}_{i\gamma}$. (We may need several such coordinate systems to cover H_γ, but a finite number will certainly suffice.) In this coordinate system γ the equation for H''_γ is

$$z''_{i\gamma} \equiv \bar{k}''_{i\gamma} = c_\gamma + it''_{i\gamma}(\bar{K}_\gamma(X_\gamma; K)/H_\gamma)$$
$$\equiv z''_{j\gamma}(X_\gamma/H_\gamma; K),$$

which, by the theorem on compositions of analytic functions,[36] is analytic in K. The meaning of X_γ/H_γ is evident.

The other edge of this boundary strip lies on $\partial U''(K_0)$, hence on the image in γ space of

$$Z''_{i\gamma}(X_\alpha/H_\gamma(K_0); K_0),$$

which is

$$\hat{Z}''_\gamma(X_\gamma/H_\gamma; K) \equiv \hat{Z}_\gamma[\bar{K}_\alpha(Z''_\alpha\{X_\alpha[\bar{K}_\gamma(X_\gamma/H_\gamma; K_0)]; K_0\}; K)].$$

The i_γ component of this equation is

$$\hat{z}''_{i\gamma}(X_\gamma/H_\gamma; K) = \bar{k}_{i\beta\alpha}(Z''_\alpha\{X_\gamma[\bar{K}_\gamma(X_\gamma/H_\gamma; K_0)]; K_0\}; K),$$

where $\bar{k}_{i\beta\alpha}(Z''_\alpha; K)$ is the $\beta\alpha = \gamma$ component of the function $\bar{K}_\alpha(Z''_\alpha; K)$. This function $\hat{z}''_{j\gamma}(X_\gamma/H_\gamma; K)$ is analytic in K, and hence so is

$$\Delta z''_{j\gamma}(X_\gamma/H_\gamma; K) \equiv z''_{i'}(X_\gamma/H_\gamma; K) - \hat{z}''_{j\gamma}(X_\gamma/H_\gamma; K).$$

Let $\mathcal{Z}''_\gamma$ represent the set of coordinates other than $z''_{j\gamma} = \bar{k}''_{i\gamma}$ in the set Z''_λ, and let $\bar{X}_\gamma$ be defined similarly. The value of $\mathcal{Z}''_\gamma$ on H''_γ has not been specified so far. The point is that the contour in $\mathcal{Z}''_\gamma$ can be slightly shifted, keeping $z''_{i\gamma}(X_\gamma/H_\gamma; K)$ fixed, without altering the value of $F_\alpha(K)$. This is because the contribution to $F_\alpha(K)$ from a piece of the contour confined to H''_γ vanishes, because of the vanishing of $dz''_{i\gamma}$. This result is familiar in simple cases, where the shifting of the contour in $\mathcal{Z}_\gamma$ space is justified by the Cauchy theorem in $\mathcal{Z}_\gamma$ space.

Since the exact value of $\mathcal{Z}''_\gamma(X_\gamma/H_\gamma; K)$ is not important, we shall leave it unspecified, except to require that the surface $\mathcal{R}''(K)$ be smooth (i.e., continuous

[36] See, e.g., Ref. 30, Theorem 5, p. 6.

with continuous first derivatives with respect to the variables of some local coordinate system).

The contribution to $F_\alpha(K)$ from the boundary strip near H_γ'' is then given (up to a sign perhaps) by

$$\int_{H_\gamma} f_\gamma(Z_\gamma''(X_\gamma/H_\gamma;\, K_0);\, K_0)\, \frac{\partial \bar{K}_\gamma(Z_\gamma''(X_\gamma/H_\gamma;\, K_0);\, K_0)}{\partial Z_\gamma''} \times \frac{\partial Z_\gamma''(X_\gamma/H_\gamma;\, K_0)}{\partial \bar{X}_\gamma} \times \Delta z_{j\gamma}''(X_\gamma/H_\gamma;\, K)\, d\bar{X}_\gamma$$

plus higher-order terms in ΔK. The dependence on K is through the analytic function $\Delta z_{j\gamma}''(X_\gamma/H_\gamma;\, K)$. Thus the limit $\Delta z_{i\gamma}''/\Delta K$ will be well defined (i.e., independent of the phase of ΔK). The analyticity of $F_\alpha(K)$ then follows by the same arguments that were used in Theorem A1.

APPENDIX B: THE PHASE FACTOR IN THE POLE-FACTORIZATION THEOREM

The phase factor of some particular contribution "a" in the fourth term of Fig. 3 is $\tilde{\alpha}_{a1}\tilde{\alpha}_{a2}^*\tilde{\alpha}_{a3}$. Here $\tilde{\alpha}_{ai}$ is the phase α_p of the particular partition of the ith factor that leads to the contribution a. For the first term of Fig. 6 we have, instead, $\alpha_{a1}\alpha_{a2}^*\alpha_{a3}$. If one considers various contributions $c = (a, b, \cdots)$ that correspond to a fixed set of internal lines of Fig. 3, then it follows from Postulate *E2* that the ratios $\tilde{\alpha}_{ci}/\alpha_{ci}$ are the same for all these contributions. In this particular application of *E2* the "remaining set of variables R" is just the single variable p, and it follows from Eq. (2.30h) that

$$\frac{\tilde{\alpha}_{ai}}{\alpha_{ai}} = \frac{\tilde{\alpha}_{bi}}{\alpha_{bi}} = \cdots.$$

Thus one can evaluate $r = \tilde{\alpha}_{a1}\tilde{\alpha}_{a2}^*\tilde{\alpha}_{a3}/\alpha_{a1}\alpha_{a2}^*\alpha_{a3}$ by choosing any convenient partitions for the three factors in Fig. 3.

According to the above result, one can calculate r by choosing for the three partitions in Fig. 3 just the first partitions, which give simply the connected parts. That is, the ratio r can be expressed as

$$r = \frac{(\alpha_1\alpha_2^*\alpha_3)_{B1}}{(\alpha_1\alpha_2^*\alpha_3)_{B2}},$$

where the α_i's in the numerator and denominator are the α_p's for the three factors in the functions represented by the diagrams in Figs. B1 and B2, respectively.

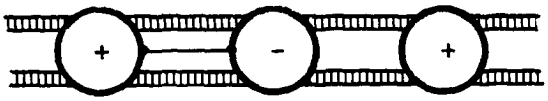


FIG. B1. A diagram related to the fourth term of Fig. 3. The shaded strips represent sets of freely moving lines.

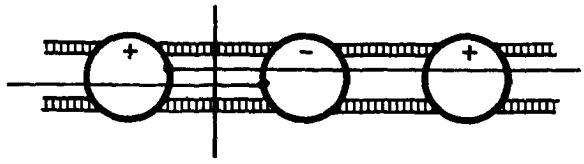


FIG. B2. A diagram related to the first term in Fig. 6.

If one interchanges the two identical initial particles of the M function represented by the part of Fig. B2 lying left of the vertical line, one obtains the function represented by Fig. B3.

This interchange changes the M function by a factor σ_p. Thus $(\alpha_1\alpha_2^*\alpha_3)_{B2} = \sigma_p(\alpha_1\alpha_2^*\alpha_3)_{B3}$ and one can write

$$r = (\alpha_1\alpha_2^*\alpha_3)_{B1}/\sigma_p(\alpha_1\alpha_2^*\alpha_3)_{B3}.$$

Using *E3* one finds that $(\alpha_1\alpha_2^*\alpha_3)_{B3}$ is equal to α_{B4}, where α_{B4} is the α_p of the decomposition shown in Fig. B4.

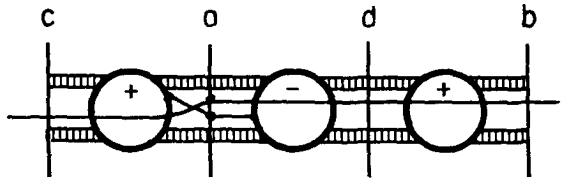


FIG. B3. Diagram representing functions obtained by interchanging the two identical initial particles of the M function represented by the part of Fig. B2 lying left of the vertical line.

The equality

$$(\alpha_1\alpha_2^*\alpha_3)_{B3} = \alpha_{B4}$$

is obtained by writing it in the form

$$\alpha_{ac}\alpha_{ad}^{-1}\alpha_{bd} = \alpha_{bc},$$

which is a special case of Eq. (2.30i). For this special case n is 2, and the first subsets of a, b, c and d consist of just p itself.

By virtue of Eq. (2.30d), one has $(\alpha_1\alpha_2^*\alpha_3)_{B1} = 1$. The $\alpha_p = \alpha_{B4}$ is just the α_b of the pole-factorization theorem.

In the proof of spin and statistics, this result for the phase of the residue was not used. Rather, it argued

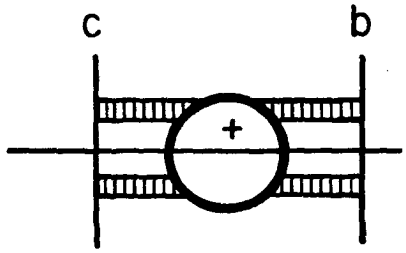


FIG. B4. A diagram with the same connectedness structure as the diagram in Fig. B3.

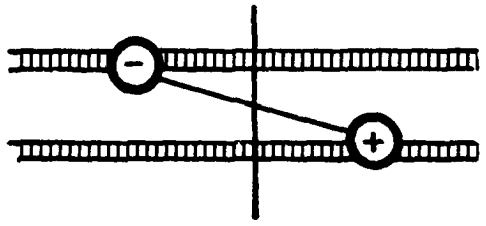


FIG. B5. The term in the unitarity equation corresponding to a pole term.

that the phase must in fact be precisely i times the phase of the corresponding term in the unitarity equation. This corresponding term in the unitarity equation is shown in Fig. B5.

The phase of this term is $\alpha = \alpha_1\alpha_2^*$, where α_1 and α_2^* are the phases α_p of the cluster decomposition equation for the two factors of Fig. B5.

By virtue of Postulates *E2* and *E3*, the over-all phase that multiplies the fourth term in Fig. 3, when the α_p are taken into account, is just $\alpha = \alpha_1\alpha_2^*$; aside from this one over-all phase, the function is given by the indicated product of M functions, each of which is defined as if acted alone. The point is that *E2* insures that the relative phases of different contributions associated with any given set of internal lines are the same as they would be if the parts acted alone. And *E3* ensures that the relative phases of contributions associated with different sets of internal lines are just as they would be if the various parts acted alone.

This over-all phase factor α multiplies the function represented by the fourth term of Fig. 3. Since the boxes represent the M functions with the phases they would have if they acted alone, one may invoke the equation represented by Fig. 4 to obtain the second term of Fig. 5, but now with the phase α.

In order that the fifth and six terms of Fig. 3 cancel it is necessary that over-all phases that multiply them be equal. This follows from Postulates *E2* and *E3*. The relevant diagrams are shown in Fig. B6.

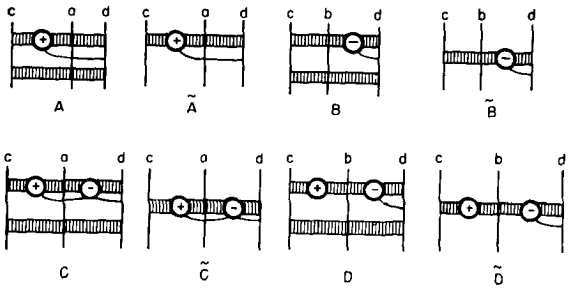


FIG. B6. Diagrams used to show that the phase $(\alpha_1\alpha_2^*)_A = (\alpha_1\alpha_2^*)_B$.

The proof is as follows: By virtue of *E2*, we have

$$\frac{(\alpha_1\alpha_2^*)_A}{(\alpha_1\alpha_2^*)_{\tilde{A}}} = \frac{(\alpha_1\alpha_2^*)_C}{(\alpha_1\alpha_2^*)_{\tilde{C}}}$$

and

$$\frac{(\alpha_1\alpha_2^*)_B}{(\alpha_1\alpha_2^*)_{\tilde{B}}} = \frac{(\alpha_1\alpha_2^*)_D}{(\alpha_1\alpha_2^*)_{\tilde{D}}}.$$

But from Postulate *E3*, we have

$$(\alpha_1\alpha_2^*)_C = (\alpha_1\alpha_2^*)_D.$$

On the other hand, Eq. (2.30d) gives

$$(\alpha_1\alpha_2^*)_{\tilde{C}} = (\alpha_1\alpha_2^*)_{\tilde{D}} = 1.$$

Inserting this into the above equation, one obtains

$$\frac{(\alpha_1\alpha_2^*)_A}{(\alpha_1\alpha_2^*)_B} = \frac{(\alpha_1\alpha_2^*)_{\tilde{A}}}{(\alpha_1\alpha_2^*)_{\tilde{B}}}.$$

If one takes $a = d$ and $c = d$, then the right-hand side of this equation is unity, by virtue of Eq. (2.30d).

APPENDIX C

In Ref. 13, $M_c(K)$ is shown to consist of a sum of terms $M_c^i(K)$, each of which is a limit to physical points of a function analytic in a region containing physical points as boundary points. In particular, we have

$$M_c^i[\varphi] = \int M_c^i(K)\varphi(K)\,dK$$
$$= \lim_{\epsilon\to 0^+} \int M_c^i(K + i\in\Delta^i(K))\varphi(K)\,dK,$$

where K is here considered a point in a real vector space of dimension $(3N-4)$, and $\Delta^i(K)$ is a point in the same space (i.e., local linear coordinates are introduced). Certain properties of such functions will now be derived.

Lemma C1: Let $\mathfrak{D}_R$ be the space of C^∞ test functions with support confined to R, the closure of the bounded open set R' in the space R_n of n real runbers. Let $f[\varphi]$ be a functional of φ such that for any φ in $\mathfrak{D}_R$

$$f[\varphi] = \lim_{\epsilon\to 0^+} \int f(x + i\in\Delta(x))\varphi(x)\,dx, \qquad \text{(C1)}$$

where x and $\Delta(x)$ are elements of R_n, the components of $\Delta(z)$ are entire functions of $z = x + iy$, and $f(z)$ is analytic in the strip

$$S = \{z: z = x + i\in\Delta(x), x\in R, 0 < \epsilon < \eta > 0\}.$$

If $\varphi(x)$ in $\mathfrak{D}_R$ is, at points x of R', the restriction to R' of a function $\varphi(z)$ analytic and uniformly bounded in

$$S'' = \{z: \mathrm{Re}\, z\in R', |\mathrm{Im}\, z| < \rho(x)\},$$

where $\rho(x)$ is continuous and positive for x in R', then

$$f[\varphi] = \lim_{\epsilon \to 0^+} \int_{R-R''} f(x + iy(x) + i\epsilon\Delta(z)) \times \varphi(x + iy(x))\, d(x + iy(x)) + \int_{R''} f(x + iy(x))\varphi(x + iy(x))\, d(x + iy(x)), \tag{C2}$$

where the closure of R'' is in R' and $y(x)$ is any continuous real function that is zero for x not in R', and for x in R' gives a $z = x + iy(x)$ that is in $S \cap S'$. The set of points $x + iy(x)$ can be considered a contour C lying over the real points x. Then Eq. (C2) can be written in the more compact form

$$f[\varphi] = \lim_{\epsilon \to 0^+} \int_{C(R-R'')} f(z + i\epsilon\Delta(z))\varphi(z)\, dz + \int_{C(R'')} f(z,\varphi(z)\, dz, \tag{C2'}$$

where $C(T)$ is the part of C lying over T.

Proof: By Cauchy's theorem, generalized to several variables,[34] Eq. (C1) is equivalent to

$$f[\varphi] = \lim_{\epsilon \to 0^+} \int_C f(z + i\epsilon\Delta(z))\varphi(z)\, dz, \tag{C3}$$

since, for some $\epsilon' > 0$ and every fixed $0 < \epsilon < \epsilon'$, the functions are analytic in the region through which the contour is shifted. The points on the boundary of R', where the $\varphi(z)$ are not analytic, give no contribution because of the boundedness condition on $\varphi(z)$. For the part $C(R'')$ of C the limit $\epsilon \to 0$ can be taken, since the integrand is analytic, hence continuous, in ϵ at these points.

Lemma C2: If the conditions of Lemma C1 are satisfied and if $|f(z)|$ is bounded over the intersection of S with some neighborhood of $R - R'$, then

$$f[\varphi] = \int_C f(z)\varphi(z)\, dz. \tag{C2''}$$

Proof: The boundedness of $f(z)$ and $\varphi(z)$ assures that the contribution from $R' - R''$ vanishes as $R'' \to R'$.

Corollary C1: If $f[\varphi]$ and φ satisfy the conditions of Lemma C1 and C2, then $f[\varphi]$ is finite (noninfinite).

Proof: The right-hand side of Eq. (C2'') is finite.

Corollary C2: If in place of Eq. (C1) we have

$$f[\varphi] = \lim_{\epsilon \to 0^+} \int [f^+(x + i\epsilon\Delta^+(x)) - f^-(x - i\epsilon\Delta^-(x)]\varphi(x)\, dx,$$

and if $f^{\pm}(z)$ and $\varphi(x)$ satisfy the conditions of Lemmas C1 and C2, then we have, in place of Eq. (C2''), rather

$$f[\varphi] = \int_{C^+} f^+(z)\varphi(z)\, dz - \int_{C^-} f^-(z)\varphi(z)\, dz. \tag{C4}$$

Proof: The manipulations of the proof of Lemma C2 can be carried through for each term separately.

In the following theorems, x is a single real variable (i.e., $n = 1$). The set R' will be the real set $R' = \{x: |x| < a\}$.

Theorem C1: Let $f[\varphi]$ be a functional of φ such that for any φ in $\mathfrak{D}_R$

$$f[\varphi] = \lim_{\epsilon \to 0^+} \int [f^+(x + i\epsilon) - f^-(x - i\epsilon)]\varphi(x)\, dx,$$

where the $f^{\pm}$ are analytic in the strips

$$S^{\pm} = \{z: \operatorname{Re} Z \in R, 0 < \pm \operatorname{Im} z < \eta > 0\}.$$

Suppose the $|f^{\pm}(z)|$ are bounded in $S^{\pm} \cap N$, where N is a neighborhood of $R - R'$. Then the vanishing of $f[\varphi]$ for all φ in $\mathfrak{D}_R$ implies that the limits $f^{\pm}(x)$ exist, are analytic, and are equal, for all x in R'.

Proof: By virtue of Corollary C2, one has

$$f[\varphi] = \int_{C^+} f^+(z)\varphi(z)\, dz - \int_{C^-} f^-(z)\varphi(z)\, dz \tag{C5}$$

for any $\varphi(z)$ satisfying the conditions of the Lemma C1, where C^+ and C^- are certain (compact) contours from $x = -a$ to $x = a$ that lie just above and below the real axis for $|x| < a$. Let $\psi(x)$ be the function $\psi(z) = \exp[-(a^2 - z^2)^{-1}]$ and let $\varphi(z; z') \equiv \psi(z) \times (z - z')^{-1}$. Then define

$$F(z') = \int_{C^+ + \bar{C}^-} f^+(z)\varphi(z; z')\, dz - \int_{C^- + \bar{C}^-} f^-(z)\varphi(z; z')\, dz, \tag{C6}$$

where $\bar{C}^+$ is a contour from a to $-a$ that lies inside S^+ and above C^+ for $|x| < a - b$; and $\bar{C}^-$ is a contour from a to $-a$ that lies inside S^- and below C^- for $|x| < a - b$. For $a \geq |x| \geq a - b > 0$, the contours $\bar{C}^{\pm}$ are taken to coincide with $-C^{\pm}$, respectively. By Cauchy's theorem, $F(z') = 2\pi i f^+(z')\psi(z')$ if z' is between C^+ and $\bar{C}^+$, and $F(z') = 2\pi i f^-(z')\psi(z')$ if z' is between C^- and $\bar{C}^-$. By virtue of the vanishing of Eq. (C5), one also has for z' in either of these two regions

$$F(z') = \int_{C^+} \frac{f^+(z)\psi(z)}{(z - z')}\, dz - \int_{\bar{C}^-} \frac{f^-(z)\psi(z)}{(z - z')}\, dz. \tag{C7}$$

In view of the analyticity and boundedness conditions on $f^{\pm}$ and φ, Eq. (C7) implies that $F(z')$ is a single

analytic function throughout the interior of $\bar{C}^+ - \bar{C}^-$. This implies that $f^+(z')$ and $f^-(z')$ are both equal to $F(z')/2\pi i\psi(z')$, and hence are analytic, inside $\bar{C}^+ - \bar{C}^-$.

Theorem C2: The condition of boundedness on $|f^\pm|$ in the statement of Theorem C1 can be replaced by the condition that $|f^\pm(z)|$ be bounded in $S^\pm \cap N$ by $C \exp B\,|\mathrm{Im}\, z|^{-m}$ for some positive values of the constants C, B, and m.

Proof: In the proof of Theorem C1 one can replace the function $\psi(3) = \exp[-(a^2 - z^2)^{-1}]$ by

$$\exp[-A(a^2 - z^2)^{-m}].$$

And the curve $z(x)$ can be taken to approach the endpoints at $x = \mp a$ along the lines $\arg(z \pm a) = \pm\pi/4m$. Then for points on the contour sufficiently near $x = \pm a$ the function $|\psi(z)|$ is less than $\exp -A'\,|\mathrm{Im}\, z|^{-m}$, where $A' \equiv A[\sin(\pi/4m)]^m\,(2a)^{-m}/2$. If one chooses A so that $A' > B$, then $|f^\pm\varphi|$ is bounded near $x = \pm a$. The same argument also shows that $|f^\pm(z)\varphi(z; z')|$ is bounded for z near $\pm a$, if $z' \neq \pm a$. But it is the boundedness of these products, rather than of the $|f^\pm|$ themselves, that is actually needed both in Lemma C2 and its corollaries and in the proof of Theorem C1 itself. Thus the argument in that proof carries over immediately to the present case.

Remark C1: Theorems similar to C2 have been proved by other authors[37,38] under the more stringent assumption that $f[\varphi]$ is a distribution. This distribution assumption demands that $f(z)$ be bounded near $\mathrm{Im}\, z = 0$ by some negative power of $|\mathrm{Im}\, z|$.[39] It is not clear that we wish to impose such a strict requirement on the allowed functions. In fact, from the S-matrix view point it is natural to allow all functionals $f[\varphi]$ on $\mathfrak{D}_R$ that can be expressed as sums of limits of analytic functions. Theorem C2 is a step in this direction. The condition, required in this theorem, that f be bounded by an exponential of an inverse power, while already very weak, can be much further weakened by replacing the constant power m by $C' \exp B'\,|\mathrm{Im}\, z|^{-m'}$. Moreover, this new m can again be replaced in the same way, and so on. Thus the bound on $f(z)$ can be made extremely weak. Whether the boundedness condition can be removed altogether is still an open question, as far as I know.

[37] H. G. Tillmann, Math. Z. 77, 106 (1961).

[38] See R. F. Streater and A. S. Wightman, *PCT, Spin, and Statistics and All That* (W. A. Benjamin Inc., New York, 1964), Sec. 2-5. In both this reference and the preceding one the main interest is the many-dimensional case.

[39] See Ref. 37 or Theorem 2-10 of Ref. 38.

Theorem C3: If a functional $f[\varphi]$ has a finite value for all infinitely differentiable functions $\varphi(x)$ of compact support $\mathcal{R}$ that are restrictions to $\mathcal{R}$ of functions analytic in the interior of R, and if for such φ the functional is given by

$$f[\varphi] = \lim_{\epsilon \to 0} \int [f^+(x + i\epsilon) - f^-(x - i\epsilon)]\varphi(x)\,dx, \quad \text{(C8)}$$

where the functions $f^\pm(z)$ are analytic in the strips

$$S^\pm = \{z : \mathrm{Re}\, z \in R,\ 0 < \pm\,\mathrm{Im}\, z < \eta > 0\} \quad \text{(C9)}$$

and are bounded at points of $S^\pm$ near the boundary of R by $C \exp B\,|\mathrm{Im}\, z|^{-m}$, for some positive values of constants, C, B, and m, then the functions $f^\pm(z)$ are unique up to a common additive function that is analytic at interior points of R.

Proof: This follows from Theorem C2 by taking the $f^\pm$ of that theorem to be differences of possible functions $f^\pm$ of this theorem. The extra condition in this theorem that φ be analytic at interior points of $\mathcal{R}'$ does not alter the proofs, since only functions having this property were used.

Remark C2: The particular boundedness condition used in Theorem C3 can, according to the Remark C1, be greatly weakened, if the need should arise. Also, class of φ for which Eq. (C8) holds can be much further restricted, if the need should arise.

APPENDIX D

The mass-shell paths C_{ij}, connecting various crossed reaction and Hermitian conjugate points, are constructed by following the distortions of paths $C_{ij}(a_p)$ as the effective masses a_p increase from zero to μ_p^2. At the start, where all $a_p = 0$, the $C_{ij}(a_p)$ all reduce to a single common point. As the a_p increase, the various end points of the $C_{ij}(a_p)$ move along definite singularity-free paths and the interiors of the $C_{ij}(a_p)$ are distorted so as to avoid a certain set of Landau surfaces $\mathcal{S}$.

In this continuation in a_p, the various $C_{ij}(a_p)$ may be distorted in such a way that it becomes possible to find a closed loop lying in the set of C_{ij}. Since no surfaces of the specified set $\mathcal{S}$ cross this loop as it develops from a single point at $a_p = 0$ to its form at $a_p = \mu_p^2$, one can say that, in a certain sense, none of these surfaces lies "within" the loop. Thus one might expect that the mass-shell loop should be able to be shrunk (*staying within the mass shell*) to a point, without crossing any of these surfaces. This is in fact true, within limits, if the set of surfaces $\mathcal{S}$ is such that it is possible to construct some function singular on just this set $\mathcal{S}$.

This result is proved by an application of the continuity theorem for functions of several complex variables.[40] First it can be noted that the actual loop, as it grows from a point to its final form, can, at each stage, be approximated to arbitrary precision (pointwise) by a curve that is a boundary of a disc lying on an analytic manifold. In particular, if the equation for the loop at any particular value of the a_p is given in terms of a set of mass-shell variables z_i by $z_i = z_i(\theta)$, where θ is a cyclic variable, then these equations can be approximated to arbitrary pointwise precision[41] by the expansion

$$z_i \simeq \sum_{-N}^{N} (\exp i\theta)^n c_n$$

$$= \bar{z}_i(\exp i\theta).$$

The surface $\bar{z}_i(z)$ is a one (complex) dimensional analytic manifold.[30] The curve $\{\bar{z}_i(z); |z| = 1\}$ passes arbitrarily close to the original curve.

If the original curve is always confined to a bounded region, as we shall suppose, and remains at more than some finite minimum distance from S throughout the contribution, which we can suppose, then N can be held fixed over the entire journey from $a_p = 0$ to $a_p = \mu_p$. Since the boundary curve $\{z_i(z); |z| = 1\}$ crosses no singularity of S, neither can the interior points $[z_i(z), |z| < 1]$, by virtue of the continuity theorem. Thus finally at $a_p = \mu_p$ we can shrink the curve to a point by the transformation $|z| \to 0$.

The above argument applies within the limit set by the requirement that a system of analytic local coordinates z_i can be found such that the loop lies within the coordinate patch corresponding to these coordinates. Although the question can be pursued further, it is simpler to restrict the paths so that no closed loop occurs within the set of paths.

For the set of analytic functions with the singularities lying on S (at least within the relevant region) we have in mind the truncated or renormalized perturbation theory functions. It remains to be shown, however, that the restriction imposed on S by the requirement that only positive-α singularities enter the physical region of the larger process actually forces S to be identical to the analogous set of singularity surfaces in perturbation theory.

---

[40] See the article by A. S. Wightman in the Les Houches Lectures, *Dispersion Relations and Elementary Particles*, R. Omnes and C. Dewitt, Eds. (John Wiley & Sons, Inc., New York, 1960), p. 260.

[41] G. H. Hardy and W. W. Rogosinski, *Fourier Series* (Cambridge University Press, London, 1950), p.21.

TCP Invariance and the Dimensionality of Space-Time*

S. P. ROSEN
Department of Physics, Purdue University, Lafayette, Indiana

(Received 19 January 1968)

The original proofs of the *TCP* theorem suggest that it may breakdown in an odd-dimensional space–time. An explicit example of this breakdown is given and a comment is made about the current × current interaction.

In relativistic quantum field theory, the *TCP* theorem[1,2] is usually proved by demonstrating the equivalence between the *TCP* operation and the combined operation of strong reflection plus Hermitian conjugation. Under strong reflection[1] (SR), all space–time coordinates are reflected through the origin

$$\mathbf{r} \to -\mathbf{r}, \quad t \to -t \tag{1}$$

and the order of field operators in any product is reversed. Any relativistically covariant Lagrangian is SR-invariant[2] as long as it is properly symmetrized in accordance with the spin-statistics theorem,[3] and it will be *TCP*-invariant if it is Hermitian.

The crux of this argument is Eq. (1). Because space–time has four dimensions, the reflection of coordinates has determinant equal to $(+1)$, and it is connected continuously, albeit by complex Lorentz transformations, with the identity.[4] Consequently, there exist no "pseudo" tensors with respect to SR, and the phase with which a tensor transforms is determined by its rank and the spin-statistics theorem.

Suppose now that we consider an odd-dimensional space–time. In this case the reflection of coordinates has determinant (-1) and it is not connected with the identity. Therefore there may exist both "pseudo" and "proper" tensors with respect to coordinate reflection, and we can expect a breakdown of the *TCP* theorem. To the best of the present writer's knowledge, there is no example of such a breakdown in the literature, and the purpose of this article is to present one.

We construct the analog of the Dirac equation for a $(2n + 1)$ space–time in which x_0 is the time coordinate and the x_k $(k = 1, \cdots, 2n)$ are the spatial coordinates. The required equations are

$$\begin{aligned} (\Gamma_\mu \partial^\mu + m)\Psi(x) &= 0, \\ \bar{\Psi}(x)(\Gamma_\mu \partial^\mu - m) &= 0, \end{aligned} \tag{2}$$

where

$$\bar{\Psi}(x) = -i\Psi^+(x)\Gamma_0 \tag{3}$$

and

$$\begin{aligned} &\Gamma_\mu \Gamma_\nu + \Gamma_\nu \Gamma_\mu = 2g_{\mu\nu}, \\ &\Gamma_0^+ = -\Gamma_0, \quad \Gamma_k^+ = +\Gamma_k \quad (k = 1, \cdots, 2n), \\ &g_{kk} = -g_{00} = 1. \end{aligned} \tag{4}$$

Because there are an even number of spatial dimensions, the parity transformation

$$\begin{aligned} x_k &\to -x_k, \\ x_0 &\to x_0 \end{aligned} \tag{5}$$

is an element of the appropriate spatial rotation group, and Eq. (2) is automatically invariant under it. The question of *TCP* invariance then reduces to one of *TC* invariance.

Following Luders,[2] we define the charge conjugation ($\mathcal{C}$) and time-reversal ($\mathcal{T}$) operations as

$$\begin{aligned} \mathcal{C}\Psi(x_0, x_k)\mathcal{C}^{-1} &= \eta_c \bar{\Psi}(x_0, x_k)C, \\ \mathcal{C}\bar{\Psi}(x_0, x_k)\mathcal{C}^{-1} &= -\eta_c C^+ \Psi(x_0, x_k), \end{aligned} \tag{6}$$

$$\begin{aligned} \mathcal{T}\Psi(x_0, x_k)\mathcal{T}^{-1} &= \eta_T T\Psi(-x_0, x_k), \\ \mathcal{T}\bar{\Psi}(x_0, x_k)\mathcal{T}^{-1} &= \eta_T^* \bar{\Psi}(-x_0, x_k)T^+, \end{aligned} \tag{7}$$

where the η_x are intrinsic phases associated with the field $\Psi(x)$. $\mathcal{C}$ is a linear transformation and $\mathcal{T}$ an antilinear one. Equation (2) will be invariant under $\mathcal{C}$ and $\mathcal{T}$ if we can find unitary matrices C and T such that[5]

$$\begin{aligned} C\tilde{\Gamma}_\mu C^{-1} &= -\Gamma_\mu, \\ T^{-1}\tilde{\Gamma}_\mu T &= \Gamma_\mu, \end{aligned} \tag{8}$$

where the tilde denotes transpose. A necessary condition for Eq. (8) is that there exist a matrix $S = CT$ with the property

$$S\Gamma_\mu S^{-1} = -\Gamma_\mu. \tag{9}$$

Clearly S is associated with strong reflections.

* Supported in part by the U.S. Atomic Energy Commission and by the U.S. Air Force.

[1] W. Pauli, *Niels Bohr and the Development of Physics* (Pergamon Press, Inc., London, 1955), p. 30.

[2] G. Luders, Ann. Phys. (N.Y.) **2**, 1 (1957).

[3] G. Luders and B. Zumino, Phys. Rev. **110**, 1450 (1958). See also R. F. Streater and A. S. Wightman, *PCT, Spin, and Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964).

[4] An explicit example of the complex Lorentz transformation is given by J. J. Sakurai, *Invariance Principles and Elementary Particles* (Princeton University Press, Princeton, N.J., 1964), Chap. 6.

[5] The Dirac field Ψ and its adjoint $\bar{\Psi}$ are assumed to anticommute with each other.

Let us consider the lowest representation of the Γ_μ. In $(2+1)$ space–time it is given by Pauli matrices

$$(\Gamma_0^{(3)}, \Gamma_1^{(3)}, \Gamma_2^{(3)}) = (i\sigma_3, \sigma_2, -\sigma_1) \tag{10}$$

and in $(4+1)$ space–time it is given by the Dirac matrices $\gamma_1, \cdots, \gamma_5$. Together with the unit matrix, the $\Gamma_\mu^{(3)}$ form a complete set, and any other (2×2) matrix must be a linear combination of them. Consequently, it is not possible to find a matrix S satisfying Eq. (9), and the $(2+1)$ Dirac equation is not *TC*-invariant. Similarly, the $(4+1)$ Dirac equation cannot be *TC*-invariant because any (4×4) matrix is a linear combination of γ_λ and $\gamma_\mu\gamma_\nu$, with $\lambda, \mu, \nu = 1, \cdots, 5$.

It is not difficult to generalize this result to the $(2n+1)$ Dirac equation. The lowest representation of the Γ_μ in $(2n+1)$ space–time can be constructed from the corresponding representation in $(2n-1)$ space–time by means of the Kronecker products[6,7]

$$\Gamma_\mu^{(2n+1)} \equiv (\Gamma_\nu^{(2n-1)} \otimes \sigma_2, I \otimes \sigma_3, I \otimes \sigma_1), \tag{11}$$

where I is a unit $(2n-2) \times (2n-2)$ matrix. As defined in Eqs. (10) and (11), the $\Gamma_\mu^{(2n+1)}$ form a maximal set,[6] and any other $(2n) \times (2n)$ matrix can be expanded as a linear combination of r-fold products of the $\Gamma_\mu^{(2n+1)}$, with $r = 1, \cdots, (n+1)$. Consequently, there exists no $(2n) \times (2n)$ matrix S which satisfies Eq. (9), and the Dirac equation cannot be *TC*-invariant.

Up to now we have taken it for granted that the mass m in Eq. (2) does not vanish. The presence of this term is responsible for the choice of signs appearing on the right-hand sides of Eqs. (8) and (9); and if it were absent, i.e., $m = 0$, we would be free to make a different choice. In particular, we could replace the matrices C and T by two others, C' and T', whose product $S' = C'T'$ obeys

$$S'\Gamma_\mu S'^{-1} = +\Gamma_\mu, \tag{12}$$

rather than Eq. (9). Since there exists one matrix which always satisfies Eq. (12), namely, the unit matrix, it would seem that the breakdown of *TCP* does not occur for the massless Dirac equation.

It can be shown, however, that the breakdown persists when we introduce interactions between different particles. Equation (12) implies that Γ_μ behaves as a pseudovector with respect to the reflection of coordinates:

$$\Gamma_\mu \to +\Gamma_\mu, \tag{13}$$

while the differential operator behaves as a proper vector:

$$\partial_\mu \to -\partial_\mu. \tag{14}$$

As a result, the currents[5]

$$J_\mu = \bar{\varphi}\Gamma_\mu\varphi'$$

and

$$K_\mu = \bar{\varphi}\overleftrightarrow{\partial}_\mu\varphi' \tag{15}$$

constructed from massless fields φ and φ' transform with opposite phases, and any interaction involving a linear combination of them, for example,

$$\{J_\mu + K_\mu, J_\mu^+ + K_\mu^+\}, \tag{16}$$

is not *TCP*-invariant. Other examples of noninvariant interactions are easily constructed.

It is interesting to compare the current × current interaction in $(2n+1)$ space–time [Eqs. (15) and (16)] with its analog in $(3+1)$ space–time. In the latter case, the produce of SR and Hermitian conjugation transforms a current $\mathfrak{J}_\mu(\mathbf{r}, t)$ according to the rule

$$\mathfrak{J}_\mu(\mathbf{r}, t) \to -\mathfrak{J}_\mu^+(-\mathbf{r}, -t), \tag{17}$$

no matter how the current may be constructed.[8] The sign on the right-hand side of Eq. (17) is fixed because the reflection of coordinates [Eq. (1)] belongs to the complex extension of the proper Lorentz group, and it implies that the symmetrized product

$$\{\mathfrak{J}_\mu, \mathfrak{J}_\mu^+\} \tag{18}$$

is always invariant under Eq. (17). Thus we see that the *TCP* invariance of the usual current × current theory of weak interactions is a consequence of the even dimensionality of our space–time.[9]

ACKNOWLEDGMENTS

Much of this work was carried out at the Summer Symposium of the Institute for Theoretical Physics, State University of New York at Stony Brook. The author would like to thank Professor C. N. Yang and Professor M. Dresden for their kind hospitality at the Symposium and for a number of stimulating conversations. He is also indebted to Dr. W. A. Bardeen and other members of the Institute for their interest and helpful comments.

[6] For a discussion of the Clifford algebra [Eq. (4)], see H. Boerner, *Representation of Groups* (North-Holland Publishing Co., Amsterdam, 1963), Chap. 8.

[7] A. Pais, J. Math. Phys. **3**, 1135 (1962).

[8] It is assumed, of course, that the usual connection between spin and statistics holds.

[9] G. Rajasekaran [Phys. Rev. **160**, 1427 (1967)] has discussed the converse of this result; namely, that if Eq. (18) is to be *TCP*-invariant, then $\mathfrak{J}_\mu$ must transform according to Eq. (17). His result does not depend upon the dimensionality of space–time.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Null and Pseudonull Data for Scalar Fields*

EMILE GRGIN
Belfer Graduate School of Science, Yeshiva University, New York

(Received 18 January 1968)

In this paper, a projective geometric formalism for the description of the conformal compactification of Minkowski space and of its invariance groups is developed. It is then modified to include the description of Minkowskian vectors, in particular, the momentum space. The d'Alembert equation is solved by constructing its solutions from global null data, completely arbitrary numbers assigned to the null cone at infinity. The Klein–Gordon equation, solved by the same method, leads to the concept of pseudonull data. Pseudonull data are also arbitrary numbers, but they are assigned to hyperboloids at a suitably defined infinity outside the conformal compactification of the Minkowski space.

I. INTRODUCTION

It has been shown by Penrose [1] that noninteracting spinor fields of zero rest mass and any spin are relatively invariant under conformal transformations; i.e., they are invariant up to some power of the conformal factor. In particular, one can compactify the Minkowski space with a conformal factor, singular at infinity. The new space is of constant curvature, and it is homeomorphic to the product of a circle with a three-dimensional sphere. Its invariance group is the 15-parametric conformal group, isomorphic with $SO(4, 2)$. Consequently, free spinor fields are covariant with respect to this group.

A global study of these fields yields some interesting results, e.g., that it is possible to obtain them from global null data. The concept of null data was also introduced by Penrose[2] (in the more general background of Riemannian spaces, but locally). Unlike Cauchy data, which may not be given on a characteristic hypersurface, null data are given on null cones. In this paper, the null cone at infinity is used. The global null data for zero-rest-mass free-spinor fields are complex numbers, one complex number being arbitrarily assigned to each point of the null cone at infinity (which reminds one of the Bondi–Sachs news functions).

Once the formalism of global null data is developed for zero-rest-mass fields, it is most natural to inquire whether a similar geometric approach to fields of nonvanishing mass exists. In this paper, that problem is solved for scalar fields, solutions of the Klein–Gordon equation.

Since this equation is not conformally invariant, the conformal compactification of the Minkowski space is not a natural setting for this problem. One must be able to deal with infinity without having to bring it into the space itself. However, this compactification must exist as a limit within the formalism, since it must appear in the limit of vanishing mass. Klein's Erlangen program provides a general solution to this kind of problem. It consists in taking, for an original n-dimensional pseudo-Euclidean space of signature s, an $(n + 1)$-dimensional real projective space, and selecting in it an invariant n-dimensional quadric of signature s. This quadric represents the conformal compactification of the original space, whose conformal group is isomorphic with the subgroup of the projective group under which the quadric is invariant. By requiring that one point of the quadric be fixed (puncturing), one recovers the original pseudo-Euclidean space and its invariance group

$$SO[\tfrac{1}{2}(n + s), \tfrac{1}{2}(n - s)].$$

Given an invariant quadric and a fixed point, the polar hyperplane of this point with respect to the quadric is also invariant, and so is the intersection of this hyperplane with the quadric. If the metric of the original space is positive-definite, this intersection consists of a single real point—the fixed point of contact itself (e.g., the north pole of the sphere in the elementary case of stereographic projection). If the metric is indefinite, the intersection is a real $(n - 1)$-dimensional quadric in the polar hyperplane (e.g., the null cone at infinity in the case of the Minkowski space).

The original pseudo-Euclidean space studied in this paper is the Minkowski space M^4. The corresponding real projective space $P^5(R)$ is built over a real six-dimensional linear space $L^6(R)$. The conformal compactification $\mathcal{M}^4$ of M^4 is a quadric $Q^4 \subset P^5(R)$, and M^4 itself is the intersection of Q^4 with a five-dimensional affine space $A^5 \subset L^6$. The complement of M^4 in Q^4 is the null cone at infinity, $\mathfrak{I}$. This null cone is also the intersection of Q^4 with a five-dimensional

* Supported in part by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR 816-67.

[1] R. Penrose, in *Relativity, Groups, and Topology* (Gordon and Breach Science Publishers, Inc., New York, 1964).

[2] R. Penrose, "Null Hypersurface Initial Data," in P. G. Bergman *et al.*, *Quantization of Generally Covariant Field Theories* (Syracuse Univ., preprint ARL 63-56, 1963).

linear subspace I^5 of L^6. It is shown that I^5, called the linear space at infinity, is related to momentum space. Null momenta correspond to points of $\mathfrak{I} \subset I^5$, nonnull momenta of given mass m to points of some hypersurface in I^5, $M_c^4(m^2) \subset I^5$. Global null data for the d'Alembert equation are given on $\mathfrak{I}$, i.e., within the conformal compactification of the Minkowski space, those for the Klein–Gordon equation, on some three-dimensional subsets (mass shells) in I^5, outside the conformal compactification of M^4.

This paper is logically self-contained, but familiarity with Penrose's description of the conformal compactification of M^4 is assumed.

II. THE SIX-DIMENSIONAL FORMALISM

Let L^6 be a real six-dimensional linear space. If $x \in L^6$, its six coordinates will be labeled by x^A, where capital Latin indices run from zero to five. A real projective space P^5 is built over L^6 by identifying its points with the one-dimensional linear subspaces (i.e., rays) of L^6. If $x^A \in L^6$, the same index attached to the corresponding capital letter in the kernel represents the corresponding point in P^5, i.e.,

$$X^A \overset{\text{DEF}}{=} (x^0 : \cdots : x^5) \in P^5.$$

Fundamental quadric: $Q^4 \subset P^5$ of signature -2, e.g., $\eta_{AB} = \text{diag}\ (+1, -1, -1, -1, -1, +1)$, is kept invariant. The corresponding linear structure $C^5(0) \subset L^6$ is called the *fundamental cone*. Their respective definitions are:

$$Q^4 \overset{\text{DEF}}{=} \{X^A \in P^5 \mid \eta_{AB} X^A X^B = 0\},$$
$$C^5(0) \overset{\text{DEF}}{=} \{x^A \in L^6 \mid \eta_{AB} x^A x^B = 0\}$$

The bracket symbol is used for scalar products, both in L^6 and P^5, e.g.,

$$(x, y) \overset{\text{DEF}}{=} \eta_{AB} x^A x^B = x_A x^A.$$

Two systems of basis vectors in L^6 are needed.

Orthogonal Basis: It consists of the six vectors, $e_C{}^A \in L^6$. In this notation, the first index names the vector, the second is the component. These vectors satisfy the following orthogonality relations:

$$(e_C, e_D) = \eta_{CD}, \quad \eta^{CD} e_C{}^A e_D{}^B = \eta^{AB}. \tag{1}$$

Quasi-orthogonal Basis: This system provides the most straightforward transition to M^4. It consists of the first four vectors of the orthogonal basis $e_a{}^A$ (lower case Latin indices run from zero to three) and of two null vectors o^A, i^A, defined by

$$o^A = e_4{}^A + e_5{}^A, \quad i^A = e_4{}^A - e_5{}^A. \tag{2a}$$

The corresponding orthogonality relations are:

$$(e_a, e_b) = \eta_{ab},$$
$$(o, i) = -2, \quad (o, o) = (i, i) = (o, e_a) = (i, e_a) = 0,$$
$$\eta_{AB} = \eta^{ab} e_{aA} e_{bB} - o_{(A} i_{B)}, \tag{2b}$$

where η_{ab} is of signature -2, e.g., $\eta_{ab} = \text{diag}\ (+1, -1, -1, -1)$.

The Null Objects: A number of geometric objects in, or related to, the quadric Q^4 [or cone $C^5(0)$] will have interpretations in M^4. It is convenient to call them null objects. The null objects needed in this paper are now defined. Their definitions are given in terms of projective coordinates, but they are used either in P^5 or L^6, as needed.

The points of Q^4 are called *null points*. They are defined by the *null condition*

$$(X, X) = 0. \tag{3a}$$

For any point $U^A \in P^5$, the linear equation $(U, X) = 0$ represents a hyperplane in P^5 (the polar hyperplane of U^A with respect to Q^4). If $(U, U) = 0$, this hyperplane is tangent to Q^4, and is called a *null hyperplane*. For any two points R^A, $S^A \in P^5$, the parametric equations $X^A = \lambda R^A + \mu S^A$ represent a projective line in P^5, $\lambda : \mu$ being the projective coordinates on this line. If every point of this line is a null point, the line is called a *null line*. The necessary and sufficient condition for this is

$$(R, R) = (R, S) = (S, S) = 0. \tag{3b}$$

If only one point of a line is null, but if the line is located in a null hyperplane, it is called a *pseudonull line*. The necessary and sufficient condition for this is that there exist a point R^A on the line (the point of contact) such that

$$(R, R) = (R, S) = 0, \quad (S, S) \neq 0. \tag{3c}$$

The set of all pseudonull lines through a point consists of two distinct parts: the timelike and the space-like pseudonull lines. They are defined, respectively, by $(S, S) > 0$, $(S, S) < 0$. The set of all null lines through a null point R^A is the *null cone* $\mathcal{C}(R^A)$ with vertex at that point. The points X^A of the null cone are simultaneous solutions of the equations

$$(R, X) = (X, X) = 0. \tag{3d}$$

The null cone is the common boundary of the two sets of pseudonull lines through its vertex and, with them, it fills the null hyperplane at the vertex.

Minkowski Space M^4: The points $X^A \in P^5$ not incident with the null hyperplane at i^A, i.e., satisfying

the inequality $(i, X) \neq 0$, can be uniquely represented by affine coordinates:

$$x^A = X^A/(i, X). \tag{4a}$$

These variables represent the points of an affine space $A^5 \subset L^6$, defined by the *affine condition:*

$$(i, x) = 1. \tag{4b}$$

One can project A^5 onto a four-dimensional space of parameters x^a (soon to be identified with Minkowski coordinates) by the relations

$$x^a = (e^a, x). \tag{5}$$

Let a point $X^A \in P^5$ be expressed in terms of the vectors of the quasi-orthogonal basis (2):

$$X^A = a^a e_a{}^A + bi^A + co^A, \tag{6}$$

where only the ratios $a^a{:}b{:}c$ of the coefficients are defined. The null condition (3a) implies $\eta_{ab}a^a a^b - 4bc = 0$. The affine condition (4b) removes the homogeneity from the coefficients and implies $c = -\frac{1}{2}$. If both conditions are applied simultaneously, one obtains the following relations:

$$x^A = x^a e_a{}^A - \tfrac{1}{2}x^2 i^A - \tfrac{1}{2}o^A, \tag{7a}$$

$$(o, x) = x^2, \tag{7b}$$

where $x^2 \overset{\text{DEF}}{=} \eta_{ab}x^a x^b$, x^a being defined by (5).

If one identifies the space of parameters x^a with M^4, relations (5) and (7a) establish a homeomorphism between M^4 and the intersection of the fundamental cone with the affine space. This intersection is denoted by $M^4(0)$:

$$M^4(0) \overset{\text{DEF}}{=} C^5 \cap A^5. \tag{8a}$$

From the projective point of view, this space is also the complement in Q^4 of the null cone at infinity, $\mathfrak{I}$ [defined by (3d), with the substitution $R^A = i^A$]:

$$M^4(0) = Q^4 - \mathfrak{I}. \tag{8b}$$

The quadric Q^4 itself is homeomorphic with the conformal compactification $\mathcal{M}^4$ of M^4:

$$M^4 = Q^4. \tag{8c}$$

Linear coordinates x^A, simultaneously satisfying the null and the affine conditions, i.e., $x^A \in M^4(0)$, are called *affine null coordinates* to distinguish them from *Minkowski* coordinates $x^a \in M^4$. The symbol "$x^a \leftrightarrow x^A$" means that they are related by equations (5) and (7a).

The Minkowski Distance: If $x, y \in M^4$ are any two points of M^4, one obtains, using (2), (5), and (7), and the relations $x^a \leftrightarrow x^A$, $y^a \leftrightarrow y^A$ the following expression for the square of their Minkowski distance:

$$d^2(x, y) \overset{\text{DEF}}{=} (x^a - y^a)^2 = -2(x, y). \tag{9}$$

One notices that a quadratic expression in the Minkowski variables is represented by a bilinear one in the affine null coordinates, the quadratic dependence having been absorbed by the fundamental cone.

Hyperspheres in M^4: For a given point $x^a \in M^4$, the set of all points $y^a \in M^4$ satisfying the relation

$$d^2(x, y) = q^2 = \text{const} \tag{10a}$$

is called a *hypersphere* in M^4. It follows from (9) that it is represented by a linear equation in affine null coordinates

$$(u, y) = 0. \tag{10b}$$

From (9) and (10b) one obtains the expression for u^A in terms of the five parameters defining the hypersphere (the center $x^a \leftrightarrow x^A$ and the radius squared, q^2):

$$u^A = x^A + \tfrac{1}{2}q^2 i^A. \tag{11a}$$

It follows that u^A satisfies the affine condition. The inverse relations are

$$q^2 = (u, u), \tag{11b}$$

$$x^A = u^A - \tfrac{1}{2}(u, u)i^A. \tag{11c}$$

The condition for a hyperplane $(U, X) = 0$ to represent a null cone, both in the projective sense (3d) and in the Minkowski sense (10a), is $(U, U) = 0$, so that the two concepts are equivalent in M^4. The projective definition is more encompassing, however, since it applies as well to the case $(i, U) = 0$. The Minkowski interpretation of this case must be found separately.

Let I^5 denote the five-dimensional linear subspace of L^6 defined by the equation $(i, x) = 0$. This space $I^5 \subset L^6$ will be called *the linear space at infinity*. If a point, expressed in the form (6), belongs to I^5, it follows from (2) that the coefficient c of o^A vanishes:

$$u^A = u^a e_a{}^A - vi^A. \tag{12a}$$

From (7a), (10b), and (12a) one obtains

$$u_a x^a - v = 0, \tag{12b}$$

which represents a hyperplane in M^4.

Thus, in the projective formalism, the intersections of the hyperplanes of P^5 with Q^4 represent the two-sheeted hyperboloids in M^4 ($q^2 > 0$); the null cones ($q^2 = 0$); the one-sheeted hyperboloids ($q^2 < 0$); and the spacelike ($u^2 > 0$), null ($u^2 = 0$), and timelike

($u^2 < 0$) hyperplanes. This is an instance of the general theorem about the conformal equivalence of hyperspheres and hyperplanes. In the transition to metric geometry, which is done by selecting a hyperplane in the projective space and calling it the hyperplane at infinity, a distinction appears between hyperspheres whose centers are in the finite region and those with centers at infinity. The latter represent hyperplanes in the metric space.

III. THE TRANSFORMATION GROUPS

The continuous groups of transformations in M^4 are isomorphic with the proper quadric group in P^5. The quadric group in P^5 is the subgroup of projective transformations under which the quadric Q^4 is invariant. The discrete symmetries of M^4 correspond to a subgroup of improper involutions in P^5.

The Proper Quadric Group and its Subgroups: The proper quadric group in P^5 is an $SO(4, 2)$. Its general element is

$$E^A{}_B = \exp(\eta^{AC}C_{CB}), \tag{13a}$$

or, for infinitesimal transformations,

$$E^A{}_B = \delta^A{}_B + C^A{}_B, \cdots, C^A{}_C C^C{}_B \simeq 0, \tag{13b}$$

where C_{AB} is an arbitrary skew-symmetric matrix.

The special 15-parametric conformal group in M^4 is isomorphic with $SO(4, 2)$. The relations establishing this isomorphism are now derived. Equations (14) are definitions of the generators of the quadric group, expressed in terms of the quasi-orthogonal basis:

$$M_{[ab]AB} \overset{\text{DEF}}{=} e_{a[A}e_{b|B]} \quad \cdots \text{ 6 bivectors,} \tag{14a}$$

$$A_{aAB} \overset{\text{DEF}}{=} 2e_{a[A}i_{B]} \quad \cdots \text{ 4 bivectors,} \tag{14b}$$

$$B_{aAB} \overset{\text{DEF}}{=} 2e_{a[A}o_{B]} \quad \cdots \text{ 4 bivectors,} \tag{14c}$$

$$K_{AB} \overset{\text{DEF}}{=} o_{[A}i_{B]} \quad \cdots \text{ 1 bivector.} \tag{14d}$$

The following list of isomorphisms follows from (2), (5), (7), (13), and (14). The relations $x^a \leftarrow X^A$ and $x^{a'} \leftarrow X^{A'}$ are to be understood throughout.

The homogeneous Lorentz group (6 parameters $m^{[ab]}$):

$$M^A{}_B \overset{\text{DEF}}{=} \exp(m^{ab}M_{ab}{}^A{}_B) = M^a{}_b e_a{}^A e^b{}_B, \tag{15a}$$

where $M^a{}_b \overset{\text{DEF}}{=} \exp(m^a{}_b)$. The isomorphism is

$$X^{A'} = M^A{}_B X^B \Leftrightarrow x^{a'} = M^a{}_b x^b.$$

The group of translations (4 parameters a^a):

$$A^A{}_B = \exp(a^a A_a{}^A{}_B) = \delta^A{}_B + a^a A_a{}^A{}_B - \tfrac{1}{2}a^2 i^A i_B,$$
$$X^{A'} = A^A{}_B X^B \Leftrightarrow x^{a'} = x^a + a^a. \tag{15b}$$

The group of uniform accelerations (4 parameters b^a):

$$B^A{}_B = \exp(b^a B_a{}^A{}_B) = \delta^A{}_B + b^A B_a{}^A{}_B - \tfrac{1}{2}b^2 o^A o_B,$$
$$X^{A'} = B^A{}_B X^B \Leftrightarrow x^{a'} = (x^a + x^2 b^a)/(1 + 2b_b x^b + b^2 x^2). \tag{15c}$$

The group of dilations (one parameter k):

$$D^A{}_B = \exp(kK^A{}_B) = \delta^A{}_B + \tfrac{1}{2}(1 - e^{-k})o^A i_B + \tfrac{1}{2}(1 - e^k)i^A o_B,$$
$$X^{A'} = D^A{}_B X^B \Leftrightarrow x^{a'} = e^k x^a. \tag{15d}$$

One sees from these relations that o^A and i^A play reciprocal roles. A uniform acceleration is equivalent to a translation of the point at infinity i^A [cf. Eq. (16a)], and what is interpreted as a dilatation with respect to one of these points, is a contraction with respect to the other.

The Poincaré group is characterized by the fixed point i^A. Indeed, from (2) and (15) one obtains $M^A{}_B i^B = i^A$ and $A^A{}_B i^B = i^A$, while

$$i^{A'} = B^A{}_B i^B = -2b^2(c^a e_a{}^A - \tfrac{1}{2}c^2 i^A - \tfrac{1}{2}o^A), \tag{16a}$$

where $c^a = b^a/b^2$ and

$$i^{A'} = D^A{}_B i^B = e^k i^A. \tag{16b}$$

Symmetries: Let u^A be any point of A^5 or I^5, not on the quadric Q^4. The transformation

$$(u)^A{}_B \overset{\text{DEF}}{=} \delta^A{}_B - \frac{2}{(u, u)} u^A u_B \tag{17}$$

is an improper involution in P^5. It represents a mirror reflection in the nonnull hyperplane defined by $(u, X) = 0$. It leaves the quadric Q^4 invariant, so it can be interpreted in M^4. By taking for u^A the vectors of the orthogonal basis, one obtains the following transformations of M^4:

$$T^A{}_B \overset{\text{DEF}}{=} (e^0)^A{}_B, \tag{18a}$$

which represents time reversal: $x^0 \rightarrow -x^0$.

$$P^A{}_B \overset{\text{DEF}}{=} (e^1)^A{}_C (e^2)^C{}_D (e^3)^D{}_B \tag{18b}$$

gives the parity transformation: $x^a \rightarrow -x^a$.

$$S_+{}^A{}_B \overset{\text{DEF}}{=} (e^5)^A{}_B \tag{18c}$$

represents the inversion on the two-sheeted unit hyperboloid in M^4, i.e., $x^a \rightarrow x^a/x^2$. Similarly,

$$S_-{}^A{}_B \overset{\text{DEF}}{=} (e^4)^A{}_B \tag{18d}$$

is the inversion on the one-sheeted unit hyperboloid: $x^a \rightarrow -x^a/x^2$.

The transformations T and S_+, on the one hand, and P and S_-, on the other, are essentially the same in the space $\mathcal{M}^4$, since they differ only by transformations from the proper quadric group. However, while i^A is a fixed point of the transformations T and P, it is not invariant under inversions. Under inversions it transforms according to

$$i^{A'} = S_+{}^A{}_B i^B = o^A, \tag{19a}$$

$$i^{A'} = S_-{}^A{}_B i^B = -o^A. \tag{19b}$$

Inversions are not symmetries of M^4. They interchange the null cones at infinity and at the origin. With the Poincaré group, they build the full conformal group of $\mathcal{M}^4$ (Liouville's theorem).

IV. MINKOWSKI VECTORS

From now on, the invariance group of interest is the Poincaré group, so that the point i^A is fixed. Consequently, the linear space I^5 is invariant. Since the spaces I^5 and $M^4(0)$ are disjoint, the points of I^5 are available for some useful interpretation consistent with their invariance properties. It has been shown earlier that their ratios represent hyperplanes in M^4. However, the points of I^5 themselves contain more information than necessary to determine a hyperplane. One can interpret them as representing hyperplanes with Poincaré invariants attached to them. It is now shown that they can also be interpreted as Minkowski vectors with noninvariant numbers (phases) attached to them. A particularly important application of this interpretation is the representation of momentum space by I^5.

Minkowski Vectors: Minkowski vectors are the points of a four-dimensional linear space, to be denoted by M_c^4 (for "centered M^4") provided with a Minkowski metric. In the transition from M^4 to the six-dimensional formalism, the centering restriction, which eliminates translations, can be interpreted as a requirement of invariance under the translation operator (15b).

The necessary and sufficient condition for a point $v^A \in L^6$ to be invariant under translations, i.e.,

$$A^A{}_B v^B = v^A,$$

is that the coefficient of o^A in (6) should vanish:

$$v^A = v^a e_a{}^A - \omega i^A, \tag{20}$$

which is equivalent to the requirement $(i, v) = 0$, i.e., $v^A \in I^5$.

Relation (20) establishes the interpretation of Minkowski vectors in L^6. The variable ω, to be called the phase, specifies the location of a hyperplane in M^4. It is an extra piece of information, not contained in the Minkowski vectors themselves. The projection of I^5 onto a linear Minkowski space M_c^4 is $v^a = (e^a, v)$. The symbol $v^a \leftarrow v^A$ (or $v^a \leftrightarrow v^A$ if the phase is given) means that $v^a \in M_c^4$ and $v^A \in I^5$ are related by relations (20).

If a faithful representation of M_c^4 in I^5 is desired, one must eliminate the freedom of the phase. A simple way of doing this for vector fields $v^a(x^b)$ in M^4 consists in requiring that the hyperplane in M^4, defined by $v^a(x^b)$, be incident with x^b, i.e., $x^A v_A(x^B) = 0$, where $x^a \leftrightarrow x^A$ and $v^a \leftrightarrow v^A$. This *incidence condition* determines the phase:

$$\omega = v_a x^a. \tag{21}$$

Tangent Vectors: Differentiation of (7a) yields

$$dx^A = (\partial_a{}^A - x_a i^A)\, dx^a. \tag{22}$$

Obviously, it is a vector, since $(i, dx) = 0$. It also identically satisfies the incidence relation: $(x, dx) \equiv 0$. Inversely: $dx^a = (e^a, dx)$. It is easy to verify that the infinitesimal translations by dx^a, defined by (13b) and (14b), yield the same expression (22). One finds the expression

$$dx^2 = (dx, dx) \tag{23}$$

for the Minkowski metric in infinitesimal form.

Cotangent Vectors: Let f be a scalar test function, to be understood as $f(x^a)$ or $f(x^A)$, and let $\nabla_A \in I^5$ be the vector differential operator corresponding to the gradient $\partial_a \in M_c^4$. According to (20), ∇_A must be of the form

$$\nabla_A \equiv e^a{}_A \partial_a - \omega i_A. \tag{24}$$

The identity means that the relation should yield an equality when applied to any test function.

From the null and affine conditions (3a) and (4b) and from the requirement $\nabla_A(\text{const}) = 0$ follows

$$x_B \nabla_A x^B = 0, \quad i_B \nabla_A x^B = 0. \tag{25}$$

With (24) and (7a), these equations yield, respectively, $x_B \omega x^B = 0$, $i_B \omega x^B = 0$. Consequently, ω must either vanish, or be a scalar differential operator. Since ∇_A represents the gradient, ω should be linear and of the first order. The second of the above equations, which implies $\omega(i, x) \equiv \omega = 0$, requires that it be homogeneous; hence $\omega = S^a \partial_a$. With this form of ω, both conditions are satisfied for any S^a. Due to the isotropy of M^4, S^a may not be a fixed vector. The only solution for S^a which does not introduce new structure into the space is $S^a = Sx^a$. It is not possible to determine the value of the coefficient S from the geometry

of the space. It is essentially arbitrary. In the following it is restricted to unity by imposing the incidence condition (21). Thus,

$$\nabla_A = (e^a{}_A - x^a i_A)\partial_a. \tag{26}$$

From (7a) and (26) follows the identity:

$$\nabla_A x^B = \delta_A{}^B - i_A x^B - x_A i^B, \tag{27}$$

which must be taken into account in the application of ∇_A.

The expression

$$\Box = (\nabla, \nabla) \tag{28}$$

for the d'Alembertian follows from (26).

V. NULL AND PSEUDO-NULL DATA

If p^A is a constant vector, $p^A \in I^5$, it follows from (27) that the scalar function $z = (p, x)$ satisfies the d'Alembert equation $\Box z = 0$, and has the eikonal $(\nabla z, \nabla z) = (p, p)$. Hence, for an arbitrary function $f(z)$,

$$\Box f(z) = (p, p) f''(z). \tag{29}$$

The d'Alembert Field: If $(p, p) = 0$, i.e., if $p^A \in \mathfrak{I}$, $f(z)$ identically satisfies the d'Alembert equation. The same is true for the sum of all arbitrary functions $f(p^A; z)$, one arbitrary function of z being assigned to every point $p^A \in \mathfrak{I}$:

$$f(x^A) = \mathop{\mathsf{S}}_{p^A \in \mathfrak{I}} f(p^A; z) \tag{30}$$

(the summation symbol S stands for discrete sums and/or integrals). This summation over the three-dimensional set $\mathfrak{I}$ can be reduced to a sum over a two-dimensional sphere in the following way: Consider an arbitrary but fixed point $a^A \in M^4(0)$. The intersection of the null cone $\mathcal{C}(a^A)$ with vertex at a^A and of the null cone $\mathfrak{I}$ is a two-dimensional sphere (see Fig. 1):

$$S^2(a^A) \overset{\text{DEF}}{=} \mathfrak{I} \cap \mathcal{C}(a^A).$$

The arbitrary point p^A can be expressed in the form

$$p^A = e^A + u i^A, \tag{31}$$

where $e^A \in S^2(a^A)$ is on the generator of $\mathfrak{I}$ incident with p^A. The parameter u locates the point p^A on this generator.

From (31) and the definition of z one obtains $z = u + y$, where $y = (e, x)$. This expression for z and the expression (31) for p^A are now introduced into (30). The three-dimensional summation splits into a summation over the two-dimensional sphere

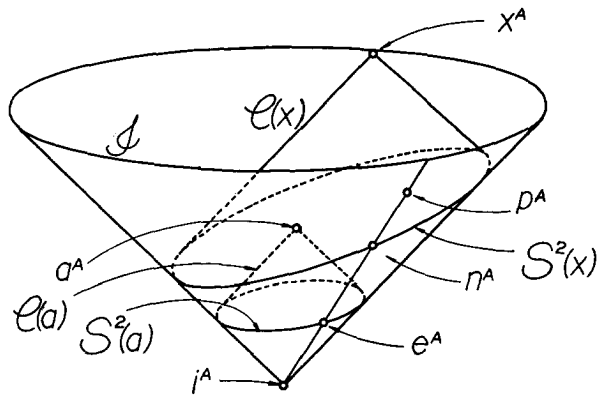


Fig. 1. Summation of null data.

$S^2(a^A)$ and an integral over u. With the notation

$$\int_{-\infty}^{\infty} du\, f(e^A + u i^A; u + y) \overset{\text{DEF}}{\equiv} g(e^A; y),$$

where g is an arbitrary function, since f is arbitrary, the expression (30) becomes

$$f(x^A) = \mathop{\mathsf{S}}_{e^A \in S^2(a^A)} g(e^A, y). \tag{32a}$$

This expression can be reinterpreted. Let the point $n^A \in S^2(a^A)$ be defined by the relation $n^A = e^A - y i^A$ and introduce the notation $g(e^A; y) \overset{\text{DEF}}{\equiv} c(n^A)$; then

$$f(x^A) = \mathop{\mathsf{S}}_{n^A \in S^2(x^A)} c(n^A). \tag{32b}$$

The numbers c, arbitrarily assigned to the points of $\mathfrak{I}$, are the *global null data* for the zero-rest-mass field $f(x^a)$. At any point x^A, this field is the sum of the null data its null cone collects at infinity, where the data are given.

The expression (32b) is the generalization to M^4 of the solution $f(x, t) = g(x - t) + h(x + t)$ of the one-dimensional wave equation. The functions g and h can be thought of as the null data assigned to the null cone at infinity of M^2. In this case, the null cones are degenerate, consisting only of two null lines. The parameters $x - t$ and $x + t$ determine the points n^A on the degenerate null cone at infinity.

The Klein–Gordon Equation: Let $M^4{}_C(m^2)$ denote the representation in I^5 of the mass shell in momentum space:

$$M^4{}_C(m^2) \overset{\text{DEF}}{\equiv} \{p^A \in I^5 \mid (p, p) = m^2\}. \tag{33}$$

The space $M^4{}_C(m^2)$ is the Cartesian product $M^4{}_C(m^2) = R' \times M^3(m^2)$. Here, R' is the range of the arbitrary real phase ω, while $M^3(m^2)$ represents, essentially, the mass shell $(p, p) = m^2$ in momentum space.

If $p^A \in M^4{}_C(m^2)$, the function $f(z)$ in (29) must be of the form

$$f(p^A; z) = A(p^A)e^{iz},$$

if it is to satisfy the Klein–Gordon equation

$$(\Box + m^2)f = 0.$$

The sum of all such functions over $M^4{}_C(m^2)$ is also a solution.

As in the mass-zero case, let a^A be an arbitrary but fixed point in $M^4(0)$. Let $H^3(a^A; m^2)$ denote the intersection of $M^4{}_C(m^2)$ with the hyperplane $T^4(a^A)$ tangent to C^5 at a^A; i.e.,

$$T^4(a^A) \overset{\text{DEF}}{\equiv} \{x^A \in A^5 \mid (a, x) = 0\}$$

and

$$H^3(a^A; m^2) \overset{\text{DEF}}{=} T^4(a^A) \cap M^4{}_C(m^2).$$

The point p^A can again be written in the form (31), provided

$$e^A \in H^3(a^A; m^2).$$

The four-dimensional sum in the expression

$$f(x^A) = \mathop{\mathsf{S}}_{p^A \in M^4{}_C(m^2)} A(p^A)e^{iz}$$

splits into a three-dimensional sum and an integral. With the notation

$$g(e^A) \overset{\text{DEF}}{=} \int_{-\infty}^{\infty} du\, A(e^A + ui^A)e^{iu},$$

where g is an arbitrary function over $H^3(a^A, m^2)$ (since the integrand is arbitrary), the solution $f(x^A)$ becomes

$$f(x^A) = \mathop{\mathsf{S}}_{e^A \in H^3(a^A; m^2)} g(e^A)e^{iv}. \tag{34a}$$

The arbitrary numbers g assigned to $H^3(a^A, m^2)$ are called *pseudonull data*, or m^2 data.

Given the global pseudonull data on $H^3(a^A, m^2)$, one can assign the numbers $C(n^A) \overset{\text{DEF}}{=} g(e^A)e^{iv}$ to the points $n^A = e^A - yi^A$ of $M^4{}_C(m^2)$. Then, the solution $f(x^A)$ becomes

$$f(x^A) = \mathop{\mathsf{S}}_{n^A \in H^3(x^A; m^2)} C(n^A). \tag{34b}$$

This is similar to the form (32b) for the mass-zero field, but it is not as economical.

In the zero-rest-mass case, the null data are arbitrarily given over the three-dimensional set $\mathfrak{I}$. They are then collected over a two-dimensional subset of $\mathfrak{I}$ by all the null line through the point x^A. Their sum is the field at this point.

In the case of nonvanishing mass, the null data are arbitrarily given over a three-dimensional subset of $M^4{}_C(m^2)$. They are then spread over the entire $M^4{}_C(m^2)$, and collected over one of its three-dimensional subsets by all pseudonull lines through the point x^A. Their sum is the field at that point. If $m^2 > 0$, only the timelike pseudonull lines collect the pseudonull data.

ACKNOWLEDGMENT

I should like to thank Professor Arthur Komar for his encouragement, many discussions, and suggestions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Green's Functions for the Ising Chain*

I. BIALYNICKI-BIRULA AND J. PIASECKI
Department of Mathematics and Physics, Warsaw University, Warsaw

(Received 13 March 1967)

The Ising chain is represented as a system of fermions interacting through two-particle forces. All thermodynamic Green's functions are calculated exactly in this model. They exhibit many properties expected on the basis of general theory for Green's functions of the interacting particles.

1. INTRODUCTION

In statistical physics, as in other many particle theories, no realistic exact solutions are known. Therefore, several exactly soluble models have been invented to study various properties of real physical systems. Among such models the Ising model for a system of interacting spins plays an important role. In this paper, we present exact thermodynamic Green's functions for the Ising chain (one-dimensional Ising model). These Green's functions contain all the information about not only static but also kinetic properties of the Ising chain. As a result of conservation of individual spins in the model, these kinetic properties are very simple. Nevertheless, we feel that the complete set of exact Green's functions for a model with interaction is worth mentioning. In this paper we investigate a system of N identical fermions whose Hamiltonian H has the form

$$H = -\sum_{i=1}^{N}(Bs_i + \tfrac{1}{2}\epsilon s_i s_{i+1}), \tag{1.1}$$

where

$$s_i = 2a_i^+ a_i - 1,$$
$$a_{N+1} = a_1, \quad a_{N+1}^+ = a_1^+,$$

and a_i^+, a_i are the creation and annihilation operators of a fermion in the ith state. The basis in the space of state vectors of the system contains 2^N vectors, which can be formed by acting with products of creation operators on the vacuum-state vector. These basic vectors can be put into a one-to-one correspondence with the states of the classical Ising chain, the energies of the corresponding states being equal. The existence of this correspondence enables us to describe the properties of interacting spins with the use of a many-fermion formalism.

In Sec. 2 we define the Green's functions and calculate them exactly. In Sec. 3 we discuss briefly their properties and we analyze the information which they contain.

* A preliminary report on this work has been published in Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys. **15**, 211, 357 (1967).

2. CALCULATION OF GREEN'S FUNCTIONS

We define the thermodynamic Green's functions in the usual manner[1]:

$$G(12\cdots k, k'\cdots 2'1') = \langle T[a(1)a(2)\cdots a(k)a^+(k')\cdots a^+(2')a^+(1')]\rangle. \tag{2.1}$$

Here T represents the Wick time-ordering operation and the brackets $\langle\ \rangle$ denote the grand canonical ensemble average. The time dependence of the creation and annihilation operators is given by

$$a^+(r) = a_{n_r}^+(t_r) = e^{iHt_r}a_{n_r}^+e^{-iHt_r} = e^{-it_r 2B}e^{-it_r\epsilon(s_{n_{r+1}}+s_{n_{r-1}})}a_{n_r}^+,$$
$$a(r) = a_{n_r}(t_r) = e^{iHt_r}a_{n_r}e^{-iHt_r} = e^{it_r 2B}e^{it_r\epsilon(s_{n_{r+1}}+s_{n_{r-1}})}a_{n_r}.$$

Since $[H, s_r]_- = 0$ [see (1.1)], we can write the Green's function (2.1) as a sum of $k!$ terms in the form

$$G(12\cdots k, k'\cdots 2'1') = \sum_{\substack{\text{perm}\\(i_1,\cdots,i_k)}} \epsilon^P \delta_{n_1 n_{i_1'}}\cdots\delta_{n_k n_{i_k'}} g(1\cdots k, i_k'\cdots i_1'), \tag{2.2}$$

where

$$g(1\cdots k, i_k'\cdots i_1') = \langle T(a_{n_1}(t_1)\cdots a_{n_k}(t_k)a_{n_k}^+(t_{i_k}')\cdots a_{n_1}^+(t_{i_1}'))\rangle,$$

and ϵ^P equals $+1$ or -1, depending on whether the permutation $(i_1, i_2, \cdots, i_k)$ of the numbers $1, 2, \cdots, k$ is even or odd. The summation in (2.2) is extended over all inequivalent permutations. The calculation of the Green's functions (2.1) is especially simple if the condition

$$|n_i - n_j| \geq 2 \quad (i \neq j) \tag{2.3}$$

is fulfilled. The case of $|n_i - n_j| < 2$ is discussed later on. Apart from the condition (2.3), we assume, for the sake of convenience, that

$$n_1 < n_2 < \cdots < n_k.$$

[1] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (W. A. Benjamin, Inc., New York, 1962), p. 3.

Before engaging in any calculations, we notice that the following equality holds:

$$g(1\cdots k, k'\cdots 1') = \langle T(a_{n_1}(t_1)a^+_{n_1}(t'_1))\cdots T(a_{n_k}(t_k)a^+_{n_k}(t'_k))\rangle. \quad (2.4)$$

This follows from the fact that if the condition (2.3) is satisfied and if $i \neq j$, we have

$$[a(i), a(j)]_+ = [a(i), a^+(j)]_+ = [a^+(i), a^+(j)]_+ = 0. \quad (2.5)$$

The chronological product of the operators $a_n(t)$ and $a^+_n(t')$ is given by

$$T(a_n(t)a^+_n(t')) = \tfrac{1}{2}(\operatorname{sgn}\Delta - s_n)\exp(i2B\Delta)\exp(i\epsilon(s_{n+1}+s_{n-1})\Delta),$$

where $\Delta = t - t'$. In order to evaluate (2.4) we make use of the transfer-matrix method.[2] Let $x = \frac{1}{2}\beta\epsilon$ and $y = \beta(B + \mu/2)$. ($\beta = 1/kT$ and μ is the chemical potential.)

We introduce the matrix

$$\mathsf{P} = \begin{pmatrix} e^{x+y}, & e^{-x} \\ e^{-x}, & e^{x-y} \end{pmatrix}$$

and the matrix

$$\mathsf{S} = \begin{pmatrix} u_1, & u_2 \\ u_2, & -u_1 \end{pmatrix} = S^{-1},$$

which diagonalizes P. The elements of S satisfy the relations

$$u_1^2 = 1 - u_2^2 = \tfrac{1}{2}(1 + \sinh y(\sinh^2 y + e^{-4x})^{-\frac{1}{2}}),$$
$$2u_1u_2 = e^{-2x}(\sinh^2 y + e^{-4x})^{-\frac{1}{2}}.$$

The eigenvalues of P, λ_+, and λ_- are

$$\lambda_\pm = e^x(\cosh y \pm (\sinh^2 y + e^{-4x})^{\frac{1}{2}}).$$

Let us define three additional matrices

$$\mathsf{F}^> = \begin{pmatrix} 0, & 0 \\ 0, & 1 \end{pmatrix}, \quad \mathsf{F}^< = \begin{pmatrix} -1, & 0 \\ 0, & 0 \end{pmatrix},$$

and

$$\mathsf{W}_r = \begin{pmatrix} e^{i\epsilon\Delta_r}, & 0 \\ 0, & e^{-i\epsilon\Delta_r} \end{pmatrix}, \quad \Delta_r = t_r - t'_r.$$

For the infinite Ising chain with the aid of the transfer-matrix method we obtain

$$g(1\cdots k, k'\cdots 1') = \operatorname{Tr}\left\{\begin{pmatrix} 1, & 0 \\ 0, & 0 \end{pmatrix}\begin{pmatrix} g(1,1'), & f(1,1') \\ f(1,1'), & h(1,1') \end{pmatrix}\begin{pmatrix} 1, & 0 \\ 0, & \gamma^{n_2-n_1-2} \end{pmatrix}\right.$$
$$\times \begin{pmatrix} g(2,2'), & f(2,2') \\ f(2,2'), & h(2,2') \end{pmatrix}\cdots\begin{pmatrix} 1, & 0 \\ 0, & \gamma^{n_k-n_{k-1}-2} \end{pmatrix}$$
$$\left.\times \begin{pmatrix} g(k,k'), & f(k,k') \\ f(k,k'), & h(k,k') \end{pmatrix}\right\},$$

where

$$\gamma = \lambda_-/\lambda_+, \quad |\gamma| < 1. \quad (2.6)$$

The functions g, f, and h appearing in (2.6) can be found from the relation

$$\begin{pmatrix} g(r,r'), & h(r,r') \\ h(r,r'), & f(r,r') \end{pmatrix} = e^{i2B\Delta_r}\lambda_+^{-2}(\mathsf{SW}_r\mathsf{PF}^>\mathsf{PW}_r\mathsf{S}^{-1}\theta(\Delta_r) + \mathsf{SW}_r\mathsf{PF}^<\mathsf{PW}_r\mathsf{S}^{-1}\theta(-\Delta_r)).$$

Let us write $g(r, r')$ in the form

$$g(r,r') = g^>(\Delta_r)\theta(\Delta_r) + g^<(\Delta_r)\theta(-\Delta_r).$$

We define $f^>, f^<$, $h^>$, and $h^<$ in a similar way. After some calculations we obtain

$$g^>(\Delta) = \lambda_+^{-2}e^{i2B\Delta}\pi^2(\Delta),$$
$$f^>(\Delta) = \lambda_+^{-2}e^{i2B\Delta}\pi(\Delta)\rho(\Delta),$$
$$h^>(\Delta) = \lambda_+^{-2}e^{i2B\Delta}\rho^2(\Delta),$$

where

$$\pi(\Delta) = u_1e^{-x}e^{i\epsilon\Delta} + u_2e^{x-y}e^{-i\epsilon\Delta}$$

and

$$\rho(\Delta) = u_2e^{-x}e^{i\epsilon\Delta} - u_1e^{x-y}e^{-i\epsilon\Delta}.$$

The function $g^<$ is related to $g^>$ by the equality

$$g^<(\Delta) = g^<(t - t') = -e^{\beta\mu}\cdot g^>((t - i\beta) - t').$$

The similar equalities hold for $f^<$ and $h^<$. From (2.6) we obtain

$$g(12, 2'1') = g(1,1')g(2,2') + \gamma^{n_2-n_1-2}f(1,1')f(2,2'),$$

$$g(123, 3'2'1') = g(1,1')g(2,2')g(3,3') + \gamma^{n_2-n_1-2}f(1,1')f(2,2')g(3,3') + \gamma^{n_3-n_2-2}g(1,1')f(2,2')f(3,3') + \gamma^{n_3-n_1-4}f(1,1')h(2,2')f(3,3').$$

In general,

$$g(12\cdots k, k'\cdots 2'1') = g(1,1')g(2,2')\cdots g(k,k') + \gamma^{n_2-n_1-2}f(1,1')f(2,2')g(3,3')\cdots g(k,k') + \cdots + \gamma^{n_k-n_{k-1}-2}g(1,1')\cdots g(k-2,(k-2)')\cdot f(k-1,(k-1)')f(k,k') + \cdots + \gamma^{n_k-n_1-2(k-1)} \times f(1,1')h(2,2')\cdots h(k-1,(k-1)')f(k,k').$$

The structure of this sum is the following: Each term contains a factor γ^α. All possible values of α, except $\alpha = 0$, are obtained by taking all the sums of the quantities

$$n_2 - n_1 - 2, \quad n_3 - n_2 - 2, \cdots, n_k - n_{k-1} - 2$$

under the condition that none of these quantities appears more than once in the sum. The factor γ^α is multiplied by a product of k functions g, h, and f

[2] K. Huang, *Statistical Mechanics* (John Wiley & Sons, Inc., London, 1963), p. 346.

according to the following rule:

	function
(i) n_r appears in α	$f(r, r')$
(ii) n_r has canceled in α	$h(r, r')$
(iii) neither (i) nor (ii)	$g(r, r')$.

After having calculated $g(12\cdots k,\ k'\cdots 2'1')$, we obtain the Green's function with the help of (2.2).

We can also compute the Green's functions with the use of the transfer matrix method when the condition (2.3) is not fulfilled. However, for $|n_i - n_j| = 1$ and $n_i - n_j = 0$, the results are much more complicated because the anticommutation relations (2.5) do not hold in these cases. For example, if $k = 2$, $n_1 = n$ and $n_2 = n + 1$, the Green's function (2.1) has the form

$$G(12, 2'1') = \delta_{nn_1'}\delta_{n+1n_2'}A(t_1 - t_1', t_2 - t_2') - \delta_{nn_2'}\delta_{n+1n_1'}A(t_1 - t_2', t_2 - t_1'),$$

where

$$A(\Delta_1, \Delta_2) = \lambda_+^{-3}e^{i2B(\Delta_1+\Delta_2)}e^{-i\epsilon(\Delta_1 \operatorname{sgn}\Delta_2+\Delta_2 \operatorname{sgn}\Delta_1)}e^{x \operatorname{sgn}(\Delta_1\Delta_2)-y} \times f(\Delta_1)f(\Delta_2)(1 + \sum{}' \theta(t - t')\,\theta(t' - t'') \times \theta(t'' - t''')[e^{i2\epsilon(t'-t'')\operatorname{sgn}(\Delta_1\Delta_2)} - 1]).$$

Here

$$f(\Delta) = \operatorname{sgn}\Delta e^{-y \operatorname{sgn}\Delta}(u_1e^{i\epsilon\Delta}e^{y-x \operatorname{sgn}\Delta} + u_2e^{-i\epsilon\Delta}e^{x \operatorname{sgn}\Delta})$$

and the sum $\sum'$ is extended over all the permutations (t, t', t'', t''') of the parameters t_1, t_2, t_2', t_1', excluding the eight permutations in which simultaneously t_1, t_1' and t_2, t_2' are neighbors.

3. PROPERTIES OF GREEN'S FUNCTIONS

Let us consider one-particle Green's function $G(1, 1') = \delta_{n_1n_1'}g(1, 1')$. We write it in the form

$$G(1, 1') = \delta_{n_1n_1'}(g^{>}(\Delta_1)\theta(\Delta_1) + g^{<}(\Delta_1)\theta(-\Delta_1)) = \delta_{n_1n_1'}\lambda_+^{-2}e^{i2B\Delta_1}((u_1e^{-x}e^{i\epsilon\Delta_1} + u_2e^{x-y}e^{-i\epsilon\Delta_1})^2\theta(\Delta_1) - (u_1e^{x+y}e^{i\epsilon\Delta_1} + u_2e^{-x}e^{-i\epsilon\Delta_1})^2\theta(-\Delta_1)).$$

In order to analyze the structure of $G(1, 1')$ it is convenient to introduce the spectral function

$$A(\omega) = \int_{-\infty}^{+\infty} dt e^{i\omega t}(g^{>}(t) - g^{<}(t)) = 2\pi\lambda_+^{-2}(u_1^2e^{2(x+y)}(1 + e^{\beta(\omega-\mu)})\delta(\omega + 2B + 2\epsilon) + 4u_1u_2\cosh y\,\delta(\omega + 2B) + u_2^2e^{2(x-y)}(1 + e^{-\beta(\omega-\mu)})\delta(\omega + 2B - 2\epsilon)).$$

The spectral function A satisfies the sum rule

$$\int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}A(\omega) = 1$$

and determines the Fourier transform of $G(1, 1')$:

$$G(1, 1') = i\delta_{n_1n_1'}\int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}e^{-i\omega\Delta_1}\int_{-\infty}^{\infty}\frac{d\omega'}{2\pi} \times \left[PP\frac{1}{\omega - \omega'} - \pi i\delta(\omega - \omega')\frac{1 - e^{-\beta(\omega'-\mu)}}{1 + e^{-\beta(\omega'-\mu)}}\right]A(\omega').$$

Using the interpretation of one-particle Green's function as a particle propagator, we conclude that a fermion added to our many-fermion system can propagate in three states with energies

$$\omega_1 = -2(B + \epsilon),$$
$$\omega_2 = -2B,$$
$$\omega_3 = -2(B - \epsilon).$$

These energies are in fact, up to the factor of 2, three possible energies of a spin in the classical Ising chain.

In the limit of zero temperature, the one-particle Green's function has the form[3]

$$G(1, 1')|_{T=0} = \delta_{n_1n_1'}g^0(1, 1'),$$

where

$$g^0(1, 1') = \begin{cases} \mp e^{i2\Delta_1(B\pm\epsilon)}\theta(\mp\Delta_1) \\ \text{for } \pm(B + \mu/2) > 0,\ \epsilon + |B + \mu/2| > 0, \\ \tfrac{1}{2}e^{i2B\Delta_1}(\theta(\Delta_1)e^{i2\epsilon\Delta_1} - e^{-i2\epsilon\Delta_1}\theta(-\Delta_1)) \\ \text{for } \epsilon + |B + \mu/2| < 0. \end{cases}$$

It turns out that all the Green's functions at $T = 0$ can be expressed in terms of functions $g^{>}$ and $g^{<}$. We notice that in the case $\epsilon + |B + \mu/2| < 0$ at $T = 0$ the Green's functions do not have the cluster property. In all other cases (zero or nonzero temperatures) all the Green's functions factorize when the "distance"

$$d(n_1\cdots n_r, n_{r+1}\cdots n_k) = \min_{\substack{i=1,\cdots,r \\ j=r+1,\cdots,k}}|n_i - n_j|$$

between two groups of indices $n_1\cdots n_r$ and $n_{r+1}\cdots n_k$ tends to infinity:

$$\lim_{d\to\infty} g(1\cdots k, k'\cdots 1') = g(1\cdots r, r'\cdots 1')g(r + 1\cdots k, k'\cdots (r + 1)').$$

The last property follows directly from the Eq. (2.6), if we take into account that $|\gamma| < 1$.

The Green's functions taken at $t_1 = t_2 = \cdots =$

[3] The special case $B + \mu/2 = -\epsilon > 0$ gives rise to some peculiar form of the Green's function, which will not be given here.

$t_k = t'_k = \cdots = t'_1$ give a complete statistical description of the Ising chain. In particular, we can obtain in this manner the spin-correlation functions calculated earlier by Marsh.[4] We find a complete agreement between his and our results.

As we have already mentioned in the introduction, the Ising chain does not exhibit any nontrivial kinetic behavior. Our investigation of the time-dependent Green's functions serves therefore as an illustration of the general features of the Green's functions formalism, rather than as a description of any new properties of the Ising chain.

Note added in proof: After having submitted this paper, we came across a publication by C. Blomberg ["Some fundamental aspects of many-body problems in statistical thermodynamics" (Stockholm, 1966)], where some of our results have been obtained. We should also mention that some general properties of Green's functions for the Ising model have been discussed in the following series of papers: F. Gorodetzky *et al.*, Phys. Letters **2**, 14 (1962); B. G. S. Doman, *ibid.* **4**, 156 (1963); A. H. Muir, Jr., E. Kankeleit, and F. Boehm, *ibid.* **5**, 161 (1963).

[4] J. S. Marsh, Phys. Rev. **145**, 251 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Possible Kinematics

HENRI BACRY*
The Institute for Advanced Study, Princeton, New Jersey
AND
JEAN-MARC LÉVY-LEBLOND
Physique Théorique, Faculté des Sciences, 06-Nice, France

(Received 27 May 1967)

The kinematical groups are classified; they include, besides space–time translations and spatial rotations, "inertial transformations" connecting different inertial frames of reference. When parity and time-reversal are required to be automorphisms of the groups, and when a weak hypothesis on causality is made, the only possible groups are found to consist of the de Sitter groups and their rotation-invariant contractions. The scheme of the contractions connecting these groups enables one to discuss their physical meaning. Beside the de Sitter, Poincaré, and Galilei groups, two types of groups are found to present some interest. The "static group" applies to the static models, with infinitely massive particles. The other type, halfway between the de Sitter and the Galilei groups, contains two nonrelativistic cosmological groups describing a nonrelativistic curved space–time.

I. INTRODUCTION

It is almost a fact of everyday life that the laws of physics are invariant under space and time "translations"[1] as well as under spatial rotations. The homogeneity of space and time and the isotropy of space were probably recognized a very long time before being explicitly stated. But there exists another type of kinematical invariance, expressing the equivalence of a larger class of frames of reference, the so-called inertial frames. This fact, much less obvious, was clearly realized for the first time by Galilei.[2] In addition to time "translations," space "translations," and rotations, there is thus an additional set of transformations, which we will call, in general, *inertial transformations*, leaving invariant the laws of nature. Due to the isotropy of space (valid at nuclear as well as astronomical dimensions), these inertial transformations form a three-parameter set, so that on the whole there is a ten-parameter group of kinematical transformations, the relativity group of nature, or *kinematical group*.[3]

But the existence of these inertial transformations does not suffice to determine the precise structure of the kinematical group. One still has to tell (i) how the laws of physics transform under inertial transformations; (ii) if there exists some natural curvature of space, or, more generally, of space–time. These two aspects of kinematics are described by the group law.

* Present address: Faculté des Sciences, Marseilles, France. Work supported by the National Science Foundation.

[1] "Translations" are understood here in a generalized sense; they do not necessarily commute when large distances are involved.

[2] G. Galilei, *Dialogo sopra i due massimi sistemi del mondo* (Edizione Nazionale, Firenze), Vol. 7, J.II, pp. 212–214.

[3] Of course, it might be that such an invariance is only approximate. In the context of general relativity, for instance, only very special models of universes possess such a large invariance group. Indeed, a Riemannian four-dimensional manifold has an isometry group with *at most* ten parameters, and this only when its curvature is a constant.

$t_k = t'_k = \cdots = t'_1$ give a complete statistical description of the Ising chain. In particular, we can obtain in this manner the spin-correlation functions calculated earlier by Marsh.[4] We find a complete agreement between his and our results.

As we have already mentioned in the introduction, the Ising chain does not exhibit any nontrivial kinetic behavior. Our investigation of the time-dependent Green's functions serves therefore as an illustration of the general features of the Green's functions formalism, rather than as a description of any new properties of the Ising chain.

Note added in proof: After having submitted this paper, we came across a publication by C. Blomberg ["Some fundamental aspects of many-body problems in statistical thermodynamics" (Stockholm, 1966)], where some of our results have been obtained. We should also mention that some general properties of Green's functions for the Ising model have been discussed in the following series of papers: F. Gorodetzky *et al.*, Phys. Letters **2**, 14 (1962); B. G. S. Doman, *ibid.* **4**, 156 (1963); A. H. Muir, Jr., E. Kankeleit, and F. Boehm, *ibid.* **5**, 161 (1963).

[4] J. S. Marsh, Phys. Rev. **145**, 251 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Possible Kinematics

HENRI BACRY*
The Institute for Advanced Study, Princeton, New Jersey
AND
JEAN-MARC LÉVY-LEBLOND
Physique Théorique, Faculté des Sciences, 06-Nice, France

(Received 27 May 1967)

The kinematical groups are classified; they include, besides space–time translations and spatial rotations, "inertial transformations" connecting different inertial frames of reference. When parity and time-reversal are required to be automorphisms of the groups, and when a weak hypothesis on causality is made, the only possible groups are found to consist of the de Sitter groups and their rotation-invariant contractions. The scheme of the contractions connecting these groups enables one to discuss their physical meaning. Beside the de Sitter, Poincaré, and Galilei groups, two types of groups are found to present some interest. The "static group" applies to the static models, with infinitely massive particles. The other type, halfway between the de Sitter and the Galilei groups, contains two nonrelativistic cosmological groups describing a nonrelativistic curved space–time.

I. INTRODUCTION

It is almost a fact of everyday life that the laws of physics are invariant under space and time "translations"[1] as well as under spatial rotations. The homogeneity of space and time and the isotropy of space were probably recognized a very long time before being explicitly stated. But there exists another type of kinematical invariance, expressing the equivalence of a larger class of frames of reference, the so-called inertial frames. This fact, much less obvious, was clearly realized for the first time by Galilei.[2] In addition to time "translations," space "translations," and rotations, there is thus an additional set of transformations, which we will call, in general, *inertial transformations*, leaving invariant the laws of nature. Due to the isotropy of space (valid at nuclear as well as astronomical dimensions), these inertial transformations form a three-parameter set, so that on the whole there is a ten-parameter group of kinematical transformations, the relativity group of nature, or *kinematical group*.[3]

But the existence of these inertial transformations does not suffice to determine the precise structure of the kinematical group. One still has to tell (i) how the laws of physics transform under inertial transformations; (ii) if there exists some natural curvature of space, or, more generally, of space–time. These two aspects of kinematics are described by the group law.

* Present address: Faculté des Sciences, Marseilles, France. Work supported by the National Science Foundation.

[1] "Translations" are understood here in a generalized sense; they do not necessarily commute when large distances are involved.

[2] G. Galilei, *Dialogo sopra i due massimi sistemi del mondo* (Edizione Nazionale, Firenze), Vol. 7, J.II, pp. 212–214.

[3] Of course, it might be that such an invariance is only approximate. In the context of general relativity, for instance, only very special models of universes possess such a large invariance group. Indeed, a Riemannian four-dimensional manifold has an isometry group with *at most* ten parameters, and this only when its curvature is a constant.

That this is not a trivial question is proved by the fact that inertial transformations have been thought for centuries to be pure Galilean transformations, although we now believe them to be pure Lorentz transformations.[4] Because the Lorentz group recently has been brought in question,[5] it is natural to ask, Are there any other possibilities? We intend here to answer this question under very general assumptions.[6]

II. ASSUMPTIONS AND RESULTS

Our investigation is based upon the use of Lie-algebraic methods. Supposing the kinematical group to be a Lie group is a very natural physical requirement, expressing the supposed continuity of space–time. We call H, P_i, J_i, and K_i ($i = 1, 2, 3$), respectively, the generators of time "translations," space "translations," spatial rotations, and inertial transformations along the i axis. We now assume:

(1) Space is isotropic, meaning that the infinitesimal generators transform correctly under rotations, i.e., H is a scalar and $\mathbf{P}$, $\mathbf{J}$, $\mathbf{K}$ are vectors. In other words, all the Lie brackets involving the angular momentum $\mathbf{J}$ have their standard form:

$$[\mathbf{J}, H] = 0, \tag{1a}$$

$$[\mathbf{J}, \mathbf{J}] = \mathbf{J}, \tag{1b}$$

$$[\mathbf{J}, \mathbf{P}] = \mathbf{P}, \tag{1c}$$

$$[\mathbf{J}, \mathbf{K}] = \mathbf{K}, \tag{1d}$$

where the notation $[\mathbf{A}, \mathbf{B}] = \mathbf{C}$ is a shorthand for $[A_i, B_j] = \epsilon_{ijk} C_k$.

(2) Parity and time-reversal are automorphisms of the kinematical group. This is a natural assumption which greatly simplifies the final result, but is by no means compelling, especially in view of the failure of parity invariance, and probably, although to a much smaller extent, of time-reversal, in weak interactions. However, we stick to this conservative approach throughout the present paper. From the interpretation of inertial transformations as "boosts" in a certain direction, we conclude that their generator $\mathbf{K}$ must be odd under the parity operation Π and the time-reversal Θ, while H, $\mathbf{P}$, and $\mathbf{J}$ transform in the standard way. To sum up, we require the two transformations

$$\Pi: \{H \to H, \mathbf{P} \to -\mathbf{P}, \mathbf{J} \to \mathbf{J}, \mathbf{K} \to -\mathbf{K}\}, \tag{2}$$

$$\Theta: \{H \to -H, \mathbf{P} \to \mathbf{P}, \mathbf{J} \to \mathbf{J}, \mathbf{K} \to -\mathbf{K}\} \tag{3}$$

to leave invariant the Lie brackets defining the Lie algebra of the kinematical group. Let us notice that the combined operation $\Gamma = \Pi\Theta$, which may replace Π and is given by

$$\Gamma: \{H \to -H, \mathbf{P} \to -\mathbf{P}, \mathbf{J} \to \mathbf{J}, \mathbf{K} \to \mathbf{K}\}, \tag{4}$$

is exchanged with Θ when $\mathbf{P}$ and $\mathbf{K}$ are interchanged.

(3) The one-dimensional subgroups generated by the K_i's are noncompact. If this were not the case, the boost parameter u of a general boost $\exp(uK_i)$ would be defined modulo some u_0 such that $\exp(u_0 K_i)$ is the identity transformation. In other words, a sufficiently large boost would be no boost at all. This is an utterly unphysical situation. Notice that we do not require the same compacity for the one-dimensional translation subgroups: Indeed the universe might very well be a closed one. The present hypothesis also may be seen to be equivalent to the very weak causality requirement that if, in a certain reference frame, two events take place at the same point, no kinematical transformation should alter the temporal order of these events.[6] However, the equivalence of these two conditions can only be proved when hypotheses (1) and (2) have been exploited and a suitable notion of space–time has been introduced.

Theorem: Under the assumptions that:

(1) *space is isotropic (rotation invariance);*

(2) *parity and time-reversal are automorphisms of the kinematical groups;*

(3) *inertial transformations in any given direction form a noncompact subgroup,*

then there are eight types of Lie algebras for kinematical groups corresponding to eleven possible kinematics. These algebras are:

R1. The two de Sitter Lie algebras[7] *isomorphic, respectively, to the Lie algebras of $SO(4,1)$ and $SO(3,2)$;*

R2. The Poincaré Lie algebra;[8]

R3. Two "para-Poincaré" Lie algebras, of which one is isomorphic to the ordinary Poincaré Lie algebra but

[4] A. Einstein, Ann. Physik **17**, 891 (1905).

[5] M. Dresden, in *Noncompact Groups in Particle Physics*, Y. Chow, Ed. (W. A. Benjamin, Inc., New York, 1966), and private communication.

[6] V. Lalan [Compt. Rend. Acad. Sci. **203**, 1491 (1936); and Bull. Soc. Math. France **65**, 83 (1937)] has led an investigation in the same spirit. However, his hypotheses are very different and much stronger than ours, since he assumes the relativity group to operate linearly on space–time and the space–time translations to be an invariant subgroup.

[7] W. de Sitter, Amsterdam Proc. **19**, 1217 (1917); **20**, 229 (1917); C. Møller, *The Theory of Relativity* (Oxford Univ. Press, London, 1952), Sec. 134; F. Gürsey, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon & Breach Science Publishers, New York, 1964), p. 365 and additional references therein.

[8] E. P. Wigner, Ann. Math. **40**, 149 (1939). A. S. Wightman, "Lectures on Relativistic Invariance" in *Relations de dispersion et particules elémentaires, Les Houches, 1960* (Hermann & Cie., Paris, 1960).

physically different[9] *and the other is the Lie algebra of an inhomogeneous* $SO(4)$ *group;*

*R*4. *The Carroll Lie algebra*[10]*;*

*A*1. *The two "nonrelativistic cosmological" Lie algebras;*

*A*2. *The Galilei Lie algebra*[11]*;*

*A*3. *The "para-Galilei" Lie algebra*[9]*;*

*A*4. *The "static" Lie algebra.*

While the Lie algebras of class R have no nontrivial central extensions by a one-parameter Lie algebra, those of class A each have one class of such extensions.

We start by giving the proof of this theorem. In the following section, we discuss the physical meaning of the various Lie algebras by considering their contraction scheme, that is, the limiting processes leading from one to the other of these algebras.

Proof: The very simple and straightforward proof consists merely in:

(i) writing the unknown Lie brackets as linear forms on the Lie algebra in our standard basis H, $\mathbf{P}$, $\mathbf{J}$, $\mathbf{K}$;
(ii) taking into account Assumptions 1 and 2;
(iii) imposing Jacobi identity;
(iv) using Assumption 3;
(v) looking for nontrivial central extensions.

We have to examine the five types of Lie brackets not involving $\mathbf{J}$, that is, $[H, \mathbf{P}]$, $[\mathbf{P}, \mathbf{P}]$, $[H, \mathbf{K}]$, $[\mathbf{K}, \mathbf{K}]$, and $[\mathbf{P}, \mathbf{K}]$. Consider, for instance, $[H, \mathbf{P}]$. This bracket is odd under Π and Θ. According to Eqs. (2) and (3), it can only be a linear function of $\mathbf{K}$ (Assumption 2). Since it has a vector character, the bracket $[H, P_i]$ will be proportional to K_i (Assumption 1), a property which may be written in the form $[H, \mathbf{P}] = \alpha\mathbf{K}$. More generally, when Assumption 2 is used, one term at most survives for each bracket, and Assumption 1 requires the admissible linear combinations of generators to be formed only with the covariant tensors I, δ_{ij}, ϵ_{ijk}. We may finally write

$$\begin{aligned} [H, \mathbf{P}] &= \alpha\mathbf{K}, \quad [H, \mathbf{K}] = \lambda\mathbf{P}, \\ [\mathbf{P}, \mathbf{P}] &= \beta\mathbf{J}, \quad [\mathbf{K}, \mathbf{K}] = \mu\mathbf{J}, \\ [\mathbf{P}, \mathbf{K}] &= \rho H, \end{aligned} \tag{5}$$

where the structure constants α, β, λ, μ, ρ are real numbers and $[\mathbf{P}, \mathbf{K}] = \rho H$ is a shorthand for $[P_i, K_j] = \rho\delta_{ij}H$.

Corresponding to the symmetry already noticed of the automorphisms Θ and $\Gamma = \Pi\Theta$ under the exchange $\mathbf{P} \leftrightarrow \mathbf{K}$, we remark that the expressions (5) are invariant under the symmetry S defined by

$$S: \{\mathbf{P} \leftrightarrow \mathbf{K}, \alpha \leftrightarrow \lambda, \beta \leftrightarrow \mu, \rho \leftrightarrow -\rho\}. \tag{6}$$

We now require the Jacobi identity to be satisfied by all the triples of distinct basis elements of what must be a Lie algebra. Due to the fact that we already took into account rotation invariance, the Jacobi identity is automatically obeyed by all the triples containing at least one component of $\mathbf{J}$. Of the seven types of remaining triples [**PPP**], [**PPK**], [**PKK**], [**KKK**], and [H**PP**], [H**PK**], [H**KK**], only four need to be considered if we use the symmetry S. In fact, only three types of triples give rise to constraints on the structure constants. The Jacobi identity for [H**PK**] requires

$$\alpha\mu + \beta\lambda = 0, \tag{7}$$

while for [**PPK**] and [**PKK**], respectively, it requires

$$\beta - \alpha\rho = 0, \tag{8a}$$

$$\mu + \lambda\rho = 0. \tag{8b}$$

Since the constraints (8) imply (7), we are left with only these two conditions (8) to analyze. Due to the peculiar form of the brackets (5), namely, the homogeneity of their left-hand sides with respect to each structure constant, it is seen that any nonzero structure constant may be normalized arbitrarily by a scale change, in particular, to unit absolute value. Sign changes, however, are not always permissible since they could give different real forms of the same complex algebra and consequently lead to different groups. Disregarding the question of signs leads to a classification of the physical algebras in the following eight types (characterized by their null structure constants):

Class R ("relative-time" Lie algebras): $\rho \neq 0$. We have the following types:

*R*1. $\alpha \neq 0$, $\lambda \neq 0$ [hence, by (8), $\beta \neq 0$, $\mu \neq 0$]. One recognizes the Lie algebras of the groups $SO(5)$, $SO(4, 1)$, and $SO(3, 2)$. More precisely, one obtains *two* groups isomorphic to $SO(4, 1)$, depending on whether the subgroup $SO(4)$ is generated by $\mathbf{J}$ and $\mathbf{P}$ or by $\mathbf{J}$ and $\mathbf{K}$. This last group, as well as $SO(5)$, must be rejected in virtue of Assumption 3. Consequently, one is left with the Lie algebras of the two de Sitter groups which will be denoted (dS).

*R*2. $\alpha = 0$, $\lambda \neq 0$ (hence $\beta = 0$, $\mu \neq 0$).

[9] We emphasize that two relativity groups may be isomorphic but completely different from the physical point of view, since the basis elements (H, $\mathbf{P}$, $\mathbf{J}$, $\mathbf{K}$) of their Lie algebras have a well-defined physical meaning and cannot be transformed arbitrarily, except by scale changes. Moreover, even a given real Lie algebra may have many interpretations by a change of normalization (for the para-Galilei groups for instance) or by making the corresponding group acting on a space which is not a homogeneous space of the group.

[10] J.-M. Lévy-Leblond, Ann. Inst. Henri Poincaré **3**, 1 (1965). See also N. D. Sen Gupta, Nuovo Cimento **44**, 512 (1966).

[11] M. Hamermesh, Ann. Phys. (N.Y.) **9**, 512 (1960); J.-M. Lévy-Leblond, J. Math. Phys. **4**, 776 (1963).

TABLE I. Characteristic Lie brackets for the Lie algebras of the various kinematical groups. The terms inside parentheses refer to the possible central extensions of the algebras. They vanish for the nonextended Lie algebras. To each Lie algebra corresponds one physical interpretation except for the para-Galilei algebra endowed with two physical interpretations.

	(*dS*)1		(*P*)2	(*P'*)3		(*C*)4	(*N*)1		(*G*)2	(*G'*)3	(*St*)4
	de Sitter			Inh.	Para-		Expanding	Oscillating		Para-	
	SO(4, 1)	*SO*(3, 2)	Poincaré	*SO*(4)	Poincaré	Carroll	Universe	Universe	Galilei	Galilei	Static
[*H*, **P**]	**K**	−**K**	0	**K**	−**K**	0	**K**	−**K**	0	**K**	0
[*H*, **K**]	**P**	**P**	**P**	0	0	0	**P**	**P**	**P**	0	0
[**P**, **P**]	**J**	−**J**	0	**J**	−**J**	0	0	0	0	0	0
[**K**, **K**]	−**J**	−**J**	−**J**	0	0	0	0	0	0	0	0
[**P**, **K**]	*H*	*H*	*H*	*H*	*H*	*H*	(*I*)	(*I*)	(*I*)	(*I*)	(*I*)

They are the Lie algebras of the Poincaré group (*P*)[8] and of inhomogeneous *SO*(4). This last algebra does not satisfy Assumption 3 and must be discarded.

*R*3. $\alpha \neq 0$, $\lambda = 0$ (hence $\beta \neq 0$, $\mu = 0$).

One obtains in that case two Lie algebras isomorphic respectively to those of *R*2, but where the roles of space translations and inertial transformations have been exchanged. In opposition with the case *R*2, both algebras agree with Assumption 3. We will denote these groups by (*P'*).

*R*4. $\alpha = 0$, $\lambda = 0$ (hence $\beta = 0$, $\mu = 0$).

One recognizes the Lie algebra of the Carroll group (*C*)[9] which satisfies Assumption 3.

Class A ("*absolute-time*" *Lie algebras*): $\rho = 0$ (hence $\beta = 0$, $\mu = 0$). We have also four types in this class, all of them satisfying Assumption 3.

*A*1. $\alpha = 0$, $\lambda \neq 0$.

It is shown below that the two Lie algebras of this type are those of two "nonrelativistic cosmological groups," which we feel appropriate to call "Newton groups" (*N*).

*A*2. $\alpha \neq 0$, $\lambda \neq 0$.

This is the Lie algebra of the Galilei group (*G*).[10]

*A*3. $\alpha \neq 0$, $\lambda = 0$.

One recognizes a Lie algebra isomorphic to the Galilei Lie algebra. The situation is similar to cases *R*2 and *R*3 above. The group associated with this Lie algebra will be called para-Galilei group (*G'*).[11]

*A*4. $\alpha = 0$, $\lambda = 0$.

This is the Lie algebra of what we call the static group (*St*) for reasons to be discussed below.

The above classification is summarized in Table I, which clearly exhibits the Lie brackets defining the various Lie algebras.

We must now give the proof of the last part of the theorem concerning the possible central extensions of these Lie algebras by a one-dimensional Lie algebra. In fact, this is necessary for constructing the projective representations of the kinematical groups, representations which are of interest for the quantum-mechanical applications.[8] Using standard methods,[12] one easily shows that the four algebras of Class *A* each have one type of central extension. They are very simply characterized: the central element *I* of the extended Lie algebra (i.e., the one-dimensional extending Lie algebra) only appears in the Lie bracket [**P**, **K**] which is zero for the nonextended Lie algebras ($\rho = 0$ defines the Class *A*). This element *I* thus takes in this bracket the place occupied by the Hamiltonian *H* in cases *R* (see Table I). This property justifies our classification of possible Lie algebras in two classes even from a physical point of view since the nonrelativistic groups will be characterized by the property of additivity of masses. Another justification is provided by the fact that time will have an absolute character for every group of Class *A* (see below).

III. GENERAL PHYSICAL DISCUSSION

There is no need to comment on the appearance in our classification of the de Sitter, Poincaré, and Galilei groups. These are well known and have been thoroughly studied,[7,8,10] especially from the point of view of their role in quantum physics, for the last two ones at least. The surprising fact is perhaps that there are other possibilities, which we discuss now.

The Carroll group has already been introduced as a second nonrelativistic limit of the Poincaré group,[9] describing low-velocity transformations of large spacelike intervals. This explains why we devote a whole section to the static group and another one to the Newton groups.

In this section, we intend to examine the relationships between all the Lie algebras listed in Sec. II from the point of view of their physical interpretation. Such an approach is required by the fact that the real group of symmetry is either the Poincaré group or a group from which the Poincaré group is an approximation. The mathematical aspect of the relationships we are looking for is described by the process of

[12] V. Bargmann, Ann. Math. **59**, 1 (1954).

contraction, a concept which has been defined by Inönu and Wigner[13] in order to give a precise meaning to some singular limits encountered when trying to replace "exact" kinematical groups by "approximate" ones. For instance, the Galilei group is obtained by contracting the Poincaré group with respect to the direct product of the rotation and time-translation subgroups; leaving unchanged the generators $\mathbf{J}$ and H of these subgroups, one substitutes $\mathbf{P} \to \epsilon\mathbf{P}$ and $\mathbf{K} \to \epsilon\mathbf{K}$ in the Lie algebra and considers the (singular) limit $\epsilon \to 0$ of the Lie brackets. One is left with the Lie algebra of the Galilei group. The physical meaning of the contraction is very simple: the factor ϵ has affected the generators $\mathbf{K}$ and $\mathbf{P}$, so that the contracted (Galilei) group will describe a situation where velocities (parameters associated to $\mathbf{K}$) and space translations (parameters associated to $\mathbf{P}$) are small. More precisely, taking the light speed as the unit speed, velocities have to be small compared to the unit, and spacelike intervals small compared to timelike intervals. This is why we call such a contraction a *speed–space contraction*. On this particular example, one sees immediately that a given kinematical group may have, beside its "intrinsic" interpretation (for instance Newtonian mechanics for the Galilei group), one or more "approximate" interpretations (the Galilei group is an approximate symmetry of special relativity but also of a de Sitter universe).

The contraction of a group always being defined with respect to a particular subgroup, one can consider four types of physical contractions for general kinematical groups, since there are four possible rotation-invariant subgroups of any kinematical group; indeed, it is easily seen, looking at Table I, that the only rotation-invariant Lie subalgebras common to all eight types of kinematical Lie algebras are generated by $(\mathbf{J}, H)$, $(\mathbf{J}, \mathbf{P})$, $(\mathbf{J}, \mathbf{K})$, and $(\mathbf{J})$. Let us discuss in turn the corresponding four types of contraction:

Speed–Space Contraction. It is defined by

$$\mathbf{P} \to \epsilon\mathbf{P}, \quad \mathbf{K} \to \epsilon\mathbf{K}, \quad \epsilon \to 0. \tag{9a}$$

It is the contraction with respect to the rotation and time-translation subgroups which leads from the Poincaré group to the Galilei group.[13] This contraction process corresponds to the passage from relative to absolute time and maps in a one-to-one correspondence the groups of Class R onto the groups of Class A:

$$R1 \to A1, \quad R2 \to A2, \quad R3 \to A3, \quad R4 \to A4. \tag{9b}$$

[13] I. Segal, Duke Math. J. **18**, 221 (1951); E. Inönu and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953); **40**, 119 (1954); E. Saletan, J. Math. Phys. **2**, 1 (1961).

Speed–Time Contractions. One puts

$$H \to \epsilon H, \quad \mathbf{K} \to \epsilon\mathbf{K}, \quad \epsilon \to 0, \tag{10a}$$

contracting with respect to the three-dimensional Euclidean group. The physical interpretation of the contracted group is obtained in considering only low velocities and large spacelike intervals. This contraction leads from the Poincaré group to the Carroll group.[9] According to its very definition, such a process only yields groups describing intervals connecting events without any causal connection, hence without much physical applications. The mapping corresponding to this contraction is

$$R1 \to R3, \quad R2 \to R4, \quad A1 \to A3, \quad A2 \to A4. \tag{10b}$$

Let us notice that under a speed–time contraction, space intervals become absolute.

Space–Time Contraction. By choosing

$$\mathbf{P} \to \epsilon\mathbf{P}, \quad J \to \epsilon H, \quad \epsilon \to 0, \tag{11a}$$

one contracts with respect to the group of rotations and inertial transformations (Lorentz group). Physically, this means that we consider very small units of space and time. A group obtained in this way describes *local* transformation properties, but, contrary to the preceding contractions, for arbitrarily large inertial transformations. The space–time contraction realizes the following mapping:

$$R1 \to R2, \quad R3 \to R4, \quad A1 \to A2, \quad A3 \to A4. \tag{11b}$$

Because of the local character of the contracted groups, it seems natural to call $R1$, $R3$, $A1$, and $A3$ *cosmological groups* and their contracted groups $R2$, $R4$, $A2$, and $A4$ *local groups*.

Before examining the fourth kind of contraction, we will refer the reader to Fig. 1, where the eight types of Lie algebras have been put at the eight vertices of a cube, the edges of which describe the contraction processes we have just discussed. Each upper face of the cube is transformed under a given type of contraction into the opposite face.

General Contraction. Here we put

$$\mathbf{P} \to \epsilon\mathbf{P}, \quad H \to \epsilon H, \quad \mathbf{K} \to \epsilon\mathbf{K}, \quad \epsilon \to 0, \tag{12}$$

contracting with respect to the rotation group which alone is unchanged.

This contraction combines the features of the preceding ones and will lead to a description of local and small inertial transformations. As we will see, it is a rather drastic operation from a physical point of view.

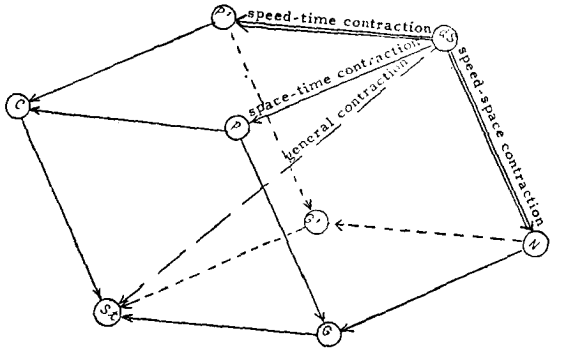


FIG. 1. The contraction scheme for the relativity groups.

Relative-time groups:	(*dS*),	(*P'*),	(*P*),	(*C*).
Absolute-time groups:	(*N*),	(*G'*),	(*G*),	(*St*).
Relative-space groups:	(*dS*),	(*N*),	(*P*),	(*G*).
Absolute-space groups:	(*P'*),	(*G'*),	(*C*),	(*St*).
Cosmological groups:	(*dS*),	(*N*),	(*P'*),	(*G'*).
Local groups:	(*P*),	(*G*),	(*C*),	(*St*).

Remark: The effect of the symmetry S of Eq. (6) is equivalent to a symmetry of the cube with respect to the plane containing the vertices (*dS*), (*N*), (*C*), *and* (*St*).

All groups are contracted in this way into the static group.

We notice that geometrical concepts are naturally associated to each kind of contractions. For instance, the group obtained via a speed–space contraction describes space–time properties in the neighborhood of a worldline. This is due to the fact that the contraction is made with respect to the subgroup which leaves invariant a given worldline. On the other hand, space–time contractions are associated with events, points in space–time.

We may now state the main result of this section: the eight possible types of kinematical Lie algebras may all be obtained from one among them (*R*1) by a suitable sequence of contractions[14] (i.e., a contraction in the sense given by Saletan[13]). The detailed contraction scheme is much better understood by looking at Figs. 1 and 2 than by reading a verbal description. We note immediately that the Poincaré group is a contracted group of the two de Sitter groups only; this property is essential, since it follows that the only possible *exact* symmetry groups compatible with our assumptions are the Poincaré and the de Sitter groups. We examine in detail this statement in our conclusion.

The cube of Fig. 1 enables us to understand the relationships between all the groups we obtained and to find their interpretations as approximated symmetries of the Poincaré or of the de Sitter symmetries.

(a) The groups (*P'*) and (*C*) are obtained respectively from (*dS*) and (*P*) via a speed–time contraction. Thus they describe the transformation properties of large spacelike intervals, forbidding causal relations between events. The physical interest of these groups then is very much reduced. One may notice that, for these two groups, $[H, \mathbf{K}] = 0$ (see Table I), which means that inertial transformations do not act effectively on time translations; this is consistent with the large spacelike character of the considered intervals, but underlines the unphysical nature of the groups, since ordinarily an inertial transformation followed by a time translation yields a space translation (waiting a certain time after boosting an object, one finds it displaced!). The same considerations could seem to apply to the para-Galilei and static groups. However, we show that these groups are endowed with special significance.

(b) The para-Galilei group (*G'*) can be obtained from (*dS*) in the following way:

$$\mathbf{P} \to \epsilon\mathbf{P}, \quad H \to \epsilon H, \quad \mathbf{K} \to \epsilon^2\mathbf{K}, \quad \epsilon \to 0,$$

and so describes a situation where timelike and spacelike intervals are reduced but the velocities are completely negligible, that is to say, a situation of a static "de Sitter" universe. Looking at the cube of Fig. 1, one can expect an analogous interpretation for the static group *relative to the Poincaré group* but the fact that this group can be reached directly through a general contraction provides it with even more general properties.

(c) The groups (*N*) acquire an obvious meaning: they are obtained from the de Sitter groups by performing a speed–space contraction. Thus they describe the low-velocity transformations of the universe at large endowed with an absolute time, hence their denomination as "nonrelativistic cosmological groups."

(d) It is a remarkable fact that, without any additional hypotheses, each of our groups can be considered as a group of transformations acting on space–time coordinates. Indeed, it results from our basic assumptions that rotations and inertial transformations automatically form a subgroup of any kinematical group; this is seen on Table I, where $\mathbf{J}$ and $\mathbf{K}$ clearly form a Lie subalgebra. As a consequence, there exists a four-dimensional homogeneous space for each kinematical group, namely, its quotient by the six-dimensional subgroup of rotations plus inertial transformations. This inhomogeneous space may be identified with the manifold of the space–time translations[15] and gives a purely group-theoretical definition of space–time itself. From the group law, one immediately obtains the formulas defining the action of the

[14] Conversely, the eight possible kinematical Lie algebras may be considered as "deformations" of the static Lie algebra. For a discussion of this concept see M. Lévy-Nahas, J. Math. Phys. **8**, 1211 (1967).

[15] Note that the space–time translations do not form a subgroup in general. It is a subgroup in the Poincaré, Carroll, Galilei, and static cases. This subgroup is, in fact, an invariant subgroup.

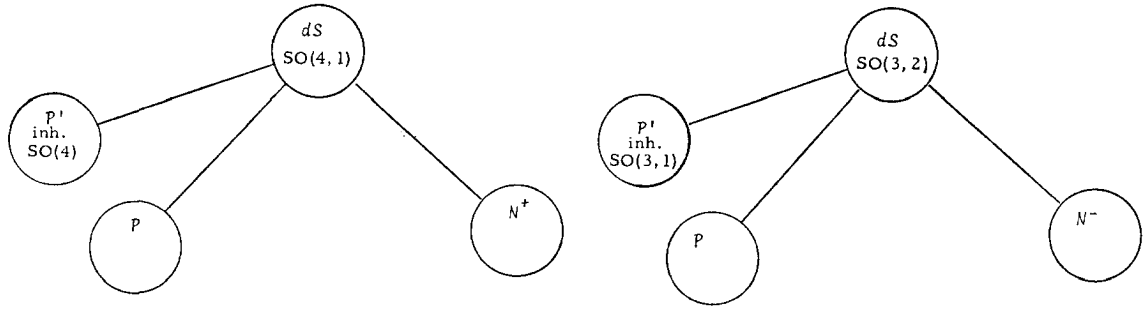


FIG. 2. Top of the cube of Fig. 1 (detail).

group on space–time coordinates. This is where we see that time intervals are not invariant under inertial transformations for type R groups, while they are for type A groups. Indeed, this results directly from the presence or absence of the time-translation generator H in the derived algebra, i.e., in the "right-hand side" of the Lie brackets (see Table I). Of course, we have no guarantee that the group acts effectively on space–time, i.e., that every group element induces a nontrivial transformation. As a matter of fact, for the static group (and the para-Galilei group as well), inertial transformations do not act on space–time; this results from the commutation relations $[H, \mathbf{K}] = 0$, $[\mathbf{P}, \mathbf{K}] = 0$ (Table I). But it is an advantage of our approach not to have considered beforehand the kinematical groups as effectively acting on space–time, since this would have eliminated the static group, for instance, despite its real physical content.

(e) By identifying space–time with a homogeneous space of the Lie algebra, we have excluded some other possibilities offered in making the group acting *on the closure* of a homogeneous space. The group proposed by Dresden[5] in his model is claimed to fulfill all our three assumptions. In any case space–time cannot be identified as a homogeneous space of the Dresden group.

IV. THE STATIC GROUP

We saw that the static group is reached from any of the other kinematical groups by performing the whole sequence of physical contractions in any order, or the general contraction directly. This somewhat obscures its possible physical meaning.

The interpretation of the static group appears most clearly when we relate it to its closest physical neighbor, the Galilei group, by a slight extension of our considerations on the contraction scheme. It has been mentioned that, in quantum physics, the state space of a system invariant under a group of transformations is a representation space for a projective (i.e., up to a factor) representation of the group, that is, a representation of a central extension of the group by a one-parameter "phase group." Contrarily to the de Sitter and Poincaré groups, the Galilei and static groups have nontrivial extensions. The Lie algebras of the extended groups differ from those of the original group by the addition in the Lie brackets of scalar terms, multiples of the phase-group generator I. As indicated in Table I, for kinematical groups, such terms can always be reduced so as to appear only in the commutator $[\mathbf{P}, \mathbf{K}]$. For the Galilei group, it is most important that only the nontrivial projective representations have a physical interpretation; the vector representations do not allow for any sensible notion of localizability.[16] The phase generator I in this case physically appears as the mass operator. Since the mass obeys a superselection rule,[10,12] we must consider systems with a definite mass value, say m. Within their state space, one has

$$I = m\mathbf{1}, \tag{13}$$

so that the commutation relation

$$[P_i, K_j] = \delta_{ij} I \tag{14}$$

is nothing but the canonical commutation relation

$$[P_i, Q_j] = \delta_{ij}\mathbf{1} \tag{15}$$

between the momentum $\mathbf{P}$ and the center-of-mass position operator $\mathbf{Q} = (1/m)\mathbf{K}$. Taking $\mathbf{Q}$ as a basis vector of our Lie algebra instead of $\mathbf{K}$, we have to rewrite the Lie bracket involving the Hamiltonian

$$[H, \mathbf{Q}] = \frac{1}{m}\mathbf{P}, \tag{16}$$

which is just the velocity operator. But then we see that by taking the limit $m \to \infty$ we obtain a new Lie algebra with the property (15) and

$$[H, \mathbf{Q}] = 0. \tag{17}$$

This is just the Lie algebra of the extended static group! In more technical terms, we have contracted

[16] E. Inönu and E. P. Wigner, Nuovo Cimento **9**, 705 (1952); A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).

a central extension of the Galilei group into a central extension of the static group with respect to the rotation and space–time translations subgroup; that is taking the limit

$$\mathbf{K} \to \epsilon\mathbf{K}, \quad I \to \epsilon I, \quad \epsilon \to 0. \tag{18}$$

If we now point out that, for the same reasons as in the Galilean case (lack of localizability), the vector representations of the static group have no physical meaning, the interpretation of the group is quite clear. We have just shown that this group represents "infinitely massive" systems. One sees how, due to the $m \to \infty$ limit, the position operator of the system now commutes with the Hamiltonian (17), instead of giving the velocity operator (16). This, of course, means that the system cannot move—whence the name of the group. The key point is that such a static system is invariant not only under the Euclidean motion group and time translations; the additional invariance under pure Galilean transformations, although broken by the infinite-mass limit, leaves its imprint by ensuring the existence of a position operator, endowing the system with nice localization properties.

It may be interesting to give the group law. Calling

$$(b, \mathbf{a}, \mathbf{u}, R) = e^{bH} e^{\mathbf{a}\cdot\mathbf{P}} e^{\mathbf{u}\cdot\mathbf{K}} e^{\mathbf{n}\cdot\mathbf{J}} \tag{19}$$

the generic element with b a time translation, $\mathbf{a}$ a space translation, $\mathbf{u}$ a pure static inertial transformation,[17] and R a rotation, one easily computes the multiplication law

$$(b', \mathbf{a}', \mathbf{u}', R')(b, \mathbf{a}, \mathbf{u}, R) = (b' + b, \mathbf{a}' + R\mathbf{a}, \mathbf{u}' + R'\mathbf{u}, R'R). \tag{20}$$

The peculiar feature of this law, as opposed to its Poincaré and Galilei counterpart, is that the pure inertial transformations do not act on the space–time (identified to the space–time translations subgroup).[18] Any system is at rest in any inertial reference frame and the inertial motions are no motion at all.

Finally, we can construct the irreducible unitary projective representations of the static group, using the standard "little group" technique.[8] Taking as our representation space the space of $(2s + 1)$-component square-integrable functions in momentum space, the Lie-algebra representation is

$$\begin{aligned} H &= E = \text{const}, \\ P &= \mathbf{p}, \\ \mathbf{K} &= i\nabla_p, \\ \mathbf{J} &= -i\mathbf{p} \times \nabla_p + \mathbf{S}, \end{aligned} \tag{21}$$

where the three matrices S_i $(i = 1, 2, 3)$ generate a $(2s + 1)$-dimensional representation of the rotation group. In (21), the generators have been redefined so as to be Hermitian, having been multiplied by (i). We see that:

(1) a "static elementary system" can possess a spin, as can Galilean and relativistic systems;

(2) the energy E of a static system is a constant, independent of its momentum $\mathbf{p}$;

(3) under a pure inertial transformation ($\mathbf{u}$), the momentum of a static system changes according to $\mathbf{p} \to \mathbf{p} + \mathbf{u}$.

All this is perfectly in keeping with the intuitive picture of a static system as having a constant energy, rotational degrees of freedom, and being able to exchange momentum in arbitrary amounts while its energy does not change. Such results obviously would not obtain if one only considered Euclidean invariance.

As an application, it is amusing to consider the various static models of quantum field theories, such as the Chew–Low model.[19] One easily sees that they do have the static group as an invariance group. It may be noted that Feynman had proposed to modify the Chew–Low model so as to take into account Galilean invariance.[20] It has been pointed out elsewhere[21] that this cannot be done consistently because of the Galilean mass superselection rule. Accepting, however, the intrinsically static nature of the model, one need not introduce additional interaction terms, as proposed by Feynman, to have a model with some remnant of Galilean invariance, namely static invariance. Anyway, the new terms had the effect of badly spoiling the agreement between the predictions of the standard model and the experimental results.[22]

V. THE NEWTON GROUPS

The Newton groups are characterized by a nonvanishing Lie bracket $[H, \mathbf{P}]$, proportional to the

[17] Before the contraction, the parameter $\mathbf{u}$ is linked to the velocity $\mathbf{v}$ of the pure Galilean transformation by $\mathbf{u} = m\mathbf{v}$. The contraction process amounts to taking the limits $m \to \infty$, $\mathbf{v} \to 0$ with $\mathbf{u}$ fixed.

[18] However, for the nontrivial central extensions of the static group, the inertial transformations act on the phase group. We have

$$(\theta', b', \mathbf{a}', \mathbf{u}', R')(\theta, b, \mathbf{a}, \mathbf{u}, R) = (\theta' + \theta + \mathbf{u}' \cdot R'\mathbf{a}, b' + b, a' + R'a, \mathbf{u}' + R'\mathbf{u}, R'R),$$

where θ is the generic element of the phase group (a real number). It is very amusing to notice that such an extension of the static group is isomorphic to the (trivial) extension of the Carroll group. The isomorphism is simply realized by exchanging the roles of the phase group and the time-translation group ($H \leftrightarrow I$).

[19] S. S. Schweber, *Introduction to Relativistic Quantum Field Theory* (Row and Peterson, Evanston, 1961), Chap. 12 and original references therein.

[20] R. P. Feynman, *Proceedings of the Third Annual Conference on High Energy Nuclear Physics, Rochester, 1952* (University of Rochester, 1953), p. 87.

[21] J.-M. Lévy-Leblond, Commun. Math. Phys. **4**, 157 (1967).

[22] E. M. Henley and M. A. Ruderman, Phys. Rev. **90**, 719 (1952).

boost generator $\mathbf{K}$. Depending on the sign of the coefficient α [see Eq. (5)], we have one of two possible real forms for the Lie algebra. We choose to write

$$[H, \mathbf{P}] = \pm(1/\tau^2)\mathbf{K}, \tag{22}$$

where τ is a characteristic time, which might be chosen as the natural time unit, but which we prefer to keep apparent in our expressions. Depending on the sign in (22), we will have a group N_+ or N_-.

In order to acquire some familiarity with the new groups, let us first compute the group law by direct exponentiation of the Lie brackets. We denote the generic element by

$$(b, \mathbf{a}, \mathbf{v}, R) = e^{bH}e^{\mathbf{a}\cdot\mathbf{P}}e^{\mathbf{v}\cdot\mathbf{K}}e^{\mathbf{n}\cdot\mathbf{J}} \tag{23}$$

(the vector $\mathbf{n}$ defines the axis and angle of the rotation R). Multiplying together two such elements, reordering the product in the same normal form (23) by repeatedly using Baker–Hausdorff formula and the Lie-algebra properties of the infinitesimal generators, we obtain

$$(b', \mathbf{a}', \mathbf{v}', R')(b, \mathbf{a}, \mathbf{v}, R)$$
$$= \left(b' + b, \cosh\frac{b}{\tau}\mathbf{a}' + \tau\sinh\frac{b}{\tau}\mathbf{v}' + R'\mathbf{a},\right.$$
$$\left.\cosh\frac{b}{\tau}\mathbf{v}' + \frac{1}{\tau}\sinh\frac{b}{\tau}\mathbf{a}' + R'\mathbf{v}, R'R\right) \quad \text{for} \quad N_+, \tag{24a}$$

$$\left(b' + b, \cos\frac{b}{\tau}\mathbf{a}' + \tau\sin\frac{b}{\tau}\mathbf{v}' + R'\mathbf{a},\right.$$
$$\left.\cos\frac{b}{\tau}\mathbf{v}' + \frac{1}{\tau}\sin\frac{b}{\tau}\mathbf{a}' + R'\mathbf{v}, R'R\right) \quad \text{for} \quad N_-. \tag{24b}$$

The group can be considered as a group of transformations of the space–time manifold. Under a general transformation $(b, \mathbf{a}, \mathbf{v}, R)$, the coordinates $(\mathbf{x}, t)$ of a physical event are transformed according to

$$\begin{cases}\mathbf{x}' = R\mathbf{x} + \mathbf{v}\tau\sinh\dfrac{t}{\tau} + \mathbf{a}\cosh\dfrac{t}{\tau},\\ t' = t + b,\end{cases} \quad \text{for} \quad N_+, \tag{25a}$$

or

$$\begin{cases}\mathbf{x}' = R\mathbf{x} + \mathbf{v}\tau\sin\dfrac{t}{\tau} + \mathbf{a}\cos\dfrac{t}{\tau},\\ t' = t + b,\end{cases} \quad \text{for} \quad N_-. \tag{25b}$$

Notice that, as in the de Sitter case, these are nonlinear transformations; more precisely, the new spatial coordinates do not depend linearly on the time. This is in direct contrast to the Poincaré and Galilei groups. Physically, it means that inertial motions, obtained by subjecting a motionless material point (located at the origin $\mathbf{x} = 0$ for definiteness) to a pure inertial transformation, are no longer uniform motions, but are given by

$$\mathbf{x} = \mathbf{v}\tau\sinh\frac{t}{\tau} \quad \text{for} \quad N_+ \tag{26a}$$

or

$$\mathbf{x} = \mathbf{v}\tau\sin\frac{t}{\tau} \quad \text{for} \quad N_-. \tag{26b}$$

Inertial motions thus are either exponentially accelerated (N_+) or oscillating (N_-). Correspondingly, N_+ and N_-, respectively, describe expanding and oscillating universes.

It seems, then, that both groups have features very much like the de Sitter groups[7] from which they derive by speed–space contraction, $SO(4, 1)$ leading to N_+ and $SO(3, 2)$ to N_-. In other words, despite the nonrelativistic limit, the effects of space–time curvature are present, due to the fact that we are still considering the universe on a large scale of time. These effects disappear when we take the local limit, i.e., $\tau \to \infty$, which immediately yields the Galilean formulas, as expected, for both N_+ and N_-. Notice that in such a universe there is an absolute time: as in the Galilean case, simultaneity of two events is preserved by an inertial transformation. We see, then, that the groups N may be interpreted as describing the kinematics of nonrelativistic universes at large.[23] We propose to call "Newton groups" these nonrelativistic cosmological groups which bear to the Galilei group the same relation as the de Sitter groups to the Poincaré group.

It is of interest for possible applications to quantum physics to construct the unitary irreducible representations of the Newton group. As in the Galilei and static cases, only the projective representations may have a physical meaning, for lack of a proper localizability notion for vector representations. There are two invariants of the extended Lie algebra, which we write:

$$Q_1 = 2IH - \mathbf{P}^2 \pm (1/\tau^2)\mathbf{K}^2 \equiv 2IU \quad \text{for} \quad N_\pm, \tag{27a}$$
$$Q_2 = (I\mathbf{J} - \mathbf{P}\times\mathbf{K})^2 \equiv I^2\mathbf{S}^2. \tag{27b}$$

The interpretation is very similar to the Galilean case: I appears as a "mass," obeying a superselection rule; from Q_1 one defines an "internal energy" U and from Q_2 a "spin" S. The most novel feature of a "particle" described by such a representation is that its energy is

[23] A perfectly consistent, nonrelativistic, Newtonian model of expanding universe has been described by C. Callan, R. H. Dicke, and P. J. E. Peebles, Am. J. Phys. 33, 105 (1965). We do not expect it, however, to be invariant under our nonrelativistic cosmological group, since this group, as the de Sitter group from which it comes, certainly describes an empty universe, contrarily to the cited model,

no longer invariant under a translation, as one sees directly in Eq. (22). Also it is amusing to notice that in the N_- case the kinetic energy of the particle, that is,

$$T = H - U = 1/2I[\mathbf{P}^2 + (1/\tau^2)\mathbf{K}^2], \tag{28}$$

is quantized, which is not surprising in view of the "compactness" of the corresponding universe. The oscillator levels have a separation

$$\delta E = \hbar/\tau, \tag{29}$$

in agreement with the uncertainty principle, since τ may be thought of as the "lifetime" of this oscillating universe. Of course, δE is absurdly small ($\delta E \simeq 2 \times 10^{-33}$ eV for $\tau \simeq 10^{10}$ yr).

VI. CONCLUSIONS

We think it rather remarkable that the very simple conditions which we require so severely constrain the possible relativity groups that we found only a small number of more or less degenerate cousins of the de Sitter group. Besides the known de Sitter, Poincaré, and Galilei groups, there are two physically interesting cases. The static group is appropriate to the description of the various static models where the studied particles are supposed to have infinite masses. The Newton groups present features quite close to those of the de Sitter groups, but in a nonrelativistic situation, so that they are perhaps closer to one's intuition and may be used as a simple model to understand some effects of space–time curvature.

The present work could be extended by relaxing the conditions that parity and time reversal be automorphisms of the relativity groups. This can be done, but the number of possible cases become very large and many of them seem difficult to interpret in physical terms.[24] This is why in this paper we have presented only the simplest approach.

It seems important, in our opinion, to recall in the conclusion one of the main consequences of our theorem: With the requirements of kinematical rotation, parity, and time-reversal invariance, there exists only one way to "deform" the Poincaré group, namely, in endowing space–time with a certain curvature.[25]

ACKNOWLEDGMENTS

The authors are grateful to F. Lurçat and J. Nuyts for stimulating discussions and their interest in this work. One of us (H. B.) wishes to acknowledge the hospitality extended to him at The Institute for Advanced Study. Part of this work was done while the other author (J.-M. L.-L.) benefitted from the hospitality of the University of Rochester.

[24] H. Bacry, J.-M. Lévy-Leblond, and J. Nuyts (unpublished).

[25] This result is, of course, related to the statement that "the only deformations of the Poincaré group are the two de Sitter groups" (see Ref. 14), but the quoted result concerns *abstract* groups, whereas our work is based on concrete considerations. It is not difficult to prove, for instance, that the Poincaré group can be "deformed" in groups isomorphic to itself. An example can be found in H. Bacry and N. P. Chang "Kinematics at Infinite Momentum" [Ann. Phys. (N.Y.) (to be published)]. See also Ref. 9.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Unitary Irreducible Representations of $SU(2, 2)$. II

TSU YAO
Department of Physics, Rutgers, The State University, New Brunswick, New Jersey

(Received 31 July 1967; revised manuscript received 17 October 1967)

Paper II of this series [Paper I in J. Math. Phys. **8**, 1931 (1967)] is concerned with a general study of the degenerate representations. The explicit expressions for the "raising" and "lowering" functions $a_i^3(p, q, \lambda)$, and $b_i^3(p, q, \lambda)$, $i = 1, 2, 3, 4$ are found. The three Casimir operators C_2, C_3, and C_4 depend on only two complex parameters A and B, a fact reflecting the degenerate nature of the representations under study here. The finite representations are studied first, and thus provide a proof for the degenerate part of Theorem 2, Paper I. The unitary representations are studied next, and we find that there are fourteen classes of degenerate unitary irreducible representations. There are two continuous series, ten discrete series, and two series which depend on one discrete and one continuous parameter. The degenerate part of the $D^{\pm}$ series is studied, and thus provides an explicit demonstration of Harish-Chandra's theorem.

1. DEGENERATE REPRESENTATIONS

In Paper II of this series we present a general study of the degenerate representations. We find that there are fourteen classes of degenerate unitary irreducible representations: ten are in the discrete series, two are in the continuous series, and two belong to neither. We adhere to the notations used in Paper I.[1]

We begin by making some general remarks about irreducible representations of $SU(2, 2)$. These representations may be degenerate or nondegenerate, finite or infinite. We classify them according to the shape of their p–λ diagrams which were introduced in Paper I. There are five cases to be considered.

Let $p = j + k$, $q = j - k$, and $|p, q, \lambda)$ be an irreducible representation of $SU(2) \times SU(2) \times U(1)$ contained in an irreducible representation of $SU(2, 2)$. For fixed p, we have five cases, as shown in Figs. 1–5.

Figures 1(a), 1(b), and 1(c) are finite-dimensional, and, from Theorem 1,[2] 1(a) and 1(b) are degenerate, while 1(c) is nondegenerate. Figure 1(d) is infinite-dimensional, since p is not bounded from above. Cases 2–5 are all infinite-dimensional. From Theorems I3 and I4, we see that unitary irreducible representations in the D^- series are in 2(a) and 2(b), and those in the D^+ series are in 3(a) and 3(b). Figures 2(a) and 3(a) are degenerate, while 2(b) and 3(b) are nondegenerate.

We next consider an arbitrary degenerate representation, i.e., not necessarily a degenerate unitary irreducible representation in the discrete series. Equations (I4.10) and (I4.11) are now recursion relations, since we may drop the labels α, ρ. These equations can be solved in a straightforward manner.

Let $p = j + k, q = j - k$, and we write $a_i(p, q, \lambda) = a_i(j, k, \lambda)$, $b_i(p, q, \lambda) = b_i(j, k, \lambda)$, $i = 1, 2, 3, 4$. Let p_0 = minimum value of p in an irreducible representation of $SU(2, 2)$. Then

$$a_4(p_0, q, \lambda) = b_4(p_0, q, \lambda) = 0. \tag{1.1}$$

Equation (I4.10) now gives ($p_0 > 0$)

$$\begin{aligned} &2p_0(p_0 + q + 1)(p_0 - q + 1)(p_0 + q + 2) \\ &\quad \times (p_0 - q + 2)a_1^2(p_0, q, \lambda) \\ &= p_0(p_0 + 1)C_2 - 2qC_3 \\ &\quad - p_0(p_0 + 1)\{(p_0 + q + \lambda + 2)(p_0 - q + \lambda + 2) \\ &\quad + 2(p_0 - 1)(p_0 + 2)\}. \end{aligned} \tag{1.2}$$

For $p_0 > 0$, let $C_2 = 2(p_0 + 2)(p_0 - 1) + A^2$, $C_3 = p_0(p_0 + 1)A$; then Eq. (1.2) becomes

$$a_1^2(p_0, q, \lambda) = \frac{(p_0 + 1)[(q - A)^2 - (p_0 + \lambda + 2)^2]}{2(p_0 - q + 1)(p_0 - q + 2)(p_0 + q + 1)(p_0 + q + 2)}. \tag{1.3}$$

Similarly, from Eq. (I4.11), we have

$$b_1^2(p_0, q, \lambda) = \frac{(p_0 + 1)[(q + A)^2 - (p_0 - \lambda + 2)^2]}{2(p_0 - q + 1)(p_0 - q + 2)(p_0 + q + 1)(p_0 + q + 2)}, \tag{1.4}$$

where $|q| \leq p_0$ and $\lambda = \pm$integer if $p_0 =$ integer, or $\lambda = \pm$half-integer if $p_0 =$ half-integer. We see also that, starting with Eqs. (I4.1) and (I4.2), Eqs. (1.3) and (1.4) are also valid for $p_0 = 0$.

For p–λ diagrams of Figs. 1(a), 2(a), 3(a), and 4, the solutions of Eqs. (I4.10) and (I4.11) are particularly simple. All the $a_1^2(p, q, \lambda)$ and $b_1^2(p, q, \lambda)$ for $p \geq p_0$ are uniquely determined by Eqs. (1.3) and (1.4),

[1] T. Yao, J. Math. Phys. **8**, 1931 (1967). Hereafter referred to as Paper I.

[2] Hereafter, the equations and theorems of Paper I are prefixed by a Roman numeral I.

and we have

$$a_1^2(p,q,\lambda) = \frac{(p-p_0+1)(p+p_0+2)[(q-A)^2-(p+\lambda+2)^2]}{4(p-q+1)(p-q+2)(p+q+1)(p+q+2)}, \tag{1.5}$$

$$b_1^2(p,q,\lambda) = \frac{(p-p_0+1)(p+p_0+2)[(q+A)^2-(p-\lambda+2)^2]}{4(p-q+1)(p-q+2)(p+q+1)(p+q+2)}. \tag{1.6}$$

Equations (1.5) and (1.6) were first obtained by Murai.[3] They are valid *only* for a subclass of degenerate representations, namely, those whose p–λ diagrams are as in Figs. 1(a), 2(a), 3(a), and 4. We know, for instance, that for degenerate representations of the type of Figs. 1(b) and 1(d), $a_1^2(p,q,\lambda)$ and $b_1^2(p,q,\lambda)$ in general do not have the form (1.5) and (1.6), since the left and right boundaries of these representations give additional conditions, i.e.,

$$a_i^2(p,q,\lambda) = 0, \quad i = 2,3,4,$$

on the right boundary, and $b_i^2(p,q,\lambda) = 0, i = 2,3,4$, on the left boundary.

However, we observe (or through a direct lengthy calculation) that if we replace p_0 by B, where B is a complex number, in Eqs. (1.5) and (1.6), then, together with Eqs. (I4.6)–(I4.9), we may write

$$a_1^2(p,q,\lambda) = \frac{(p+1-B)(p+2+B)[(q-A)^2-(p+\lambda+2)^2]}{4(p-q+1)(p-q+2)(p+q+1)(p+q+2)}, \tag{1.7a}$$

$$a_2^2(p,q,\lambda) = \frac{(B-q)(B+q+1)[(q+\lambda+1)^2-(p+1-A)^2]}{4(p-q)(p-q+1)(p+q+1)(p+q+2)}, \tag{1.7b}$$

$$a_3^2(p,q,\lambda) = \frac{(B+q)(B-q+1)[(q-\lambda-1)^2-(p+1+A)^2]}{4(p+q)(p+q+1)(p-q+1)(p-q+2)}, \tag{1.7c}$$

$$a_4^2(p,q,\lambda) = \frac{(p-B)(p+B+1)[(q+A)^2-(p-\lambda)^2]}{4(p-q)(p-q+1)(p+q)(p+q+1)}, \tag{1.7d}$$

$$b_1^2(p,q,\lambda) = \frac{(p+1-B)(p+2+B)[(q+A)^2-(p-\lambda+2)^2]}{4(p-q+1)(p-q+2)(p+q+1)(p+q+2)}, \tag{1.8a}$$

$$b_2^2(p,q,\lambda) = \frac{(B-q)(B+q+1)[(q-\lambda+1)^2-(p+1+A)^2]}{4(p+q+1)(p+q+2)(p-q)(p-q+1)}, \tag{1.8b}$$

$$b_3^2(p,q,\lambda) = \frac{(B+q)(B-q+1)[(q+\lambda-1)^2-(p+1-A)^2]}{4(p+q)(p+q+1)(p-q+1)(p-q+2)}, \tag{1.8c}$$

$$b_4^2(p,q,\lambda) = \frac{(p-B)(p+B+1)[(q-A)^2-(p+\lambda)^2]}{4(p-q)(p-q+1)(p+q)(p+q+1)}. \tag{1.8d}$$

Substituting Eqs. (1.7) and (1.8) into Eqs. (I4.1)–(I4.3), we find the three Casimir operators (which are independent of p, q, and λ)

$$C_2 = 2(B-1)(B+2) + A^2, \tag{1.9}$$

$$C_3 = AB(B+1), \tag{1.10}$$

$$C_4 = \tfrac{1}{4}A^4 - A^2(B^2+B+1). \tag{1.11}$$

It is interesting to note that C_2, C_3, and C_4 depend on only two parameters A and B, a fact reflecting the degenerate nature of the representations considered in this paper.

For completeness we also give the eigenvalues of F_3, although for degenerate representations they are really redundant for a complete labeling of states. From Eq. (I4.5) we find

$$\alpha = Aq(p+1) - \lambda B(B+1) + \tfrac{1}{2}\lambda[(p+1)^2+q^2-1], \tag{1.12}$$

where α is the eigenvalue of F_3 for the state $|p,q,\lambda)$.

Finally, in conclusion to this section, it can be shown explicitly that Eqs. (1.7) and (1.8) satisfy the 27 relations in Appendix A of Paper I.

2. DEGENERATE FINITE IRREDUCIBLE REPRESENTATIONS

We turn our attention first to the degenerate finite irreducible representations. Comparing Eqs. (1.9)–(1.11) with Eqs. (I5.2)–(I5.4), we see that there are two solutions:

(a) $k_m = 0, \quad B = p_0 = j_m, \quad A = -(\lambda_m + 2),$

or

$$j_m = 0, \quad B = p_0 = k_m, \quad A = \lambda_m + 2.$$

To be specific, we consider the case $k_m = 0$ (the case $j_m = 0$ is similar). Then, on the right boundary $p + \lambda = p_0 + \lambda_m = j_m + \lambda_m$, $q = j_m$, Eqs. (1.7a)–(1.7c) give $a_i^2(p,q,\lambda) = 0$, $i = 1, 2, 3$, and Eq. (1.8a)

[3] Y. Murai, Progr. Theoret. Phys. (Kyoto) 9, 147 (1953).

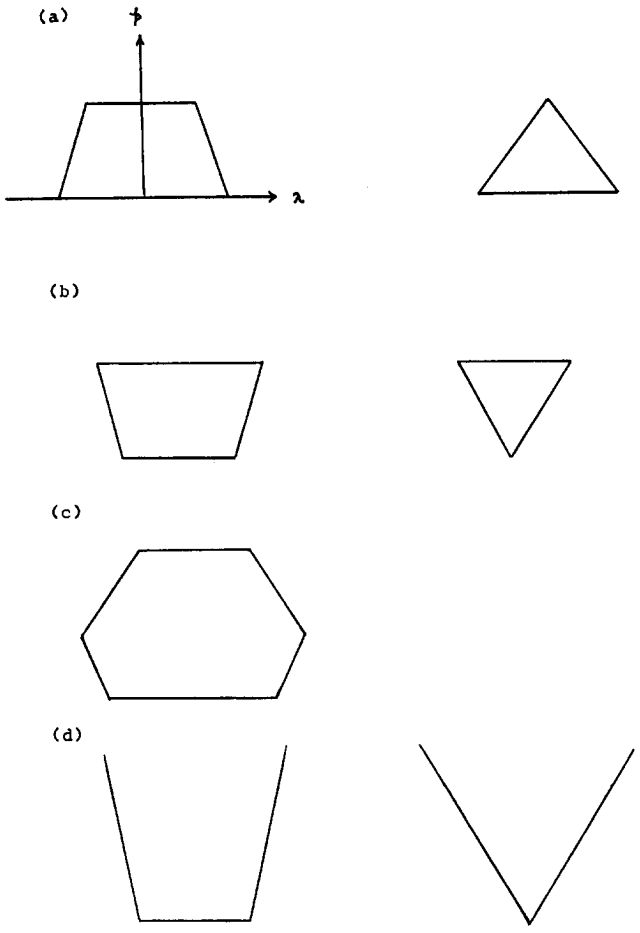


FIG. 1. λ is bounded on both sides.

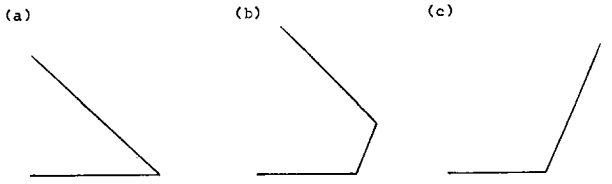


FIG. 2. λ is bounded on the right.

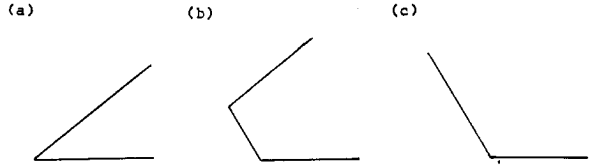


FIG. 3. λ is bounded on the left.

FIG. 4. λ is bounded: $-\infty < \lambda < +\infty$.

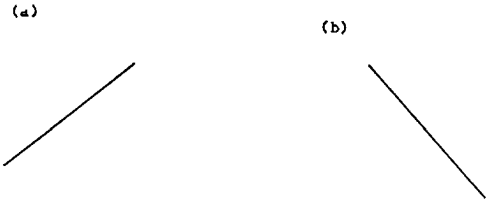


FIG. 5. λ is fixed. Figure 5(a): $p - \lambda = \text{const}$; Fig. 5(b): $p + \lambda = \text{const}$.

gives

$$b_1^2\,(p = p_0 + s_m, q, \lambda = \lambda_m - s_m) = b_1^2\,(p = \lambda_m, q = j_m, \lambda = j_m) = 0,$$

where $s_m = \lambda_m - j_m - k_m$. Similarly, on the left boundary $p - \lambda = p_0 - \lambda_m = j_m - \lambda_m$, $q = -j_m$, Eqs. (1.8a)–(1.8c) give $b_i^2(p, q, \lambda) = 0$, $i = 1, 2, 3$, and Eq. (1.7a) gives

$$a_1^2\,(p = p_0 + s_m, q, \lambda = -\lambda_m + s_m) = a_1^2\,(p = \lambda_m, q = -j_m, = -j_m) = 0.$$

The allowed values of p are $p = p_0$, $p_0 + 1$, $p_0 + 2, \cdots, p_0 + s_m = \lambda_m$. The p–λ diagram of this case is given by Fig. 1(a).

(b) $\lambda_m = j_m + k_m$ $\quad (s_m = \lambda_m - j_m - k_m = 0)$,
$B = j_m + k_m + 1$, $\quad A = -(j_m - k_m)$.

Here on the right boundary $p - \lambda = p_m - \lambda_m = j_m + k_m - \lambda_m = 0$, $q = j_m - k_m$, Eqs. (1.7b)–(1.7d) give $a_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$, and Eq. (1.7a) gives $a_1^2\,(p = j_m + k_m, q, \lambda = \lambda_m) = 0$. Similarly, on the left boundary, $p + \lambda = p_m - \lambda_m = 0$, $q = -(j_m - k_m)$, Eqs. (1.8b)–(1.8d) give $b_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$, and Eq. (1.8a) gives

$$b_1^2\,(p = j_m + k_m, q, \lambda = -\lambda_m) = 0.$$

The allowed values of p are

$$p = |j_m - k_m|, |j_m - k_m| + 1, \cdots, j_m + k_m.$$

The p–λ diagram of this case is given by Fig. 1(b).

The Young diagrams associated with the finite degenerate representations have already been given in Theorem I1. The problem of decomposition of a finite degenerate representation of $SU(2, 2)$ into irreducible representations of

$$SU(2) \times SU(2) \times U(1)$$

was discussed in Theorem I2.

3. DEGENERATE UNITARY IRREDUCIBLE REPRESENTATIONS

We now study degenerate unitary irreducible representations. Equations (1.7) and (1.8) must

satisfy the unitary conditions Eqs. (I6.4) and (I6.5):

$$a_1^2(p, q, \lambda),\quad a_4^2(p, q, \lambda),\quad b_1^2(p, q, \lambda),\quad b_4^2(p, q, \lambda) \leq 0, \tag{3.1}$$

$$a_2^2(p, q, \lambda),\quad a_3^2(p, q, \lambda),\quad b_2^2(p, q, \lambda),\quad b_3^2(p, q, \lambda) \geq 0. \tag{3.2}$$

It is most convenient to proceed systematically following the p–λ diagrams listed in Sec. 1, since otherwise it is rather easy to miss some of the representations.

(1) Here we consider those unitary representations whose p–λ diagrams are of the form given in Fig. 1(d).

On the right boundary $p - \lambda = p_0 - \lambda_0$, $q = q_0$, $a_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$. Equations (1.7b)–(1.7d) give us three relations:

$$(q_0 + \lambda_0 + s + 1)^2 = (p_0 + s + 1 - A)^2, \tag{3.3}$$

$$(q_0 - \lambda_0 - s - 1)^2 = (p_0 + s + 1 + A)^2, \tag{3.4}$$

$$(q_0 + A)^2 = (p_0 - \lambda_0)^2, \tag{3.5}$$

where $p = p_0 + s$, $\lambda = \lambda_0 + s$, from which we get

$$\lambda_0 = p_0, \tag{3.6}$$

$$q_0 = -A. \tag{3.7}$$

By considering Eqs. (1.7b) and (1.7c) for fixed q and λ but increasing p, we see that the unitary conditions (I6.5) require

$$(B - q)(B + q + 1) \leq 0, \tag{3.8}$$

$$(B + q)(B - q + 1) \leq 0, \tag{3.9}$$

from which we obtain

$$(B + \tfrac{1}{2})^2 \leq (q \pm \tfrac{1}{2})^2. \tag{3.10}$$

There are now four cases to be considered:

(i) $B = -\frac{1}{2} \pm i\rho$, where ρ is real, $\rho > 0$. Here p_0 = integer or half-integer.

(ii) $B = \sigma$, $-1 < \sigma < 0$, and p_0 = integer only.

(iii) $B = -\frac{1}{2}$, p_0 = half-integer [p_0 = integer has already been included in (ii)].

(iv) $B = 0$, or $B = -1$, and p_0 = integer.

Following the terminology used in the homogeneous Lorentz group, we call case (i) the principal series, and case (ii) the complementary series. Cases (iii) and (iv) are in the most degenerate discrete series.

Referring back to Eqs. (1.7d) and (1.8d), which gives us

$$a_4^2(p_0, q, \lambda) = b_4^2(p_0, q, \lambda) = 0,$$

$$(q + A)^2 = (p_0 - \lambda)^2, \tag{3.11}$$

$$(q - A)^2 = (p_0 + \lambda)^2. \tag{3.12}$$

There are two solutions:

$$A = p_0, \qquad q = -\lambda, \tag{3.13a}$$

or

$$A = -p_0, \quad q = \lambda. \tag{3.13b}$$

Now we are ready to list all the degenerate unitary irreducible representations whose p–λ diagrams are of the form given in Fig. 1(d).

(i) For the principal series we get

$$A^2 = p_0^2, \quad p_0 = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots,$$
$$B = -\tfrac{1}{2} \pm i\rho, \quad \rho > 0.$$

(a)
$$A = p_0,$$
$$C_2 = p_0^2 - 2\rho^2 - \tfrac{9}{2},$$
$$C_3 = -p_0(\rho^2 + \tfrac{1}{4}),$$
$$C_4 = \tfrac{1}{4}p_0^4 + p_0^2(\rho^2 - \tfrac{3}{4}),$$

$$\begin{matrix} p = p_0 \\ -2p_0 \leq q - \lambda \leq 2p_0 \\ q + \lambda = 0 \end{matrix}\Bigg\} \quad \begin{matrix} p = p_0 + 1 \\ -2p_0 - 1 \leq q - \lambda \leq 2p_0 + 1 \\ -1 \leq q + \lambda \leq 1 \end{matrix}\Bigg\} \cdots \Bigg\} \quad \begin{matrix} p = p_0 + s \\ -2p_0 - s \leq q - \lambda \leq 2p_0 + s \\ -s \leq q + \lambda \leq s \end{matrix}\Bigg\} \cdots\Bigg\}, \tag{3.14}$$

where $s = 0, 1, 2, 3, \cdots$; $-2p_0 - s \leq q - \lambda \leq 2p_0 + s$ means $q - \lambda = -2p_0 - s, -2p_0 - s + 2, -2p_0 - s + 4, \cdots, 2p_0 + s$; $-s \leq q + \lambda \leq s$ means $q + \lambda = -s, -s + 2, \cdots, s$. Of course, $-p \leq q \leq p$ always.

(b)
$$A = -p_0,$$
$$C_2 = p_0^2 - 2\rho^2 - \tfrac{9}{2},$$
$$C_3 = p_0(\rho^2 + \tfrac{1}{4}),$$
$$C_4 = \tfrac{1}{4}p_0^4 + p_0^2(\rho^2 - \tfrac{3}{4}),$$

$$\begin{matrix} p = p_0 \\ -2p_0 \leq q + \lambda \leq 2p_0 \\ q - \lambda = 0 \end{matrix}\Bigg\} \cdots\Bigg\} \quad \begin{matrix} p = p_0 + s \\ -2p_0 - s \leq q + \lambda \leq 2p_0 + s \\ -s \leq q - \lambda \leq s \end{matrix}\Bigg\} \cdots\Bigg\}. \tag{3.15}$$

(ii) For the complementary series we get

$$A^2 = p_0^2, \quad p_0 = 0, 1, 2, 3, \cdots,$$
$$B = \sigma, \quad -1 < \sigma < 0.$$

(a)
$$A = p_0,$$
$$C_2 = p_0^2 + 2(\sigma - 1)(\sigma + 2),$$
$$C_3 = p_0\sigma(\sigma + 1),$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2(\sigma^2 + \sigma + 1).$$

The decomposition of this unitary representation into irreducible representations of $SU(2) \times SU(2) \times U(1)$ is the same as in (3.14), the only difference is that here p_0 = integer.

(b)
$$A = -p_0,$$
$$C_2 = p_0^2 + 2(\sigma - 1)(\sigma + 2),$$
$$C_3 = -p_0\sigma(\sigma + 1),$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2(\sigma^2 + \sigma + 1).$$

The decomposition is the same as in (3.15).

(iii) For the most degenerate discrete series (p = half-integer) we get

$$A^2 = p_0^2, \quad p_0 = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots,$$
$$B = -\tfrac{1}{2}.$$

There are four cases:

(a)
$$A = p_0, \quad p_0 = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots,$$
$$C_2 = p_0^2 - \tfrac{9}{2},$$
$$C_3 = -\tfrac{1}{4}p_0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - \tfrac{3}{4}p_0^2.$$

The three Casimir operators depend on one discrete parameter p_0 only, and the unitary representations under study are said to belong to the most degenerate discrete series.

We already know that on the right boundary $p - \lambda = p_0 - \lambda_0 = 0$, $q = q_0 = -A = -p_0$, $a_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$. Similarly, on the left boundary $p + \lambda = p_0 + \tfrac{1}{2}$, $q = -\tfrac{1}{2}$, and Eqs. (1.8b)–(1.8d) give $b_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$. Therefore, q is always bounded from above by $-\tfrac{1}{2}$, $q = -p, -p + 1, -p + 2, \cdots, -\tfrac{1}{2}$, where $p = p_0 + s$ is a half-integer.

$$\left.\begin{array}{l} p = p_0 \\ q = -\gamma \\ \lambda = \gamma \end{array}\right\} \left.\begin{array}{cc} & p = p_0 + 1 \\ q = -\gamma & , \quad q = -(p_0 + 1) \\ \gamma - 1 \le \lambda \le \gamma + 1 & \lambda = p_0 \end{array}\right\} \cdots \Bigg\}$$

$$\left.\begin{array}{cc} & p = p_0 + s \\ q = -\gamma & , \quad q = -(p_0 + t) \\ -s + \gamma \le \lambda \le s + \gamma & -s + p_0 + t \le \lambda \le s + p_0 - t \end{array}\right\} \cdots \Bigg\}, \tag{3.16}$$

where $\gamma = \tfrac{1}{2}, \tfrac{3}{2}, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, 3, \cdots$.

(b)
$$A = p_0, \quad p_0 = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots.$$

This case is the mirror image of (a); we merely change $q \to -q$, $\lambda \to -\lambda$. On the left boundary $p + \lambda = p_0 - \lambda_0 = 0$, $q = -q_0 = p_0$, $b_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$. On the right boundary $p - \lambda = p_0 + \tfrac{1}{2}$, $q = \tfrac{1}{2}$, and $a_i^2(p, q, \lambda) = 0$, $i = 2, 3, 4$. Here q is always bounded from below by $\tfrac{1}{2}$, $q = \tfrac{1}{2}, \tfrac{3}{2}, \cdots, p$:

$$\left.\begin{array}{l} p = p_0 \\ q = \gamma \\ \lambda = -\gamma \end{array}\right\} \left.\begin{array}{cc} & p = p_0 + 1 \\ q = \gamma & , \quad q = p_0 + 1 \\ -\gamma - 1 \le \lambda \le -\gamma + 1 & \lambda = -p_0 \end{array}\right\} \cdots \Bigg\}$$

$$\left.\begin{array}{cc} & p = p_0 + s \\ q = \gamma & , \quad q = p_0 + t \\ -s - \gamma \le \lambda \le s - \gamma & -s - p_0 + t \le \lambda \le s - p_0 - t \end{array}\right\} \cdots \Bigg\}, \tag{3.17}$$

where $\gamma = \tfrac{1}{2}, \tfrac{3}{2}, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, 3, \cdots$.

(c)
$$A = -p_0, \quad p_0 = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots,$$
$$C_2 = p_0^2 - \tfrac{9}{2},$$
$$C_3 = \tfrac{1}{4}p_0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - \tfrac{3}{4}p_0^2.$$

This is similar to (a) with $q \to -q$.

$$\left.\begin{matrix} p = p_0 \\ q = \gamma \\ \lambda = \gamma \end{matrix}\right\} \cdots \left.\begin{matrix} q = \gamma \\ -s + \gamma \le \lambda \le s + \gamma \end{matrix}\right\} \cdots \left.\begin{matrix} p = p_0 + s \\ , \\ q = p_0 + t \\ -s + p_0 + t \le \lambda \le s + p_0 - t \end{matrix}\right\} \cdots \Bigg\}, \tag{3.18}$$

where $\gamma = \frac{1}{2}, \frac{3}{2}, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, 3, \cdots$.

(d)
$$A = -p_0, \quad p_0 = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots.$$

This is similar to (b2) with $q \to -q$.

$$\left.\begin{matrix} p = p_0 \\ q = -\gamma \\ \lambda = -\gamma \end{matrix}\right\} \cdots \left.\begin{matrix} q = -\gamma \\ -s - \gamma \le \lambda \le s - \gamma \end{matrix}\right\} \cdots \left.\begin{matrix} p = p_0 + s \\ , \\ q = -(p_0 + t) \\ -s - p_0 + t \le \lambda \le s - p_0 - t \end{matrix}\right\} \cdots \Bigg\}, \tag{3.19}$$

where $\gamma = \frac{1}{2}, \frac{3}{2}, \cdots, p_0$, $t = 1, 2, \cdots, s$. $s = 1, 2, 3, \cdots$.

We note that cases (a) and (b) have the same eigenvalues for the three Casimir operators. However, since the composition of each unitary irreducible representation is different, the two representations with the same C_2, C_3, and C_4 are unitarily inequivalent. Similarly, the representations in (c) are unitarily inequivalent to those in (d).

(iv) For most degenerate discrete series (p = integer) we get

$$A^2 = p_0^2, \quad p_0 = 0, 1, 2, 3, \cdots,$$
$$B = 0, \quad \text{or} \quad B = -1,$$
$$C_2 = p_0^2 - 4,$$
$$C_3 = 0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2.$$

There are six cases:

(a)
$$A = p_0 = 0.$$

From Eqs. (1.7b), (1.7c), (1.8b), and (1.8c), we see that q is always 0:

$$\left.\begin{matrix} p = p_0 = 0 \\ \lambda = 0 \end{matrix}\right\} \left.\begin{matrix} p = 1 \\ -1 \le \lambda \le 1 \end{matrix}\right\} \cdots \left.\begin{matrix} p = s \\ -s \le \lambda \le s \end{matrix}\right\} \cdots \Bigg\}, \tag{3.20}$$

where $s = 0, 1, 2, \cdots$; $-s \le \lambda \le s$ means $\lambda = -s, -s + 2, -s + 4, \cdots, s$.

(b)
$$A = p_0, \quad p_0 = 1, 2, 3, \cdots.$$

This case is similar to (iiia). The only difference is that here p is integral, and $q = -p, -p + 1, \cdots, -1$:

$$\left.\begin{matrix} p = p_0 \\ q = -\gamma \\ \lambda = \gamma \end{matrix}\right\} \cdots \left.\begin{matrix} q = -\gamma \\ -s + \gamma \le \lambda \le s + \gamma \end{matrix}\right\} \cdots \left.\begin{matrix} p = p_0 + s \\ , \\ q = -(p + t) \\ -s + p_0 + t \le \lambda \le s + p_0 - t \end{matrix}\right\} \cdots \Bigg\}, \tag{3.21}$$

where $\gamma = 1, 2, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, 3, \cdots$.

(c)
$$A = p_0, \quad p_0 = 1, 2, 3, \cdots.$$

This case is similar to (iiib). Here, $q = 1, 2, \cdots, p$:

$$\left.\begin{matrix} p = p_0 \\ q = \gamma \\ \lambda = -\gamma \end{matrix}\right\} \cdots \left\} \begin{matrix} p = p_0 + s \\ q = \gamma \\ -s - \gamma \le \lambda \le s - \gamma \end{matrix} \right. , \quad \left. \begin{matrix} p = p_0 + s \\ q = p_0 + t \\ -s - p_0 + t \le \lambda \le s - p_0 - t \end{matrix} \right\} \cdots \Bigg\}, \tag{3.22}$$

where $\gamma = 1, 2, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, 3, \cdots$.

(d) $$A = -p_0, \quad p_0 = 1, 2, 3, \cdots.$$

This case is similar to (ivb) with $q \to -q$:

$$\left.\begin{matrix} p = p_0 \\ q = \gamma \\ \lambda = \gamma \end{matrix}\right\} \cdots \left\} \begin{matrix} p = p_0 + s \\ q = \gamma \\ -s + \gamma \le \lambda \le s + \gamma \end{matrix} \right. , \quad \left. \begin{matrix} p = p_0 + s \\ q = p_0 + t \\ -s + p_0 + t \le \lambda \le s + p_0 - t \end{matrix} \right\} \cdots \Bigg\}, \tag{3.23}$$

where $\gamma = 1, 2, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, \cdots$.

(e) $$A = -p_0, \quad p_0 = 1, 2, 3, \cdots.$$

This case is similar to (ivc) with $q \to -q$:

$$\left.\begin{matrix} p = p_0 \\ q = -\gamma \\ \lambda = -\gamma \end{matrix}\right\} \cdots \left\} \begin{matrix} p = p_0 + s \\ q = -\gamma \\ -s - \gamma \le \lambda \le s - \gamma \end{matrix} \right. , \quad \left. \begin{matrix} p = p_0 + s \\ q = -(p_0 + t) \\ -s - p_0 + t \le \lambda \le s - p_0 - t \end{matrix} \right\} \cdots \Bigg\}, \tag{3.24}$$

where $\gamma = 1, 2, \cdots, p_0$, $t = 1, 2, \cdots, s$, $s = 1, 2, \cdots$.

(f) $$A = \pm p_0, \quad p_0 = 1, 2, 3, \cdots.$$

Cases (b) and (d) include $q \le -1$, while cases (c) and (e) include $q \ge 1$. Here in (f) $q = 0$ always. Equations (1.7b), (1.7c), (1.8b), and (1.8c) give $a_2^2(p, q = 0, \lambda) = a_3^2(p, q = 0, \lambda) = b_2^2(p, q = 0, \lambda) = b_3^2(p, q = 0, \lambda) = 0$ for all p and λ. On the right boundary $p - \lambda = p_0$, $a_4^2(p, q = 0, \lambda) = 0$, while on the left boundary $p + \lambda = p_0$, $b_4^2(p, q = 0, \lambda) = 0$:

$$\left.\begin{matrix} p = p_0 \\ \lambda = 0 \end{matrix}\right\} \left\{ \begin{matrix} p = p_0 + 1 \\ -1 \le \lambda \le 1 \end{matrix} \right. \cdots \left\} \begin{matrix} p = p_0 + s \\ -s \le \lambda \le s \end{matrix} \right. \cdots \Bigg\}, \tag{3.25}$$

where $s = 0, 1, 2, \cdots$; $-s \le \lambda \le s$ means $\lambda = -s, -s\ 2, -s + 4, \cdots, s$.

Since only A^2 appears in Eqs. (1.7) and (1.8), $A = \pm p_0$ gives the same unitary irreducible representation.

Here we note that the irreducible representations of (ivb)–(ivf), all have the same eigenvalues for C_2, C_3, and C_4, but they are all unitarily inequivalent.

(2) Next we consider those unitary representations whose p–λ diagrams are of the form given in Fig. 2(a). These representations are in the D^- series.

From Eqs. (1.7d) and (1.8d) we note that, for $p = p_0$ but arbitrary q and λ,

$$a_4^2(p_0, q, \lambda) = b_4^2(p_0, q, \lambda) = 0$$

requires

$$B = p_0. \tag{3.26}$$

On the right boundary

$$p + \lambda = p_0 + \lambda_0, \quad q = q_0,$$
$$a_i^2(p, q, \lambda) = 0, \quad i = 1, 2, 3.$$

Equations (1.7a)–(1.7c) give us three relations:

$$(q_0 - A)^2 = (p_0 + \lambda_0 + 2)^2 \tag{3.27}$$

and either

$$(q_0 + \lambda_0 - s + 1)^2 = (p_0 + s + 1 - A)^2, \tag{3.28a}$$
$$B = -q_0 = p_0, \tag{3.29a}$$

or

$$B = q_0 = p_0, \tag{3.28b}$$
$$(q_0 - \lambda_0 + s - 1)^2 = (p_0 + s + 1 + A)^2. \tag{3.29b}$$

The solution to Eqs. (3.27), (3.28a), and (3.29a) is

$$A = \lambda_0 + 2, \tag{3.30a}$$
$$q_0 = -p_0, \tag{3.31a}$$

and the solution to Eqs. (3.27), (3.28b), and (3.29b) is

$$A = -(\lambda_0 + 2), \tag{3.30b}$$
$$q_0 = p_0. \tag{3.31b}$$

From Eqs. (1.7b) and (1.7c), we have

$$a_2^2(p, q = p_0, \lambda) = a_3^2(p, q = -p_0, \lambda) = 0;$$

therefore, $-p_0 \leq q \leq p_0$ always.

Now we shall determine the allowed values for λ_0 such that Eqs. (1.7) and (1.8) give us unitary irreducible representations. The unitary conditions Eqs. (I6.4) and (I6.5) give eight inequalities (we need to consider $A = \lambda_0 + 2$ only):

$$(q - \lambda_0 + p + \lambda)(q - \lambda_0 - p - \lambda - 4) \leq 0, \tag{3.32a}$$

$$(q + \lambda + p - \lambda_0)(q + \lambda - p + \lambda_0 + 2) \geq 0, \tag{3.32b}$$

$$(q - \lambda + p + \lambda_0 + 2)(q - \lambda - p - \lambda_0 - 4) \geq 0, \tag{3.32c}$$

$$(q + \lambda_0 + p - \lambda + 2)(q + \lambda_0 - p + \lambda + 2) \leq 0, \tag{3.32d}$$

$$(q + \lambda_0 + p - \lambda + 4)(q + \lambda_0 - p + \lambda) \leq 0, \tag{3.33a}$$

$$(q - \lambda + p + \lambda_0 + 4)(q - \lambda - p - \lambda_0 - 2) \geq 0, \tag{3.33b}$$

$$(q + \lambda + p - \lambda_0 - 2)(q + \lambda - p + \lambda_0) \geq 0, \tag{3.33c}$$

$$(q - \lambda_0 + p + \lambda - 2)(q - \lambda_0 - p - \lambda - 2) \leq 0. \tag{3.33d}$$

Equations (3.32) and (3.33) have two solutions:

(i)

$$q - \lambda_0 + p + \lambda \leq 0, \tag{3.34a}$$

$$q - \lambda_0 - p - \lambda - 4 \geq 0, \tag{3.34b}$$

$$q + \lambda - p + \lambda_0 + 2 \leq 0, \tag{3.34c}$$

$$q - \lambda + p + \lambda_0 + 2 \geq 0, \tag{3.34d}$$

(ii)

$$q + \lambda_0 + p - \lambda + 4 \leq 0, \tag{3.35a}$$

$$q + \lambda_0 - p + \lambda \geq 0, \tag{3.35b}$$

$$q - \lambda - p - \lambda_0 - 2 \leq 0, \tag{3.35c}$$

$$q + \lambda + p - \lambda_0 - 2 \geq 0. \tag{3.35d}$$

Since λ is bounded on the right, $\lambda_0 - \lambda \geq 0$, solution (ii) is ruled out. Solution (i) is possible, and Eq. (3.34b) gives

$$\lambda_0 \leq -p_0 - 2. \tag{3.36}$$

When we compare Eq. (3.36) with the expression $\lambda_0 = \Lambda_m = -J_m - K_m - S_m - 4$ in Theorem I3, we see that ($p_0 = J_m + K_m$) the allowed values of S_m are now $S_m = -2, -1, 0, 1, 2, 3, \cdots$.

The decomposition of a unitary irreducible representation in the D^- series has already been given in Theorem I3, in which Case (A.1), with $J_m = 0$, $K_m = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, corresponds to Eqs. (3.30a)–(3.31a) with $p_0 = K_m$, $q_0 = -K_m$; Case (A.2), with $K_m = 0$, $J_m = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, corresponds to Eqs. (3.30b)–(3.31b) with $p_0 = J_m$, $q_0 = J_m$; and Case (A.3), with $J_m = K_m = 0$, corresponds to either (3.30a) or (3.30b) with $p_0 = 0$, since only A^2 appears in Eqs. (1.7)and (1.8). Since $S_m = -2, -1, 0, 1, 2, 3, \cdots$, we see that, for every finite representation $((j_m, k_m, \lambda_m))$, there exists a unitary representation $D^-(J_m, K_m, \Lambda_m)$, but the converse is not true. (There are no finite representations with $S_m = -2, -1$.) It is trivial to check that expression (I6.15) satisfies Eqs. (3.34).

The three Casimir operators are

$$C_2 = 2(p_0 - 1)(p_0 + 2) + (\lambda_0 + 2)^2, \tag{3.37}$$

$$C_3 = \pm(\lambda_0 + 2)p_0(p_0 + 1), \tag{3.38}$$

$$C_4 = \tfrac{1}{4}(\lambda_0 + 2)^4 - (\lambda_0 + 2)^2(p_0^2 + p_0 + 1), \tag{3.39}$$

where

$$p_0 = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots,$$

$$\lambda_0 = -p_0 - 2, -p_0 - 3, -p_0 - 4, \cdots,$$

which are the same as that given by Eqs. (I6.8)–(I6.10) for the degenerate representations.

In conclusion we remark on the p–λ diagrams of the form given in Fig. 2(c). From Eq. (1.7b) we see that for fixed λ and q but large p, $(q + \lambda + 1)^2 - (p + 1 - A)^2$ is negative, while for fixed p and q but large negative λ, $(q + \lambda + 1)^2 - (p + 1 - A)^2$ is positive. Hence, the unitary condition (I6.5) cannot always be satisfied, and there do not exist unitary degenerate representations whose p–λ diagrams are of the form given by Fig. 2(c).

(3) We do not have to consider the unitary representations whose p–λ diagrams are of the form given in Fig. 3(a) in detail, since they are mirror images of the representations in the D^- series. These representations are in the D^+ series and have been discussed in Theorem I4.

(4) Now we come to those unitary representations whose p–λ diagrams are such that λ is unbounded, $-\infty < \lambda < \infty$, for $p \geq p_0$.

From Eqs. (1.7d) and (1.8d),

$$a_4^2(p_0, q, \lambda) = b_4^2(p_0, q, \lambda) = 0$$

and $B = p_0$. But from Eq. (1.7b) we see that for fixed $(q + \lambda + 1)^2$ and increasing p, $a_2^2(p, q, \lambda)$ cannot remain positive. Therefore, $p_0 = 0$ and $q = 0$ always, and

$$a_2^2(p, q, \lambda) = a_3^2(p, q, \lambda) = b_2^2(p, q, \lambda) = b_3^2(p, q, \lambda) = 0. \tag{3.40}$$

There are two cases to be considered:

(i) The most degenerate principal continuous series.

Equations (1.7a), (1.7d), (1.8a), and (1.8d), together with the unitary condition (I6.4), require

$$A^2 < 0. \tag{3.41}$$

The limiting case $A^2 = 0$ has already been included in the D^- and D^+ series, since the boundaries $p + \lambda = -2$ [on which $a_1^2(p, 0, \lambda) = 0$] and $p - \lambda = -2$ [on which $b_1^2(p, 0, \lambda) = 0$] exist.

Let $A = i\rho$, with $\rho > 0$, the three Casimir operators are

$$C_2 = -4 - \rho^2, \tag{3.42}$$

$$C_3 = 0, \tag{3.43}$$

$$C_4 = \tfrac{1}{4}\rho^4 + \rho^2. \tag{3.44}$$

The three Casimir operators depend on one continuous parameter ρ only, and the unitary representations under study are said to belong to the most degenerate principal continuous series.

Again there are two cases.

(a) $p + \lambda = \pm$even integer.

$$\left.\begin{matrix} p = 0 \\ \lambda = 0, \pm 2, \pm 4, \cdots \end{matrix}\right\} \left.\begin{matrix} p = 1 \\ \lambda = \pm 1, \pm 3, \cdots \end{matrix}\right\} \cdots\Big\},$$

$$\left.\begin{matrix} p = 2s \\ \lambda = 0, \pm 2, \pm 4, \cdots \end{matrix}\right\} \left.\begin{matrix} p = 2s + 1 \\ \lambda = \pm 1, \pm 3, \cdots \end{matrix}\right\} \cdots\Big\}, \tag{3.45}$$

where $s = 0, 1, 2, \cdots$, and $q = 0$ always.

(b) $p + \lambda = \pm$odd integer.

$$\left.\begin{matrix} p = 0 \\ \lambda = \pm 1, \pm 3, \cdots \end{matrix}\right\} \left.\begin{matrix} p = 1 \\ \lambda = 0, \pm 2, \pm 4, \cdots \end{matrix}\right\} \cdots\Big\},$$

$$\left.\begin{matrix} p = 2s \\ \lambda = \pm 1, \pm 3, \cdots \end{matrix}\right\} \left.\begin{matrix} p = 2s + 1 \\ \lambda = 0, \pm 2, \pm 4, \cdots \end{matrix}\right\} \cdots\Big\}, \tag{3.46}$$

where $s = 0, 1, 2, \cdots$, and $q = 0$ always.

(ii) The most degenerate complementary continuous series.

This series occurs for $p + \lambda = \pm$odd integer only. From Eq. (1.7a) we see that

$$0 \leq A^2 < 1 \tag{3.47}$$

is also allowed.

Let $A = \sigma$, $0 \leq \sigma < 1$. The three Casimir operators are

$$C_2 = -4 + \sigma^2, \tag{3.48}$$

$$C_3 = 0, \tag{3.49}$$

$$C_4 = \tfrac{1}{4}\sigma^4 - \sigma^2. \tag{3.50}$$

$$\left.\begin{matrix} p = 0 \\ \lambda = \pm 1, \pm 3, \cdots \end{matrix}\right\} \left.\begin{matrix} p = 1 \\ \lambda = 0, \pm 2, \pm 4, \cdots \end{matrix}\right\} \cdots\Big\},$$

$$\left.\begin{matrix} p = 2s \\ \lambda = \pm 1, \pm 3, \cdots \end{matrix}\right\} \left.\begin{matrix} p = 2s + 1 \\ \lambda = 0, \pm 2, \pm 4, \cdots \end{matrix}\right\} \cdots\Big\}, \tag{3.51}$$

where $s = 0, 1, 2, \cdots$, and $q = 0$ always.

(5) Finally, we study the unitary representations whose p–λ diagrams have degenerated into straight lines. There are two cases to be considered:

(i) p–$\lambda = \gamma$ (an integer).

Here we have

$$a_2(p, q, \lambda) = a_3(p, q, \lambda) = a_4(p, q, \lambda) = b_1(p, q, \lambda) = b_2(p, q, \lambda) = b_3(p, q, \lambda) = 0. \tag{3.52}$$

There are again two cases:

(a) $q = p_0$.

From Eq. (1.7d), we get

$$p_0 + A = \pm\gamma, \tag{3.53}$$

which, when taken together with Eq. (1.8a),

$$(p_0 + A)^2 - (p - \lambda + 2)^2 = 0, \tag{3.54}$$

gives us

$$\gamma = -1$$

or

$$\lambda = p + 1. \tag{3.55}$$

Equations (1.7b) and (1.7c) give us two possible solutions: either

$$(B - p_0)(B + p_0 + 1) = 0, \tag{3.56a}$$

$$(p_0 - \lambda - 1)^2 - (p + 1 + A)^2 = 0, \tag{3.56b}$$

which result in

$$A = -p_0 + 1,$$
$$B = p_0, \quad \text{or} \quad B = -p_0 - 1, \tag{3.57}$$

or

$$(p_0 + \lambda + 1)^2 - (p + 1 - A)^2 = 0, \tag{3.58a}$$

$$(B + p_0)(B - p_0 + 1) = 0, \tag{3.58b}$$

which result in

$$A = -p_0 - 1,$$
$$B = -p_0 \quad \text{or} \quad B = p_0 - 1. \tag{3.59}$$

Equations (3.57) and (3.59) give the same expression for the three Casimir operators,

$$C_2 = 3(p_0^2 - 1), \tag{3.60a}$$

$$C_3 = -p_0(p_0^2 - 1), \tag{3.60b}$$

$$C_4 = -\tfrac{3}{4}(p_0^2 - 1)^2, \tag{3.60c}$$

where

$$p_0 = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots.$$

Equations (1.7a) and (1.8a) now become very simple.

$$a_1^2(p, q = p_0, \lambda = p + 1) = -1 \quad (3.61)$$

for $p = p_0, p_0 + 1, p_0 + 2, \cdots$, and

$$b_4^2(p, q = p_0, \lambda = p + 1) = -1 \quad (3.62)$$

for $p = p_0 + 1, p_0 + 2, \cdots$. Therefore, they are constants and independent of p_0!

Similarly, for

(b) $$q = -p_0$$

we simply replace p_0 by $-p_0$ in Eqs. (3.53)–(3.59), and the three Casimir operators are

$$C_2 = 3(p_0^2 - 1), \quad (3.63a)$$
$$C_3 = p_0(p_0^2 - 1), \quad (3.63b)$$
$$C_4 = -\tfrac{3}{4}(p_0^2 - 1)^2. \quad (3.63c)$$

Again, we have

$$a_1^2(p, q = -p_0, \lambda = p + 1) = -1 \quad (3.64)$$

for $p = p_0, p_0 + 1, p_0 + 2, \cdots$, and

$$b_4^2(p, q = -p_0, \lambda = p + 1) = -1 \quad (3.65)$$

for $p = p_0 + 1, p_0 + 2, \cdots$.

Next, we study

(ii) $$p + \lambda = \gamma \text{ (an integer).}$$

Here we have

$$a_1(p, q, \lambda) = a_2(p, q, \lambda) = a_3(p, q, \lambda) = b_2(p, q, \lambda) = b_3(p, q, \lambda) = b_4(p, q, \lambda) = 0. \quad (3.66)$$

There are again two cases:

(a) $$q = p_0.$$

From Eq. (1.8d), we get

$$p_0 - A = \pm\gamma \quad (3.67)$$

which, when taken together with Eq. (1.7a),

$$(p_0 - A)^2 - (p + \lambda + 2)^2 = 0, \quad (3.68)$$

gives us again

$$\gamma = -1$$

or

$$\lambda = -p - 1. \quad (3.69)$$

This is just the "mirror image" of Eq. (3.55), as expected. Equations (1.7b) and (1.7c) give us either

$$(B - p_0)(B + p_0 + 1) = 0, \quad (3.70a)$$
$$(p_0 - \lambda - 1)^2 - (p + 1 + A)^2 = 0, \quad (3.70b)$$

which result in

$$A = p_0 - 1,$$
$$B = p_0, \quad \text{or} \quad B = -p_0 - 1, \quad (3.71)$$

or

$$(p_0 + \lambda + 1)^2 - (p + 1 - A)^2 = 0, \quad (3.72a)$$
$$(B + p_0)(B - p_0 + 1) = 0, \quad (3.72b)$$

which result in

$$A = p_0 + 1,$$
$$B = -p_0, \quad \text{or} \quad B = p_0 - 1. \quad (3.73)$$

The three Casimir operators are

$$C_2 = 3(p_0^2 - 1), \quad (3.74a)$$
$$C_3 = p_0(p_0^2 - 1), \quad (3.74b)$$
$$C_4 = -\tfrac{3}{4}(p_0^2 - 1)^2. \quad (3.74c)$$

Here we have

$$b_1^2(p, q = p_0, \lambda = -p - 1) = -1 \quad (3.75)$$

for $p = p_0, p_0 + 1, p_0 + 2, \cdots$, and

$$a_4^2(p, q = p_0, \lambda = -p - 1) = -1 \quad (3.76)$$

for $p = p_0 + 1, p_0 + 2, \cdots$.

Similarly, for

(b) $$q = p_0$$

we simply replace p_0 by $-p_0$ in Eqs. (3.67)–(3.73), and the three Casimir operators are

$$C_2 = 3(p_0^2 - 1), \quad (3.77a)$$
$$C_3 = -p_0(p_0^2 - 1), \quad (3.77b)$$
$$C_4 = -\tfrac{3}{4}(p_0^2 - 1)^2. \quad (3.77c)$$

Again, we have

$$b_1^2(p, q = -p_0, \lambda = -p - 1) = -1 \quad (3.78)$$

for $p = p_0, p_0 + 1, p_0 + 2, \cdots$, and

$$a_4^2(p, q = -p_0, \lambda = -p - 1) = -1 \quad (3.79)$$

for $p = p_0 + 1, p_0 + 2, \cdots$.

Since the functions $a_i^2(p, q, \lambda)$, and $b_i^2(p, q, \lambda)$ are either 0 or -1, for $i = 1, 2, 3, 4$, we say these unitary irreducible representations belong to the exceptional degenerate discrete series, the $E^{\pm}$ series.[4]

4. SUMMARY OF DEGENERATE UNITARY IRREDUCIBLE REPRESENTATIONS

I. The Most Degenerate Principal Continuous Series

$$C_2 = -4 - \rho^2,$$
$$C_3 = 0,$$
$$C_4 = \tfrac{1}{4}\rho^4 + \rho^2,$$

where $\rho > 0$, and $A = i\rho$, $B = 0$. $q = 0$ always.

[4] The exceptional degenerate discrete series (the E^+ series) have recently been used in the study of infinite-component wave equations [Y. Nambu, Phys. Rev. **160**, 1171 (1967)].

(a) $p + \lambda = \pm$even integer [see Eq. (3.45)],
(b) $p + \lambda = \pm$odd integer [see Eq. (3.46)],
with $p = 0, 1, 2, \cdots$.

II. The Most Degenerate Complementary Continuous Series

$$C_2 = -4 + \sigma^2,$$
$$C_3 = 0,$$
$$C_4 = \tfrac{1}{4}\sigma^4 - \sigma^2,$$

where $0 \leq \sigma < 1$, and $A = \sigma$, $B = 0$, $q = 0$ always. $p + \lambda = \pm$odd integer with $p = 0, 1, 2, \cdots$ [see Eq. (3.51)].

III. The Discrete Series (the D^- Series)

$$C_2 = 2(p_0 - 1)(p_0 + 2) + (\lambda_0 + 2)^2,$$
$$C_3 = \pm(\lambda_0 + 2)p_0(p_0 + 1),$$
$$C_4 = \tfrac{1}{4}(\lambda_0 + 2)^4 - (\lambda_0 + 2)^2(p_0^2 + p_0 + 1),$$

where

$$p_0 = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots,$$
$$\lambda_0 = -p_0 - 2, \quad -p_0 - 3, \quad -p_0 - 4, \cdots,$$

$A = \pm(\lambda_0 + 2)$, $B = p_0$ [see Theorem I3, Eq. (I6.15)].

IV. The Discrete Series (the D^+ Series)

$$C_2 = 2(p_0 - 1)(p_0 + 2) + (\lambda_0 + 2)^2,$$
$$C_3 = \pm(\lambda_0 + 2)p_0(p_0 + 1),$$
$$C_4 = \tfrac{1}{4}(\lambda_0 + 2)^4 - (\lambda_0 + 2)^2(p_0^2 + p_0 + 1),$$

where

$$p_0 = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots,$$
$$\lambda_0 = p_0 + 2, \quad p_0 + 3, \quad p_0 + 4, \cdots,$$
$$A = \pm(\lambda_0 + 2), \quad B = p_0 \text{ [see Theorem I4]}.$$

V. The Most Degenerate Discrete Series (p = Half-Integer)

$$C_2 = p_0^2 - \tfrac{9}{2},$$
$$C_3 = \pm\tfrac{1}{4}p_0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - \tfrac{3}{4}p_0^2,$$

where $p_0 = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$. $p - \lambda \geq 0$, and $p + \lambda \geq p_0 + \frac{1}{2}$. $A = \mp p_0$, $B = -\frac{1}{2}$ [See Eqs. (3.16) and (3.18)].

VI. The Most Degenerate Discrete Series (p = Half-Integer)

$$C_2 = p_0^2 - \tfrac{9}{2},$$
$$C_3 = \pm\tfrac{1}{4}p_0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - \tfrac{3}{4}p_0^2,$$

where $p_0 = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$. $p + \lambda \geq 0$, and $p - \lambda \geq p_0 + \frac{1}{2}$. $A = \mp p_0$, $B = -\frac{1}{2}$ [see Eqs. (3.17) and (3.19)].

VII. The Most Degenerate Discrete Series (p = Integer)

$$C_2 = p_0^2 - 4,$$
$$C_3 = 0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2,$$

where $p_0 = 1, 2, 3, \cdots$. $p - \lambda \geq 0$, and $p + \lambda \geq p_0 + 1$. $A = \pm p_0$, $B = 0$, or $B = -1$ [see Eqs. (3.21) and (3.23)].

VIII. The Most Degenerate Discrete Series (p = Integer)

$$C_2 = p_0^2 - 4,$$
$$C_3 = 0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2,$$

where $p_0 = 1, 2, 3, \cdots$. $p + \lambda \geq 0$, and $p - \lambda \geq p_0 + 1$. $A = \pm p_0$, $B = 0$, or $B = -1$ [see Eqs. (3.22) and (3.24)].

IX. The Most Degenerate Discrete Series (p = Integer)

$$C_2 = p_0^2 - 4,$$
$$C_3 = 0,$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2,$$

where $p_0 = 1, 2, 3, \cdots$, $q = 0$ always. $p - \lambda \geq p_0$, and $p + \lambda \geq p_0$. $A = \pm p_0$, $B = 0$, or $B = -1$ [see Eq. (3.25)].

X. The Most Degenerate Discrete Representation ($p_0 = 0$)

This is an isolated representation; it could have been included in Class IX:

$$C_2 = -4,$$
$$C_3 = 0,$$
$$C_4 = 0,$$

where $A = 0$, $B = 0$, or $B = -1$ [see Eq. (3.20)].

XI. The Exceptional Degenerate Discrete Series (the E^+ Series)

$$C_2 = 3(p_0^2 - 1),$$
$$C_3 = \mp p_0(p_0^2 - 1),$$
$$C_4 = -\tfrac{3}{4}(p_0^2 - 1)^2,$$

where $p_0 = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots, q = \pm p_0$, and $\lambda = p + 1$ always [see Eqs. (3.61) and (3.64)].

XII. The Exceptional Degenerate Discrete Series (the E^- Series)

$$C_2 = 3(p_0^2 - 1),$$
$$C_3 = \pm p_0(p_0^2 - 1),$$
$$C_4 = -\tfrac{3}{4}(p_0^2 - 1)^2,$$

where $p_0 = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots, q = \pm p_0$, and $\lambda = -p - 1$ always [see Eqs. (3.75) and (3.78)].

XIII. The Principal Series

$$C_2 = p_0^2 - 2\rho^2 - \tfrac{9}{2},$$
$$C_3 = \pm p_0(\rho^2 + \tfrac{1}{4}),$$
$$C_4 = \tfrac{1}{4}p_0^4 + p_0^2(\rho^2 - \tfrac{3}{4}),$$

where $p_0 = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots, \rho > 0$. $p - \lambda \geq 0$ and $p + \lambda \geq 0$. $A = \mp p_0$, $B = -\frac{1}{2} \pm i\rho$ [see Eqs. (3.14) and (3.15)].

XIV. The Complementary Series

$$C_2 = p_0^2 + 2(\sigma - 1)(\sigma + 2),$$
$$C_3 = \pm p_0\sigma(\sigma + 1),$$
$$C_4 = \tfrac{1}{4}p_0^4 - p_0^2(\sigma^2 + \sigma + 1),$$

where $p_0 = 0, 1, 2, 3, \cdots, -1 < \sigma < 0$. $p - \lambda \geq 0$, and $p + \lambda \geq 0$. $A = \pm p_0$, $B = \sigma$ [see Eqs. (3.14) and (3.15)].

There are 14 classes of degenerate unitary irreducible representations. The first four classes are the exact analogs of the four classes found by Bargmann[5] in his study of $SU(1, 1)$. In both cases there are two continuous series. (1) an even series: for $SU(2, 2)$, $C_2 < -4$, $C_3 = 0$, $C_4 = \frac{1}{4}C_2(C_2 + 4)$, $p + \lambda = \pm$ even integer; for $SU(1, 1)$, C_q^0, $q > 0$, $m = \pm$integer. (2) an odd series: for $SU(2, 2)$, $C_2 < -3$, $C_3 = 0$, $C_4 = \frac{1}{4}C_2(C_2 + 4)$, $p + \lambda = \pm$odd integer; for $SU(1, 1)$, $C_q^{\frac{1}{2}}$, $q > \frac{1}{4}$, $m = \pm$half-integer. There are also two discrete series in both cases: the $D^{\pm}$ series. Here the similarity is due of course to Harish-Chandra's theorem connecting the finite representations with the unitary representations in the discrete series. Furthermore, in $SU(2, 2)$, we find that [following Eq. (3.36)] $S_m = -2, -1, 0, 1, 2, \cdots$. The values $S_m = -2, -1$ have no analog in the finite representations. In $SU(1, 1)$ the same situation occurs, $k = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, and the value $k = \frac{1}{2}$ has no analog in the finite representations.

Finally, we comment briefly on the unitary representations found by Murai.[3] His classes I and II are essentially the same as our classes I and II. However, there is a slight mistake in his class I. From Eqs. (3.45) and (3.46), together with Eqs. (1.7) and (1.8), we see that the representation with $p + \lambda = \pm$even integer and the representation with $p + \lambda = \pm$odd integer are disjoint; there are two irreducible representations. Murai's classes III and V together are exactly the same as our class IV, the D^+ series, while his classes III′ and V′ together are the same as our class III, the D^- series. His class IV is identical to our class X, the isolated most degenerate discrete representation. Our classes V–IX and XI–XIV are all absent in Murai's study for reasons already mentioned above.

[5] V. Bargmann, Ann. Math. **48**, 568 (1947).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Inverse Functions of the Products of Two Bessel Functions and Applications to Potential Scattering

H. A. MAVROMATIS AND K. SCHILCHER
Department of Physics, American University of Beirut, Beirut, Lebanon

(Received 26 December 1967)

Inverse functions of products of two Bessel functions $j_l(xy)j_m(xy)$ are determined for the cases $m = l$, $l + 1$, and $l + 2$. Integral representations for these inverse functions in terms of Neumann functions are given, and some of the simplest ones are expressed in terms of trigonometric functions.

We show how one may obtain an integral representation for any well-behaved function in terms of products of two Bessel functions, with the help of these inverse functions and also outline some of their applications to potential scattering. In particular, we demonstrate the usefulness of the inverse functions in determining the potential explicitly from the phase shifts in the Born approximation.

I. INTRODUCTION

In some recent nuclear calculations[1] a particular integral representation of $r^{2(l+p)}e^{-r^2}$ was used. The crucial characteristic of the integral representation was that it contained $j_l^2(kr)$ as part of the integrand. With the help of this representation, integrals of the nuclear two-body potential $V_l(r)$ weighted by the products of two harmonic oscillator wavefunctions $= \sum_p a_p r^{2(l+p)} e^{-r^2}$ were obtained directly from the Born approximation.

In other words, this representation was used to establish a simple connection between

$$\frac{\delta_l(k)}{k} \sim \int_0^\infty j_l^2(kr) V_l(r) r^2 \, dr$$

and

$$\sum_p a_p \int_0^\infty r^{2(l+p)} e^{-r^2} V_l(r) r^2 \, dr.$$

It is because the Born approximation gives integrals of $V_l(r)$ weighted by $j_l^2(kr)$ that this function appears in the integrand of the above-mentioned representation.

Expression (1) is the representation that was obtained in the case $p = 0$. A direct derivation of this result is given in Appendix A.

$$r^{2l}e^{-r^2} = \int_0^\infty j_l^2(kr) \left[-\frac{2\Gamma(l+\frac{1}{2})}{\pi} k^2 \times {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2} - l; -k^2) \right] dk. \quad (1)$$

In this expression $j_l^2(kr)$ is the usual spherical Bessel function[2] and ${}_1F_1(\frac{3}{2}; \frac{1}{2} - l; -k^2)$ is a generalized hypergeometric series.[3]

The function ${}_1F_1(\frac{3}{2}; \frac{1}{2} - l; -k^2)$ in expression (1) can be written with the help of Kummer's formula (see Appendix A) as $e^{-k^2} {}_1F_1(-l - 1; \frac{1}{2} - l; k^2)$ where ${}_1F_1(-l - 1; \frac{1}{2} - l; k^2)$ is an $(l + 2)$-term polynomial in k^2. The generalization of expression (1) to the case where $p \neq 0$ is carried out in Sec. II [see expression (14)]. Section II also concerns itself with the determination of the inverse function of $j_l^2(\rho)$ and some of its other mathematical and physical applications. In Secs. III and IV the inverse functions for $j_l(\rho)j_{l+1}(\rho)$ and $j_l(\rho)j_{l+2}(\rho)$ are calculated.

II. THE INVERSE FUNCTION OF $j_l^2(xy)$

One can write expression (1) as follows:

$$\psi_l(r) = \int_0^\infty j_l^2(kr) \phi_l(k) \, dk, \quad (2)$$

where $\psi_l(r) \equiv r^{2l}e^{-r^2}$ and

$$\phi_l(k) \equiv -\frac{2\Gamma(l+\frac{1}{2})}{\pi} k^2 {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2} - l; -k^2).$$

All three functions in Eq. (2) are even, $\psi_l(r) = \psi_l(-r)$, $\phi_l(k) = \phi_l(-k)$, and $j_l^2(kr) = j_l^2(-kr)$. When cast into this form, expression (1) is reminiscent of Fourier sine and cosine transforms, and of Fourier–Bessel (or Hankel) transforms.[4] What is necessary for an analogous transform in $j_l^2(kr)$ is the inverse of expression (2), namely,

$$\phi_l(k) = \int_0^\infty g_l(kr) \psi_l(r) \, dr, \quad (3)$$

where $g_l(kr)$ is the "inverse" function of $j_l^2(kr)$.

[1] J. P. Elliott, H. A. Mavromatis, and E. A. Sanderson, Phys. Letters **24B**, 358 (1967).

[2] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Co., Inc., New York, 1955), p. 77.

[3] I. N. Sneddon, *Special Functions of Mathematical Physics and Chemistry* (Oliver and Boyd, London, 1961), p. 38.

[4] G. Goertzel and N. Tralli, *Some Mathematical Methods of Physics* (McGraw-Hill Book Co., Inc., New York, 1960), pp. 266–279.

A general formula for obtaining $g_l(kr)$ after the appropriate identifications is[5]

$$\int_0^\infty {}_pF_q(\alpha_1, \alpha_2, \cdots, \alpha_p; \beta_1, \beta_2, \cdots, \beta_q; \pm(kr)^2) \times \frac{e^{-r^2/2b^2}}{(2b^2)^{\mu/2}} r^{\mu-1}\, dr$$
$$= \tfrac{1}{2}\Gamma\left(\frac{\mu}{2}\right) \times {}_{p+1}F_q\left(\alpha_1, \alpha_2, \cdots, \alpha_p, \frac{\mu}{2}; \beta_1, \beta_2, \cdots, \beta_q; \pm 2b^2k^2\right). \quad (4)$$

With the identifications $\mu = 2l + 3$, $p = 1$, $q = 2$, $2b^2 = 1$, $\alpha_1 = \frac{3}{2}$, $\beta_1 = \frac{1}{2} - l$, $\beta_2 = l + \frac{3}{2}$, expression (4) reduces to

$$-\frac{2}{\pi}\Gamma(l + \tfrac{1}{2})k^2\, {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2} - l; -k^2) = \int_0^\infty \left\{\frac{-4(kr)^2}{\pi(l + \frac{1}{2})}\, {}_1F_2(\tfrac{3}{2}; \tfrac{1}{2} - l, l + \tfrac{3}{2}; (kr)^2)\right\} \times [r^{2l}e^{-r^2}]\, dr. \quad (5)$$

Comparison of Eq. (5) with Eq. (2) shows that the required inverse function is

$$g_l(kr) = -\frac{4(kr)^2}{\pi(l + \frac{1}{2})}\, {}_1F_2(\tfrac{3}{2}; \tfrac{1}{2} - l, l + \tfrac{3}{2}; -(kr)^2). \quad (6)$$

As $j_l^2(kr)$, $g_l(kr)$ is also an even function, $g_l(kr) = g_l(-kr)$. Substituting expressions (2) and (3) into one another and rearranging the order of integration leads to

$$\psi_l(r) = \int_0^\infty \int_0^\infty j_l^2(kr) g_l(kr')\, dk \psi_l(r')\, dr',$$
$$\phi_l(k) = \int_0^\infty \int_0^\infty j_l^2(kr) g_l(k'r)\, dr \phi_l(k')\, dk'.$$

Hence

$$\int_0^\infty j_l^2(xy) g_l(xy')\, dx = \delta(y - y'), \quad x, y = r, k \text{ or } k, r. \quad (7)$$

Particular functions $\psi_l(r)$, $\phi_l(k)$ have been used to obtain the inverse function of $j_l^2(kr)$. These increase in complexity as l increases. An integral representation for $g_l(\rho)$ is readily derived (for $l \neq 0$) from the general expression[6]:

$$\int_0^1 x^{n-1}(1 - x)^{l-1} \times {}_pF_q(\alpha_1, \alpha_2, \cdots, \alpha_p; \beta_1, \beta_2, \cdots, \beta_q; -\rho^2 x)\, dx = B(n, l)\, {}_{p+1}F_{q+1}(\alpha_1, \alpha_2, \cdots, \alpha_p, n; \beta_1, \beta_2, \cdots, \beta_q, n + l; -\rho^2). \quad (8)$$

[5] Reference 3, p. 48.

[6] Reference 3, p. 47. A misprint appears in Sneddon's expression 16(*i*) where $B(1, m)$ appears in print instead of the correct $B(l, m)$.

After using some elementary identities between Bessel functions and hypergeometric series[2,7] and the following identifications in expression (8): $p = 0$, $q = 1$, $n = \frac{3}{2}$, $\beta_1 = \frac{1}{2} - l$, one obtains

$$g_l(\rho) = \frac{(-1)^l 16\rho^{l+3}\Gamma(\frac{1}{2} - l)}{\pi^{\frac{3}{2}}(l + \frac{1}{2})B(\frac{3}{2}, l)} \times \int_0^1 y^{l+3}(1 - y^2)^{l-1} n_l(2\rho y)\, dy, \quad l \neq 0. \quad (9)$$

Thus, $g_l(\rho)$ is a particular weighted integral over a spherical Neumann function of order l. The three simplest $g_l(\rho)$ (for $l = 0, 1, 2$) are listed in Eq. (10):

$$g_0(\rho) = -\frac{8}{\pi}\rho^2 \cos 2\rho = \frac{16}{\pi}\rho^3 n_0(2\rho),$$
$$g_1(\rho) = \frac{8}{\pi}\left\{(\rho^2 - 2)\cos 2\rho + \left(\frac{1}{\rho} - 2\rho\right)\sin 2\rho\right\},$$
$$g_2(\rho) = -\frac{8}{\pi}\left\{\left(\rho^2 - 18 + \frac{36}{\rho^2}\right)\cos 2\rho + \left(-6\rho + \frac{33}{\rho} - \frac{18}{\rho^3}\right)\sin 2\rho\right\}. \quad (10)$$

We have verified Eq. (7) for the cases $l = 0$, 1 by using the explicit expression (10) for $g_l(\rho)$ and the expression for $j_l^2(\rho)$ in terms of trigonometric functions. Already for the case $l = 1$ the verification becomes rather lengthy.

It appears that the leading terms of $g_l(\rho)$ for large ρ is $(-1)^{l+1}(8/\pi)\rho^2 \cos 2\rho$. It is interesting to note that this term and the leading term of $j_l^2(\rho)$ for large ρ, namely, $\sin^2(\rho - l\pi/2)/\rho^2$ by themselves satisfy an orthonormality condition analogous to Eq. (7) (where $x, y, y' \geq 0$):

$$\int_0^\infty \frac{\sin^2(xy - l\pi/2)}{(xy)^2}\left\{(-1)^{l+1}\frac{8}{\pi}(xy')^2 \cos(2xy')\right\} dx = \delta(y - y'). \quad (11)$$

This implies that what is left over of the functions $j_l^2(\rho)$, $g_l(\rho)$ constitutes functions whose overlap is zero:

$$\int_0^\infty \left\{j_l^2(xy) - \frac{\sin^2(xy - l\pi/2)}{(xy)^2}\right\} \times \left\{g_l(xy') - (-1)^{l+1}\frac{8}{\pi}(xy')^2 \cos(2xy')\right\} dx = 0. \quad (12)$$

With the help of $g_l(kr)$ one can find the "j_l^2 transform" of any well-behaved function $P(r)$ which goes to zero at least as $1/r^2$ for $r \to \infty$:

$$P(r) = \int_0^\infty j_l^2(kr) F(k)\, dk, \quad F(k) = \int_0^\infty g_l(kr) P(r)\, dr. \quad (13)$$

[7] Reference 3, p. 116.

Using these integral representations one may also extract from the Born expression integrals of $V_l(r)$ weighted with any well-behaved function $P(r)$.

We illustrate the usefulness of the inverse functions of $j_l^2(kr)$ with two specific examples.

Suppose we wish to generalize expression (1) and obtain an integral expression for $P(r) = r^{2(l+p)}e^{-r^2}$. Then from expression (13),

$$F(k) = \int_0^\infty \left\{ -\frac{4(kr)^2}{\pi(l+\frac{1}{2})} \times {}_1F_2(\tfrac{3}{2}; \tfrac{1}{2} - l, l + \tfrac{3}{2}; -(kr)^2) \right\} r^{2(l+p)}e^{-r^2}\, dr,$$

which may be evaluated with the help of expression (4) and the identifications $p = 1$, $q = 2$, $\alpha_1 = \frac{3}{2}$, $\beta_1 = \frac{1}{2} - l$, $\beta_2 = l + \frac{3}{2}$, $\mu = 2l + 2p + 3$, $2b^2 = 1$. Thus

$$r^{2(l+p)}e^{-r^2} = \int_0^\infty j_l^2(kr) \left\{ -\frac{2\Gamma(l+p+\frac{3}{2})}{\pi(l+\frac{1}{2})} k^2 \times {}_2F_2(\tfrac{3}{2}, l + p + \tfrac{3}{2}; \tfrac{1}{2} - l, l + \tfrac{3}{2}; -k^2) \right\} dk. \quad (14)$$

Expression (14) simplifies to expression (1) for $p = 0$ and is also quite simple for $l = 0$ and arbitrary p since then, as in the case $l = 0$, the ${}_2F_2$ in the integral reduces to an ${}_1F_1$ which is just e^{-k^2} multiplied by a $(p + 2)$-term polynomial in k^2.

Another more practical application of the inverse function $g_l(kr)$ occurs in potential theory. In the Born approximation the phase shift for the lth partial wave is given by

$$-\frac{\hbar^2}{m}\frac{\delta_l(k)}{k} = \int_0^\infty j_l^2(kr)V_l(r)r^2\, dr.$$

If the phase shifts are known from experiment and the Born approximation is assumed to be valid, one can in fact obtain the potential $V_l(r)$ directly from $\delta_l(k)$, since multiplying the above expression by $g_l(kr')$, integrating over k,

$$\int_0^\infty dk g_l(kr') \left(-\frac{\hbar^2}{m}\frac{\delta_l(k)}{k} \right) = \int_0^\infty dk g_l(kr') \int_0^\infty dr j_l^2(kr)V_l(r)r^2,$$

interchanging the order of integration, and using Eq. (7), one obtains

$$V_l(r)r^2 = \int_0^\infty g_l(kr) \left(-\frac{\hbar^2}{m}\frac{\delta_l(k)}{k} \right) dk. \quad (15)$$

For a simple illustration of (15) consider the $l = 0$ partial wave and assume $V_0(r)$ is the square well $V_0(r) = V_0$, $r < b$; $V_0(r) = 0$, $r > b$. One can calculate the phase shifts directly:

$$-\frac{\hbar^2}{m}\frac{\delta_0(k)}{k} = V_0 \int_0^b j_0^2(kr)r^2\, dr = V_0 \left\{ \frac{b^2}{2k^2} - \frac{\sin 2kb}{4k^3} \right\}.$$

If this expression for $\delta_0(k)$ is substituted into (15), one regains the square-well potential:

$$V_0(r) = V_0 \frac{1}{\pi} \left\{ \int_0^\infty \frac{\sin 2(b + r)k}{k}\, dk + \int_0^\infty \frac{\sin 2(b - r)k}{k}\, dk \right\}$$
$$= V_0, \quad \text{for} \quad r < b,$$
$$= 0, \quad \text{for} \quad r > b,$$

since

$$\int_0^\infty \frac{\sin \alpha k}{k}\, dk = \frac{\pi}{2}, \quad \text{if} \quad \alpha > 0,$$
$$= -\frac{\pi}{2}, \quad \text{if} \quad \alpha < 0.$$

III. THE INVERSE FUNCTION OF $j_l(xy)j_{l+1}(xy)$

To obtain an inverse function for $j_l(\rho)j_{l+1}(\rho)$ analogous to the inverse function for $j_l^2(\rho)$ of the previous section, the following approach may be adopted. Divide expression (1) by r^{2l} and differentiate both sides with respect to r using the relationship[2]

$$\frac{d}{d\rho}\{\rho^{-l}j_l(\rho)\} = -\rho^{-l}j_{l+1}(\rho).$$

The result is

$$r^{2l+1}e^{-r^2} = \int_0^\infty j_l(kr)j_{l+1}(kr) \times \left\{ -\frac{2}{\pi}\Gamma(l + \tfrac{1}{2})k^3 {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2} - l; -k^2) \right\} dk. \quad (16)$$

An expression involving the inverse function of $j_l(\rho)j_{l+1}(\rho)$ may be obtained exactly as before with the help of Eq. (4). The identifications are $\mu = 2l + 5$, $p = 1$, $q = 2$, $2b^2 = 1$, $\alpha_1 = \frac{3}{2}$, $\beta_1 = \frac{1}{2} - l$, $\beta_2 = l + \frac{5}{2}$:

$$\left\{ -\frac{2}{\pi}\Gamma(l + \tfrac{1}{2})k^3 {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2} - l; -k^2) \right\} = \int_0^\infty \frac{-4(kr)^3}{\pi(l + \frac{1}{2})(l + \frac{3}{2})} \times {}_1F_2(\tfrac{3}{2}; \tfrac{1}{2} - l, l + \tfrac{5}{2}; -(kr)^2)(r^{2l+1}e^{-r^2})\, dr. \quad (17)$$

The inverse function of $j_l(\rho)j_{l+1}(\rho)$ is

$$t_l(\rho) = -\frac{4\rho^3}{\pi(l + \frac{1}{2})(l + \frac{3}{2})} {}_1F_2(\tfrac{3}{2}; \tfrac{1}{2} - l, l + \tfrac{5}{2}; -\rho^2) \quad (18)$$

or

$$\int_0^\infty j_l(xy)j_{l+1}(xy)t_l(xy')\,dx = \delta(y - y'). \quad (19)$$

A difference from the previous section is that both $j_l(\rho)j_{l+1}(\rho)$ and $t_l(\rho)$ are odd functions

$$j_l(-\rho)j_{l+1}(-\rho) = -j_l(\rho)j_{l+1}(\rho), \quad t_l(-\rho) = -t_l(\rho).$$

For $l = 0$ we evaluated $t_l(\rho)$ explicitly in terms of trigonometric functions and verified expression (19) directly:

$$t_0(p) = -\frac{8}{\pi}\{(\rho^2 - \tfrac{1}{2})\sin 2\rho + \rho\cos 2\rho\}. \quad (20)$$

Similar considerations apply as before. The functions $t_l(\rho)$ can be expressed with the help of expression (8) as integrals over a particular weighted integral of a spherical Neumann function.

The leading term in $t_l(\rho)$ as $\rho \to \infty$ appears to be $(-1)^{l+1}(8/\pi)\rho^2 \sin 2\rho$. Multiplied by the leading terms of $j_l(\rho)j_{l+1}(\rho)$ as $\rho \to \infty$, namely

$$\frac{1}{\rho^2}\sin\left(\rho - \frac{l\pi}{2}\right)\sin\left(\rho - \frac{(l+1)\pi}{2}\right),$$

this product of leading terms again satisfies by itself an orthonormality relation analogous to Eq. (19).

Further partial differentiations of expression (16) with respect to r will yield inverse functions of sums of products of two Bessel functions, for instance of

$$j_{l+1}^2(\rho) + j_l(\rho)j_{l+2}(\rho).$$

Physical applications of such functions are not immediate.

IV. INVERSE FUNCTION OF $j_l(xy)j_{l+2}(xy)$

A tensor force couples states of equal total angular momentum but differing by two units of orbital angular momentum. The coupling parameter $\rho_{l+1}(k)$ is in the Born approximation proportional to an integral of the potential weighted by $j_l(kr)j_{l+2}(kr)$. Hence the physical interest in the inverse function of $j_l(kr)j_{l+2}(kr)$.

In this section we make use of an expression derived by Sanderson.[8] His procedure is outlined in Appendix B. The result is the following:

$$r^{2l}/b^3\frac{\partial}{\partial b}\{e^{-r^2/2b^2}/b^{2l+1}\} = \int_0^\infty j_{l-1}(kr)j_{l+1}(kr)\left\{\left[\frac{2^{l+9/2}3\Gamma(l+\frac{1}{2})}{\pi(1-2l)}\right]k^4 \times {}_1F_1(\tfrac{5}{2};\tfrac{3}{2}-l;-2b^2k^2)\right\}dk. \quad (21)$$

By applying two judicious partial differentiations to expression (4), once with the respect to b and once with respect to k^2, it can be cast into the useful form

$$\int_0^\infty\left[\frac{1}{r^2}\frac{\partial}{\partial k^2}\left\{\frac{1}{k^2r^2}\,{}_pF_q(\alpha_1,\alpha_2,\cdots,\alpha_p;\beta_1,\beta_2,\cdots,\beta_q;-(rk)^2)\right\}\right]\frac{1}{b^3}\frac{\partial}{\partial b}\left[\frac{e^{-r^2/2b^2}}{b^\mu}\right]r^{\mu+3}\,dr = \frac{2^{(\mu/2)+2}\Gamma(\frac{1}{2}\mu)(\alpha_1)(\alpha_2)\cdots(\alpha_p)(\frac{1}{2}\mu)(\alpha_1+1)\cdots(\alpha_p+1)(\frac{1}{2}\mu+1)}{(\beta_1)(\beta_2)\cdots(\beta_q)(\beta_1+1)(\beta_2+1)\cdots(\beta_q+1)}\,{}_{p+1}F_q(\alpha_1+2,\cdots,\alpha_p+2,\tfrac{1}{2}\mu+2;\beta_1+2,\cdots,\beta_q+2;-2b^2k^2). \quad (22)$$

With the identifications $\alpha_1 + 2 = \frac{5}{2}$, $\beta_1 + 2 = \frac{3}{2} - l$, $\beta_2 + 2 = l + \frac{5}{2}$, $\mu = 2l + 1$, one obtains from Eq. (22) the expression

$$\int_0^\infty\left[-\frac{2^3(l+\frac{1}{2})}{\pi}k^4 \times\frac{\partial}{\partial k^2}\left\{\frac{1}{k^2}\,{}_1F_2(\tfrac{1}{2};-l-\tfrac{1}{2},l+\tfrac{1}{2};-(kr)^2)\right\}\right]\times\left(\frac{1}{b^3}\frac{\partial}{\partial b}\left[\frac{e^{-r^2/2b^2}}{b^{2l+1}}\right]r^{2l}\right)dr = \frac{2^{l+9/2}3\Gamma(l+\frac{1}{2})}{\pi(1-2l)}k^4\,{}_1F_1(\tfrac{5}{2};\tfrac{3}{2}-l;-2b^2k^2). \quad (23)$$

By comparing expression (21) and (23) it follows that

$$q_l(\rho) = -\frac{2^3(l+\frac{3}{2})}{\pi}\rho^4 \times\frac{d}{d\rho^2}\left\{\frac{1}{\rho^2}\,{}_1F_2(\tfrac{1}{2};-l-\tfrac{3}{2},l+\tfrac{3}{2};-\rho^2)\right\} \quad (24)$$

is the inverse function of $j_l(\rho)j_{l+2}(\rho)$, or

$$\int_0^\infty j_l(xy)j_{l+2}(xy)q_l(xy')\,dx = \delta(y - y'). \quad (25)$$

For $l = 0$,

$$q_0(\rho) = \frac{8}{\pi}\left\{\left(\rho^2 - \frac{9}{2}\right)\cos 2\rho + 3\left(\frac{1}{\rho} - \rho\right)\sin 2\rho\right\}. \quad (26)$$

A direct verification of Eq. (25) was carried out for $l = 0$. As in Sec. II, $j_l(\rho)j_{l+2}(\rho)$ and $q_l(\rho)$ are both even functions. $q_l(\rho)$ can be written as a sum of two ${}_1F_2$ series or as a sum of two integrals of weighted spherical Neumann functions. We use Eq. (25) to obtain $G(k)$ in

$$r^p e^{-r^2} = \int_0^\infty j_{l-1}(kr)j_{l+1}(kr)G(k)\,dk, \quad (27)$$

[8] E. A. Sanderson (private communication).

an expression similar to Eqs. (1), and (16), and slightly more useful than expression (21).

By arguments similar to those leading to expression (13),

$$G(k) = \int_0^\infty q_{l-1}(kr) r^p e^{-r^2}\, dr. \tag{28}$$

This integral is evaluated with the help of Eq. (4):

$$G(k) = \frac{(2l+1)2}{\pi}\left[\Gamma\left(\frac{p+1}{2}\right) \times {}_2F_2\left(\tfrac{1}{2}, \frac{(p+1)}{2}; -l-\tfrac{1}{2}, l+\tfrac{1}{2}; -k^2\right) - \frac{2k^2}{(2l+1)^2}\Gamma\left(\frac{p+3}{2}\right) \times {}_2F_2\left(\tfrac{3}{2}, \frac{(p+3)}{2}; -l+\tfrac{1}{2}, l+\tfrac{3}{2}; -k^2\right)\right]. \tag{29}$$

If $p = 2l$ the ${}_2F_2$'s in expression (29) reduce to ${}_1F_1$'s and expression (27) becomes

$$r^{2l}e^{-r^2} = \int_0^\infty j_{l-1}(kr) j_{l+1}(kr) \times \left\{\frac{2\Gamma(l+\frac{1}{2})}{\pi} e^{-k^2}[-k^2\, {}_1F_1(-l-1; \tfrac{1}{2}-l; -k^2) + (2l+1)\, {}_1F_1(-l-1; -l-\tfrac{1}{2}; k^2)]\right\} dk, \tag{30}$$

where the term in square brackets is an $(l+2)$-term polynomial in k^2.

V. CONCLUSION

A conjecture from this work appears to be that inverse functions for any $j_l(xy)j_m(xy)$ (and not only the cases $m = l, l+1, l+2$ to which we restricted our attention) could be obtained if one utilizes the procedures and techniques outlined in this paper. As to physical applications, some of the expressions in this paper have already been used in nuclear calculations.[1,9] We are presently using the inverse functions to obtain the two-body potential explicitly from the experimental phase shifts. For this purpose we employ Eq. (15) and related expressions.

ACKNOWLEDGMENTS

One of the authors (H. A. M.) wishes to express his gratitude to Professor R. J. Blin-Stoyle for a post-doctoral fellowship at the University of Sussex where this work was begun with the help and encouragement of Dr. J. P. Elliott and Dr. E. A. Sanderson. We would like to thank Dr. R. Acharya for his interest and a number of helpful discussions.

APPENDIX A

In this appendix we derive expression (1). Using Kummer's relation,[3]

$${}_1F_1(\alpha; \beta; x) = e^x\, {}_1F_1(\beta-\alpha; \beta; -x),$$

the right-hand side of Eq. (1) becomes

$$\int_0^\infty j_l^2(kr)\left[-\frac{2}{\pi}\Gamma(l+\tfrac{1}{2}) \times e^{-k^2}k^2\, {}_1F_1(-l-1; \tfrac{1}{2}-l; k^2)\right] dk = -\Gamma(l+\tfrac{1}{2})\sum_{n=0}^{l+1}\frac{(-l-1)_n}{(\frac{1}{2}-l)_n n!}\int_0^\infty J_{l+\frac{1}{2}}(kr)^2 e^{-k^2}k^{2n+1}\, dk, \tag{A1}$$

where we have written out ${}_1F_1(-l-1; \frac{1}{2}-l; k^2)$ as an explicit sum, interchanged the order of summation and integration, and converted from spherical to cylindrical Bessel functions[2]

$$\left(j_l(\rho) = \sqrt{\frac{\pi}{2\rho}}\, J_{l+\frac{1}{2}}(\rho)\right).$$

The integrals of expression (A1) can be evaluated with the help of the expression[10]

$$\int_0^\infty J_\mu(kr)J_\nu(kr)e^{-k^2}k^{\lambda-1}\, dk = \frac{r^{\mu+\nu}\Gamma\left(\frac{\lambda+\mu+\nu}{2}\right)}{2^{\mu+\nu+1}\Gamma(\mu+1)\Gamma(\nu+1)} \times {}_3F_3\left(\frac{\mu+\nu+1}{2}, \frac{\mu+\nu+2}{2}, \frac{\lambda+\mu+\nu}{2}; \mu+1, \nu+1, \mu+\nu+1; -r^2\right), \quad \lambda+\mu+\nu > 0, \tag{A2}$$

and the identifications $\mu = \nu = l+\frac{1}{2}$, $\lambda = 2n+2$. Thus expression (A1) is equal to:

$$-\Gamma(l+\tfrac{1}{2})\sum_{n=0}^{l+1}\frac{(-l-1)_n r^{2l}\Gamma(n+l+\frac{3}{2})}{(\frac{1}{2}-l)_n n!\, 2^{2l+2}\Gamma(l+\frac{3}{2})^2}\, {}_2F_2(l+1, n+l+\tfrac{3}{2}; l+\tfrac{3}{2}, 2l+2; -r^2) = \frac{-\Gamma(l+\frac{1}{2})r^{2l}}{2^{2l+2}\Gamma(l+\frac{3}{2})^2}\sum_{n=0}^{l+1}\sum_{m=0}^{\infty}\frac{(-l-1)_n\Gamma(n+l+\frac{3}{2})(l+1)_m \times (n+l+\frac{3}{2})_m(-r^2)^m}{(\frac{1}{2}-l)_n n!\,(l+\frac{3}{2})_m(2l+2)_m m!}. \tag{A3}$$

[9] J. P. Elliott, H. A. Mavromatis, E. A. Sanderson, and B. Singh Nucl. Phys. (to be published).

[10] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, London, 1966), p. 396. A factor of $\frac{1}{2}$ is missing from the right-hand side of this expression in Watson. This was kindly pointed out by E. A. Sanderson.

However,

$$\frac{\Gamma(n+l+\frac{3}{2})(n+l+\frac{3}{2})_m}{\Gamma(l+\frac{3}{2})(l+\frac{3}{2})_m} = (m+l+\tfrac{3}{2})_n.$$

We use this identity in expression (A3) and interchange the order of the summations.

$$= \frac{-\Gamma(l+\frac{1}{2})r^{2l}}{\Gamma(l+\frac{3}{2})2^{2l+2}} \sum_{m=0}^{\infty} \frac{(l+1)_m(-r^2)^m}{(2l+2)_m m!} \times \left\{ \sum_{n=0}^{l+1} \frac{(-l-1)_n(m+l+\frac{3}{2})_n}{(\frac{1}{2}-l)_n n!} \right\}$$

$$= \frac{-\Gamma(l+\frac{1}{2})r^{2l}}{\Gamma(l+\frac{3}{2})2^{2l+2}} \sum_{m=0}^{\infty} \frac{(l+1)_m(-r^2)^m}{(2l+2)_m m!} \times {}_2F_1(-l-1, m+l+\tfrac{3}{2}; \tfrac{1}{2}-l; 1). \quad \text{(A4)}$$

But by Vandermondes' Theorem,[11,12]

$${}_2F_1(-(l+1), m+l+\tfrac{3}{2}; \tfrac{1}{2}-l; 1) = \frac{(-1)^{l+1}(m+l+1)_{l+1}}{(\frac{1}{2}-l)_{l+1}},$$

which can be written as

$$= \frac{(-1)^{l+1}(2l+2)_m\Gamma(2l+2)(-1)^l\Gamma(\frac{1}{2})^2}{\Gamma(\frac{3}{2})(l+1)_m\Gamma(l+1)\Gamma(l+\frac{1}{2})} = \frac{(-1)^{2l+1}(2l+1)_m 2^{2l+2}\Gamma(l+\frac{3}{2})}{(l+1)_m\Gamma(\frac{1}{2}+l)}.$$

Upon making this substitution, Eq. (A4) reduces to

$$r^{2l}\sum_{m=0}^{\infty}\frac{(-r^2)^m}{m!} = r^{2l}e^{-r^2},$$

assuming l to be integral. Hence

$$r^{2l}e^{-r^2} = \int_0^{\infty} j_l^2(kr)\left[-\frac{2\Gamma(l+\frac{1}{2})}{\pi}k^2 \times {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2}-l; -k^2)\right] dk.$$

[11] Reference 3, pp. 11, 17, 18, and 23.

[12] Y. L. Luke, *Integrals of Bessel Functions* (McGraw-Hill Book Co., Inc., New York, 1962), p. 18.

APPENDIX B

For completeness here we reproduce Sanderson's derivation of expression (21). In Eq. (1) let $r^2 \to r^2/2b^2$, $k^2 \to 2b^2k^2$; then

$$\frac{r^{2l}}{b^{2l}}e^{-r^2/2b^2} = Ab^3\int_0^{\infty} {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2}-l; -2b^2k^2) j_l^2(kr)k^2\, dk, \quad \text{(B1)}$$

where

$$A = \Gamma(l+\tfrac{1}{2})2^{l+\frac{5}{2}}/\pi.$$

Now use the fact[13] that

$$j_l^2(kr)k^2 = \frac{d}{dk}\{\tfrac{1}{2}k^3(j_l^2(kr) - j_{l-1}(kr)j_{l+1}(kr))\}, \quad l \geq 1,$$

and integrate (B1) by parts. The surface term vanishes, leaving

$$\frac{r^{2l}e^{-r^2/2b^2}}{b^{2l+5}} = 2A\int_0^{\infty} k^4\left[\frac{d}{d(-2b^2k^2)} {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2}-l; -2b^2k^2)\right] \times \{j_l^2(kr) - j_{l-1}(kr)j_{l+1}(kr)\}\, dk. \quad \text{(B2)}$$

Differentiating Eq. (B1) with respect to b on both sides,

$$\frac{e^{-r^2/2b^2}r^{2l}}{b^{2l+5}}\left\{-(2l+3)+\frac{r^2}{b^2}\right\} = -4A\int_0^{\infty} k^4\left[\frac{d}{d(-2b^2k^2)} {}_1F_1(\tfrac{3}{2}; \tfrac{1}{2}-l; -2b^2k^2)\right] \times j_l^2(kr)\, dk, \quad \text{(B3)}$$

and adding Eq. (B3) to two times Eq. (B2) yields the desired result,

$$\frac{r^{2l}}{b^3}\frac{\partial}{\partial b}\left\{\frac{e^{-r^2/2b^2}}{b^{2l+1}}\right\} = \int_0^{\infty} j_{l-1}(kr)j_{l+1}(kr) \times \left[\frac{2^{l+9/2}3\Gamma(l+\frac{1}{2})}{\pi(1-2l)}\right]k^4 {}_1F_1(\tfrac{5}{2}; \tfrac{3}{2}-l; -2b^2k^2)\, dk. \quad \text{(B4)}$$

[13] Reference 3, p. 116.

High-Energy Fixed-Angle Potential Scattering*

PORTER JOHNSON
Department of Physics, Case Western Reserve University, Cleveland, Ohio

(Received 18 December 1967)

We continue the study of the asymptotic behavior of the scattering amplitude $f(k, \theta)$ at high-momenta k for fixed (nonforward) scattering angles θ. The considerations here are limited to the class of potentials for which the momentum-space representations $V(q)$ decrease less rapidly than some inverse power of q as $q \to \infty$ through positive real values. We place some additional relatively simple conditions upon $V(q)$ for $q > 0$ which are sufficient to guarantee that the asymptotic limit of f for large k with fixed θ is, in fact, the first Born approximation $f_1(k, \theta)$.

I. INTRODUCTION

In this paper we limit our considerations to scattering from central potentials $V(r)$ and study the asymptotic behavior of the corresponding scattering amplitude $f(k, \theta)$ in the limit as the momentum $k \to \infty$ while the angle θ is fixed away from the forward direction. In particular, f is compared with the first Born approximation $f_1(k, \theta)$ in this asymptotic limit.

One may give plausibility arguments to indicate that f_1 is the asymptotic limit of f.[1] These plausibility arguments are misleading and lead to incorrect conclusions in certain cases. The purpose of this paper is to establish that f_1 is the rigorous high-energy, fixed-angle limit of f for a certain class of potentials.

Let us adopt the following convention for the Fourier transform of the potential $V(q)$:

$$V(q) = \frac{1}{(2\pi)^3} \int d\mathbf{x} V(x) e^{-i\mathbf{q}\cdot\mathbf{x}}. \tag{1}$$

At the outset $V(q)$ is restricted by the requirement that there exists a number $p > 3$ such that[2]

$$\lim_{q\to\infty} q^n |V(q)| = 0, \quad \text{for } n < p, \tag{2a}$$

$$\lim_{q\to\infty} \frac{1}{q^n |V(q)|} = 0, \quad \text{for } n > p. \tag{2b}$$

We have in mind to prove the following limit involving f and f_1 for certain potentials which are subject to (2)[3]:

$$\lim_{\substack{k\to\infty \\ \theta\neq 0 \text{ fixed}}} \frac{f(k, \theta)}{f_1(k, \theta)} = 1. \tag{3}$$

It has been established that $f \to f_1$ in the limit as $k \to \infty$ with the momentum transfer Δ fixed. A particularly elegant proof of this has been given by Hunzicher.[4] Hunzicher obtained an expansion of $f(k, \Delta)$ for fixed Δ in inverse powers of k, which is of the form

$$f(k, \Delta) = f_1(\Delta) + \frac{1}{k} G_1(\Delta) + \frac{1}{k^2} G_2(\Delta) + \cdots$$

[One can easily check that $f_1(\Delta)$ does dominate the first two terms of this expansion in our fixed angle limit for a wide class of potentials subject to (2), so that the limit (3) is at least plausible.]

We now indicate a procedure for constructing a proof of (3) for a given potential which is subject to (2). The first step is to write the potential $V(x)$ in the form

$$V(x) = v_1(x) v_2(x).$$

It is convenient to restrict v_1 and v_2 to being square-integrable functions of $\mathbf{x}$. [This automatically restricts $V(x)$ to being absolutely integrable and square-integrable over $\mathbf{x}$.] As a consequence, the Fourier transforms to v_1, v_2, and V are well defined. [The convention (1) is used for $v_1(q)$ and $v_2(q)$ as well.]

Let us require that $v_1(q)$ and $v_2(q)$ be subject to the following restrictions:

There exist positive numbers C_1, C_2, and b and a number $s > \frac{3}{2}$ such that

$$|v_1(q)| \leq C_1 \frac{1}{(1 + q^2 b^2)^s}, \tag{4a}$$

$$\left| \frac{d}{dq} v_1(q) \right| \leq C_1 b \frac{1}{(1 + q^2 b^2)^s}. \tag{4b}$$

* Supported in part by the Atomic Energy Commission.

[1] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London 1965), p. 110ff.

[2] The results of this paper can be extended to cases for which the limit (2b) does not exist, such as the example

$$V(q) = \frac{\sin qa}{1 + (qa)^4},$$

for which neither (a) nor (b) exists for $n \geq 4$. In such cases one can define an alternative limit, namely

$$A(n) = \lim_{q\to\infty} \frac{n - 1}{q^{n-1} \int_q^\infty dp\, |V(p)|}.$$

For our example, $A(n) = 0$ for $n > 4$. Note that $A(n)$ is identical with (2b) whenever (2b) exists. We can extend the results of this paper by considering limits defined in this way. Details will not be given.

[3] In a previous paper the limit (3) is considered for an entirely different class of potentials. [In fact, the limit (2a) would be zero for all those potentials.] The references are P. Johnson [J. Math. Phys. **9**, 712 1968]; and Ph.D. thesis, Princeton University (1967).

[4] W. Hunzicher, Helv. Phys. Acta **36**, 868 (1964). Hunzicher's result is sufficient to guarantee (3) for the case $\theta = 0$.

Under the assumption of these restrictions upon the potential factors, one can establish the following order relation involving $f - f_1$ as $k \to \infty$ for $\theta \neq 0$ fixed:

$$f(k, \theta) - f_1(k, \theta) = O\left(\frac{1}{k^{2s+\epsilon}}\right) \tag{5}$$

for $0 \leq \epsilon < \frac{1}{2}$.

Let us make the further restriction that $2s + \frac{1}{2} > p$, where p is defined in (1). Then one can conclude that the following order relation is valid as $k \to \infty$ with $\theta \leq 0$ fixed:

$$f(k, \theta) - f_1(k, \theta) = o[f_1(k, \theta)].$$

This relation is equivalent to (3), so that the assumptions (4) are sufficient to establish the asymptotic limit (3).

Let us note that the exponential potential $V(r) = \lambda e^{-\mu r}$ can be factored subject to (3). In particular, if we write $v(r) = v_1(r) = v_2(r) = (\lambda)^{\frac{1}{2}} e^{-\mu r/2}$, we can show that $v(q)$ is subject to (4) with $s = 2$, and that $V(q)$ is subject to (1) with $p = 4$. It is a consequence of the results of this paper that (3) is valid for the exponential potential.

In Sec. II the limit (5) is established under conditions (4). In Sec. III we discuss a rather universal factorization of V and will indicate how the conditions for the validity of (3) may be made weaker.

II. BOUNDING THE BORN SERIES

With the restriction (4) we are justified in applying the factorization $V = v_1 v_2$ to the terms of the Born series and writing the following expression:

$$f - f_1 = v_1 W v_2 - v_1 W^2 v_2 + \cdots, \tag{6}$$

where $W = v_2 G_k v_1$.[5]

We may explicitly write the kernel W as follows:

$$W(q_1, q_2; k) = \int d\mathbf{p} \frac{1}{p^2 - k^2 - i\epsilon} \times v_1(q_1 - p) v_2(p - q_2). \tag{7}$$

One wishes to obtain an upper bound upon $|W|$ for use in (6). It is convenient for this purpose to consider the function I, defined by

$$I(q_1, q_2; p) = \int d\Omega_p \frac{1}{[1 + b^2(q_1 - p)^2]^s} \times \frac{1}{[1 + b^2(p - q_2)^2]^s}. \tag{8}$$

In Appendix A it is shown that, for $s > \frac{3}{2}$, there exists a number $N(s)$ such that[6]

$$I(q_1, q_2; p) \leq \frac{N(s)}{p^2 b^2} f_s(\Delta), \tag{9}$$

where, for convenience, we have written

$$f_s(\Delta) = 1/(1 + \Delta^2 b^2)^s.$$

One can now use conditions (4) upon v_1 and v_2 along with the inequality (9) to derive the following bound upon $|W|$:

For every number ϵ such that $0 \leq \epsilon < \frac{1}{2}$, there exists a number $M(\epsilon)$ such that

$$|W(q_1, q_2; k)| \leq \frac{M(\epsilon)}{b} \frac{1}{(kb)^\epsilon} f_s(\Delta). \tag{10}$$

The details of this derivation are given in Appendix B. For notational convenience let us define

$$\lambda = \frac{M(\epsilon)}{b} \frac{1}{(kb)^\epsilon}.$$

We may now use (10) along with condition (4a) to bound (6) term-by-term by the following:

$$|f(k, \Delta) - f_1(\Delta)| \leq 2\pi^2 C_1 C_2 \left[\sum_{n=2}^{\infty} \lambda^{n-1} g_n(\Delta)\right], \tag{11}$$

where

$$g_n(\Delta) = \int d\mathbf{p}_1 \cdots d\mathbf{p}_n f_s(q_1 - p_1) \times f_s(p_1 - p_2) \cdots f_s(p_{n-1} - p_n) f_s(p_n - q_2).$$

In Appendix C the following crucial result is shown to be valid:

For every number $s > \frac{3}{2}$, there exists a number $M(s)$ such that

$$\int d\mathbf{p} f_s(q - p) f_s(p - r) \leq M(s) f_s(q - r). \tag{12}$$

As a consequence one can conclude that $g_n(\Delta) \leq [M(s)]^n f_s(\Delta)$. It is then possible to establish the following order relation in the limit $k \to \infty$ for fixed angles:

$$f(k, \theta) - f_1(k, \theta) = O\left(\frac{1}{k^{2s+\epsilon}}\right) \tag{5}$$

for $\epsilon < \frac{1}{2}$. We have thus established the desired result.

[5] The motivation for this factorization of V in this context is that the kernel W can be made square-integrable with appropriate choices of v_1 and v_2 for a wide class of potentials V. The scattering amplitude can be formally written

$$f = -v_1 \frac{1}{1 + W} v_2.$$

The usual procedure is to write the amplitude as $f = -[1/(1 + VG_k)]V$. The point is that VG_k is too singular to be square-integrable in most cases. As a consequence, relatively simple bounds upon W will lead to more economical bounds upon f than are obtained through bounds upon the usual resolvent.

[6] $\Delta = q_2 - q_1$.

III. DISCUSSION

We have shown that if V is factored subject to (4), the limit (3) involving the corresponding scattering amplitude and its first Born approximation is valid. In this section a particularly general factorization is exhibited, and a relatively simple set of sufficient conditions for (3) is obtained.

Let us formally factor $V(x)$ as follows:

$$v_1(x) = 1/(x^4 + b^4),$$
$$v_2(x) = (x^4 + b^4)V(x). \tag{13}$$

One may compute the Fourier transform $v_1(q)$ exactly, and a formal expression for $v_2(q)$ may be obtained:

$$v_1(q) = \frac{1}{4\pi b} \cdot \frac{1}{qb} \sin \frac{qb}{\sqrt{2}}\, e^{-qb/\sqrt{2}},$$
$$v_2(q) = b^4 V(q) + \frac{1}{q}\frac{d^4}{dq^4}[qV(q)]. \tag{14}$$

One can easily verify that $v_1(q)$ can be made to satisfy (4) with a suitable choice of C for any number $s > 0$. The following conditions upon $V(q)$ are sufficient to guarantee that $v_2(q)$ exists and is subject to (4) for the chosen values of s:

$$\begin{aligned} |V(q)| &\leq \frac{C}{b^4} f_s(q), \\ \left|\frac{d}{dq} V(q)\right| &\leq \frac{C}{b^3} f_s(q), \\ \left|\frac{1}{q}\frac{d^4}{dq}(qV(q))\right| &\leq C f_s(q), \\ \left|\frac{d}{dq}\left(\frac{1}{q}\frac{d^4}{dq^4}[qV(q)]\right)\right| &\leq C b f_s(q). \end{aligned} \tag{15}$$

This particular factorization is relatively convenient in providing a simple, though not conclusive, test for the validity of (3). In a given case this factorization might not be the most suitable choice to obtain bounds of this type upon the scattering amplitude.

We have considered the class of potentials which is subject to (2) and have shown that with restrictions (4) the scattering amplitude converges to its first Born approximation at high energies for fixed angles of scattering. The restrictions (4) automatically eliminate potentials $V(x)$ such that

$$\lim_{x\to\infty} x^6 |V(x)| \neq 0,$$

for example, $V(x) = 1/(x + b)^4$. There is no known, nonpathological potential $V(x)$ which is subject to (2) and for which (3) has been proved incorrect. Our restrictions (4) in a sense must be artificial and unnecessary. One can prove that the following conditions upon the suitably chosen potential factors v_1 and v_2 are also sufficient for (3):

$$\begin{aligned} |v_1(q)| &\leq C_1/q^{2\lambda} f_s(q), \\ \left|\frac{d}{dq} v_1(q)\right| &\leq C_1 b/q^{2\lambda} f_s(q), \end{aligned} \tag{16}$$

for $\lambda < 1$, $\lambda + s > \frac{3}{2}$. We will not go into details here; the techniques involved in establishing this are the same as those developed here. Let us note, however, that these conditions implicitly contain the requirements $x^4 V(x) \to 0$ as $x \to \infty$ and $q^3 V(q) \to 0$ as $q \to \infty$. These limitations should be regarded as due to the failure of a rather simple-minded technique; it should not be considered in any sense as implicit evidence that (3) is not correct.

The most distinctive feature here is that the conditions (4), for example, do not imply or require any underlying analytic properties for $V(q)$ or $V(x)$. In other words, convergence of the Born series to the first Born approximation has been established with conditions placed upon $V(q)$ only on the real q axis, so that the analytic structure of V is irrelevant in this case.

ACKNOWLEDGMENTS

I would like to thank P. Kantor and J. Pietenpol of Case Western Reserve University for pointing out several errors in the initial work. I am also grateful to L. Foldy, R. Thaler, and K. Kowalski for several interesting discussions.

APPENDIX A

In this appendix we derive the bound (9) for the function I defined by (8). For this purpose it is convenient to introduce the following integral representation:

$$f_s(q) = \int_0^\infty d\alpha \rho_s(\alpha) e^{-q^2/4\alpha}, \tag{A1}$$

where

$$\rho_s(\alpha) = \frac{4b^2}{\Gamma(s)} \frac{1}{(4\alpha b^2)^{s+1}} e^{-1/4\alpha b^2}.$$

The restriction $s > \frac{3}{2}$ is certainly sufficient to guarantee that $f_s(q)$ is square-integrable over $\mathbf{q}$, so that its Fourier transform, which we write as $\tilde{f}_s(x)$, exists and may be represented as follows:

$$\tilde{f}_s(x) = \int_0^\infty d\alpha (4\pi\alpha)^{\frac{3}{2}} \rho_s(\alpha) e^{-\alpha x^2}.$$

One is justified in changing orders of integration to

write I in the following form:

$$I(q_1, q_2; p) = \frac{1}{(2\pi)^6} \iint dx\, dy f_s(x) f_s(y) \int d\Omega_p \times \exp[-i(\mathbf{q}_1 - \mathbf{p})\cdot\mathbf{x} - i(\mathbf{p} - \mathbf{q}_2)\cdot\mathbf{y}] = \frac{4}{\pi^2 p} \int_0^\infty \int d\alpha\, d\beta (\alpha\beta)^{\frac{3}{2}} \rho_s(\alpha)\rho_s(\beta) \times J(\alpha, \beta; p; q_1, q_2), \quad \text{(A2)}$$

where

$$J(\alpha, \beta; p; q_1, q_2) = \iint d\mathbf{x}\, d\mathbf{y} \frac{\sin p|\mathbf{x} - \mathbf{y}|}{|\mathbf{x} - \mathbf{y}|} \times e^{-(\alpha x^2 + \beta y^2)} e^{-i(q_2 \cdot x - q_1 - y)}.$$

The following bound upon $|J|$ can be obtained in a straightforward manner[7]:

$$|J| \le \frac{64\pi^3}{p} \frac{1}{\alpha + \beta} \frac{1}{(\alpha\beta)^{\frac{1}{2}}} e^{-\Delta^2/4(\alpha+\beta)}.$$

We thus obtain a bound upon I of the form

$$I \le \frac{256\pi}{p^2} \frac{1}{[\Gamma(s)]^2} \frac{1}{[4b^2]^{2s}} \int_0^\infty \int \frac{d\alpha\, d\beta}{(\alpha\beta)^s} \frac{1}{\alpha + \beta} \times \exp[-(1/\alpha + 1/\beta)1/4b^2 - \Delta^2/4(\alpha + \beta)]. \quad \text{(A3)}$$

One can bound I by

$$I \le \frac{256\pi}{p^2} \frac{1}{[\Gamma(s)]^2} \frac{1}{(4b^2)^{2s}} \left[\int_0^\infty \frac{d\alpha}{\alpha^{s+\frac{1}{2}}} e^{-1/4b^2\alpha}\right]^2 = \frac{64\pi}{p^2 b^2} \left[\frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)}\right]^2 \le \frac{64\pi}{p^2 b^2}. \quad \text{(A4)}$$

Let us define new variables of integration $u = \alpha + \beta$ and $v = 4\alpha\beta$, and write

$$I \le \frac{64\pi}{p^2} \frac{1}{[\Gamma(s)]^2} \frac{1}{[4b^2]^{2s}} \int_0^\infty \frac{du}{u} e^{-\Delta^2/4u} \times \int_0^{u^2} \frac{dv}{(u^2 - v)^{\frac{1}{2}}} \frac{1}{V^s} e^{-u/b^2 v}. \quad \text{(A5)}$$

We split the internal integration into two parts:

$$\int_0^{u^2} \frac{dv}{(u^2 - v)^{\frac{1}{2}}} \frac{1}{v^s} e^{-u/b^2 v} = H + J,$$

where

$$H = \int_0^{u^2/2} \frac{dv}{(u^2 - v)^{\frac{1}{2}}} \frac{1}{v^s} e^{-u/b^2 v} \le \frac{\sqrt{2}}{u} \int_0^{u^2/2} \frac{dv}{v^s} e^{-u/b^2 v} \le \frac{\sqrt{2}}{u} \int_0^\infty dz\, z^{s-2} e^{-uz/b^2} = \frac{\sqrt{2}}{u} \Gamma(s - 1)(b^2/u)^{s-1}$$

and

$$J = \int_{u^2/2}^{u^2} \frac{dv}{(u^2 - v)^{\frac{1}{2}}} \frac{1}{v^s} e^{-u/b^2 v} \le \left(\frac{2}{u^2}\right)^s \int_{u^2/2}^{u^2} \frac{dv}{(u^2 - v)^{\frac{1}{2}}} = \frac{1}{2}\left(\frac{2}{u^2}\right)^{s-1/2}.$$

There consequently exists constants C' and C'' such that

$$I \le \frac{1}{p^2} \int_0^\infty \frac{du}{u} e^{-\Delta^2/4u} \left\{\frac{C'}{u^s} (b^2)^{s-1} + \frac{C''}{u^{2s-1}}\right\} \frac{1}{(b^2)^{2s}}.$$

If we require $\Delta b > 1$, we can find a number C such that

$$I \le \frac{C}{p^2 b^2} \frac{1}{(\Delta^2 b^2)^s}. \quad \text{(A6)}$$

Thus, as a result of (A4) and (A6), the bound (11) upon I is proved.

APPENDIX B

Here we consider the kernel $W(q_1, q_2; k)$ defined in Eq. (7), where k is restricted such that $kb > 1$. The bound (10) upon W will be established here.

From (7) one can derive the following expression for the imaginary part of W:

$$\text{Im } W = \frac{k\pi}{2} \int_{|p|=k} d\Omega_p v_1(q_1 - p) v_2(p - q_2).$$

Conditions (4a) upon v_1 and v_2 are used, along with the inequality (9), to get the bound

$$|\text{Im } W| \le \frac{\pi C_1 C_2}{2b} \frac{N(s)}{kb} f_s(\Delta). \quad \text{(B1)}$$

The quantity Re W may be expressed through a principal value integral

$$\text{Re } W = \lim_{\epsilon \to 0^+} \left\{\int_0^{k^2 - \epsilon} dp \frac{p^2}{p^2 - k^2} \times \int d\Omega_p v_1(q_1 - p) v_2(p - q_2) + \int_{k^2 + \epsilon}^\infty dp \frac{p^2}{p^2 - k^2} \times \int d\Omega_p v_1(q_1 - p) v_2(p - q_2)\right\}.$$

Through a change in variables, we can rewrite Re W in terms of the following nonsingular integral:

$$\text{Re } W = k \int_0^1 \frac{du}{u^2 - 1} \times \left\{u^2 \int d\Omega_l v_1(|\mathbf{q}_1 - uk\mathbf{l}|) v_2(|uk\mathbf{l} - \mathbf{q}_2|) - \int d\Omega_l v_1\left(\left|\mathbf{q}_2 - \frac{1}{u} k\mathbf{l}\right|\right) v_2\left(\left|\frac{1}{u} k\mathbf{l} - \mathbf{q}_2\right|\right)\right\}.$$

[7] $\Delta = \mathbf{q}_2 - \mathbf{q}_1$. For proof of the inequality, see P. Johnson, Ph.D. thesis, Princeton University, 1967 (unpublished).

(The integrand clearly vanishes for $u = 1$.) We now make the decomposition Re $W = H + J$, where

$$H = k\int_0^1 du \int d\Omega_l v_1\left(\left|\mathbf{q}_1 - \frac{1}{u}k\mathbf{l}\right|\right) v_2\left(\left|\mathbf{q}_2 - \frac{1}{u}k\mathbf{l}\right|\right)$$

and

$$J = k\int_0^1 du\, \frac{u^2}{u^2 - 1} f(u),$$

with the definition

$$f(u) = \int d\Omega_l \Big\{ v_1(|\mathbf{q}_1 - uk\mathbf{l}|) v_2(|\mathbf{q}_2 - uk\mathbf{l}|) - v_1\left(\left|\mathbf{q}_1 - \frac{1}{u}k\mathbf{l}\right|\right) v_2\left(\left|\mathbf{q}_2 - \frac{1}{u}k\mathbf{l}\right|\right)\Big\}. \quad \text{(B2)}$$

One can easily obtain a bound upon $|H|$ through the use of conditions (4a) and the inequality (9):

$$|H| \leq \frac{N(s)}{(kb)^2} C_1 C_2 f_s(\Delta).$$

We now bound $|J|$ using Hölder's inequality.[8] For this purpose a number $r > 2$ is chosen and a number s is defined by $(1/r) + (1/s) = 1$. Hölder's inequality leads to the following bound on $|J|$:

$$|J| \leq k\left[\int_0^1 du\, u^2 |f(u)|\right]^{1/r}\left[\int_0^1 du\, \frac{u^2}{(1-u^2)^s} |f(u)|\right]^{1/s}.$$

One may bound the former integral in the same manner as above:

$$\int_0^1 du\, u^2 |f(u)| \leq \frac{2N(s)}{(kb)^2} C_1 C_2 f_s(\Delta).$$

The latter integral, however, must be treated with more delicacy. Let us use the conditions (4a) and (4b) along with the inequality (9) to obtain the following bound upon $|f'(u)|$:

$$|f'(u)| \leq \frac{4N(s)}{kb} \frac{C_1 C_2}{u^3} f_s(\Delta).$$

One may now use the inequality

$$|f(u)| \leq \int_u^1 dv\, |f'(v)|$$

to obtain the upper bound

$$\int_0^1 du \frac{u^2}{(1-u^2)^s} |f(u)| \leq \frac{4N(s)}{kb} C_1 C_2 \frac{[\Gamma(2-s)]^2}{\Gamma(4-s)} f_s(\Delta).$$

As a consequence we have bounded $|J|$ as follows:

$$|J| \leq \frac{4N(s)}{b} \frac{C_1 C_2}{(kb)^{1/r}} \left|\frac{[\Gamma(2-s)]^2}{\Gamma(4-2s)}\right|^{1/s} f_s(\Delta).$$

We are now correct in concluding that (10) is valid with $\epsilon = (1/r) < (1/2)$, where $M(\epsilon)$ is given by the formula

$$M(\epsilon) = C_1 C_2 N(s)\left(3 + 4\left[\frac{\Gamma^2(\mu)}{\Gamma(2\mu)}\right]^{1-\epsilon}\right) \quad \text{(B3)}$$

with $\mu = 1 - [\epsilon/(1-\epsilon)]$.

APPENDIX C

Let us consider the function $Q_s(\Delta)$ defined as follows[9]:

$$Q_s(\Delta) = \int d\mathbf{p} f_s(q-p) f_s(p-r). \quad \text{(C1)}$$

We will establish the following result about Q_s:
For every number $s > \frac{3}{2}$ there exists a number $N(s)$ such that

$$Q_s(\Delta) \leq N(s) f_s(\Delta). \quad \text{(C2)}$$

For proof of this result it is convenient to use the following "generalized Feynman identity"[10]:

$$\frac{1}{A^s C^p} = \frac{1}{B(s,p)} \frac{1}{2^{s+p+1}} \times \int_{-1}^{+1} du\, \frac{(1+u)^{s-1}(1-u)^{p-1}}{[A(1+u/2) + C(1-u/2)]^{s+p}}.$$

Then one may write

$$Q_s(\Delta) = \frac{1}{B(s,s)} \frac{1}{2^{2s-1}} \int_{-1}^{+1} du (1-u^2)^{-1} \times \int d\mathbf{p} \Big\{1 + b^2(q-p)^2\left(\frac{1+u}{2}\right) + b^2(p-r)^2\left(\frac{1-u}{2}\right)\Big\}^{-2s}.$$

For $s > \frac{3}{2}$, the $\mathbf{p}$ integration certainly converges. It may be explicitly performed so that one obtains

$$Q_s(\Delta) = \frac{B(\frac{3}{2}, 2s - \frac{3}{2})}{B(s,s)} \frac{1}{2^{2s}} \frac{4\pi}{b^3} \times \int_0^1 \frac{du(1-u^2)^{s-1}}{[1 + (1-u^2)\Delta^2 b^2/4]^{2s-\frac{3}{2}}}. \quad \text{(C3)}$$

We see immediately from (C3) that, for $s > \frac{3}{2}$,

$$Q_s(\Delta) \leq \frac{2\pi}{b^3} B(\tfrac{3}{2}, s - \tfrac{3}{2}). \quad \text{(C4)}$$

It is convenient to use the following inequalities to bound $Q_s(\Delta)$ as given in (C3):

$$(1-u^2) \leq 2(1-u),$$
$$\left[1 + \frac{\Delta^2 b^2}{4}(1-u^2)\right]^{-1} \leq \left[1 + \frac{\Delta^2 b^2}{4}(1-u)\right]^{-1}.$$

[8] $|\int dz\, A(z)B(z)| \leq \{\int dz\, |A(z)|^r\}^{1/r} \cdot \{\int dz\, |B(z)|^s\}^{1/s}$ with $r, s > 1$ and $(1/r) + (1/s) = 1$. See G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge University Press, Cambridge, 1964), pp. 139–143.

[9] Q_s depends only upon $\Delta = q - r$, as is seen below.

[10] $B(x,y) = \int_0^1 dt\, t^{x-1}(1-t)^{y-1}$ is our definition of the Beta function. See I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic Press Inc., New York, 1965), p. 948ff.

One obtains

$$Q_s(\Delta) \leq \frac{4\pi}{b^3}\frac{1}{2^s}\frac{B(\frac{3}{2}, 2s-\frac{3}{2})}{B(s,s)}\int_0^1 \frac{dv\, v^{s-1}}{[1+v\Delta^2b^2/4]^{2s-\frac{3}{2}}}.$$

Under the restrictions $s > \frac{3}{2}$, we are allowed to write

$$Q_s(\Delta) \leq \frac{4\pi}{b^3}\left(\frac{2}{\Delta^2b^2}\right)^s\int_0^\infty \frac{dz\, z^{s-1}}{(1+z)^{2s-\frac{3}{2}}}\frac{B(\frac{3}{2}, 2s-\frac{3}{2})}{B(s,s)}.$$

or

$$Q_s(\Delta) \leq \frac{4\pi}{b^3}\left(\frac{2}{\Delta^2b^2}\right)^s\frac{B(s, s-\frac{3}{2})B(\frac{3}{2}, 2s-\frac{3}{2})}{B(s,s)}. \quad \text{(C5)}$$

As a result of (C4) and (C5), one may conclude that (C2) is valid.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Limitable Dynamical Groups in Quantum Mechanics. I. General Theory and a Spinless Model

H. D. DOEBNER* AND O. MELSHEIMER†

International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy

(Received 25 July 1967)

A pure group-theoretical description of nonrelativistic interacting systems in terms of irreducible representations $U(D)$ of a so-called dynamical group D is investigated. The description is assumed to be complete in the sense that all observable quantities of the system can be calculated from $U(D)$ in the same way as the nonrelativistic free particle can be identified with an irreducible representation $U(G_E)$ of the central extension of the inhomogeneous Galilei group G_E. D depends on the interaction. It is a noninvariance group and it contains a spectrum-generating algebra. Our problem is to connect a representation of an arbitrary abstract group with a complete description of an interacting system. This needs some physically motivated principles. Some such principles are proposed. We assume that the interaction can be turned off, which implies that $U(D)$ and the physical representation $U(G_E)$ of the free-particle group G_E can be limited into each other. If this limitation can be formulated without violating superselection rules, i.e., mass and spin conservation in nonrelativistic systems, the group D^t is called a limitable group. Properties of these groups are derived. An explicit construction of a limitable D^t is given by embedding the free-particle group G_E into a larger group. A discussion of all embeddings leads to the special choice

$$D^t \approx G_E^0(N) \supset\!\times Sp(2N, R).$$

$G_E^0(N)$ is the central extension of the pure inhomogeneous Galilei group in N dimensions and $Sp(2N, R)$ the noncompact real form of the symplectic group. A representation theory for D^t is established using the technique of Nelson extensions, together with some properties of the universal enveloping algebra of the Lie algebra of $G_E^0(N)$. Our main success is that D^t is a limitable dynamical group and that the physical system described by D^t and the physical representation can be calculated uniquely from the proposed principles. The group-theoretical description is equivalent to nonrelativistic quantum mechanics for a spinless particle in N dimensions with an arbitrary second-order polynomial in P_i, Q_i, $i = 1, \cdots, N$ as Hamiltonian. The possibility of further models is discussed.

INTRODUCTION

1. In particle physics it is reasonable to consider internal and space–time symmetry groups as a first step of a pure and complete group-theoretical description of the interaction in the sense that all observable quantities of the system can be derived from one irreducible representation of the so-called dynamical group D.[1–7] The interaction usually given by a Hamiltonian is replaced by a group, in the same way as one can try to translate the interaction into a certain singularity structure of the S matrix. A dynamical group is a noninvariance group; it does not commute with the corresponding S matrix and it also contains a spectrum-generating subgroup because it describes scattering states as well as bound states.

* Permanent address: Institut für Theoretische Physik, Universität Marburg, Federal Republic of Germany.

† Institut für Theoretische Physik, Universität Marburg, Federal Republic of Germany.

[1] A. O. Barut, in *Conference on Symmetry Principles at High Energy, 1964* (W. H. Freeman and Co., New York, 1964).

[2] A. O. Barut, Phys. Rev. **135**, B835 (1964).

[3] A. O. Barut and A. Böhm, Phys. Rev. **139**, B1107 (1965).

[4] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Letters **19**, 322 (1965).

[5] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965).

[6] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters **17**, 143 (1965).

[7] For a recent review and further references, see A. O. Barut, "Istanbul lectures, 1966"; A. Böhm, "Boulder lectures, 1966"; P. Budini, "Schladming lectures, 1967" (ICTP, Trieste report prior to publication, IC/67/18).

One obtains

$$Q_s(\Delta) \leq \frac{4\pi}{b^3}\frac{1}{2^s}\frac{B(\frac{3}{2}, 2s-\frac{3}{2})}{B(s,s)}\int_0^1 \frac{dv\, v^{s-1}}{[1+v\Delta^2b^2/4]^{2s-\frac{3}{2}}}.$$

Under the restrictions $s > \frac{3}{2}$, we are allowed to write

$$Q_s(\Delta) \leq \frac{4\pi}{b^3}\left(\frac{2}{\Delta^2b^2}\right)^s\int_0^\infty \frac{dz\, z^{s-1}}{(1+z)^{2s-\frac{3}{2}}}\frac{B(\frac{3}{2}, 2s-\frac{3}{2})}{B(s,s)}.$$

or

$$Q_s(\Delta) \leq \frac{4\pi}{b^3}\left(\frac{2}{\Delta^2b^2}\right)^s\frac{B(s, s-\frac{3}{2})B(\frac{3}{2}, 2s-\frac{3}{2})}{B(s,s)}. \quad \text{(C5)}$$

As a result of (C4) and (C5), one may conclude that (C2) is valid.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Limitable Dynamical Groups in Quantum Mechanics. I. General Theory and a Spinless Model

H. D. DOEBNER* AND O. MELSHEIMER†

International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy

(Received 25 July 1967)

A pure group-theoretical description of nonrelativistic interacting systems in terms of irreducible representations $U(D)$ of a so-called dynamical group D is investigated. The description is assumed to be complete in the sense that all observable quantities of the system can be calculated from $U(D)$ in the same way as the nonrelativistic free particle can be identified with an irreducible representation $U(G_E)$ of the central extension of the inhomogeneous Galilei group G_E. D depends on the interaction. It is a noninvariance group and it contains a spectrum-generating algebra. Our problem is to connect a representation of an arbitrary abstract group with a complete description of an interacting system. This needs some physically motivated principles. Some such principles are proposed. We assume that the interaction can be turned off, which implies that $U(D)$ and the physical representation $U(G_E)$ of the free-particle group G_E can be limited into each other. If this limitation can be formulated without violating superselection rules, i.e., mass and spin conservation in nonrelativistic systems, the group D^t is called a limitable group. Properties of these groups are derived. An explicit construction of a limitable D^t is given by embedding the free-particle group G_E into a larger group. A discussion of all embeddings leads to the special choice

$$D^t \approx G_E^0(N) \supset\!\times Sp(2N, R).$$

$G_E^0(N)$ is the central extension of the pure inhomogeneous Galilei group in N dimensions and $Sp(2N, R)$ the noncompact real form of the symplectic group. A representation theory for D^t is established using the technique of Nelson extensions, together with some properties of the universal enveloping algebra of the Lie algebra of $G_E^0(N)$. Our main success is that D^t is a limitable dynamical group and that the physical system described by D^t and the physical representation can be calculated uniquely from the proposed principles. The group-theoretical description is equivalent to nonrelativistic quantum mechanics for a spinless particle in N dimensions with an arbitrary second-order polynomial in P_i, Q_i, $i = 1, \cdots, N$ as Hamiltonian. The possibility of further models is discussed.

INTRODUCTION

1. In particle physics it is reasonable to consider internal and space–time symmetry groups as a first step of a pure and complete group-theoretical description of the interaction in the sense that all observable quantities of the system can be derived from one irreducible representation of the so-called dynamical group D.[1–7] The interaction usually given by a Hamiltonian is replaced by a group, in the same way as one can try to translate the interaction into a certain singularity structure of the S matrix. A dynamical group is a noninvariance group; it does not commute with the corresponding S matrix and it also contains a spectrum-generating subgroup because it describes scattering states as well as bound states.

* Permanent address: Institut für Theoretische Physik, Universität Marburg, Federal Republic of Germany.

† Institut für Theoretische Physik, Universität Marburg, Federal Republic of Germany.

[1] A. O. Barut, in *Conference on Symmetry Principles at High Energy, 1964* (W. H. Freeman and Co., New York, 1964).

[2] A. O. Barut, Phys. Rev. **135**, B835 (1964).

[3] A. O. Barut and A. Böhm, Phys. Rev. **139**, B1107 (1965).

[4] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Letters **19**, 322 (1965).

[5] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965).

[6] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters **17**, 143 (1965).

[7] For a recent review and further references, see A. O. Barut, "Istanbul lectures, 1966"; A. Böhm, "Boulder lectures, 1966"; P. Budini, "Schladming lectures, 1967" (ICTP, Trieste report prior to publication, IC/67/18).

For the free relativistic particle, a group-theoretical approach is familiar; its space–time properties can be calculated from one irreducible representation of the Poincaré group.

A realization of this program for interacting systems proceeds in three steps:

(1) Choose a suitable abstract group D;
(2) Determine its physical representation $U_{P\hbar}(D)$;
(3) Identify the physical system described by this representation.

There are several attempts to do this. A plausible choice for D is, for example, an embedding $\tilde{G}$ of internal and space–time symmetry groups S and P, respectively, or an algebraic enlargement $\hat{P}$ of P such that P is a group contraction of $\hat{P}$. From additional principles and from model considerations, the physical representation must be derived together with a solution of the identification problem. It has been proposed to relate a certain representation of $\hat{P}$ to the physical mass spectrum[3] (which is in good agreement with experiments) and to consider, as in field theory, suitable matrix elements of $\tilde{G}$ as vertex parts or form factors.[8]

To have some more information on the physical background of these additional principles necessary for a solution of steps 2 and 3, it is reasonable to put solvable models into this formulation. This is done in the following sections. We present some methods and nonrelativistic examples for interpreting an abstract group as a dynamical group, so that the physical system described by the group and the physical representation can be calculated from basic principles, and so that the group-theoretical description is complete and equivalent to a quantum-mechanical formulation by a Hamiltonian. Our intention is not to explore the group-theoretical structure of certain nonrelativistic systems, but to derive in the reverse direction the physical system and the physical representation belonging to a given abstract group.

2. We discuss in Sec. 1 the dynamical group of the free quantum-mechanical particle in N dimensions, which is given by a central extension G_E of the N-dimensional inhomogeneous Galilei group, and we formulate those aspects of its representation theory which are relevant for the identification problem. Physical restrictions for a dynamical group of interacting systems are obtained in Sec. 2. We translate the fact that it is possible to vary the interaction strength λ within a Hamiltonian $H[\lambda]$ and to turn off the interaction by putting $\lambda \to 0$, i.e., $H[\lambda] \xrightarrow[\lambda\to 0]{} H^0$, with H^0 being the free Hamiltonian, into a property of the dynamical group $D[\lambda]$ belonging to $H[\lambda]$. Formally, this implies a limitation between $D[\lambda]$ and G_E. If a limitation of an abstract group into G_E can be mathematically realized, the group is called *limitable* and it is reasonable to demand that dynamical groups be limitable. Because mass and spin conservation are superselection rules in nonrelativistic quantum mechanics, we restrict the limitation between $D[\lambda]$ and G_E so that mass and spin conservation holds. We assume furthermore that the Hamiltonians and the generators of the corresponding degeneracy groups of all systems described by $D[\lambda]$ are contained in the Lie algebra of $D[\lambda]$. With these principles a group-theoretical formulation for some nonrelativistic systems is given in Sec. 3. We construct (step 1) a dynamical group D^t by embedding G_E in a semidirect product

$$D^t \approx G_E^0(N) \textcircled{\times} Sp(2N, R),$$

with $G_E^0(N)$ the central extension of the inhomogeneous pure Galilei group in N dimensions and $Sp(2N, R)$ the noncompact real form of symplectic group in $2N$ dimensions. A representation theory of D^t is formulated using the method of Nelson extensions. It is shown that D^t is a limitable group. The limitation is given for $N = 1$ by a generalized group contraction and for $N > 1$ by a more formal procedure. It determines, together with mass and spin conservation and with mass and spin of the free particle, the physical representation $U_{P\hbar}(D^t)$ (step 2). The generators of $U_{P\hbar}(D^t)$ can be written uniquely—up to equivalence—as functions of momentum and position operators P_i, Q_i, $i = 1, \cdots, N$. Therefore, the identification of $U_{P\hbar}(D^t)$ is possible (step 3). It turns out that D^t is the dynamical group of all spinless systems with Hamiltonians being second-order polynomials in P_i, Q_i, $i = 1, \cdots, N$. Some generalizations of the spinless model are discussed in Sec. 3E.

3. The proposed method does not depend on the existence of a degeneracy group E which was used to construct spectrum-generating groups for the three-dimensional oscillator[9–12] with $E \approx SU(3)$ and for the hydrogen atom[4,13–15] with $E_1 \approx SO(4)$ and $E_2 \approx SO(3, 1)$ for the discrete and continuous spectrum, respectively, and for the strong coupling

[8] C. Cocho, C. Fronsdal, H. Ar Rashid, and R. White, Phys. Rev. Letters **17**, 275 (1966).

[9] A. O. Barut, Phys. Rev. **139**, B1433 (1965).
[10] R. C. Hwa and J. Nuyts, Phys. Rev. **145**, 1188 (1966).
[11] G. Bisiacchi and P. Budini, Nuovo Cimento **44**, 418 (1966); **47**, 792 (1967).
[12] S. S. Sannikov, Zh. Eksp. Teor. Fiz. **49**, 1913 (1965) [Sov. Phys.—JETP **22**, 1306 (1966)].
[13] A. O. Barut, P. Budini, and C. Fronsdal, Proc. Roy. Soc. (London) **106**, A291 (1966).
[14] A. Böhm, Nuovo Cimento **43**, 665 (1966).
[15] R. Musto, Phys. Rev. **148**, 1274 (1966); R. H. Pratt, and T. F. Jordan, Phys. Rev. **148**, 1276 (1966).

theory.[16,17] The attractive idea of this approach was to evaluate the group-theoretical structure of these systems and to find a minimal embedding $\tilde{G}$ of E such that the set $\{U(E)\}$ of irreducible representations which are physically present is contained in a representation $U(\tilde{G})$, i.e., $U(\tilde{G}) \downarrow E = \{U(E)\}$, and that the Hamiltonian is a function of Casimir operators of E such that the desired eigenvalues are obtained. The existence of $\tilde{G}$ with these properties is a well-defined mathematical problem. $\tilde{G}$ is a noninvariance group and its physical representation is calculated from the known quantum-mechanical solution. The result is even unique if an additional parity transformation is introduced. The idea of putting the Hamiltonian into the enveloping algebra is in some sense complementary to the above approach, where H is one of the generators in $U_{Ph}(D^t)$. For a complete description one needs two irreducible representations of $\tilde{G} \approx Sp(6, R)$ for the oscillator and for the hydrogen atom one irreducible representation of both spectrum-generating groups $\tilde{G}_1 \approx SO(4, 1)$ and $\tilde{G}_2 \approx SO(3, 2)$. It was shown[18] that the construction can also be extended to incorporate electromagnetic transitions if $\tilde{G}_1$ is embedded into $SO(4, 2)$. These group-theoretical recalculations of quantum-mechanical properties are useful and strong tools for a relativistic generalization of the above systems.

1. THE DYNAMICAL GROUP OF THE FREE PARTICLE

In Secs. 1A and 1B we summarize and generalize some mathematical results on projective representations of the N-dimensional in the homogeneous Galilei group $G(N)$. The physical aspects and the identification of $G(N)$ as a dynamical group for the nonrelativistic free particle are discussed in 1C.

A. Central Extension $G_E(N)$ of the Galilei Group $G(N)$

1. Let $\mathbf{u}$, $\mathbf{v}$ be any two vectors of an N-dimensional vector space R_N and let R be an orthogonal transformation $R \in SO(N)$ in R_N and let τ be a real number. The N-dimensional Galilei group $G(N)$ is given by the set of elements $g = (R, \mathbf{u}, \mathbf{v}, \tau)$ with multiplication rule

$$g \cdot g' = (R, \mathbf{u}, \mathbf{v}, \tau)(R', \mathbf{u}', \mathbf{v}', \tau') = (R \cdot R', \mathbf{u} + R\mathbf{u}' + \tau'\mathbf{v}, \mathbf{v} + R\mathbf{v}', \tau + \tau')$$

and identity $e = (1, 0, 0, 0)$. The group has an $(N + 1)$-dimensional representation in $R_{N+1} \in (\mathbf{x}, t)$; $g \cdot (\mathbf{x}, t) = (R\mathbf{x} + \mathbf{u}t + \mathbf{v}, \tau + \tau')$. The interpretation of the subgroups $\{(R, 0, 0, 0)\} \approx SO(N)$ as N-dimensional rotations, $\{(1, 0, 0, \tau)\} \approx H$ and $\{(1, \mathbf{u}, 0, 0)\} \approx P_N$ as time and space translations, respectively, and $\{(1, 0, \mathbf{v}, 0)\} \approx Q_N$ as pure Galilei transformations is obvious. The Abelian subgroup of pure inhomogeneous Galilei transformations $P_N \otimes Q_N$ is denoted by $G^0(N)$. The dimension of $G(N)$ is $d = \frac{1}{2}N(N + 3) + 1$.[19]

2. $G(N)$ possesses strongly continuous unitary projective representations[20] which are not equivalent to vector representations and which can be written at least locally, i.e., in some neighborhood of $e \in G(N)$, as representations "up to a factor" with a group relation

$$U(g)U(g') = \omega(g, g')U(g \cdot g') = \exp[\varphi(g, g')] \cdot U(g \cdot g'), \quad g, g' \in G(N),$$

including the local factor ω and the local exponent φ, respectively. The local exponent $\varphi(g, g')$ and $G(N)$ can be combined in a bigger group $G_E^{[\varphi]}(N)$, the local group of $G(N)$. Take a fixed φ and consider the set of all pairs (θ, g) with θ being any real number and introduce a multiplication for (θ, g) by

$$(\theta, g) \cdot (\theta', g') = (\theta + \theta' + \varphi(g, g'), g \cdot g').$$

Then $G_E^{[\varphi]}(N)$ is given by $\{(\theta, g)\}$ and it is a central extension of $G(N)$. The set $F_\theta = \{(\theta, e)\}$ is a one-dimensional subgroup and $G_E^{[\varphi]}(N)/F_\theta$ is locally isomorphic to $G(N)$. All projective representations of $G(N)$ can be derived in two steps[20]:

(1) Calculation of all local factors $\omega(g, g')$ or exponents $\varphi(g, g')$;
(2) Construction of all vector representations of $G_E^{[\varphi]}(N)$.

3. The determination of local factors is an algebraic problem. In general, different $\varphi(g, g')$ can lead to representations of the corresponding local groups which are equivalent. Therefore, the set of all $\varphi(g, g')$ is decomposed into equivalence classes $\mathfrak{L}$ such that $\varphi(g, g') \in \mathfrak{L}$ lead to equivalent representations of $G_E^{[\varphi]}(N)$. A parametrization of $\mathfrak{L}$ is given by a real number γ, $-\infty < \gamma < +\infty$ (Appendix A), and hence also the local groups $G_E^{[\varphi]}(N)$ are classified by γ. Take for each γ the set of all irreducible representations $U(G_E^{[\varphi]}(N))$ of $G_E^{[\varphi]}(N)$; it turns out that this set is independent of γ and it suffices to construct all irreducible representations for, e.g., $\gamma = 1$. $G_E^{[1]}(N)$ is written as $G_E(N)$. The eigenvalues of the center in

[16] T. Cook, C. J. Goebel, and B. Sakita, Phys. Rev. Letters **15**, 35 (1965).
[17] J. G. Kuriyan, Ph. D. thesis, Syracuse University (1966).
[18] A. O. Barut, Phys. Rev. **156**, 1538, (1967); A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1541 (1967).
[19] Lie groups are denoted by G, D, $SO(N)$, etc., the corresponding Lie algebra by G, D, $so(N)$, unitary faithful strongly continuous representation of G by $U(G)$, and integrable representations of G by g, etc.
[20] V. Bargmann, Ann. Math. **59**, 1 (1954).

$G(N)$ determine the equivalence classes as well as, partly, the irreducible representation. With this result a representative local factor in each equivalence class is

$$\omega(g, g') = \exp i \cdot \tfrac{1}{2}m(\mathbf{u} \cdot R\mathbf{v}' - \mathbf{v} \cdot R'\mathbf{u}' + \tau'\mathbf{v} \cdot R\mathbf{v}').$$

For the corresponding Lie algebra we find (the subgroups $SO(N)$, H, P_N, Q_N, F_θ are generated by d_{ij}; P_i; Q_i; H_0; $i = 1, \cdots, N$, C)

$$\begin{aligned}
&[d_{ij}, d_{ke}] = \delta_{jk}\, d_{ie} - \delta_{ik}\, d_{je} + \delta_{ie}\, d_{jk} - \delta_{je}\, d_{ik},\\
&[P_i, P_k] = 0; \quad [Q_i, Q_k] = 0,\\
&[d_{ij}, P_k] = \delta_{jk} P_i - \delta_{ik} P_j,\\
&[d_{ij}, Q_k] = \delta_{jk} Q_i - \delta_{ik} Q_j,\\
&[d_{ij}, H_0] = 0; \quad [P_i, Q_j] = \delta_{ij}\mathsf{C},\\
&[P_i, H_0] = 0; \quad [Q_i, H_0] = P_i.
\end{aligned}$$

To include also reducible representations, the center element is written as C. Note that the algebra $\mathsf{G}_E(N)$ coincides with $\mathsf{G}(N)$ for $\mathsf{C} \equiv 0$.

4. There are two important semidirect product decompositions of $G_E(N)$. Let $\mathsf{G}_E^0(N)$ be the algebra spanned by $(\mathsf{P}_N \oplus \mathsf{C})\dot{+}\, \mathsf{Q}_N$:

$$\begin{aligned}
&[P_i, Q_j] = \delta_{ij}\mathsf{C},\\
&[P_i, P_j] = [Q_i, Q_j] = [P_i, \mathsf{C}] = [Q_i, \mathsf{C}] = 0.
\end{aligned}$$

The simply connected group which possesses $\mathsf{G}_E^0(N)$ as Lie algebra is called central extension of the pure inhomogeneous Galilei group, and we have

$$G_E(N) \approx G_E^0(N) \dot{\times} (SO(N) \otimes H_0).$$

Consider the subgroups $G_1 = \{(\theta, 1, \mathbf{u}, 0, \tau)\}$ and $G_2 = \{(0, R, 0, \mathbf{v}, 0)\}$ with Lie algebras $\mathsf{P}_N \oplus \mathsf{H}_0 \oplus \mathsf{C}$ and $\mathsf{Q}_N + so(N)$, respectively. Then the decomposition

$$G_E(N) \approx G_1(N) \dot{\times} G_2(N)$$

is a regular semidirect product between G_1 and G_2, because G_1 is Abelian and G_2 separable and locally compact. For practical calculations with the last decomposition the choice

$$\tilde{\varphi}(g, g') = \mathbf{v}R\mathbf{u}' + \tfrac{1}{2}\tau' v^2$$

for the representative in the local factor is suitable because then the multiplication

$$\begin{aligned}
g_1 \cdot g_2 &= (\theta, 1, \mathbf{u}, 0, \tau)(0, R, 0, \mathbf{v}, 0)\\
&= (\theta, R, \mathbf{u}, \mathbf{v}, \tau)
\end{aligned}$$

is quite simple.

B Representation Theory for $G_E(N)$

1. The regular semidirect decomposition suggests the application of Mackey's theory of induced representations,[21] which is useful if the representations of the corresponding little groups are known. The little groups G_l of $G_E(N)$ with $\mathsf{C} = im \cdot 1$ are easily calculated. They depend on the character group of G_1 and on the eigenvalues m and V of the Casimir operators $\mathsf{C} = im \cdot 1$ and

$$\mathsf{C}_{H_0}^{(2)} = \mathsf{H}_0\mathsf{C} + \frac{1}{2}\sum_{i=1}^{N} \mathsf{P}_i^2,$$

i.e., $G_e^{m,V}$; it can be shown that $G_e^{m,V} \approx SO(N)$ holds independently of m, V (Appendix B). Then the calculation of all irreducible representations $U(G_E(N))$ is a technical question and already known for the three-dimensional case.[22,23] The generalization to arbitrary dimensions is straightforward. Because the representations are "induced" by a fixed irreducible representation of $SO(N)$, they are classified by the set $s = \{s_1, \cdots, s_r\}$ of $r = [N/2]$ highest weights and furthermore by those Casimir operators which are functions of generators in G_1 only, i.e., C and $\mathsf{C}_{H_0}^{(2)}$. Hence the set $[m; V; s = \{s_1, \cdots, s_r\}]$ classifies all irreducible representations $U(G_E(N))^{[m;V;s]}$ of $G_E(N)$. Unitary projective representations of $G(N)$ are given by $U(0, R, \mathbf{v}, \mathbf{u}, \tau)$ and the irreducible ones are classified by $[m; s]$ because the definition of equivalence between projective representations[24] implies that different V but equal m, s lead to equivalent representations $U(G(N))$. In Sec. 1 C.1 it is shown that the constant V in $U(G_E(N))^{[m;V;s]}$ is physically unimportant.

2. Consider the subgroup $G_E^0(N)$ with Lie algebra $\mathsf{G}_E^0(N)$. Irreducible unitary representations of $G_E^0(N)$ and irreducible integrable representations of $\mathsf{G}_E^0(N)$ are known by a theorem of von Neumann[25] and by Lemma 1 of Ref. 26, respectively. There is—up to unitary equivalence—only one irreducible representation $U(G_E^0(N))^{[m]}$ determined by m or by $\mathsf{C} = im \cdot 1$, which is furthermore equivalent to the Schrödinger representation in quantum mechanics. Take now any representation $U(G_E^0(N))$ with $\mathsf{C} = im \cdot 1$ in a separable Hilbert space; then it can be decomposed in a countable set of equivalent irreducible representations $U(G_E^0(N))^{[m]}$. We call such representation quasi-irreducible if the multiplicity $n(m)$ of $U(G_E^0(N))^{[m]}$ in $U(G_E^0(N))$ is finite and we denote this by

[21] G. W. Mackey, Am. J. Math. **73**, 193 (1951); Ann. Math. **55**, 101 (1952); Bull. Am. Math. Soc. **69**, 628 (1963); Chicago lecture notes (1955).

[22] J.-M. Lévy-Leblond, J. Math. Phys. **4**, 776 (1963); see also A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).

[23] J. Voisin, J. Math. Phys. **6**, 1519 (1965); **6**, 1822 (1965).

[24] H. Neumann, Diplom-Arbeit, University of Marburg (1966); A. Hartkämper, Diplom-Arbeit, University of Marburg (1966).

[25] J. von Neumann, Math. Ann. **104**, 570 (1931).

[26] H. D. Doebner and O. Melsheimer, Nuovo Cimento **49**, 73 (1967).

$U(G_E^0(N))^{[m,n(m)]}$. If a representation $U(G_E(N))^{[m;V;s]}$ is restricted to $G_E^0(N)$, we get a quasi-irreducible representation $U(G_E^0(N))^{[m,n(m)]}$ with $n(m) = d(s)$ if $d(s)$ denotes the dimension of the representation of the corresponding little group.

Instead of s also eigenvalues of a set of Casimir operators $\mathsf{C}_{[s]}$ of $\mathsf{G}_E(N)$ can be used. For $N = 3$ the representation is characterized, e.g., by three Casimir operators:

$$\mathsf{C}_m^{(1)} = \mathsf{C} = im \cdot 1;\quad \mathsf{C}_{H_0}^{(2)} = \mathsf{H}_0\mathsf{C} + \frac{1}{2}\sum_{i=1}^{3}\mathsf{P}_i^2 = V \cdot 1;$$

$$\mathsf{C}_S^{(2)} = \left(\mathbf{d} + \frac{1}{2m}\,\mathbf{P} \times \mathbf{Q}\right)^2 = s(s+1)\cdot 1.$$

The last one defines the representation of the little group $SO(3)$.
Collecting the results we have:

Theorem 1: Irreducible unitary strongly continuous (vector) representations $U(G_E(N))$ of the central extension of the inhomogeneous Galilei group in N dimensions $G_E(N)$ are uniquely classified by $[m; V; s = \{s_1, \cdots, s_r\}]$ with s denoting the set of highest weights for the corresponding representation of the little group $SO(N)$ of $G_E(N)$ and with $-\infty < m, V < +\infty$. The restriction to $G_E^0(N)$ leads to a quasi-irreducible representation $U(G_E^0(N))^{[m,n(m)]}$ with multiplicity $n(m)$ given by the dimension of the representation space of the little group $SO(N)$.

C. $G_E(N)$ as Dynamical Group for the Free Particle

1. The relation between $G_E(N)$ and quantum mechanics is well known for $N = 3$,[22,23,27–29] for arbitrary $U(G_E(N))^{[m;V;s]}$ leads to a complete description of a free particle with mass m and spin s. Hence $G_E(N)$ is its dynamical group with a physical representation determined by mass and spin. The derivation of this result is straightforward. Consider the case $N = 1$. We describe the free particle with mass m by its Hamiltonian $H^0 = (i/2m)P^2$ and by position and momentum operators Q and P with commutation relation $[P, Q] = im \cdot 1$. The algebra spanned by $\{P, Q, H^0, C\}$ is isomorphic to $\mathsf{G}_E(1)$. Then we construct in $L^2(-\infty, +\infty)$ the Schrödinger representation of this algebra which leads to the Schrödinger equation of the free particle. One can prove[26] that this representation of $\mathsf{G}_E(1)$ is integrable to an irreducible unitary strongly continuous representation of the group $G_E(1)$[30] and hence the solution of the free-particle problem is equivalent to the construction of an irreducible representation of $G_E(1)$. The eigenvalue V of $\mathsf{C}_{H_0}^{(2)}$ is an additive physically unimportant constant in H^0 and it suffices to consider representations $U(G_E(1))^{[m;0]}$ only.[31] Mass conservation is a superselection rule and is related to the center in $G_E(1)$.

2. For $N > 1$ the degeneracy group $SO(N)$ of

$$H^0 = \frac{i}{2m}\sum_{i=1}^{N} P_i^2$$

must also be considered and generators $d_{ij} = i(P_iQ_j - P_jQ_i)$ additional to P_i, Q_i, C, $i = 1, \cdots, N$ appear. The algebra arising is isomorphic to $\mathsf{G}_E(N)$. The Schrödinger representation is constructed as usual and the integrability in respect to $\mathsf{G}_E^0(N)$ ensures in this case also the integrability of $\mathsf{G}_E(N)$. Then $U(G_E(N))^{[m;0;0]}$ describes the N-dimensional free particle with mass m and spin 0. To introduce spin for $N > 3$ in the same way as for $N = 3$, one adds to the orbital part d_{ij} of $SO(N)$ a spin part s_{ij} given by a $d \times d$ matrix, being the corresponding generator to d_{ij} in a d-dimensional irreducible representation of $SO(N)$ characterized by s,[32] and one needs a Hilbert space over vector-valued functions as representation space. The procedure corresponds uniquely to the construction of $U(G_E(N))^{[m;0;s]}$. s is interpreted as spin for $N = 3$ and for $N > 3$ as "spin content." The case $N > 3$ is not related to a physical free particle.

2. ON THE STRUCTURE OF DYNAMICAL GROUPS IN QUANTUM MECHANICS

Dynamical groups for nonrelativistic systems are physically restricted by a limitation property and by mass and spin conservation. A precise formulation of a limiting procedure between dynamical groups for

[27] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications Inc., London, 1931).
[28] J. V. Lepore, Phys. Rev. **119**, 821 (1960).
[29] M. Hammermesh, Ann. Phys. (N.Y.) **9**, 518 (1960).

[30] A rigorous formulation of the representation problem needs a domain assumption because P and/or Q are necessarily unbounded operators. In general, the physical content of a representation $g_E(1)$ of $\mathsf{G}_E(1)$ depends on the domain $\mathscr{D}_g \subset \mathscr{G}$ on which the representation is defined. The integrability of $g_E(1)$ is necessary and sufficient for $g_E(1)$, being unitary equivalent to the Schrödinger representation Ref. 26.
[31] The same result is obtained if the free particle is defined via a unitary irreducible projective representation of $G(N)$ labeled by m, s instead of via a unitary irreducible vector representation of $G_E(N)$; see Sec. 1B.1.
[32] The reverse problem of how to identify an irreducible representation of the abstract group $G_E(N)$ with a physical system can be solved if the physical position, momentum, and angular-momentum operators are defined in the representation space of $G_E(N)$ by their transformation properties under transformations $g = (0, R, \mathbf{u}, \mathbf{v}, 0) \in G(N)$. Because there are generators in $\mathsf{G}_E(N)$ which have just these properties, they can be identified with the physical operators. The remaining generator H_0 is identified from $\mathsf{C}_{H_0}^{(2)}$. For the uniqueness of this construction we refer to Ref. 33.
[33] H. Araki, *Einführung in die Axiomatische Quantenfeldtheorie I* (Zürich, 1962).

interacting and free systems is given in subsections 2A and 2B. Consequences of mass and spin conservation and the identification problem are discussed in 2B and 2C, respectively.

A. Interaction Type and p Limitation

1. Suppose that a dynamical group D exists for a class of interacting systems with Hamiltonians H of the following type:

(α) Let H be a function of P_i, Q_i, $i = 1, \cdots, N$ and of a set of real interaction strengths $\lambda = \{\lambda_1, \cdots, \lambda_p\}$, e.g.,

$$H = \sum_{i=1}^{p} \lambda_i V_i(P_1, \cdots, Q_N),$$

with linearly independent functions V_i, $i = 1, \cdots, p$, such that $H(P_1, \cdots, Q_N, \lambda)$ is essentially self-adjoint and consider the λ dependence of H. In general, different λ will lead to different interactions, and to collect all of them we define the interaction type t of $H(P_1, \cdots, Q_N, \lambda)$ as the set of those linearly independent functions which can be obtained from $H(P_1, \cdots, Q_N, \lambda)$ by letting λ run over all "allowed" values leading to essentially self-adjoint H. It is convenient to introduce a p-dimensional Λ space spanned by the components of $\lambda = \{\lambda_1, \cdots, \lambda_p\}$ and the subset Λ_a of those points which correspond to allowed values of λ. The type t is denoted as upper index and H is written as $H^t[\lambda]$. If

$$t = \{V_i(P_1, \cdots, Q_N), i = 1, \cdots\}$$

contains a finite number of linearly independent functions, then the interaction type is called finite. We restrict ourselves to Hamiltonians of this type, e.g., to

$$H = \sum_{i=1}^{p} \lambda_i V_i(P_1, \cdots, Q_N)$$

with $t = \{V_i(P_1, \cdots, Q_N), i = 1, \cdots, p\}$. The functions V_i, $i = 1, \cdots, p$ can be considered as basis in t and there are points $\lambda^1, \cdots, \lambda^p$ in Λ space with $V_i = H^t[\lambda^i]$.

(β) Let

$$H^0 = \frac{i}{2m} \sum_{i=1}^{N} P_i^2$$

be the Hamiltonian of the free particle and assume that the interaction part in $H^t[\lambda]$ can be turned off by letting $\lambda \to \lambda^0$ with $H^0 = H^t[\lambda^0]$, i.e., that $H^t[\lambda]$ is limitable into H^0. This means in Λ space that there is a curve $\lambda(\epsilon)$ (continuous and continuous differentiable) connecting an arbitrary point $\lambda = \lambda(1) \in \Lambda_a$ with $\lambda^0 = \lambda(0)$ and $H^t[\lambda(\epsilon)] \xrightarrow[\epsilon \to 0]{} H^t[\lambda(0)] = H^0$. As one knows from quantum mechanics, this limitation is discontinuous in respect to the properties of $H^t[\lambda]$; take as example the spectrum of H. Physically, the process may be possible by some experimental device, in the sense of an asymptotic condition or merely as *gedankenexperiment*.

2. The λ and t dependence of the above Hamiltonians is formally taken over to the corresponding dynamical groups $D^t[\lambda]$. The mathematical meaning of this dependence is specified in the following. Since discrete subgroups are not interesting in $D^t[\lambda]$, we pass over to the corresponding Lie algebras $\mathsf{D}^t[\lambda]$ with finite dimension n_D.

It is reasonable that the generator $H^t[\lambda]$ of quantum-mechanical time translations is contained as generator also in $\mathsf{D}^t[\lambda]$,

$$H^t[\lambda] \in \mathsf{D}^t[\lambda], \tag{1}$$

or that a basis exists in $\mathsf{D}^t[\lambda]$ such that $H^t[\lambda]$ in the physical representation (see 2B) is equivalent to one of the generators in $\mathsf{D}^t[\lambda]$. Because a degeneracy group $E^t[\lambda]$ of $H^t[\lambda]$ is directly related to the spectrum of $H^t[\lambda]$, which can be calculated by (1) from the physical representation of $\mathsf{D}^t[\lambda]$, it is plausible that

$$E^t[\lambda] \subset \mathsf{D}^t[\lambda]. \tag{2}$$

Both postulates are fulfilled for the free-particle dynamical group.

3. To utilize the limitation $H^t[\lambda] \to H^0$ we consider the Λ space and the manifold of algebras $\mathsf{D}^t[\lambda]$ over Λ_a with fixed t. $H^t[\lambda^0] = H^0$ implies $\mathsf{D}^t[\lambda^0] \approx \mathsf{G}_E$ for any type t. Because it was supposed that $H^t[\lambda]$, $\lambda \in \Lambda_a$ is limitable and that it possesses a dynamical group $\mathsf{D}^t[\lambda]$, there is also a limitation between $\mathsf{D}^t[\lambda]$ and G_E and a set of dynamical algebras exist, defined on a curve $\lambda(\epsilon) \in \Lambda$, which approaches G_E if $\epsilon \to 0$; $\lambda(\epsilon) \to \lambda(0) = \lambda^0$, i.e.,

$$\mathsf{D}^t[\lambda(\epsilon)] \xrightarrow[\epsilon \to 0]{} \mathsf{G}_E, \quad \lambda(\epsilon) \in \Lambda_a. \tag{3}$$

Postulate (2) implies that all degeneracy algebras $E^t[\lambda]$ which can be obtained by letting λ run over Λ_a are subalgebras of $\mathsf{D}^t[\lambda]$. It was already mentioned that this limitation is not continuous and one expects that $\mathsf{D}^t[\lambda(\epsilon)]$ changes the type, i.e., the constants of structure, and also its dimension n_D along the curve $\lambda(\epsilon)$. However, a curve $\lambda(\epsilon)$ may exist such that there is up to isomorphism of $\mathsf{D}^t[\lambda]$ only a jump for $\epsilon = 0$. We define therefore the following:

The Lie-algebra of a dynamical group describing interactions of finite type t is called *limitable* into the Lie-algebra G_E of the dynamical group of the free particle if a continuous and continuous differentiable curve $\lambda(\epsilon)$ exists in the allowed region of Λ space, spanned by the interaction constants, such

that

$$\mathsf{D}^t[\lambda(\epsilon)] \to \mathsf{D}^t[\lambda(0)] \approx \mathsf{G}_E. \tag{4}$$

$\mathsf{D}^t(\lambda)$ is called *p limitable* into G_E if, furthermore,

$$\mathsf{D}^t[\lambda(\epsilon_1)] \approx \mathsf{D}^t[\lambda(\epsilon_2)]; \quad 0 < \epsilon_1, \epsilon_2 \leq 1. \tag{5}$$

The finite interaction type of $H^t[\lambda]$ is a necessary condition for $\mathsf{D}^t[\lambda]$ being finite-dimensional because there are at least p different generators in $\mathsf{D}^t[\lambda]$ corresponding to H_i, $i = 1, \cdots, p$ and therefore $n_D \geq p$. An example for p-limitable algebras is given in Sec. 5D.1.

4. We add the following remarks:

The definition of limitable or p-limitable algebras does not include a mathematical prescription to perform the limiting process. Therefore we cannot disprove any noncompact group to be a possible dynamical group. An exploration of limitation processes between Lie algebras[34] and between representations of Lie algebras would be useful and is necessary for a rigorous treatment.

A simple formulation of p limitation between two algebras L and L' is a λ dependence of the constants of structure in L, i.e., $ad\mathsf{L}[\lambda] = (ad\mathsf{L})[\lambda]$, as first mentioned by Segal[35] and later used by Inönü and Wigner[36] in a special case. Take a curve $\lambda(\epsilon)$ in the space spanned by the constants of structure of L such that $\mathsf{L}[\lambda(\epsilon_1)] \approx \mathsf{L}[\lambda(\epsilon_2)]$, $0 < \epsilon_1, \epsilon_2 \leq 1$, but $\mathsf{L}[\lambda(\epsilon)] \xrightarrow[\epsilon \to 0]{} \mathsf{L}' \not\approx \mathsf{L}$ and ask for all L' which can be obtained with an ϵ dependence of polynomial type (X_i, $i = 1, \cdots, n_L$ are generators of L)

$$(adX_i)^k_l[\epsilon] = (adX_i)^k_l[1]\epsilon^{m_i+m_l-m_k}$$

with m_i, $i = 1, \cdots, n_L$, real and fixed, the so-called p contraction.[37] The λ, ϵ dependence of the constants of structure can be taken over also to the generators and vice versa. Obviously a p contraction implies a p limitation. The method is restricted to a limitation of Lie algebras with equal dimensions.[38] From some recent results[39] we guess that there is only a small number of algebras not isomorphic to G_E which can be contracted into G_E.

The above limitation is not related to a recently discussed level model[40] based on the observation that the dynamical groups of the free relativistic and nonrelativistic particle and the group structure of classical mechanics are related by a sequence of Inönü–Wigner contractions.

If for two interaction types, with $t_2 \supset t_1$, $\mathsf{D}^{t_2} \supset \mathsf{D}^{t_1}$ also holds, then it suffices to discuss D^{t_2} or in general the maximal type t_M of a class $t_M \supset \cdots \supset t_1$ with $\mathsf{D}^{t_M} \supset \cdots \supset \mathsf{D}^{t_1}$ (see 3D.1–4). An embedding of D^{t_1} in D^{t_2} with $t_2 \supset t_1$ is possible only in special cases. The existence of an algebra for t_1 and for t_2 with $t_1 \cap t_2 \supset H^0$ does not imply an algebra for $t_1 \oplus t_2$.

B. p Limitations for Representations

1. So far, the structure of abstract dynamical algebras $\mathsf{D}^t[\lambda]$ has been analyzed. However, the physical system is described by a representation of $\mathsf{D}^t[\lambda]$ and therefore the limitation between $\mathsf{D}^t[\lambda]$ and G_E must be reformulated as a limitation between their representations. Denote irreducible representations of $\mathsf{D}^t[\lambda]$ and of G_E by $d^t[\lambda; c]$ and $g_E(m, s)$, respectively, with $c = \{c_1, \cdots, c_d\}$ being a set of numbers, say for simplicity of invariants, characterizing the representation.[41] Then the limitation is formally given by

$$d^t[\lambda(\epsilon); C] \xrightarrow[\epsilon \to 0]{} g_E(m, s); \quad m, s; \lambda(\epsilon) \in \Lambda_a \text{ fixed}. \tag{6}$$

On the right-hand side of (6) not only the algebra but also the representation is fixed[42] and one obtains by mass and spin conservation a relation between some of the invariants of $\mathsf{D}^t[\lambda]$ and Casimir operators C and $\mathsf{C}^{(2)}_{[s]}$ of G_E with eigenvalues determined by m and $s = \{s_1, \cdots, s_r\}$. Because both conservation laws are valid for any (skew-adjoint) Hamiltonian, it must be possible to calculate m and s from $d^t[\lambda; c]$ also. The result is independent of $\lambda(\epsilon)$ and we assume that m and s are given by Casimir invariants C^t_m and $\mathsf{C}^t_{[s]}$ of $\mathsf{D}^t[\lambda]$ with constant eigenvalues during the limitation. It is reasonable that C^t_m is limited into C and $\mathsf{C}^t_{[s]}$ into $\mathsf{C}^{(2)}_{[s]}$ or even that $\mathsf{C}^t_m = \mathsf{C}$ and $\mathsf{C}^t_{[s]} = \mathsf{C}^{(2)}_{[s]}$. Therefore from $g_E(m, s)$ at least some of the c_i in c are known, which is denoted by $c(m, s)$, and $d^t[\lambda(\epsilon); c(m, s)]$ can be limited into $g_E(m, s)$ without violating mass and spin conservation.

2. Casimir operators of D^t which are not determined by m and s are in general not fixed during the limitation

[34] M. Gerstenhaber, Ann. Math. **78**, 267 (1963).
[35] J. Segal, Duke Math. J. **18**, 221 (1951).
[36] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953).
[37] H. D. Doebner and O. Melsheimer, Nuovo Cimento **49**, 306 (1967).
[38] To overcome this difficulty a p contraction of L can be used: $\mathsf{L} \to \mathsf{L}' \oplus \mathsf{A}$ with A a $(n_L - n_{L'})$-dimensional Abelian algebra. A necessary and sufficient condition for this process was derived in Ref. 37.
[39] M. Levy-Nahas, J. Math. Phys. **8**, 1211 (1967); R. Hermann, Commun. Math. Phys. **3**, 53 (1966).
[40] M. Flato and D. Sternheimer, J. Math. Phys. **7**, 1932 (1966).
[41] The restriction to irreducible representations and the assumption that the representation can be labeled by Casimir operators is not essential. Note that for reducible ones an embedding $\mathsf{D}^{t'}[\lambda]$ of $\mathsf{D}^t[\lambda]$ with $t' \supset t$ may exist such that $d^{t'}[\lambda'; c']$ is irreducible and that the branching rule $d^{t'}[\lambda'; c'] \downarrow d^t[\lambda; c]$ leads to the desired $d^t[\lambda; c]$ content (see 3D.1).
[42] The limitation between representations is only loosely related to a limitation of the corresponding abstract algebras (see 3D.2).

$c = c(\lambda(\epsilon); m, s)$.[43] Then they depend uniquely on $\lambda(\epsilon)$ because there is only one *physical* representation for given $\lambda(\epsilon) \in \Lambda_a$ characterized by $c_{Ph}(\lambda; m, s)$. The same result can be derived if one starts with $g_E(m, s)$ and if an interaction with finite type t is turned on. Then the system is described uniquely by $d^t[\lambda, c_{Ph}(\lambda; m, s)]$ and one finds the relation $c = c_{Ph}(\lambda; m, s)$.

3. For a more detailed discussion of (6) it is convenient to introduce the space Γ spanned by the components of c. Then $c = c(\lambda(\epsilon); m, s)$ is a curve in Γ and to each curve $\lambda(\epsilon) \in \Lambda$ belongs a curve $c(\epsilon) \in \Gamma$. We define the following:

Let $\mathsf{D}^t[\lambda]$ be limitable or p limitable into G_E on a curve $\lambda(\epsilon)$. Take a fixed representation $g_E(m, s)$. Then $d^t[\lambda; c]$ is called limitable or p limitable into $g_E(m, s)$ if a curve $c = c(\lambda(\epsilon), m, s)$ exists such that

$$d^t[\lambda(\epsilon), c(\lambda(\epsilon); m, s)] \underset{\epsilon \to 0}{\longrightarrow} g_E(m, s); \quad m, s \text{ fixed} \qquad (6')$$

holds and if mass and spin conservation are valid. If the restriction of $d^t[\lambda(\epsilon); c(m, s)]$ by p limitation and by mass and spin conservation is already so strong that only one representation is obtained, then the p limitation is called *unique* and the corresponding value of $c = c_{Ph}(\lambda; m, s)$ is identified with the physical one. This is the case if there is for any $\lambda(\epsilon) \in \Lambda_a$ only one possible curve in Γ space.

The p limitation seems to be a natural property of physical dynamical algebras, for a small variation of the interaction strength $\lambda(\epsilon)$ in a stable system should cause only a small variation of the physical properties which is, in general, the case if the type of the corresponding dynamical group is independent of $\lambda(\epsilon)$.

4. Generally the p limitation is not sufficient to calculate the physical representation $d^t[\lambda; c_{Ph}]$, especially if t depends on $c(m, s)$, i.e., if one abstract algebra describes different interaction types $t^{(\alpha)}$, $\alpha = 1, \cdots$ depending on the irreducible physical representation $d^{t(\alpha)}[\lambda; c^{(\alpha)}(m, s)]$ chosen for $\mathsf{D}^{t(\alpha)} \approx \mathsf{D}^{t(1)}$, $\alpha = 2, \cdots$. There may be a set of curves $\lambda(\epsilon) \in \Lambda_a$ which can be divided into classes, such that the same curve $c(\lambda(\epsilon); m, s) \in \Gamma$ belongs to all elements of the class. Hence one has to identify each class with a different interaction type $t^{(\alpha)}$, $\alpha = 1, \cdots$ or one has to select the physical class of curves. In the example of Sec. 3, fortunately each curve $\lambda(\epsilon)$ belongs to the same type t and furthermore $c_{Ph}(\lambda(\epsilon), m, s)$ is independent of ϵ.

C. Physical Interpretation

1. If an irreducible p-limitable representation is known, the interaction type t must be identified. The components of t are generators of $d^t[\lambda; c_{Ph}(\lambda; m, s)]$ and unbounded operators in $\mathfrak{G}$ with a common dense domain $\mathfrak{D}_d^{an}$ of analytic vectors. Take a representation $g_E^0(m)$ of G_E^0 in a Hilbert space $\mathfrak{K}$ and assume that $\mathfrak{D}_{g_E}^{an} \cap \mathfrak{D}_d^{an} = \mathfrak{D}^{an}$ is dense. The generators P_i, Q_i, $i = 1, \cdots, N$ of $\mathsf{G}_E^0(N)$ form a complete operator system on $\mathfrak{D}^{an}$ in the sense that a reasonable class of operators A with domain $\mathfrak{D}_A$ and $\mathfrak{D}^{an} \cap \mathfrak{D}_A$ dense can be written as functions of P_i, Q_i, $i = 1, \cdots, N$.[44] The generators d_v^t of $d^t[\lambda, c_{Ph}]$ should belong to this class, i.e., we demand that d_v^t can be expressed as function of P_i, Q_i (spinless model). Different representations yield a different P_i, Q_i dependence of d_v^t. The physical interpretation now becomes trivial. The Hamiltonians can be identified as usual and all quantum-mechanical systems in t are completely described. Their properties can be calculated from $d^t[\lambda, c_{Ph}]$ and the approach contains also transition probabilities if there are Hamiltonians in t with bound states and further Hamiltonians which cause transitions between these states. The identification leads to a well-defined t and λ dependence of $D^t[\lambda]$.

2. Three steps are necessary to identify an abstract group L as a dynamical group and to describe all systems t given by $L \approx D^t$.

1. Construct a p limitation between irreducible representations $U_{\text{irr}}(L)$ and $U(G_E)^{[m,0,s]}$ such that mass and spin conservation are related to Casimir invariants in L. A representation theory for L must be known. (The realization of the p limitation depends on L.)
2. Prove that a representation $U_{\text{irr}}(L)$ can be determined by the above p limitation and by mass and spin of the corresponding free particle.
3. Calculate the interaction type.

3. A MODEL FOR p-LIMITABLE DYNAMICAL GROUPS DESCRIBING INTERACTING SYSTEMS

Following our method we construct in Sec. 3A a simple p-limitable algebra D^t and derive in 3C a representation theory for the corresponding group which needs some mathematical properties of D^t which are given in 3B. The physical discussion is contained in 3D. We realize a p limitation and derive the physical representation and the interaction type. Possible generalizations are discussed in 3E.

A. Embedding Theory for $G_E(N)$

1. To construct an algebra L which is p limitable into $\mathsf{G}_E(N)$, we consider probably the simplest case

[43] Also the number of invariants varies in general with λ except for p-limitable representations ($\epsilon \neq 0$).

[44] This connection is used in representation theory on function spaces, e.g., on $L^2(-\infty, +\infty)$ with generators being functions of ix, d/dx.

that L contains $\mathsf{G}_E(N)$ as subalgebra $\mathsf{L} \supset \mathsf{G}_E(N)$ for which a rigorous formulation of a p limitation $\mathsf{L} \to \mathsf{G}_E(N)$ exists. Decompose L and $\mathsf{G}_E(N)$ into a semidirect sum of a solvable ideal R and a semi-simple Levi-factor L_0:

$$\mathsf{L} = \mathsf{R} \dot{+} \mathsf{L}_0$$

respectively (see 1A.1–4);

$$\begin{aligned}\mathsf{G}_E(N) &= G_E^T(N) \ \dot{+}\ so(N) \\ &= [(\mathsf{C} \oplus \mathsf{H}_0 \oplus \mathsf{P}_N) \ \dot{+}\ \mathsf{Q}_N] \ \dot{+}\ so(N).\end{aligned}$$

Then a first classification of algebras $\mathsf{L} \supset \mathsf{G}_E(N)$ is obtained[45] by algebraic calculations using the Malcev–Harish Chandra theorem[46] ($\supset$ denotes proper inclusion and $\cap$ intersection).

Lemma 1: Let L with Levi decomposition $L = \mathsf{R} \ \dot{+}\ \mathsf{L}_0$ be an embedding of $\mathsf{G}_E(N) = \mathsf{G}_E^T(N) \ \dot{+}\ so(N)$. Then the following are the only possible embedding types:

(A) $so(N) \subset \mathsf{L}_0$ and

1. $\mathsf{C} \subset \mathsf{R}$ with $\mathsf{P}_N \cap \mathsf{R} = \{0\}$; $\mathsf{Q}_N \cap \mathsf{R} = \{0\}$; $\mathsf{H}_0 \cap \mathsf{R} = \{0\}$[47]:
2. $\mathsf{C} \oplus \mathsf{P}_N \subset \mathsf{R}$; $\mathsf{Q}_N \cap \mathsf{R} = \{0\}$; $\mathsf{H}_0 \cap \mathsf{R} = \{0\}$;
3. $\mathsf{C} \oplus \mathsf{H}_0 \oplus \mathsf{P}_N \subseteq \mathsf{R}$; $\mathsf{Q}_N \cap \mathsf{R} = \{0\}$;
4. $(\mathsf{C} \oplus \mathsf{P}_N) \ \dot{+}\ \mathsf{Q}_N = \mathsf{G}_E^0(N) \subseteq \mathsf{R}$; H_0 arbitrary;

(B) $G_E(N) \cap \mathsf{R} = \{0\}$.

For further results one needs some kind of minimality condition. We call R minimal if it contains generators from $\mathsf{G}_E^T(N)$ only. To list all embeddings with minimal R we note that the completion of R in A1 and A2 with further generator from $\mathsf{G}_E^T(N)$ yields case A3 or A4 and we are left with 3 types for minimal R:

$$\begin{aligned}\mathsf{R}^1 &= (\mathsf{G} \oplus \mathsf{P}_N) \ \dot{+}\ \mathsf{Q}_N = \mathsf{G}_E^0(N); \\ \mathsf{R}^2 &= (\mathsf{G} \oplus \mathsf{H}_0 \oplus \mathsf{P}_N) \ \dot{+}\ \mathsf{Q}_N = \mathsf{G}_N^T(N); \\ \mathsf{R}^3 &= \mathsf{C} \oplus \mathsf{H}_0 \oplus \mathsf{P}_N.\end{aligned}$$

The corresponding semisimple Levi factor denoted by L_0^i, $i = 1, 2, 3$, is restricted by two conditions:

(A) A semidirect sum between R^i and L_0^i exists.

(B) The following inclusions holds:

$$\begin{aligned}\mathsf{L}_0^1 &\supset \mathsf{H}_0 \oplus so(N); \\ \mathsf{L}_0^2 &\supset so(N); \\ \mathsf{L}_0^3 &\supset \mathsf{Q}_N + so(N).\end{aligned}$$

2. For a discussion of (A) we recall some definitions.[48] Let L_i be Lie algebras over vector spaces R_i, $i = 1, 2$, with generators X and Y, respectively. The automorphism group of L_i is denoted by Aut L_i. It is a subgroup of the group of all nonsingular endomorphisms of L_i. The Lie algebra $\partial(\mathsf{L}_i)$ of Aut L_i consists of all derivations of L_i. The group in-Aut L_i of inner automorphism with Lie algebra $ad(\mathsf{L}_i)$ is an invariant subgroup of Aut L_i and the outer automorphism group out-Aut L_i is given by out-Aut $\mathsf{L}_i \approx$ Aut L_i/in-Aut L_i. If L is simply connected, then Aut L and Aut L have the same Lie algebra.[49]

To define a semidirect sum $\mathsf{L}_1 \ \dot{+}\ \mathsf{L}_2$ between L_1 and L_2, take a homomorphism σ which maps L_2 in $\partial(\mathsf{L}_1)$,

$$Y \in \mathsf{L}_2 \xrightarrow{\sigma} Y_\sigma \in \partial(\mathsf{L}_1),$$

and use for the Lie brackets on the vector space $R_1 + R_2$ the expression

$$\begin{aligned}&[(X + Y), (X' + Y')] \\ &\quad = ([X, X'] + Y_\sigma(X') - Y'_\sigma(X) + [Y, Y']).\end{aligned}$$

Then $\{X + Y\}$ spans $\mathsf{L} = \mathsf{L}_1 \ \dot{+}\ \mathsf{L}_2$. If L_2 can be mapped homomorphic into $\partial(\mathsf{L}_1)$, it is given by $\mathsf{L}_2/\mathsf{N} \subseteq \partial(\mathsf{L}_1)$ with N being ideal in L_2 and kernel of σ. Therefore $[\mathsf{N}, \mathsf{L}_1] = 0$ holds in $\mathsf{L}_1 \ \dot{+}\ \mathsf{L}_2$. A semisimple algebra L_2 can be written as direct sum of simple algebras $\mathsf{L}_2 = \sum_\oplus \mathsf{L}_2^{(\alpha)}$ and any ideal N in L_2 is a direct sum of some of the $\mathsf{L}_2^{(\alpha)}$.

3. Applying this to our model $\mathsf{L}^i = \mathsf{R}^i \ \dot{+}\ \mathsf{L}_0^i$, we have $\mathsf{L}_0^i/\mathsf{N}^i \subseteq \partial(\mathsf{R}^i)$ with N^i being ideal in L_0^i and $[\mathsf{N}^i, \mathsf{R}^i] = 0$. Because L_0^i is semisimple, the solution is given by

$$\mathsf{L}_0^i \subseteq \partial(\mathsf{R}^i) \oplus \mathsf{N}^i.$$

The discussion in Sec. 2A–D4 and condition (B) suggests a "maximality" condition for L_0. If L' and L'' with $\mathsf{L}' \supset \mathsf{L}''$ are candidates for L_0, we choose L''.

The calculation of $\partial(\mathsf{R}^i)$ is straightforward. The first case is solved by Lemma 2; the proof is given in Appendix C.

Lemma 2: Let $\mathsf{G}_E^0(N)$ be the Lie algebra of the central extension of the inhomogeneous pure Galilei group. Denote by in-Aut $\mathsf{G}_E^0(N)$ and out-Aut $\mathsf{G}_E^0(N)$ the inner and outer automorphism group of $\mathsf{G}_E^0(N)$, respectively:

(i) in-Aut $\mathsf{G}_E^0(N)$ is Abelian;

(ii) out-Aut $\mathsf{G}_E^0(N) \approx (Sp(2N, R) \otimes A_\alpha) \ \dot{\times}\ S$)[50].

$Sp(2N, R)$ is the real (noncompact) symplectic group

[45] H. D. Doebner and J. Henning, University of Marburg Report of work prior to publication (1967).

[46] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962).

[47] This implies not P_N, Q_N, $H_0 \subset \mathsf{L}_0$ because the projection part of P_N or Q_N in R has to be nontrivial.

[48] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962).

[49] C. Chevalley, *Theory of Lie groups* (Princeton University Press, Princeton, N.J. 1946).

[50] The nondiscrete outer automorphism group is a result of the nilpotency of $\mathsf{G}_E^0(N)$, Ref. 47.

in $2N$-dimensions. A_α is Abelian and isomorphic to the multiplicative group of real positive numbers. S is a discrete group with two elements.

$\partial(\mathsf{G}^0_E(N))$ is independent of the discrete subgroup in-Aut G^0_E. Because L^1_0 is semisimple, a subalgebra of L^1_0 cannot be mapped into the Abelian parts in-Aut G^0_E and A_α of $\partial(\mathsf{G}^0_E(N))$ and the maximal L^1_0 is given (uniquely) by

$$\mathsf{L}^1_0 = sp(2N, R).$$

Condition (B) is obviously fulfilled for $SO(N) \subset Sp(2N, R)$. Hence the solution for case A4 with minimal R and maximal L_0 is

$$\mathsf{L}_1 = (\mathsf{G}^0_E(N) \dot{+} sp(2N, R)) \oplus \mathsf{N}^1.$$

The direct factor N^1 implies physical results independent of those derived from $\mathsf{G}^0_E(N) \dot{+} sp(2N, R)$. It can be used to describe spin properties of interacting systems. In the following, N^1 is dropped and we choose case A4 with minimal R as candidate for a dynamical algebra D^t:

$$\mathsf{G}^0_E(N) \dot{+} sp(2N, R) = \mathsf{D}^t.$$

The semidirect sum is uniquely defined.[51] For R^2 and R^3, analogous calculations can be done.[45] The structure of L with R being nonminimal is unknown and the situation is similar to that for an embedding of the Poincaré algebra.[40,52]

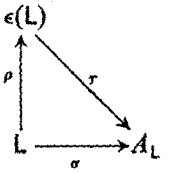
4. To derive a similar result for an embedding of the Galilei group $G_E(N)$, we note that $G^0_E(N)$ is simply connected. Hence the Lie algebra of Aut $G^0_E(N)$ is isomorphic to $\partial(\mathsf{G}^0_E(N))$ (see 3A–D) and the construction of an embedding L of $G_E(N)$ is equivalent to constructing an embedding of the corresponding Lie algebras if discrete subgroups are neglected. A case analogous to A3 leads to the unique semidirect product:

$$L^1 = G^0_E(N) \dot{\times} Sp(2N, R).$$

B. The Enveloping Algebra of the Central Extension $G^0_E(N)$ of the Inhomogeneous Pure Galilei Algebra

1. For a representation theory of D^t some properties of the enveloping algebra of $\mathsf{G}^0_E(N)$ are needed. Intuitively, the universal enveloping algebra $\epsilon(\mathsf{L})$ of a Lie algebra L is defined as a set of all polynomials built from generators of L if one identifies elements which are equal by commutation relations in L. The exact definition is the following[46]:

Let A be an associative algebra; define for $a_1, a_2 \in A$ a Lie product as $[a_1, a_2] = a_1a_2 - a_2a_1$ and obtain the infinite-dimensional Lie algebra A_L of the associative algebra A. Consider algebras $\epsilon(\mathsf{L})$ for which a homomorphism ρ exists mapping L into $\epsilon(\mathsf{L})$. We denote the homomorphism as lower index in $\epsilon_\rho(\mathsf{L})$. Take now any further associative algebra A and let σ be the homomorphism of L into A_L. Then $\epsilon_\rho(\mathsf{L})$ is defined as universal enveloping algebra of L if there exists a unique homomorphism r of $\epsilon_\rho(\mathsf{L})$ into A_L such that $\sigma = \rho \cdot r$. Drawing a diagram, this means[53] (the index ρ is suppressed in the following):



2. For the discussion of $\epsilon(\mathsf{G}^0_E(N))$ it is assumed that the center in G^0_E is represented by $im\,1$, thus restricting the results to quasi-irreducible representations of G^0_E. Consider subalgebras of $\epsilon(\mathsf{G}^0_E(N))$. Denote by $\epsilon^{(n)}_c$ and $\epsilon^{(n)}$ the set of skew-symmetric nth-order polynomials of all generators in $\mathsf{G}^0_E(N)$ ($\mathsf{C} = im \cdot 1$) and of P_N and Q_N, respectively. Obviously $\epsilon^{(n-1)}_c \subset \epsilon^{(n)}_c$ holds. We look for Lie algebras contained in a finite sum of $\epsilon^{(n)}_c$ and $\epsilon^{(n)}$, respectively. A trivial case is $\epsilon^{(1)}_c = \mathsf{G}^0_E(N)$, and $\epsilon^{(2)}_c$ and $\epsilon^{(2)}$ are further examples which are classified by:

Lemma 3: Let $\epsilon^{(2)}$ and $\epsilon^{(2)}_c$ be the set of all skew-symmetric second-order polynomials of P_i, Q_i and P_i, Q_i, C, $i = 1, \cdots, N$, respectively:
(i) $\epsilon^{(2)} \approx sp(2N, \mathsf{R})$;
(ii) $\epsilon^{(2)}_c \approx G^0_E(N) \dot{+} sp(2N, R)$.
The result is partly known (see, e.g., Ref. 54). We give a short proof using Lemma 2. It is straightforward to check that $\epsilon^{(2)}$ is a Lie algebra. Because there are $(2N^2 + N)$ linearly independent second-order polynomials of P_i, Q_i, $i = 1, \cdots, N$, its dimension is $(2N^2 + N)$. Commutators between $\epsilon^{(2)}$ and $\epsilon^{(1)}_c$ yield linear combinations of P_i, Q_i. Hence $\epsilon^{(2)}_c = \epsilon^{(1)}_c \dot{+} \epsilon^{(2)}$ and there exists a homomorphism σ mapping $\epsilon^{(2)}$ into $\partial(\mathsf{G}^0_E(N))$. Take $Y \in \epsilon^{(2)} \xrightarrow{\sigma} Y_\sigma \in \partial(\mathsf{G}^0_E(N))$. Y_σ can be decomposed in parts, Y^{in}_σ, Y^{out}_σ, lying in the Lie algebra of the inner and outer automorphism group, respectively. To derive $\epsilon^{(2)} \approx sp(2N, R)$ it suffices to discuss σ. A first property of σ is the absence of a projection part Y^{in}_σ in $ad\,\mathsf{G}^0_E(N)$. We calculate with $X' \in \mathsf{G}^0_E(N)$:

$$[Y, X'] = Y_\sigma(X') = Y^{\text{in}}_\sigma(X') + Y^{\text{out}}_\sigma(X')$$

[51] Unique up to inner automorphism of $sp(2N, R)$.
[52] L. O'Raifeartaigh, Phys. Rev. **139**, B1052 (1965).
[53] Let T be the infinite tensor algebra of L and K the ideal generated by all $[l_1, l_2] - l_1 \cdot l_2 + l_2 \cdot l_1$; $l_1, l_2 \in \mathsf{L}$. Then T/K is a realization of $\epsilon(\mathsf{L})$.
[54] H. Lipkin, *Lie Groups for Pedestrians* (North-Holland Publ. Co., Amsterdam, 1965).

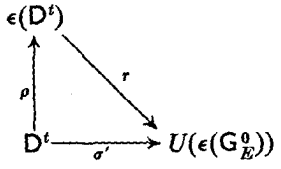
(see 3A–B) and suppose that $Y_\sigma^{\text{in}} \neq 0$. Then we choose X' such that $Y_\sigma^{\text{in}}(X') = C$, which is possible and in contradiction to the result, that commutators between $\epsilon_c^{(1)}$ and $\epsilon^{(2)}$ will never lead to a linear combination containing C. $\epsilon^{(2)}$ possesses no one-dimensional invariant subalgebra l_0, since $l_0 = \sum_i \alpha_i(P_i^2 + Q_i^2) + \sum_{ij} \beta_{ij} P_i Q_j + \sum_i \gamma_i(P_i Q_i + Q_i P_i)$ implies together with $[l_0, P_i] = [l_0, Q_i] = 0$ that $l_0 = 0$ This proves that the kernel of σ also is zero and we have from Lemma 2 that $\epsilon^{(2)}$ is mapped isomorphically into $sp(2N, R)$.

Special cases of this lemma are frequently used (e.g. Refs. 9, 54, 55), also examples with $N = 1, 2$ and the sometimes-misleading accidental isomorphism $sp(2, R) \approx so(2, 1)$, $sp(4, R) \approx so(3, 2)$. In general, only subgroups of $sp(2N, R)$ can be expressed as skew-symmetric second-order polynomials of P_i, Q_i or of formal creation and annihilation operators $a_j = (i/\sqrt{2})(P_j + iQ_j)$, $a_j = (i/\sqrt{2})(P_j - iQ_j)$. By Lemma 2 the semidirect coupling between $\mathsf{G}_E^0(N)$ and subalgebras of $sp(2N, R)$ is unique. Some examples are given in Appendix D. Note that $\epsilon_c^{(2)}$ is isomorphic to our choice of dynamical algebra D^t.

3. We apply the homomorphism σ mapping $sp(2N, R)$ into $\epsilon_c^{(2)}$. It can be extended to a homomorphism of D^t into $\epsilon_c^{(2)}$. Take a unitary quasi-irreducible representation $U(G_E^0)^{[m,n(m)]}$; denote by $U(\mathsf{G}_E^0) = g_E^0$[56] the image of G_E^0 under U, i.e., the representation of the Lie algebra derived via Stone's theorem. Because g_E^0 is integrable, the image of $\epsilon(\mathsf{G}_E^0)$ under U is a representation $U(\epsilon(\mathsf{G}_E^0))$ of $\epsilon(\mathsf{G}_E^0)$. Hence one has a homomorphism σ' of D^t into $U(\epsilon(\mathsf{G}_E^0))$:

$$\mathsf{D}^t \xrightarrow[\sigma']{} U(\epsilon(\mathsf{G}_E^0)).$$

In the defining diagram of ϵ we identify L with D^t and A, which was any algebra, with $U(\epsilon(\mathsf{G}_E^0))$. Then the following diagram holds:



From the definition of ϵ, we conclude that a unique homomorphism r of $\epsilon(\mathsf{D}^t)$ into $U(\epsilon(\mathsf{G}_E^0))$ exists and that $\sigma' = \rho r$. To derive properties of r from σ' consider $X \in \epsilon(\mathsf{G}_E^0) \subset \epsilon(\mathsf{D}^t)$. Because $\sigma' = \rho r$, we have $X_r = U(X)$ and X_r is skew-symmetric for $X \in \mathsf{D}^t$. The result is collected in the following lemma:

Lemma 4: Let $\epsilon(\mathsf{D}^t)$ and $\epsilon(\mathsf{G}_E^0)$ be the universal enveloping algebra of $\mathsf{D}^t = \mathsf{G}_E^0 \dot{+} sp(2N, R)$ and G_E^0, respectively. Let $U(\epsilon(\mathsf{G}_E^0))^{[m,n(m)]}$ be the image of $\epsilon(\mathsf{G}_E^0)$ under a unitary quasi-irreducible representation of G_E^0. Then there exists a homomorphism

$$\epsilon(D^t) \xrightarrow[r]{} U(\epsilon(\mathsf{G}_E^0))^{[m,n(m)]}$$

with

(α) $X_r = U(X)$ for $X \in \epsilon(\mathsf{G}_E^0)$,

(β) X_r skew-symmetric for $X \in \mathsf{D}^t$.

C. Representation Theory for $D^t = G_E^0(N) \textcircled{\times} Sp(2N, R)$

1. We are now prepared for a representation theory for D^t. A generalization of Mackey's theory of induced representations for nonregular semidirect products with nilpotent ideal seems to be too involved for our purposes. Therefore we present a simpler method, which is based on properties of the enveloping algebra of G_E^0. The idea is that the representations g_E^0 of G_E^0 are known by Theorem 1 and that by Lemma 3 $sp(2N, R)$ can be mapped into $U(\epsilon(\mathsf{G}_E^0))$. Therefore a representation of $\mathsf{D}^t \subset U(\epsilon(\mathsf{G}_E^0))$ can be constructed. If this representation is integrable to a unitary strongly continuous representation of the universal covering group $\bar{D}^t$ of D^t, and if a reasonable class of representations of $\bar{D}^t$ can be obtained in this way, the problem would be solved. $\bar{D}^t$ is an example of a group in which certain irreducible representations of an ideal already determines the representation of the whole group, and it is promising for $\bar{D}^t$ as dynamical group that the representations of this ideal are determined by the mass of the free particle. The Casimir invariants of D^t support this idea.

2. To realize the method, we use a theorem of Nelson[57] to construct representations of $\bar{D}^t$ and D^t.

Extension Theorem (Nelson)

(1) Let L be a simply connected Lie group with algebra L. Let L_1 be ideal in L and L_1 be the simply connected subgroup of L with Lie algebra L_1 and with a strongly continuous unitary representation $U(L_1)$ in $\mathcal{H}$. Denote by $\epsilon(\mathsf{L})$, $\epsilon(\mathsf{L}_1)$ the universal enveloping algebras of L, L_1, respectively, and by $U(\epsilon(\mathsf{L}_1))$ the image of $\epsilon(\mathsf{L}_1)$ under U.

(2) Let σ be a homomorphism

$$\epsilon(\mathsf{L}) \xrightarrow[\sigma]{} U(\epsilon(\mathsf{L}_1))$$

with

(α) $X_\sigma = U(X)$ for all $X \in \epsilon(\mathsf{L}_1)$,

(β) X_σ skew symmetric for all $X \in \mathsf{L}$.

[55] R. Santilli, Nuovo Cimento **51A**, 74 (1967).

[56] The upper index $[m, n(m)]$ is dropped.

[57] E. Nelson, Ann. Math. **70**, 3 (1959).

Then there exists in $\mathcal{H}$ a unique strongly continuous unitary representation $\tilde{U}(L)$ of L such that $\tilde{U}(X) = U(X)$ for all $X \in \mathsf{L}_1$. $\tilde{U}(L)$ is called Nelson extension of $U(L_1)$.[58]

If σ exists, the theorem states that it is possible to extend the representation of the ideal L_1 to a representation $U(L)$. The restriction of $U(L)$ to $U(L_1)$ leads back to the representation which started the construction. For the application we identify $\bar{D}^t$ with L, G_E^0 with L_1. By Lemma 4 the homomorphism σ with properties (α) and (β) exist and we have:

Theorem 2: Let $G_E^0(N)$ be the central extension of the inhomogeneous pure N-dimensional Galilei group and D^t be the semidirect product between $G_E^0(N)$ and $Sp(2N, R)$. Let $U(G_E^0(N))$ be a strongly continuous unitary representation of $G_E^0(N)$. Then $U(G_E^0(N))$ can be uniquely extended to a strongly continuous unitary representation $\tilde{U}(\bar{D}^t)$ of $\bar{D}^t$. The restriction of $\tilde{U}(\bar{D}^t)$ to $G_E^0(N)$ is given by $U(G_E^0(N))$.

3. For representations $\tilde{U}(\bar{D}_t)^{[m]}$ [59] constructed from irreducible representations $U(G_E^0(N))^{[m]}$ the result can be sharpened.

$\tilde{U}(\bar{D}^t)^{[m]}$ is irreducible. There is by definition no subspace invariant under $U(G_E^0)$ and hence also no subspace invariant under $\tilde{U}(\bar{D}^t)^{[m]}$.

$\tilde{U}(\bar{D}^t)^{[m]}$ is univalent. Let $a = \{a_1, \cdots, a_\rho\}$ be a parametrization of $Sp(2N, R)$. Denote by $\tilde{U}(Sp(2N, R)$ the restriction of $\tilde{U}(\bar{D}^t)^{[m]}$ to $Sp(2N, R)$ and consider $V(a) \in \tilde{U}(Sp(2N, R))$. For $Sp(2N, R) \subset$ out-Aut $G_E^0(N)$ an automorphism in $\mathsf{G}_E^0(N)$ is given by $V(a)$, i.e.,

$$V(a)P_iV^*(a) = \sum_j (b_{ij}^1 P_j + c_{ij}^1 Q_j),$$
$$V(a)Q_iV^*(a) = \sum_j (b_{ij}^2 P_j + c_{ij}^2 Q_j).$$

Now, $\tilde{U}(\bar{D}^t)^{[m]}$ would be univalent if there is for each automorphism (i.e., fixed $b_{ij}^{1,2}$, $c_{ij}^{1,2}$) only one operator of the form $V(a)$. Let V_1, V_2 be two such operators leading to the same automorphism. Then $sp(2N, R) \subset \epsilon^{(2)}$ implies that $V_1^{-1}V_2$ commutes with $U(\mathsf{G}_E^0)^{[m]}$, which is irreducible. Therefore $V_1^{-1}V_2 = e^{i\alpha}1$ and even $\alpha = 0$. Then the univalence is a consequence of the isomorphism between the set

$$\left\{V(\mathrm{a}) = \exp\left(\sum_v a_v \mathsf{A}_v\right), \mathsf{A}_v \in \epsilon^{(2)}\right\}$$

and the connected group $Sp(2N, R)$.

[58] Nelson extensions are denoted by $\tilde{U}(\cdot)$ in the following sections.

[59] The existence of irreducible representation of D^t which are not Nelson extensions of $U(G_E^0)^{[m]}$ will be discussed in a subsequent paper because this question is directly related to the construction of a model containing spin, Ref. 60.

[60] H. D. Doebner and O. Melsheimer, University of Marburg report prior to publication (1967).

The restriction of $\tilde{U}(\overline{D^t})^{[m]}$ to G_E is irreducible and, using Theorem 1, is given by $\tilde{U}(G_E)^{[m,0,s=0]}$. Any representation $U(G_E)^{[m,0,s=0]}$ can be enlarged uniquely to an irreducible representation of $\bar{D}^t$.

Let $U(D^t)^{[m]}$ be any unitary representation such that $U(G_E^0)^{[m]}$ is the restriction to G_E^0. Then we conclude from the preceding results that $U(D^t)^{[m]}$ is also a Nelson extension of $U(G_E^0)^{[m]}$ and therefore irreducible and univalent. Suppose that two representations $U_i(D^t)^{[m]}$, $i = 1, 2$, exist, being not unitarily equivalent, i.e., with different restrictions $U_i(Sp(2N, R))$, $i = 1, 2$, and with equivalent $U_i(G_E^0)$, $i = 1, 2$, which are connected by a unitary operator W:

$$W \cdot U_1(\mathsf{G}_E^0)^{[m]} \cdot W^* = U_2(\mathsf{G}_E^0)^{[m]}.$$

Take $g_{i\mu} \in U_i(\mathsf{G}_E^0)^{[m]}$ and $V_i(a) \in U_i(Sp(2N, R))$, $i = 1, 2$. Then, as above,

$$V_1(a) \cdot g_{1\mu} \cdot V_1^*(a) = W \cdot V_2(a)g_{2\mu} \cdot V_2^*(a) \cdot W^*$$

holds and the irreducibility of $U_i(G_E^0)^{[m]}$ implies

$$V_1(a) = e^{i\alpha}WV_2(a)W^* \quad \text{for all } V_i(a), \quad i = 1, 2,$$

in contradiction to our assumption. The statement that $U(D^t)^{[m]}$ is uniquely given by the Nelson extension is useful for later applications.

4. Representations $U(\overline{D^t})^{[m,n(m)]}$ constructed from quasi-irreducible $U(G_E^0)^{[n,n(m)]}$ will not be discussed here, because there exists no (irreducible) representation $U(\overline{D^t})$ with restriction $U(G_E^0)^{[m,n(m)]}$ $m > 0$, $n(m) < \infty$, such that the restriction to G_E is given by $U(G_E)^{[m;0;s]}$ with $d(s) = n(m)$ (see 1B.2).[60] Therefore D^t is unsuitable as dynamical groups for systems including spin.

D. $D^t \approx G_E^0(N) \textcircled{\times} Sp(2N, R)$ as p-Limitable Dynamical Group

The results of the preceding section are sufficient to derive the interaction type t as well as the physical representation of D^t. A detailed explanation is given in 3D.1 for one-dimensional systems. The N-dimensional case is solved in 3D.2.

1. One-dimensional Systems

1. The dynamical group for $N = 1$ is

$$D^t \approx G_E^0(1) \textcircled{\times} Sp(2, R) \approx G_E^0(1) \textcircled{\times} SO(2, 1).$$

It contains the free-particle group $G_E(1)$ by construction uniquely. To realize a limitation

$$\mathsf{D}^t \to \mathsf{G}_E(1)$$

we consider p contractions (see 2A–D). The number of generators in D^t and $\mathsf{G}_E(1)$ is different. Hence a p

contraction is at best a limitation of D^t into $\mathsf{G}_E \oplus \mathsf{A}_2$ with A_2 being a two-dimensional Abelian algebra. The physical interpretation of $\mathsf{G}_E(1)$ is independent of the direct factor A_2.[61] A suitable basis in $D^t\{P, Q, C; A_i = i = 1, 2, 3\}$ is given in Appendix D. We introduce a λ dependence in D^t by a nonsingular transformation of $sp(2, R)$ depending on $\lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ and acting as $A_i = \lambda_i A_i'$. Then $\mathsf{D}^t \approx \mathsf{D}^t[\lambda^1] \approx \mathsf{D}^t[\lambda^2]$ for $\lambda^1 \neq \lambda^2$ holds. The Λ space is three-dimensional. Take a curve $\lambda'(\epsilon) = \{\epsilon^{m_1}, \epsilon^{m_2}, \epsilon^{m_3}\} \in \Lambda$ with $m_1 = 0, m_2 = -2, m_3 = -1$, i.e., $\lambda_2' = \lambda_3'^2$, $\lambda_1' = 1$, and perform a p contraction along this curve by letting $\epsilon \to 0$. The result is

$$\mathsf{D}^t[\lambda'(\epsilon)] \xrightarrow[\epsilon \to 0]{} \mathsf{G}_E(1) \oplus \mathsf{A}_2,$$

which proves that the algebras are related via a p contraction which is not of Wigner–Inönü type. The curve $\lambda'(\epsilon)$ is not unique.

2. We now analyze p contractions between representations $U(D^t[\lambda])_{\text{irr}} = d^t[\lambda, c]$:

$$d^t[\lambda, c] \to U(\mathsf{G}_E(1))^{[m',0]} = g_E(m').$$

A spin-index s is superfluous for $N = 1$. The representation of the Abelian algebra is omitted. The Γ space is two-dimensional. Consider $U(D^t[\lambda])_{\text{irr}}$. Because $G_E(1)$ is not changed by the contraction, we identify $\mathsf{C} \in U(\mathsf{G}_E(1)) \subset U(\mathsf{D}^t)$ as mass operator of the interacting system with eigenvalue $+im$. By mass conservation, m is constant during the contraction and hence $m = m'$. A candidate for the physical representation is the Nelson extension $\tilde{U}(D^t[\lambda])^{[m']}$ of $U(G_E^0(1))^{[m',0]}$; obviously $\mathsf{C} = im' \cdot 1$ holds and a p contraction into $g_E(m')$ along the curve $\lambda'(\epsilon)$ also exists. There are no further irreducible representations with this property. Take $U(D^t)'_{\text{irr}}$ with $C = +im1$ not being Nelson extension (if possible). Because the restriction $U(G_E)$ of $U(D^t)_{\text{irr}}$ to G_E is irreducible only for $U(D^t)^{[m]}$, the contraction gives at least two equivalent representations $g_E(m)$, in contradiction to the postulate that the limitation leads to one representation only. Hence these $U(D^t)_{\text{irr}}$ are not limitable. The physical representation is independent of $\lambda(\epsilon)$.

3. The interaction type t is easily calculated. Take the physical representation with $m = m'$; remember that $sp(2, R) \approx \epsilon^{(2)}$. The solution is[62]

$$A_1 = (i/2m')P^2;$$
$$A_2 = (i/2m')Q^2;$$
$$A_3 = (i/2m')(PQ + QP).$$

Since all generators in D^t are interaction terms, the general Hamiltonian reads

$$iH^t[\lambda] = (\lambda_1/2m')P^2 + (\lambda_2/2m')Q^2 + (\lambda_3/2m')(PQ + QP)$$

or

$$t = \{P^2, Q^2, PQ + QP\},$$

and the group leads to spectra and eigenfunctions for one-dimensional systems with harmonic force and damping.[63] The abstract-defined λ turns out to have the meaning of a physical interaction strength and the p contraction implies switching off the harmonic force and the damping; only the free Hamiltonian $H^0 = H^t[\lambda^0]$, $\lambda^0 = \{1, 0, 0\}$, remains. Because each $H^t[\lambda]$ generates a one-dimensional subgroup in D^t, the limitation

$$H^t[\lambda] \to H^0$$

is performed by "running over a set of one-dimensional subgroups" uniquely coupled to $G_E^0(1)$. This is the group-theoretical translation of the limitation between $H^t[\lambda]$ and H^0 in quantum mechanics (see 2A.1).

4. A subalgebra $\mathsf{D}_{\text{osc}} \subset \mathsf{D}^t$ spanned by $\{P, Q, C, A_1 + A_2\}$ is the so-called oscillator algebra[64] and corresponds to $t' = \{P^2 + Q^2\} \subset t$. D_{osc} is also p limitable into $\mathsf{G}_E(1)$. Its physical representation can be determined by, e.g., embedding D_{osc} in D_t (see 2A.4). The subalgebra $sp(2, R)$ was used as spectrum-generating algebra for the energy levels of the oscillator,[9–12] which can be arranged in two irreducible $Sp(2, R)$ representations. The embedding of $sp(2, R)$ in D^t yields a description of all levels in one representation (see 2B.1). The restriction of $U(D^t)^{[m]}$ to $Sp(2, R)$ is reducible.[11]

5. For $N = 1$ the problem is solved. The approach was based on the abstract group D^t and on the information that the mass of the corresponding free particle is m'. Using our "principles" for synamical groups, we have the result as follows:

(1) Denote by $\xrightarrow[p]{}$ a p limitation or physically a switching off of the interaction. Then for algebras (the Abelian factor is omitted)

$$\mathsf{D}^t \approx \mathsf{G}_E^0(1) \dot{+} Sp(2, R) \xrightarrow[p]{} \mathsf{G}_E(1)$$

is valid, but for representations

$$U(D^t)_{\text{irr}} \xrightarrow[p]{} U(G_E(1))^{[m,0]}$$

is possible if $U(D^t)_{\text{irr}}$ is the Nelson extension of $U(G_E^0)^{[m]}$.

[61] The Abelian factor can be avoided if a limitation between representations is used.

[62] The result is unique up to unitary equivalence.

[63] In principle, the generator Q can also be added in t describing a constant force. However, switching off the interaction term μQ is involved because $G_E(1)$, has also to be transformed. The same situation already appears in $G_E(1)$, which describes not only the free particle, but also those moving under a constant force.

[64] R. F. Streater, Commun. Math. Phys. **4**, 217 (1967).

(2) p-contraction and mass conservation determine the physical representation $\tilde{U}(D^t)^{[m]}$. D^t describes one-dimensional quantum-mechanical systems with interaction type $t = \{P^2, Q^2, PQ + QP\}$. D^t is a unique p-limitable dynamical group.

2. N-Dimensional Systems

1. The discussion of D^t for arbitrary N is in principle analogous to the procedure for $N = 1$. $G_E(N)$ is uniquely contained in

$$D^t \approx G_E^0(N) \textcircled{\times} Sp(2N, R),$$

but a limitation between D^t and $\mathsf{G}_E(N)$ is formally complicated. Already in the two-dimensional case a p contraction connecting D^t and $G_E(N) \oplus \mathsf{A}_9$, with A_9 being a nine-dimensional Abelian algebra, fails, which can be proved by Lemma 1 of Ref. 37. However, the restriction of an algebra to a fixed subalgebra is mathematically well defined and we use this restriction as a nonparameter-dependent description of a limitation between D^t and $\mathsf{G}_E(N) \subset \mathsf{D}^t$, which is obviously a p limitation (p limation by restriction). A formulation of this process via a parameter dependence similar to a p contraction is at least possible in representations $U(D^t)$, e.g., if a limitation exists in $U(D^t)$ such that all matrix elements of generators $d_\nu \notin \mathsf{G}_E(N)$ vanish.[60]

2. To derive the physical representation $d^t[c_{Ph}; m, s]$ of D^t we restrict ourselves to spinless ($s' = 0$) free particles with mass m'. This implies[65]

$$d^t[c_{Ph}; m', s' = 0] \to U(G_E(N))^{[m,0,s=0]}.$$

Because $\mathsf{G}_E(N)$ is not changed by the limitation, we identify the center in D^t as mass operator and we have $m = m'$ by mass conservation. By the same arguments (see 2B.1) we identify those Casimir operators for D^t, which are also Casimir operators for $\mathsf{G}_E(N)$ (except C) with operators describing the spin-content and their eigenvalues are known by spin conservation, i.e., $s = s'$ and $d(s) = 1$. From the results in 3D.3 we find that the irreducible representation $\tilde{U}(D^t)^{[m']}$ constructed as Nelson extension of $U(G_E^0(N))^{[m]}$ has the desired properties, i.e., $C = im' \cdot 1$ and $d(s') = 1$. The restriction to $G_E(N)$ leads to $U(G_E(N))^{[m',0,s'=0]}$ and the representation is p limitable. Any other irreducible representation (if it exists) with $\mathsf{C} = im' \cdot 1$ is not p limitable by restriction (see the discussion for $N = 1$). Therefore the physical representation $d^t[c_{Ph}; m', s' = 0]$[65] is given by the Nelson extension $\tilde{U}(D^t)^{[m]}$ of $U(G_E^0(N))^{[m,0,s=0]}$. The p limitation is unique.

[65] The λ dependence in $d^t[c_{Ph}]$ is dropped. An explicit determination of c_{Ph} is not necessary; it can be derived using the results in Ref. 11.

3. Deriving the interaction type t from $m = m'$ and $sp(2N, R) \approx \epsilon^{(2)}$, we find the $(2N^2 + N)$ generators of $sp(2N, R)$ as linear combinations of the $(2N^2 + N)$ linearly independent, skew-symmetric, second-order polynomials in P_i, Q_i, $i = 1, \cdots, N$, and the interaction type contains at most

$$t = \{V_i, i = 1, \cdots, 2N^2 + N\}$$
$$= \{P_iP_j; Q_iQ_j; P_iQ_j + Q_jP_i; i, j = 1, \cdots, N\}.$$

The Hamiltonians $H^t[\lambda]$ are linear combinations of some V_i, $H^t[\lambda] = \sum_i \lambda_i V_i$. The group describes N-dimensional isotropic and anisotropic oscillators with all possible types of linear damping or N-dimensional systems with second-order polynomials in P_i, Q_i as Hamiltonians. The limitation $H^t[\lambda] \to H^0$ is performed by running over one-dimensional subgroups uniquely coupled to $G_E^0(N)$.

4. We check now whether the degeneracy groups $E^t[\lambda]$ for $H^t[\lambda]$ in t are contained in D^t. As $E^t[\lambda]$ is compact, we determine the maximal compact subgroup of D^t which is given by that of $Sp(2N, R)$. The result is partly known[10]; a short proof is given in Appendix D.

Lemma 5: The maximal compact subgroup of $Sp(2N, R)$ is $U(N)$. The $2N$-dimensional real representation of $U(N)$ in $Sp(SN, R)$ is unique up to inner automorphism in $Sp(2N, R)$.

The degeneracy groups $SO(N)$ and $SU(N)$ for the free particle and for the isotropic oscillator, respectively, are subgroups of $U(N)$ as well as the degeneracy group for all further systems in t. Hence our postulate $E^t[\lambda] \subset D^t$ is automatically fulfilled. The generators of $SU(N)$ are written formally as interaction terms in t.

One system, the isotropic oscillator, is uniquely characterized by its degeneracy group. The equation $[\hat{d}, SU(N)] = 0$, $\hat{d} \in \mathsf{D}^t$ possesses only the solution $\hat{d} = H_{\text{osc}} + \alpha 1$ or, by changing the energy scale, $\hat{d} = H_{\text{osc}}$. Furthermore D_{osc} defined by $\mathsf{D}_{\text{osc}} \approx \mathsf{G}_E^0(N) \textcircled{+} (su(N) \oplus H_{\text{osc}})$ is a subalgebra of D^t and $\mathsf{D}_{\text{osc}} \subset \mathsf{D}^t$ is the N-dimensional oscillator algebra. D_{osc} is uniquely determined by the degeneracy group $SU(N)$ (see Lemma 5). We emphasize that systems without degeneracy are also described by D^t. The degeneracy is accidental in this approach, as it is in quantum mechanics.

5. Collecting the results, we have shown that D^t is a unique p-limitable dynamical group for spinless quantum-mechanical systems with Hamiltonians being second-order polynomials in P_i, Q_i, $i = 1, \cdots, N$. The p limitation is constructed by restriction and determines, together with mass and spin conservation

and the mass of the corresponding free particle m', the physical representation which is given by the Nelson extension of $U(G_E^0(N))^{[m']}$.

E. Construction of Further Models

1. With $D^t \approx G_E^0(N) \textcircled{\times} Sp(2N, R)$ a reasonably pure group-theoretical description exists for some spinless systems. The model can be generalized to particles with spin by decomposing the angular momentum into an orbital part and a spin part, which corresponds to a dynamical group being the direct product of $SO(N)$ and D^t defined above. Spin representations of this enlarged dynamical group can be constructed and the determination of the physical representation is similar to that for the spinless case. A detailed treatment of particles with spin will be given elsewhere.[60] The extension to n-particle systems with forces of the type t discussed is possible. External forces can also be included. To show this, decompose the physical representation of the free n-particle system $U_1(G_E(N)) \otimes \cdots \otimes U_n(G_E(N))$ in irreducible representations of $G_E(N)$,[22-24], which is equivalent to splitting the motion into the center-of-mass motion and in the relative motion, and apply the above approach to both parts.

2. Further quantum-mechanical examples with limitable dynamical groups are not known.[66] A systematic construction of models would need some information on physically motivated limitation processes between group representations. Perhaps further systems with limitable dynamical groups do not exist in quantum mechanics if only finite-dimensional Lie algebras are included.

3. We remark that an approach using infinite-dimensional algebras is equivalent to the construction of a field theory and all difficulties of field theory and particle physics will appear and there are no obvious physical principles as, e.g., causality or field equations to restrict the algebraic arbitrariness and to determine the physical representation.

ACKNOWLEDGMENTS

One of the authors (H.D.D.) would like to thank Professor A. Salam and Professor P. Budini and the IAEA for the kind hospitality extended to him at the International Center for Theoretical Physics in Trieste. The authors are indebted to Professor A. O. Barut and Dr. A. Böhm for stimulating their interest in this problem and to Dr. G. C. Hegerfeldt for many valuable discussions. It is a pleasure to thank Dr. J. Henning, Dr. A. Hartkämper, Dr. H. Neumann, Dr. J. Niederle, Professor G. Ludwig, and Professor R. Raczka for interesting discussions during various stages of the work.

[66] For relativistic systems the situation is similar in spite of the fact that there is no identification problem.

APPENDIX A

A classification of equivalence classes $\mathfrak{L}$ for local exponents $\varphi(g, g')$ of $G(N)$ can be derived from $G_E^{[\varphi]}$ (the N-dependence is suppressed). Let $g_\nu^{[\varphi]}$, $\nu = 0, \cdots, d$, be a standard basis in $G_E^{[\varphi]}$. $G \subset G_E^{[\varphi]}$ is spanned by $g_\nu^{[\varphi]}$, $\nu = 1, \cdots, d$; $C^{[\varphi]}$ is generator of F_θ and center of $G_E^{[\varphi]}$. The constants of structure $C_{\nu\mu}^{\lambda}[\varphi]$ are known from G for $1 \leq \nu, \mu, \lambda \leq d$. Furthermore, $C_{0\nu}^{\lambda}[\varphi] = 0$ holds. For a calculation of the remaining type $C_{\nu\mu}^{0}[\varphi]$ we apply the method used for $N = 3$.[67] The result is

$$C_{ik}^{0}[\varphi] = \delta_{ik}\gamma \quad \text{for} \quad g_i^{[\varphi]} = P_i,$$

$$g_k^{[\varphi]} = Q_k \quad \text{and} \quad C_{\nu\mu}^{0}[\varphi] = 0 \text{ otherwise},$$

with γ being any real number. In integrable-irreducible representations of $G_E^{[\varphi]}$ the center $C^{[\varphi]}$ is given by $im \cdot 1$, m real, and $C^{[\varphi]}$ is also an invariant operator. Irreducible representations of $G_E^{[\varphi]}$ are at least classified by m. There is a basis in $G_E^{[\varphi]}$ such that $C^{[\varphi]}$ enters only in the form $\gamma \cdot C^{[\varphi]}$. Hence different γ will lead to different irreducible representations and the equivalence classes are parametrized by γ, $\mathfrak{L} = \mathfrak{L}(\gamma)$, $-\infty < \gamma < +\infty$; $G_E^{[\varphi]} = G_E^{[\gamma]}$. Take $\gamma_1 \neq \gamma_2$ and consider $G_E^{[\gamma_i]}$ $i = 1, 2$. Because $C_E^{[\gamma_i]} = +im\gamma_i 1$ holds, we conclude that irreducible representations of $G_E^{[\gamma_1]}$ exist which are equivalent to those of $G_E^{[\gamma_2]}$ and vice versa. Therefore, in discussions of representations, the index γ can be dropped. It suffices to consider representations of $G_E^{[1]} = G_E$, and m classifies not only the irreducible representations but also the equivalence classes.

APPENDIX B

Some results in Secs. 1B and 3C are based on Mackey's theory of induced representations.[21] The application of this theory is known for $N = 3$. We refer to the physically clear explanation in Refs. 22 and 23. Therefore it suffices to give a short treatment for the general case, to derive the classification of irreducible unitary representations, and to show how the classification determines the structure of the representation.[68] We emphasize that the following is a short survey.

The construction of induced representations of

[67] M. Hammermesh, *Group Theory* (Addison-Wesley Publ. Co., Reading, Massachusetts, 1962).

[68] The author's are indebted to A. Hartkämper and H. Neumann for discussions of the Mackey theory and for several useful suggestions. We refer in this connection also to the detailed treatment in Ref. 24.

$G_E(N) = G_1 \textcircled{\times} G_2$, being a regular semidirect product of G_1 and G_2 with G_1 Abelian, proceeds in 5 steps:

(1) Determination of the character group $\hat{G}_1$ and the classification of the equivalence classes of $\hat{G}_1$ under G_2;
(2) Choice of a representative element χ_0 in each equivalence class C which determines the little group G_e^C belonging to $G_E(N)$;
(3) Construction of all irreducible unitary representations of the little group G_e^C;
(4) Decomposition of $G_E(N)$ by cosets $G_E(N)/G_1 \textcircled{\times} G_e^C$;
(5) Construction of induced representations for $G_E(N)$ via a representation of $G_1 \textcircled{\times} G_e^C$.

The realization of these five steps for $G_E(N)$ is now indicated.

1. G_1 is an $(N+2)$-dimensional translation group with characters (elements of $\hat{G}_1$)

$$\chi_{m,E,\mathbf{p}}(\theta, \tau, \mathbf{u}) = \exp i(m\theta + E\tau - \mathbf{pu});$$
$$m, \mathbf{p}, \mathbf{u} \quad \text{real} \quad (m \neq 0).$$

$\hat{G}_1$ is split into equivalence classes corresponding to G_2. Let $\chi_{m,E,\mathbf{p}}$ and $\chi_{m',E',\mathbf{p}'}$ be characters of G_1. Then they are called equivalent in G_E if an element $g_2 \in G_2$ exists with

$$\chi_{m',E',\mathbf{p}'}(g) = \chi_{m,E,\mathbf{p}}(g_2 g_1 g_2^{-1}) \quad \text{for all} \quad g_1 \in G_1.$$

Take two elements of one class $\chi_{m',E',\mathbf{p}'}(g_1)$. With $g_1 = (\theta, 1, \mathbf{u}, 0, \tau)$ and $g_2 = (0, R, 0, \mathbf{v}, 0)$ the relationship

$$m'\theta + E'\tau - \mathbf{p}'\mathbf{u} = m\theta + (\tfrac{1}{2}mv^2 + E - \mathbf{p}\cdot\mathbf{v})\tau - (R^{-1}\mathbf{p} - mR^{-1}\mathbf{v})\mathbf{u}$$

holds for all $\theta, \tau, \mathbf{u}$, or

$$m = m'; \quad E' = E + \tfrac{1}{2}mv^2 - \mathbf{p}\cdot\mathbf{v};$$
$$\mathbf{p}' = R^{-1}(\mathbf{p} - m\mathbf{v}),$$

and two characters are equivalent if $m = m'$ and if they both lie on the paraboloid

$$V = E - \mathbf{p}^2/2m = E' - \mathbf{p}'^2/2m.$$

This is also sufficient.

Proposition 1: The equivalence classes C of $\hat{G}_1$ in G_E are classified by two real numbers m and V. There is only one character for each $\mathbf{p}$ in C given by $\chi_{m,(\mathbf{p}^2/2m)+V,\mathbf{p}}$.

2. Let C be any equivalence class characterized by m, V, i.e., $C(m, V)$. Consider the group $G_e^{m,V} \subset G_2$ of all transformations leaving invariant one character $\chi \in C(m, V)$. It can be shown that different characters within the same equivalence class $C(m, V)$ lead to isomorphic groups. $G_e^{m,V}$ is therefore independent of χ and is usually called a "little group" for G_E. To construct $G_e^{m,V}$ we choose a special element χ_0 in $C(m, V)$:

$$\chi_0 = \chi_{m,V,0}$$

(i.e., $\mathbf{p} = 0$); then $G_e^{m,V}$ is given by

$$G_e^{m,V} = \{g_2 \mid g_2 \in G_2;\ \chi_{m,V,0}(g_2 g_1 g_2^{-1}) = \chi_{m,V,0}(g_1) \quad \text{for all} \quad g_1 \in G_1\},$$

$$G_e^{m,V} = \{g_2 \mid g_2 = (0, R, 0, \mathbf{v}, 0),\ V = V + \tfrac{1}{2}m\mathbf{v}^2;\ 0 = R^{-1}(m\mathbf{v} - 0)\} \approx SO(N).$$

Proposition 2: The little group $G_e^{m,V}$ for each equivalence class $C(m, V)$ is isomorphic to the orthogonal group $SO(N)$.

3. The little group $SO(N)$ is a simple group of rank $r = [N/2]$. The irreducible representations can be characterized uniquely by a set $s = \{s_1, \cdots, s_r\}$ of r highest weights, i.e., exactly one representation belongs to a set s. The corresponding representation space is denoted by $\mathfrak{K}^{[s]}$ with basis $|\mu\rangle$ and the matrix representation is written $L^{[s]}_{\mu\mu'}(g_0)$ for $g_0 \in G_e^{m,V}$; the mapping

$$g_1 \cdot g_0 \to \chi_{m,V,0}(g_1) \cdot L^{[s]}_{\mu\mu'}(g_0) = M^{[m,V,s]}_{\mu\mu'}(g_1 \cdot g_0)$$

defines a matrix representation $M^{[m,V,s]}_{\mu\mu'}(g_1 \cdot g_0)$ of $G_1 \textcircled{\times} G_e^{m,V}$ in $\mathfrak{K}^{[s]}$.

According to Mackey, all inequivalent irreducible unitary representations of a regular semidirect product $G_E = G_1 \textcircled{\times} G_2$ can be classified by combining the classification of the equivalence classes C of $\hat{G}_1$ and the classification of the irreducible unitary representations of the little group. $U_1(G_E)$ and $U_2(G_E)$ are equivalent if and only if the corresponding equivalence classes are equal and the corresponding irreducible unitary representations of the little group are equivalent. For all classes C and all irreducible unitary representations of the little group there exists a unitary representation $U(G_E)$.

4. A discussion of the remaining two steps is added to indicate the structure of the resulting representation $U(G_E)$.

To decompose G_E by cosets $G_E/G_1 \textcircled{\times} G_l^{m,V}$ modulo $G_1 \textcircled{\times} G_e^{m,V}$, consider the cosets

$$G_1 \textcircled{\times} G_e^{m,V} \cdot g = G_1 \textcircled{\times} G_e^{m,V} \cdot g_1 \cdot g_2 = G_1 \textcircled{\times} G_e^{m,V} \cdot g_2 \in G_E/G_1 \textcircled{\times} G_e^{m,V},$$

which are determined by

$$G_e^{m,V} \cdot g_2.$$

It can be shown that there is a one-to-one correspondence between $G_e^{m,V}g_2$ and characters in $C(m, V)$:

$$G_e^{m,V} \cdot g_2 \leftrightarrow \chi_{m,V,0}(g_2 \cdot g_1 \cdot g_2^{-1}).$$

With Proposition 1, $G_E/G_1 \textcircled{\times} G_e^{m,V}$ can be mapped on to points $\mathbf{p}$ of an N-dimensional Euclidean space. Take in $C(m, V)$ the character corresponding to $\mathbf{p}$, i.e.,

$$\chi_{m,(p^2/2m)+V,\mathbf{p}}(g_1).$$

Then $g_2 = (0, R, 0, \mathbf{v}, 0)$ is element of the coset characterized by $\mathbf{p}$, i.e.,

$$\chi_{m,V,0}(g_2 \cdot g_1' \cdot g_2^{-1}) = \chi_{m,V,0}(g \cdot g_1' \cdot g^{-1}) = \chi_{m,(p^2/2m)+V,\mathbf{p}}(g_1')$$

holds if

$$-mR^{-1}\mathbf{v} = \mathbf{p}.$$

The transformation of the coset characterized by $\mathbf{p}$ under $g \in G_E$ leads to a coset characterized by $\mathbf{p}' = R^{-1}(\mathbf{p} - m\mathbf{v})$. Hence the measure on the coset which is mapped into the measure $d_p^N = dp_1 \cdots dp_N$ of the Euclidean space is invariant under G_E.

5. For the construction of an irreducible representation $U(G_E)$ let $\mathcal{H}_g^{[m,V,s]}$ be the space of vector functions $f(g)$ over G_E with values in $\mathcal{H}^{[s]}$. Denote by $\langle \cdot \mid \cdot \rangle$ an inner product and by $\langle \cdot \mid \cdot \rangle_{[s]}$ its restriction to $\mathcal{H}^{[s]}$ and assume

(α) $f(g_1 \cdot g_0 \cdot g) = M^{[m,V,s]}(g_1 \cdot g_0) f(g)$ for all $g_1 \cdot g_0 \in G_1 \textcircled{\times} G_e^{m,V}$; $\langle \cdot \mid \cdot \rangle_{[s]} < \infty$.

$\langle \cdot \mid \cdot \rangle_{[s]}$ is constant on the cosets $G_E/G_1 \textcircled{\times} G_e^{m,V}$ which can be mapped isomorphically onto the N-dimensional p space. Hence it is a function of $\mathbf{p}$ only. If we assume furthermore

(β) $$\int_g \langle f(g) \mid f(g) \rangle_{[s]} \, d^N p < \infty,$$

then $\mathcal{H}_g^{[m,V,s]}$ becomes a Hilbert space, and a representation $U(G_E)$ is given by the set of operators $U(G_1)$, $g_1 \in G_E$ acting as $U(g_1)f(g) = f(g \cdot g_1)$.

It is convenient to consider only functions $f(g)$ over cosets $G_E/G_1 \textcircled{\times} G_e^{m,V}$ as elements of $\mathcal{H}_g^{[m,V,s]}$. This is possible because $\mathcal{H}_g^{[m,V,s]}$ can be mapped isometrically onto $\mathcal{H}_p^{[m,V,s]}$ spanned by $f(g_\mathbf{p})$, $g_\mathbf{p} \in G_E/G_1 \textcircled{\times} G_e^{m,V}$ with $f(g_\mathbf{p}) = f(\mathbf{p})$ and with inner product

$$\int_g \langle f(\mathbf{p}) \mid h(\mathbf{p}) \rangle_{[s]} \, d^N p = \langle f \mid h \rangle.$$

To perform this mapping one has to choose in each coset characterized by $\mathbf{p}$ one representative element $g_\mathbf{p}$, and different choices of $g_\mathbf{p}$ will lead to different basis systems in $\mathcal{H}_p^{[m,V,s]}$ because $U(G_e^{m,V})$ is unitary.

For the explicit construction of the representation in $\mathcal{H}_p^{[m,V,s]}$, let $f_\nu(\mathbf{p})$ be components of $f(\mathbf{p})$ and introduce an (improper) basis spanned by $|\mathbf{p}, \mu\rangle$ such that

$$|f\rangle = \sum_\nu \int d^N p f_\nu(\mathbf{p}) \, |\mathbf{p}, \nu\rangle.$$

Let $r \in G_E$ and calculate the matrix elements of $U(r)$; $(r = r_1 r_2)$:

$$\langle \mathbf{p}, \nu \mid U(r)f\rangle = U(r)f_\nu(\mathbf{p}) = f_\nu(\mathbf{p} \cdot r) = f_\nu(g_\mathbf{p} \cdot r) = f_\nu(g_\mathbf{p} \cdot r \cdot g_\mathbf{p} g_\mathbf{p}^{-1} \cdot g_{\mathbf{p}r}^{-1} g_{\mathbf{p}r})$$

with $\mathbf{p}_r = R^{-1}(\mathbf{p} - m\mathbf{v})$ denoting the coset containing $g_\mathbf{p} \cdot r$. Because

$$g_\mathbf{p} r_1 g_\mathbf{p}^{-1} \in G_1; \quad g_\mathbf{p} r_2 g_\mathbf{r}^{-1} \in G_2; \quad g_\mathbf{p} r g_\mathbf{p}^{-1} \in G_1 \textcircled{\times} G_e^{m,V},$$

we can apply (α) and have, with the result of step 3,

$$\langle \mathbf{p}, \mu \mid U(r)f\rangle = \sum_{\mu'} \chi_{m,V,0}(g_\mathbf{p} r_1 g_\mathbf{p}^{-1}) L_{\mu\mu'}^{[s]}(g_\mathbf{p} r_2 g_{\mathbf{p}r}^{-1}) f_\mu(\mathbf{p}_r)$$

or

$$\langle \mathbf{p}, \mu \mid U(r)f\rangle = \sum_{\mu'} \exp \cdot i\left(m\theta + \left(\frac{\mathbf{p}^2}{2m} + V\right)\tau - \mathbf{p} \cdot \mathbf{u}\right) \times L_{\mu\mu'}^{[s]}(g_\mathbf{p} r_2 g_{\mathbf{p}r}^{-1}) \langle R^{-1}(\mathbf{p} - m\mathbf{v}), \mu' \mid f\rangle.$$

The matrix $L_{\mu\mu'}^{[s]}$ still depends on the choice of the representative $\mathbf{p}_r$ or on the basis in the Hilbert space $\mathcal{H}_p^{[m,V,s]}$, as pointed out above. One can prove that the representation is unitary and irreducible and that all unitary representations $U(G_E)$ are obtained by this method. The number of components of the vector functions is directly related to the representation of the little group $G_e^{m,V}$.

APPENDIX C: PROOF OF LEMMA 2

(a) Aut $\mathsf{G}_E^0(N)$ is subgroup of $GL(d, R)$, $d = 2N + 1$ defined by a set of algebraic conditions which are consequences of the invariance of the Lie brackets in $\mathsf{G}_E^0(N)$. To derive them, choose in $\mathsf{G}_E^0(N)$ a basis

$$\{X_{2i-1} = P_i; \quad X_{2i} = Q_i, i = 1, \cdots N; \quad X_{2N+1} = \mathsf{C}\}$$

and consider a matrix $A = (a_{i,k}) \in GL(d, R)$ and $X, X' \in \mathsf{G}_E^0(n)$ with

$$X = \sum_{i=1}^{d} X_i \quad \text{and} \quad X' = AX.$$

If

$$[P_i', Q_j'] = \mathsf{C}'\delta_{ij}; \quad [P_i', P_j'] = [Q_i', Q_j'] = 0,$$
$$[P_j', \mathsf{C}'] = [Q_i', \mathsf{C}'] = 0, \quad i, j = 1, \cdots, N$$

holds, then $A \in$ Aut $\mathsf{G}_E(N)$ and we find $2N^2 - N$ and $2N$ conditions, respectively, for A:

$$\sum_{\nu=1}^{2N-1} \det \begin{vmatrix} a_{k,\nu} & a_{k,\nu+1} \\ a_{j,\nu} & a_{j,\nu+1} \end{vmatrix} = \delta_{k,j-1} a_{2N+1,2N+1}, \quad k, j = 1, \cdots, 2N, \quad \text{(C1)}$$

$$a_{2N+1,k} = 0, \quad k = 1, \cdots, 2N. \quad \text{(C1')}$$

(b) To extract int-Aut $G_E^0(N)$ we calculate its Lie algebra ad $G_E^0(N)$; after a straightforward discussion, one shows that $A_{\text{int}} \in$ int-Aut $G_E^0(N)$ has the form

$$A_{\text{int}} = \begin{pmatrix} 1 & & & & \lambda_1 \\ & \ddots & & \iddots & \\ & & 0 & & \\ & \iddots & & \ddots & \\ 0 & & & & \lambda_{2n} \\ & & & & 1 \end{pmatrix}.$$

Hence it is Abelian. From (1) we have for any $A_{\text{out}} \in$ Aut $G_E^0(N)$/int-Aut $G_E^0(N) \approx$ out-Aut $G_E^0(N)$ the form ($\lambda = a_{2N+1,2N+1}$)

$$A_{\text{out}} = \begin{pmatrix} B & 0 \\ 0 & \lambda \end{pmatrix} \in \text{out-Aut } G_E^0(N)$$

with B being a $2N \times 2N$ matrix. λ is arbitrary.

(c) To decompose out-Aut $G_E^0(N)$ consider two special outer automorphisms ($\lambda > 0$):

$$A_\lambda = \begin{pmatrix} \lambda^{\frac{1}{2}} 1_{2N} & 0 \\ 0 & \lambda \end{pmatrix}$$

with 1_{2N} being a $2N \times 2N$ unit matrix and a kind of inversion J acting as

$$JP_i = Q_i, \quad JQ_i = P_i, \quad JC = -C, \quad i = 1, \cdots, N.$$

A_λ and J fulfil condition (C1) and A_λ, $J \in$ out-Aut $G_E^0(N)$. We define, furthermore, $\hat{A}_{\text{out}} \in$ out-Aut $G_E^0(N)$

$$\hat{A}_{\text{out}} = \begin{pmatrix} B' & 0 \\ 0 & 1 \end{pmatrix}$$

restricted by condition (C1) only and have

$$A_{\text{out}} = \begin{cases} A_\lambda \hat{A}_{\text{out}}, \\ J A_\lambda \hat{A}_{\text{out}}. \end{cases}$$

A_λ and $\hat{A}_{\text{out}}$ commute. Therefore out-Aut $G_E^0(N)$ is of the form

$$\text{out-Aut } G_E^0(N) \approx (\{\hat{A}_{\text{out}}\} \otimes \{A_\lambda\}) \textcircled{\times} S,$$

S being a discrete group with elements $(J, 1)$. A_λ is Abelian and isomorphic to the multiplicative group of the real positive numbers.

(d) To derive $\hat{A}_{\text{out}} \in Sp(2N, R)$ we show that $\hat{A}_{\text{out}}$ possesses a skew-symmetric invariant form $\varphi(x, y)$ with x, y being vectors in a $2N$-dimensional real space. Choose a basis $\{e_{-N}, \cdots, e_{+N}\}$ in this space such that $\varphi(x, y) = (x, S_0 y)$ and

$$S_0 = \begin{pmatrix} & & & & & & -1 \\ & & & & & \iddots & \\ & & & & -1 & & \\ & & & +1 & & & \\ & & \iddots & & & & \\ & \iddots & & & & & \\ +1 & & & & & & \end{pmatrix}.$$

$\varphi(x, y)$ is invariant under $B \in GL(2N, R)$, i.e., $(Bx, S_0\, By) = (x, S_0 y)$. $B \in Sp(2N, R)$ holds iff

$$\text{sign } (j \cdot k) b_{ie} \cdot b_{-j-k} = \delta_{kl}. \qquad \text{(C2)}$$

Arrange now the basis in $G_E^0(N)$ as $\{P_1, Q_2, P_3, Q_4, \cdots, P_2, Q_1; C\}$ and identify P_i, Q_i, $i = 1, \cdots, N$ with the corresponding basis vectors in $\{e_{-N}, \cdots, e_{+N}\}$, i.e., $e_{-N} = P_1, \cdots, e_{+N} = Q_1$. Then the conditions (C2) for $Sp(2N, R)$ and the set of conditions (C1) for $\hat{A}_{\text{out}}$ are of the same form and $\hat{A}_{\text{out}} \in Sp(2N, R)$ holds.

(e) Because any element of $Sp(2N, R)$ also fulfils condition (C1), $\{\hat{A}_{\text{out}}\} \approx Sp(2N, R)$; $Sp(2N, R)$ occurs in its lowest-dimensional representation, which is unique up to equivalence. The dimension of $Sp(2n, R)$ is $2N^2 + N$. We remark that the conditions (C1) are linearly independent because this is the case for the equivalent formulation in (C2). This concludes the proof.

APPENDIX D

In some cases the identification of a basis in $\epsilon^{(2)}$ in terms of skew-symmetric second-order polynomials in P_i, Q_i, $i = 1, \cdots, N$ is given explicitly and the semidirect sum $\epsilon_c^{(1)} \textcircled{+} \epsilon^{(2)}$ is described by commutation relations ($C = im1$).

(a) $N = 1$. $\epsilon^{(2)}$ is three dimensional. The basis

$$L_{12} = \frac{i}{4m}(P^2 + Q^2), \quad L_{13} = \frac{i}{4m}(PQ + QP),$$
$$L_{23} = \frac{i}{4m}(P^2 - Q^2)$$

span $so(2, 1) \approx sp(2, R)$ with Lie algebra

$$[L_{\mu\nu}, L_{\mu\lambda}] = g_{\mu\mu} L_{\nu\lambda}, \; L_{\mu\nu} = -L_{\nu\mu},$$
$$g_{11} = g_{22} = -g_{33} = -1,$$

A further useful basis in $\epsilon^{(2)}$ is $A_1 = L_{12} + L_{23}$; $A_2 = L_{12} - L_{23}$; $A_3 = 2L_{13}$, i.e., $A_1 = (i/2m)P^2$; $A_2 = (i/2m)Q^2$. $\epsilon_c^{(1)} \textcircled{+} \epsilon^{(2)}$ is spanned by six generators P, Q, C, A_i and the semidirect sum is defined by

$$[P, A_1] = 0, \quad [P, A_2] = -Q, \quad [P, A_3] = -P,$$
$$[Q, A_1] = +P, \quad [Q, A_2] = 0, \quad [Q, A_3] = Q,$$
$$[C, A_i] = 0, \quad i = 1, 2, 3.$$

(b) $N = 2$. $\epsilon^{(2)}$ is ten-dimensional. The following bases span $so(3, 2) \approx sp(4, R)$:

$$M_{54} = L_{12} = \frac{i}{4m} \sum_{i=1}^{2} (P_i^2 + Q_i^2);$$

$$M_{35} = L_{13} = \frac{i}{4m} \sum_{i=1}^{2} (P_i Q_i + Q_i P_i);$$

$$M_{34} = L_{23} = \frac{i}{4m} \sum_{i=1}^{2} (P_i^2 - Q_i^2);$$

$$M_{21} = L_1 = \frac{i}{2m} (Q_2 P_1 - Q_1 P_2);$$

$$M_{13} = L_2 = \frac{i}{4m} (P_1^2 + Q_1^2) - \frac{i}{4m} (P_2^2 + Q_2^2);$$

$$M_{23} = L_3 = \frac{i}{2m} (P_1 P_2 + Q_1 Q_2);$$

$$M_{15} = T_1 = \frac{i}{4m} (P_1^2 - Q_1^2) - \frac{i}{4m} (P_2^2 - Q_2^2);$$

$$M_{42} = T_4 = \frac{i}{2m} (P_2 Q_1 + Q_2 P_1);$$

$$M_{25} = T_2 = \frac{i}{2m} (P_1 P_2 - Q_1 Q_2);$$

$$M_{41} = T_3 = \frac{i}{4m} (Q_1 P_1 + P_1 Q_1) - \frac{i}{4m} (P_2 Q_2 + Q_2 P_2);$$

$$[M_{\mu\nu}, M_{\varkappa\lambda}] = g_{\mu\lambda} M_{\nu\varkappa} - g_{\mu\varkappa} M_{\nu\lambda} + g_{\nu\varkappa} M_{\mu\lambda} - g_{\nu\lambda} M_{\mu\varkappa} (\mu, \varkappa, \nu, \lambda = 1, 2, 3, 4, 5);$$

$$g_{11} = g_{22} = g_{33} = -1, \quad g_{44} = g_{55} = 1;$$

$$g_{\mu\nu} = 0 \quad \text{for} \quad \mu \neq \nu.$$

The subalgebras $so(2, 1)$ and $su(2)$ are spanned by $\{L_{12}, L_{13}, L_{23}\}$ and by $\{L_1, L_2, L_3\}$, respectively.

(c) $N = 3$. $\epsilon^{(2)}$ is 21-dimensional. A suitable basis can be calculated following Ref. 11. For example, the subalgebra $su(3)$ is identified. We denote by $\mathbf{M} = (\mathbf{Q} \times \mathbf{P})$, $A_{ij} = P_i P_j$, $i, j = 1, 2, 3$. A standard basis of $su(3)$ is

$$H_1 = \frac{i}{2m} (A_{11} - A_{22}),$$

$$H_2 = -\frac{i}{4m} A_{33},$$

$$E_{\pm 1} = \mp \frac{i}{2(6)^{\frac{1}{2}} m} (M_3 \pm A_{12}),$$

$$E_{\pm 2} = \mp \frac{i}{2(6)^{\frac{1}{2}} m} (M_2 \mp A_{13}),$$

$$E_{\pm 3} = \mp \frac{i}{2(6)^{\frac{1}{2}} m} (M_1 \pm A_{23}).$$

H_i and $E_{\pm\alpha}$ commute with

$$E = \sum_{i=1}^{3} A_{ii}.$$

Complex linear combinations are not suitable for an identification because the difference between compact and noncompact groups is lost.

APPENDIX E: PROOF OF LEMMA 5

(a) It suffices to prove Lemma 5 for $Sp(2N, R)$ and $U(N)$ being considered as subgroups of $GL(2N, R)$. Let $U(2N) \subset GL(2N, R)$, being the group of unitary unimodular matrices. Then $Sp(2N, R) \cap U(2N) \approx U(N)$[48] and hence $U(N) \subset Sp(2N, R)$. $U(n)$ is compact.

(b) The stronger statement that $U(N)$ is even maximal subgroup is a consequence of the criterion[69] that a compact subgroup of $Sp(2N, R)$ is maximal if there exists an N-dimensional subspace R_N of R_{2N} which is invariant under $U(N) \subset Sp(2N, R)$. We show that $U(N)$ fulfils this criterion. The $2N$-dimensional fundamental representation of $Sp(2N, R)$ is a real one and unique up to equivalence; the representation $U_{(2N)}(N)$ of $U(N)$ in $Sp(2N, R)$ is real too. $U_{(2N)}(N)$ is equal to its complex conjugate representation $\bar{U}_{(2N)}(N)$ and, in the reduction of $U_{(2N)}(N)$ into irreducible k-dimensional parts $U_k(N)$, there appears with each $U_k(N)$ also $\bar{U}_k(N)$. The dimensions of the irreducible representations of $U(N)$ are 1, N, $d(N)$ with $d(N) > 2N$ for $N > 2$. Hence for $N > 2$ only 1- or N-dimensional representations $U_1(N)$ and $U_n(N)$,[70] respectively, can appear. At least one $U_k(N)$ is N-dimensional; otherwise the representation of $U_{(2N)}(N)$ is not faithful. Therefore $Sp(2N, R)$ contains $U(N)$ in the form

$$U_{(2N)}(N) \approx \begin{pmatrix} U_N(N) & 0 \\ 0 & U_N(N) \end{pmatrix}. \tag{E1}$$

For $N = 2$ this is also true, as is easily checked. The case $N = 1$ is trivial. The invariance of an N-dimensional subspace R_N is obvious from (E1).

(c) The form (E1) for $U_{(2N)}(N)$ is unique to equivalence, i.e., up to transformations $AU_{(2N)}(N)A^{-1}$ with $A \in L(2N, R)$. By a theorem of Malcev[71]—proved for complex algebras but valid also for their real forms—two equivalent subgroups of $L(2N, R)$ are already equivalent in $Sp(2N, R)$. Therefore $A \in Sp(2N, R)$ holds and $U_{(2N)}(N)$ is unique up to inner automorphism of $Sp(2N, R)$, for $Sp(2N, R)$ is simple. This concludes the proof.

[69] E. B. Dynkin, Am. Math. Soc. Transl. No. 33 (1950).

[70] The general N-dimensional representation $U_N(N)$ of $U(N)$ maps $U \in U(N)$ onto $(\det U)^{\lambda_1} U$ or onto $(\det U)^{\lambda_2} \bar{U}$, $\lambda_1, \lambda_2 = 0 \pm 1, \pm 2, \cdots$ up to equivalence and $U \to (\det U)^{\lambda_1} U$ is an automorphism of $U(N)$ since $(\det U)^{\lambda_1} U$ is again unitary.

[71] A. J. Malcev, Am. Math. Soc. Transl. Ser. 1, Vol. 9 (1957).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Almost Symmetric Spaces and Gravitational Radiation*

RICHARD A. MATZNER†
Department of Applied Mathematics and Theoretical Physics, University of Cambridge
and
Department of Physics and Astronomy, University of Maryland, College Park, Maryland

(Received 3 November 1967)

A generalization of the idea of Killing fields to spaces which are not symmetric is given. The field so defined specifies coordinate lines along which the variation of the metric tensor is the slowest possible in a global sense. Thus it generalizes the Killing fields in spaces with a symmetry where the metric tensor does not change along Killing trajectories. Several examples are given, and the method is then applied to spaces containing gravitational radiation of the type considered by Issacson.

For spaces containing radiation, it is shown that a real functional $\lambda[\xi]$, associated with every vector field ξ, measures some parameters associated with the radiation. In the simplest case this parameter is the "energy density" of the radiation, but, if a sufficient number of vector fields can be invariantly defined in the background, the average gravitational "stress" associated with the wave may also be measured. We conclude with some conjectures about further application of these ideas to the theory of gravitational radiation.

I. INTRODUCTION AND OUTLINE

Riemannian spaces which possess a symmetry are those in which a coordinate system may be found with the metric tensor independent of one of the coordinates. They are equivalently characterized by the fact that they admit a solution ξ to Killing's equation[1,2]

$$\mathcal{L}_\xi g_{AB} \equiv 2\xi_{(A\|B)} = 0.$$

The preferred coordinate system mentioned above is obtained by picking coordinates such that $\xi^A \equiv (\xi)^A = \delta^A_{(0)}$. The metric is then clearly independent[3] of x^0.

Practically every calculation is simplified when the space admits a Killing field; and, correspondingly, when there is no Killing field, the sheer calculational difficulties multiply. For instance, calculation of the effects of small deviations from exact symmetry in cosmological solutions must often be treated in an approximate manner. Heretofore, even a quantitative measure of the lack of symmetry in a Riemannian manifold has been lacking.

The purpose of this paper is to suggest one plausible generalization of the notion of a Killing field—a definition which simultaneously gives a quantitative measure of the degree of symmetry in the space—and to point out some properties of this definition. The hope is that this work will lead to simple covariant ways of characterizing spaces which are "almost symmetric." This would help to remove the clutter of coordinate effects which attend the usual approximation methods. Such applications are, however, left for later investigation.

The discussion is initially in terms of positive-definite manifolds, but we indicate the generalization to spaces of Minkowski signature. The treatment is general in the sense that we do not have to assume the deviation from symmetry is small, although we may do so at times to make interpretation easier.

Our definition for spaces which are not symmetric and hence have no nontrivial solution to the equation $\xi_{(A\|B)} = 0$ is the following. We characterize the amount of symmetry in a (positive-definite) Riemannian space $\mathcal{M}$ by considering the minimum possible value of the expression

$$0 \leq \lambda[\xi] = \frac{\int \xi^{(A\|B)}\xi_{(A\|B)}\, dV}{\int \xi^A \xi_A\, dV}; \quad dV \equiv \sqrt{g}\, d^n x. \quad (1)$$

Here ξ is an arbitrary vector field, and the quantity λ is the ratio of integrals of scalar fields over the space. Since the metric is positive-definite, λ is zero iff ξ is Killing. We have imposed the normalization condition in (1), dividing by the integral of the squared length of the vector, to exclude zero fields which are always solutions of Killing's equation. We shall take as our

* Supported in part by NASA Grant NSG-436 and by the Aerospace Research Lab., Office of Aerospace Research, U.S. Air Force. This work is based on a doctoral dissertation by the author submitted to the University of Maryland.

† Presently NSF Faculty Associate at the University of Texas, Austin, Texas.

[1] Capital Latin indices run and sum over the range of components in a general (positive-definite) space. Lower case Greek (0–3) and Latin (1–3) are reserved for four-dimensional Minkowskian spaces V_4 (signature $+\cdots$). Round brackets mean symmetric part, square brackets, antisymmetric. The double vertical bar (‖) means covariant derivative in a general Riemannian space, the semicolon (;) means covariant derivative in V_4, the single vertical bar (|) means covariant derivative in a 3-dimensional positive-definite subset of V_4. Ordinary derivative is indicated by a comma.

[2] K. Yano and S. Bochner, *Curvature and Betti Numbers* (Princeton University Press, Princeton, 1953).

[3] A. Trautman, "Foundations and Current Problems in General Relativity" in *Lectures on General Relativity, Brandeis Summer Institute in Theoretical Physics* (Prentice-Hall, Inc., Englewood Cliffs, N.J. 1964).

criterion for the "almost Killing" field that it minimize λ compared to all other choices of the vector field. Objects like the right side of (1) may have more than one stationary point, so we emphasize that the most interesting value of λ and the correspondingly most interesting vector field associated with it are ${}_0\lambda$, the smallest stationary value, and ${}_0\xi$, the "ground-state" vector field.

By standard arguments, assuming the compactness of $\mathcal{M}$, or restricting the class of vector fields to allow neglect of surface terms at infinity, the variational problem defined in Eq. (1) is the same as the problem of finding the eigenvalues ${}_m\lambda$ in the equation

$$\xi^{(A\|B)}{}_{\|B} + {}_m\lambda\xi^A = 0, \tag{2}$$

and ${}_0\lambda$ corresponds to the smallest (for positive-definite spaces) of these ${}_m\lambda$.

The ground-state vector field ${}_0\xi$ may be characterized in the following way. In a coordinate system in which ${}_0\xi = \delta^A_{(0)}$, then $g_{00} = {}_0\xi \cdot {}_0\xi$, and

$$g_{00}^{-\frac{1}{2}} \frac{dg_{AC}}{dx^0} \equiv \frac{dg_{AC}}{ds} = \frac{{}_0\xi_{(A\|C)}}{({}_0\xi \cdot {}_0\xi)^{\frac{1}{2}}}. \tag{3}$$

Here ds is an element of proper length along ${}_0\xi$. It is apparent that (1) is an integral "average square" of this quantity (3), but the "average" of the ratio is given by the ratio of the averages of the numerator and denominator. Even though we have a small eigenvalue ${}_0\lambda$, it is difficult to use ${}_0\lambda$ to put bounds on the quantity (3), since ${}_0\xi$ may vanish at some points for global topological reasons (for instance, nonsingular vector fields on a sphere must vanish somewhere).[4] The integral average (1) thus forces us to accept behavior which is locally rapid (e.g., schematically, g_{AB} a smooth curve with a few kinks) as being smooth in a global sense. On the other hand, the vector field defined by Eq. (1) obviously chooses the coordinate lines for x^0 which give the slowest dependence of g_{AB} on x^0 in a global sense.

For the moment assuming the existence of solutions, we can get an upper bound for the quantity ${}_0\lambda$. By definition, ${}_0\lambda$ is the minimum value of the integral (1), so any test function gives a bound. By considering a geodesic patch of radius L at some point in $\mathcal{M}$, and by taking a test field

$$\psi_A = \delta^{(0)}_A[1 - 2L^{-1}(\textstyle\sum x^A x^A)^{\frac{1}{2}}], \quad \text{for} \quad \textstyle\sum x^A x^A \leq \frac{1}{4}L^2$$
$$\psi_A = 0, \quad \text{otherwise}$$

it is easy to estimate that $0 \leq \lambda[\psi] \leq 2L^{-2}(n+1) \times (n+2)$, where n is the dimension of the space. [Thus, on the surface of a cube of edge length l for instance, the maximum possible L is $2^{-\frac{1}{2}}l$ and ${}_0\lambda(\text{cube}) \leq 2 \cdot 2l^{-2} \cdot 3 \cdot 4 = 48l^{-2}$.] The size of a geodesic coordinate patch is roughly given by $L^{-2} \sim R....$ (the Riemann tensor), so we have a rough bound for ${}_0\lambda$ in terms of the curvature.

It is important to note that estimates of this type hold for the eigenvalue ${}_0\lambda$ in *any* space. The idea of "almost symmetric" enters when it turns out that ${}_0\lambda \ll L^{-2}$, where L is a typical length of the problem. We present some examples of this type of behavior in Secs. III and IV below, but a simple example is the unit 2-sphere, where ${}_0\lambda = 0$, while the only available length is the radius ($= 1$).

In the following section (Sec. II) we state our mathematical assumptions and give the (completely standard) derivation of Eq. (2). We quote a theorem guaranteeing differentiable solutions for closed (positive-definite) Riemannian manifolds. We point out a close analogy between the problem considered here and the problem of elasticity in non-Euclidean space, and we give a "physical" meaning to the eigenvalue ${}_0\lambda$ (essentially ω_0^2 where ω_0 is the lowest natural frequency) for 2-dimensional surfaces which can be imagined embedded in flat 3-space.

We then prove a theorem which generalizes a theorem of Yano,[2] that there are no Killing vectors on compact positive-definite spaces of negative-definite curvature. We give here a criterion ${}_0\lambda$ of the departure from possible symmetry and give a *lower* bound for ${}_0\lambda$.

We then consider one peculiarity of spaces of Minkowskian signature. In some cases the quantity $\xi^{\alpha;\beta}\xi_{(\alpha;\beta)} = 0$, even though $\xi_{(\alpha;\beta)} \neq 0$. A particular example is the Kerr solution, where ξ is Killing in the flat background. We discuss the behavior of the square of the derivative for the class of metrics $g_{\alpha\beta} = b_{\alpha\beta} + k_\alpha k_\beta$, with $k_\beta k^\beta = 0$, and $b_{\alpha\beta}$ a background metric which has a symmetry. We see that all the Killing vectors for $b_{\alpha\beta}$ have null Killing tensors in $g_{\alpha\beta}$.

We show also that a slightly modified integral definition of ${}_0\lambda$ is still available in some cases where ξ and $\xi^{(\alpha;\beta)}$ become null together.

Section III gives some examples to show the dependence of the parameter ${}_0\lambda$ on the departure from an

[4] We expect, however, that the ground-state vector field will have few zeros. It is known that the nodes of this vector field cannot be separating hypersurfaces (that is, they cannot separate the domain into disjoint parts). See R. Courant and D. Hilbert, *Methods of Theoretical Physics* (Interscience Publ., Inc., New York, 1953), Vol. 1 p. 452. If we assumed that $\mathcal{M}$ and g_{AB} were analytic instead of merely C^∞, then there could be no zeros which are $(n-1)$-dimensional subsets of hypersurfaces. In any case, this theorem simplifies the construction of coordinate systems which utilize the ground-state vector field as one congruence of coordinate lines. Such a coordinate system will fail only on the at most $(n-1)$-dimensional regions where the vector field vanishes, and so "patching" over the nodes with geodesic coordinates should be simple.

exactly symmetrical situation. The first of these is a perturbed flat 2-dimensional torus, which has the advantage that the ground state of one class of solutions to Eq. (2) can be found explicitly. It shows behavior which we find to be characteristic of short gravitational radiation in general. We also give a calculation of ${}_0\lambda$ in a flat space with a linearized gravitational wave pulse. This shows how the deviation from a symmetric situation with a null Killing vector behaves. In this case the minimizing vector becomes timelike, and the eigenvalue ${}_0\lambda \sim h^2(\Delta k_\perp)^2(k_0\Delta k_\perp)^2$, where h is the amplitude, $\mathbf{k}_0$ the center momentum, and $\Delta k_\perp$ the *transverse* spread in momentum of the weak pulse.

Section IV is perhaps the most important part of this work. In this section we apply the idea of symmetry to spaces containing short gravitational radiation of the type considered by Issacson.[5] We find that we are able to characterize such spaces (which may be completely empty) by a number which is roughly the energy density of the gravitational radiation. We may in certain cases in fact specify at least as many parameters for the gravitational radiation as there are Killing vectors in the background metric (on which the radiation is superimposed), thereby giving a measure of the "stress" associated with the gravitational wave.

Finally, Sec. V mentions some of the applications and problems which should be investigated by this method, and points out some of the possibilities inherent in it.

II. MATHEMATICS

A. Existence and Differentiability for Closed Positive-Definite Metric Spaces

For the rest of this paper, except where noted otherwise, we assume that we are working with a positive-definite Riemannian C^∞ manifold $\mathcal{M}$, and we assume that $\mathcal{M}$ is compact or that boundary conditions are chosen so that integration by parts is possible with neglect of surface terms. We denote the Hilbert space of all square-integrable vector fields $\boldsymbol{\zeta}$ on $\mathcal{M}$ by $\mathrm{L}(\mathcal{M})$. The norm is

$$\|\boldsymbol{\zeta}\|_m = \left(\int dV\, \boldsymbol{\zeta}\cdot\boldsymbol{\zeta}\right)^{\frac{1}{2}}.$$

The demand that $\lambda[\boldsymbol{\xi}]$ be stationary yields, as usual, a second-order equation:

$$\delta\left[\lambda\int \xi^A\xi_A\, dV\right] = \delta\left[\int \xi^{(A\|B)}\xi_{(A\|B)}\, dV\right],$$

$$2\lambda\int \xi^A\delta\xi_A\, dV = 2\int \xi^{(A\|B)}\delta\xi_{A\|B}\, dV$$
$$= 2\int\{[\xi^{(A\|B)}\delta\xi_A]_{\|B} - \xi^{(A\|B)}{}_{\|B}\delta\xi_A\}\, dV.$$

The compactness of $\mathcal{M}$ (or the boundary condition at infinity) makes the first term on the right vanish. Then, since $\delta\boldsymbol{\xi}$ is an arbitrary variation, we find

$$\xi^{(A\|B)}{}_{\|B} + \lambda\xi^A = 0. \tag{2}$$

We take the definitions (1) and (2) to be the defining equations for the preferred vector fields in $\mathcal{M}$. Equation (2) is the generalization (because of the λ term) of the second-order equation equivalent to Killing's equation given by Yano and Bochner (Ref. 2, p. 57). It is clear that a solution to (2) with $\lambda = 0$ is Killing and vice versa. In Minkowskian-signature metrics, the stationarity of (1) still implies (2), but the equivalence of (2), for $\lambda = 0$, to Killing's equation no longer holds.

The derivation of Eq. (2) shows that the operator

$$-\frac{D}{Dx^B}\, g^{F(B}\delta^{A)}{}_C\frac{D}{Dx^F} = -\mathfrak{D}^A{}_C \tag{4}$$

is positive. (The notation is $D\xi^C/Dx^B \equiv \xi^C{}_{\|B}$.) It is positive-definite if there are no Killing vectors or if we exclude them. Also, because of the compactness of $\mathcal{M}$, or the boundary conditions at infinity, $\mathfrak{D}$ is self-adjoint on $\mathrm{L}(\mathcal{M})$.

It is clear that there are at least as many solutions to (2) as there are Killing vectors. We are of course interested in the case where there are solutions which are not Killing vectors. Consider only the subspace $\mathrm{L}'(\mathcal{M}) \subset \mathrm{L}(\mathcal{M})$ which is orthogonal to the finite number of Killing vectors in $\mathcal{M}$. The operator $-\mathfrak{D}$ is then positive-definite and in fact is strongly elliptic.[6] We may then apply the theorem quoted by Kodaira and Spencer (Ref. 6, Theorem I) for compact $\mathcal{M}$ to find that $-\mathfrak{D}$ has a complete countable set of *differentiable* eigenfunctions $\mathbf{e}_h$ with real eigenvalues whose only accumulation point is $+\infty$. [The completeness means, if $\boldsymbol{\psi}$ differentiable, $\boldsymbol{\psi} \in \mathrm{L}'(\mathcal{M})$, then

$$\boldsymbol{\psi} = \sum_{h=0}^{\infty} a_h\mathbf{e}_h,$$

where $a_h = \int dV\boldsymbol{\psi}\cdot\mathbf{e}_h$, and the series converges in $\mathrm{L}'(\mathcal{M})$.]

Thus we have all the expected "nice" properties of the operator $-\mathfrak{D}$ on the compact manifold $\mathcal{M}$. In particular, we know that a differentiable ground-state solution $\boldsymbol{\xi}_0$ exists. On compact manifolds, then, there

[5] R. A. Isaacson, Phys. Rev. **166**, 1263, 1272 (1968).

[6] K. Kodaira and D. C. Spencer, Ann. Math. **71**, 43 (1960).

will be uniform bounds for the quantities ${}_0\xi \cdot {}_0\xi$ and ${}_0\xi_{A\|B}\,{}_0\xi^{(A\|B)}$ and for all the other derivatives of ${}_0\xi$.

The quadratic form in (1) may be written

$$\xi^{(A\|B)}\xi_{(A\|B)} = C^{ABMN}\xi_{A\|B}\xi_{M\|N},$$

where

$$C^{ABMN} = \tfrac{1}{2}(g^{AM}g^{BN} + g^{AN}g^{BM}). \qquad (5)$$

The most general positive-definite form with these symmetries which depends only on the metric is

$$C^{ABMN} + \mu g^{AB}g^{MN}, \quad \mu \geq -n^{-1}.$$

where n is the dimension of the space. Although the μ addition is nonnegative for $\mu \geq 0$, even in spaces of Minkowski signature, the equation analogous to (2) for nonzero μ,

$$\tfrac{1}{2}\xi^{A\|B}{}_{\|B} + \xi^{B}{}_{\|B}{}^{\|A}(\tfrac{1}{2} + \mu) + (\tfrac{1}{2}R^{A}{}_{C} + \lambda\delta^{A}{}_{C})\xi^{C} = 0, \qquad (6)$$

is not qualitatively different from (2). Dealing only with the μ term gives conditions only on the divergence of $\boldsymbol{\xi}$ and allows too many solutions. If we have any C^{ABMN} contribution, then the equation is qualitatively like (2), and the only criterion for the choice of μ seems to be aesthetics, which suggests $\mu = 0$, as we take here.

We note the following point, however. If, in a Minkowski-signature space, $\boldsymbol{\xi}$ is taken to be the unit timelike normal to a set of spacelike 3-surfaces, then the integrand in the numerator of (1) is always nonnegative and is in fact the quantity $K^{\alpha}{}_{\beta}K^{\beta}{}_{\alpha}$, where $K^{\alpha}{}_{\beta}$ is the second fundamental form of the spacelike 3-surfaces. If we set the tensor[7] $C^{\alpha\beta\mu\nu} = 0$, but set $\mu \neq 0$, the quantity in the integrand is K^2, where K is the trace of $K^{\alpha}{}_{\beta}$. When $K = 0$, the corresponding surface is *minimal*. This has relevance to a suggestion of Komar[8] to use normals to minimal hypersurfaces as the substitute for the timelike Killing field in stationary situations. It is not obvious that a set of such minimal surfaces exists in general situations. However, the minimum principle yields them if they exist, and otherwise gives the most minimal surfaces (globally) available.

B. An Analogue to Elasticity Theory

The tensor C^{ABMN}, defined in (5), is formally similar to the elasticity-strain coefficients given by Green and Zerna[9] for isotropic elasticity in a uniform medium (with Poisson ratio identically zero because we set $\mu = 0$). The similarity of the equations to an elasticity theory is no accident. In elasticity, the strain components $u_{(i|k)}$ measure the Lie derivative of the metric along the displacement field $\mathbf{u}$. This can be seen physically in a coordinate system such that there is no relative coordinate velocity between particles. (This means that the field u^i must have constant components in these coordinates.) Then the metric gives the distance between particles, and the strain tensor is $g_{ij,l}u^l$, which is $\pounds_u g_{ij}$ in this coordinate system.

The minimization problem set here is in fact completely analogous to the eigenvalue problem for vibrations of closed elastic shells under the boundary conditions of sliding rigid contact (the type of boundary condition at the interface between a turning shaft and immobile bearing). In the limit as a shell of material becomes very thin, the bending modes become negligible, and the low frequency eigensolutions (with sliding-contact boundary conditions) become solutions $\boldsymbol{\xi}$, where the vector $\boldsymbol{\xi}$ lies entirely in the surface being considered.[9–11] An example of the situation we visualize is a closed 2-surface whose symmetry we wish to measure. We form a frictionless elastic shell over the surface (with 2-dimensional Poisson ratio $\equiv 0$ since $\mu = 0$), with the shell initially unstrained so that it resists both compression and expansion. Then the asymmetry of the object is measured by the square of the fundamental-oscillation frequency if we perturb the shell. If it has a neutral mode, the surface has a Killing vector.

C. A Theorem of Yano

After the qualitative discussion of the preceding section, we give a precise result. In positive-definite metric spaces the eigenvalue ${}_0\lambda$ is clearly nonnegative. But it is possible to obtain a better lower bound in some cases by noting the following.

We have

$$\int \xi^{(A\|B)}\xi_{(A\|B)}\,dV = -\int \xi^{(A\|B)}{}_{\|B}\xi_A\,dV + \int (\xi^{(A\|B)}\xi_A)_{\|B}\,dV.$$

The second term on the right vanishes by the compactness of the space or by the boundary conditions at infinity. Further, the coefficient of ξ_A in the integrand of the first term on the right is

$$\tfrac{1}{2}\xi^{A\|B}{}_{\|B} + \tfrac{1}{2}\xi^{B\|A}{}_{\|B} = \tfrac{1}{2}\xi^{A\|B}{}_{\|B} + \tfrac{1}{2}\xi^{B}{}_{\|B}{}^{\|A} + \tfrac{1}{2}R^{Ac}\xi^{C}.$$

[7] This was suggested by D. Lyndon-Bell, Monthly Notices Roy. Astron. Soc. **135**, 413 (1967).

[8] A. Komar, Phys. Rev. **127**, 1411 (1962); **129**, 1873 (1963).

[9] A. E. Green and W. Zerna, *Theoretical Elasticity* (Oxford University Press, London, 1954), p. 162.

[10] A. E. Green and J. E. Adkins, *Large Elastic Deformations* (Oxford University Press, London, 1960).

[11] S. C. Mikhlin, *The Problems of the Minimum of a Quadratic Functional* (Holden-Day, San Francisco, 1965).

Thus, if the space is compact or if we impose stronger than usual conditions on the vanishing of $\xi^A_{\|B}$ at infinity (note this is an unsymmetrized derivative),

$$\int \xi^{(A\|B)}\xi_{(A\|B)}\,dV = \tfrac{1}{2}\int \xi^{A\|B}\xi_{A\|B}\,dV + \tfrac{1}{2}\int \xi^{A}{}_{\|A}\xi^{B}{}_{\|B}\,dV - \tfrac{1}{2}\int R^{A}{}_{C}\xi^{C}\xi_{A}\,dV. \quad (7)$$

Consequently, in positive-definite metric spaces,

$$\int (\lambda\delta^{A}{}_{C} + \tfrac{1}{2}R^{A}{}_{C})\xi^{C}\xi_{A}\,dV \geq 0.$$

This holds for any vector ξ and the associated $\lambda[\xi]$. In particular,

$${}_0\lambda\int {}_0\xi\cdot{}_0\xi\,dV \geq -\tfrac{1}{2}\int R^{A}{}_{C}\xi^{C}\xi_{A}\,dV.$$

This is an improved bound in those cases where $R^A{}_C$ is a negative-definite quadratic form on the manifold:

$$\xi^{A}R_{A}{}^{C}\xi_{C} \leq -2\lambda_{\text{Ricci}}\xi^{A}\xi_{A},$$

for some positive number λ_{Ricci} and for all vectors ξ and all points of the manifold. (With this sign convention a hyperboloid has constant negative-definite $R_A{}^C$.)

Thus we have a lower bound for ${}_0\lambda$:

$${}_0\lambda \geq \lambda_{\text{Ricci}}.$$

This derivation is a generalization of that of Yano (Ref. 2, p. 39) to prove that there are no Killing vectors on compact manifolds if $R_A{}^C$ is a negative-definite quadratic form. The advantage of the present formulation is that it gives a criterion ${}_0\lambda$ of the deviation from symmetry.[12]

D. Null Killing Tensors

In spaces of Minkowski signature, the quantity $\xi^{(\alpha;\beta)}\xi_{(\alpha;\beta)}$ may become negative or may be zero even when $\xi_{(\alpha;\beta)}$ is not zero. This complicates the application of the methods described here to simple exact solutions to Einstein's equations, such as those solutions due to Schwarzschild, Kerr,[13] and Vaidya.[14,15]

In each of these solutions, there are vectors which are not null but whose symmetrized derivative is a null tensor. They are the Killing vectors in the flat space $b_{\alpha\beta}$ which is a base metric for each of these solutions, in the sense of Eq. (8) below.

In fact, each of these solutions is a member of a general class of metrics which can be written

$$ds^2 = (b_{\alpha\beta} + k_\alpha k_\beta)\,dx^\alpha\,dx^\beta \equiv g_{\alpha\beta}\,dx^\alpha\,dx^\beta, \quad (8)$$

where $b_{\alpha\beta}$ is some background and k_α is null in the background: $b^{\alpha\beta}k_\alpha k_\beta = 0$, and hence is also null in the full metric. Suppose ζ is Killing in the background. Then there is a coordinate frame such that ζ has constant (contravariant) components and $b_{\alpha\beta,\rho}\zeta^\rho = 0$. Then in this coordinate system,

$$2\zeta_{(\alpha;\beta)} = \pounds_\zeta k_\alpha k_\beta = (k_\alpha k_\beta)_{,\rho}\zeta^\rho,$$

and, since

$$g^{\alpha\gamma} = b^{\alpha\gamma} - k^\alpha k^\gamma$$

and

$$\begin{aligned} k^{\alpha}{}_{,\rho}k_\alpha\zeta^\rho &= (b^{\alpha\sigma}k_{\sigma,\rho}k_\alpha)\zeta^\rho \\ &= (k_{\sigma,\rho}k^\sigma)\zeta^\rho \\ &= \tfrac{1}{2}(k^\alpha k_\alpha)_{,\rho}\zeta^\rho, \end{aligned}$$

the square of $2\zeta_{(\alpha;\beta)}$ is

$$b^{\alpha\gamma}b^{\beta\lambda}(k_\alpha k_\beta)_{,\rho}\zeta^\rho(k_\gamma k_\lambda)_{,\sigma}\zeta^\sigma = 0.$$

It is easy to see that $\zeta_{(\alpha;\beta)}$ has vanishing trace also, so the possibility suggested in Sec. II, that adding a positive μ term would lead to nonnegative results, is seen to be inapplicable here. Metrics like (8) but with $b_{\alpha\beta}$ = flat have been studied by Kerr and Schild.[13]

In the case where $\xi_{(\alpha;\beta)}$ becomes a null tensor for some null vector ξ, if ξ has the same orientation (i.e., future or past directed) everywhere, we may define a limiting process. Let γ^μ be a (so far arbitrary) future-directed timelike vector field. Then define

$$\eta^\mu = \xi^\mu + \epsilon\gamma^\mu, \quad \epsilon > 0;$$

$$\eta_{(\alpha;\beta)}\eta^{(\alpha;\beta)} = \xi_{(\alpha;\beta)}\xi^{(\alpha;\beta)} + 2\epsilon\gamma_{(\alpha;\beta)}\xi^{(\alpha;\beta)} + \epsilon^2\gamma_{(\alpha;\beta)}\gamma^{(\alpha;\beta)}.$$

The first term is zero by hypothesis, as is $\xi_\alpha\xi^\alpha$. Thus

$$\lim_{\epsilon\to 0}\frac{\int \eta_{(\alpha;\beta)}\eta^{(\alpha;\beta)}\,dV}{\int \eta^\alpha\eta_\alpha\,dV} = \frac{\int \gamma_{(\alpha;\beta)}\xi^{(\alpha;\beta)}\,dV}{\int \gamma_\alpha\xi^\alpha\,dV}.$$

Since a future-directed timelike vector is never orthogonal to a null vector, the denominator is always positive.

It is not, in general, clear whether the limit is unique.

[12] The boundary conditions demanded for noncompact spaces for this derivation are rather strict, so this does *not* constitute a proof of the nonexistence of Killing vectors on negative-definite *open* surfaces. In particular, a spacelike 3-hyperboloid has six Killing vectors, each of which would give surface-integral term in Eq. (7).

[13] R. P. Kerr and A. Schild, in *Convengo Sulla Relativita Generale; Problemi Dell' Energia e Onde Gravitazionali, Proceedings of Conference in Honour of the Fourth Centanary of the Birth of Galileo* (Comitato Nationale per le Manifestazioni Celebrative, Rome, 1964).

[14] P. C. Vaidya, Proc. Indian Acad. Sci. A33, 264 (1951); Curr. Science 21, 96 (1952).

[15] R. W. Lindquist, C. W. Misner, and R. A. Schwartz, Phys. Rev. 137, B1364 (1965).

However, if γ^α tends to zero sufficiently fast at infinity,

$$\frac{\int \gamma_{(\alpha;\beta)}\xi^{(\alpha;\beta)}\,dV}{\int \gamma^\alpha \xi_\alpha\,dV} = -\frac{\int \gamma_\alpha \xi^{(\alpha;\beta)}{}_{;\beta}\,dV}{\int \gamma^\alpha \xi_\alpha\,dV}. \tag{9}$$

If ξ^α is a solution to $\mathfrak{D}\xi + \lambda\xi = 0$, for some eigenvalue λ, then even if ξ is null, we can define the characteristic integrals by $\lambda =$ lhs of (9) for an arbitrary timelike vector γ^μ which goes to zero sufficiently fast at infinity. In this case, the limit in the integral definition is independent of the timelike vector used in the limiting process, and the integral definition agrees with the eigenvalue given by the differential equation.

On the other hand, if the limit is λ, independent of the timelike vector γ (so long as γ vanishes sufficiently fast at infinity), then ξ is clearly a solution of the differential equation, by (9). So, for even-oriented null vectors, we can define a modified integral which is equivalent to the differential equation.

III. EXAMPLES

A. Exact Calculation on Torus for ${}_0\lambda$, ${}_0\xi$

The metric for a flat torus is $ds^2 = dx^2 + dy^2$, where the points $(x+1, y)$, (x, y), and $(x, y+1)$ are identified. Differentiable functions on the torus must be doubly periodic with unit period.

We consider a curved torus with metric

$$ds^2 = e^{4\sigma(y)}\,dx^2 + dy^2.$$

Here points are again identified as above by their coordinates, and $\sigma(y)$ is periodic in y with unit period.

This space still has a Killing vector giving translation in the x direction. We consider vectors which are orthogonal to the x Killing ground state, and to simplify the algebra we consider in fact only vectors pointwise-orthogonal. Then it is clear the ground-state vector itself does not depend on x and we write

$$(\xi)^i \equiv (\xi^x, \xi^y) = (0, e^{-\sigma(y)}u[y]).$$

The eigenvalue equation $\mathfrak{D}\xi + \lambda\xi = 0$ becomes

$$f^{-\frac{1}{5}}\left[\frac{d^2u}{dy^2} + \left(\lambda - \frac{1}{5f}\frac{d^2f}{dy^2}\right)u\right] = 0, \tag{10}$$

with $f \equiv e^{5\sigma}$.

We desire a "wavelike" perturbation which makes (10) explicitly manageable. We thus pick f so that (10) is a Mathieu equation. This requires that f itself be a Mathieu function; since $f^{\frac{4}{5}}$ is g_{xx}, we must take f to be the only nowhere-zero periodic solution to Mathieu's equation, namely,[16]

$$f = ce_0(ky;\, 5\epsilon/2) = 1 - (5\epsilon/4)\cos 2ky + \cdots.$$

Here k is $m\pi$, since the basic period is unity, and ϵ, which is assumed small for this expansion, is the amplitude of the small waves in the metric component.

Now f is inserted into (10), which itself becomes a Mathieu equation. By a general theorem on the Sturm–Liouville equation[17] we know that if there is a solution with no zeros, it must be the ground state ${}_0\lambda$, which we seek to find. We are again forced to take a solution ce_0 and obtain

$$u = ce_0(ky;\, \epsilon/40),$$

so that

$${}_0\xi^y = uf^{-\frac{1}{5}} \simeq 1 + \epsilon \cdot \tfrac{19}{80}\cos 2ky + \cdots$$

and

$${}_0\lambda \simeq \frac{\epsilon^2 k^2}{2} + \cdots$$

are the solutions found from the metric perturbation

$$g_{xx} \simeq 1 - \epsilon\cos 2ky + \cdots.$$

From our arguments in Sec. I, we expect ${}_0\lambda^{\frac{1}{2}}$ to be some sort of average derivative of the metric. We see that this is the case, since $g_{xx,y} \simeq -2\epsilon k\cos 2ky$, so that $\langle (g_{xx,y})^2\rangle = 2k^2\epsilon^2$.

The ϵ^2 factor in ${}_0\lambda$ is to be expected, because the flat torus is a space which minimizes λ, so any deviation would be like ϵ^2. The term k^2 suggests that it is the "energy content" of the waves which determines the size of λ. This ties in with our estimates of λ made above in terms of the Riemann tensor. It is interesting to calculate the scalar R, which completely characterizes the curvature:

$$R = -4\left[\frac{d^2\sigma}{dy^2} + 2\left(\frac{d\sigma}{dy}\right)^2\right].$$

For small amplitude waves,

$$\sigma \simeq -\frac{\epsilon}{4}\cos 2ky$$

and

$$R \simeq -4(\epsilon k^2\cos 2ky + \tfrac{1}{2}\epsilon^2 k^2 \sin^2 2ky).$$

The ground state thus gives a much smaller value of ${}_0\lambda$ (by a factor ϵ) than our previous crude estimates yielded. On the other hand, if we *average* R, we have $|\langle R\rangle| = \epsilon^2 k^2$, which is twice the eigenvalue ${}_0\lambda$. In this case, at least, the ground-state eigenfunction is

[16] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Dover Publications, Inc., New York, 1964). We have taken his notation.

[17] E. A. Coddington and N. Levenson, *Theory of Ordinary Differential Equations* (McGraw-Hill Book Co., New York, 1955), p. 212.

sampling the average (over many wavelengths) of the disturbance in the space. We shall see this is a general phenomenon, and shall meet it again in the application of these ideas to spaces containing gravitational radiation, which we take up in Sec. IV.

B. Linearized Gravitational Waves

One example of a space with a null Killing vector is flat space with a plane-weak (linearized) gravitational wave on it. We give a schematic derivation here which shows what happens to the Killing vector as we go from an idealized plane pulse to a situation where there is a spread of directions in a wave packet.

The metric in this situation can be gauged[18] so that

$$ds^2 = (\eta_{\alpha\beta} + h_{\alpha\beta})\, dx^\alpha\, dx^\beta,$$

where

$$h_{0i} = 0, \quad h_{ab} = \int h_{ab}(\mathbf{k}) e^{\mathbf{k}\cdot\mathbf{x} - i|k|t}\, d^3k.$$

We suppose this wave packet is fairly well localized at $t = 0$, and that it has center momentum $\mathbf{k}_0 = |k_0|\,\hat{z}$.

The coefficient of t in the exponent can be expanded about $\mathbf{k}_0$:

$$|\mathbf{k}| = |\mathbf{k}_0| + \Delta\mathbf{k}\cdot\frac{\mathbf{k}_0}{|k_0|} + \frac{\Delta\mathbf{k}}{|k_0|}\cdot\left(1 - \frac{\mathbf{k}_0\mathbf{k}_0}{|k_0|^2}\right)\cdot\Delta\mathbf{k} + \cdots. \tag{11}$$

The coefficient of $\hat{z}$ in this expansion stops at the linear term Δk_z, because $|\mathbf{k}|$ is linear in $\Delta\mathbf{k}$ if $\Delta\mathbf{k}$ is parallel to $\mathbf{k}_0$; the wave packet thus does not spread front-to-back.

Since $\Delta k_x = k_x$, $\Delta k_y = k_y$, and $d\Delta k_z = dk_z$ (because $k_0 = |k_0|\,\hat{z}$), we find to the order written out in (11):

$$h(x, t) = \int dk_x\, dk_y e^{i(k_x x + k_y y)} \times e^{-i(k_x^2 + k_y^2)t/|k_0|}\, {}_0\tilde{h}(k_x, k_y, z - t). \tag{12}$$

Here h is a typical h_{ab} and ${}_0\tilde{h}(k_x, k_y, z)$ is the two-dimensional Fourier transform of $h(\mathbf{x}, 0)$. The integration can be carried further to show that the lateral growth is via diffraction Green's functions. However, the form (12) is more physically transparent. We find

$$\frac{\partial h}{\partial z} = \int dk_x\, dk_y \times e^{ik_x x + ik_y y} e^{-i(k_x^2 + k_y^2)t/|k_0|} \frac{\partial}{\partial z}\, {}_0\tilde{h}(k_x, k_y, z - t) \equiv ik_1 h,$$

$$\frac{\partial h}{\partial t} = -\frac{\partial h}{\partial z} - i\int \frac{(k_x^2 + k_y^2)}{|k_0|}\, dk_x\, dk_y \times e^{ik_x x + ik_y y} e^{-i(k_x^2 + k_y^2)t/|k_0|}\, {}_0\tilde{h}(k_x, k_y, z - t) \equiv -\frac{\partial h}{\partial z} - i\frac{(\Delta k_\perp)^2}{|k_0|} h \quad \text{(defining } k_1 \text{ and } \Delta k_\perp\text{)}.$$

[18] L. D. Landau and E. M. Lifschitz, *Classical Theory of Fields* (Pergamon Press, Oxford, 1962), 2nd ed.

Although k_1 ($\sim |k_0|$) and $\Delta k_\perp$ are functions of position and time, we assume they are some representative constant values, and we take a test vector with constant components $\xi^t = a$, $\xi^z = b$. We then have

$$\mathfrak{L}_\xi h = ai[-k_1 + (\Delta k_\perp)^2/|k_0|]h + ibk_1 h$$

and

$$|\mathfrak{L}_\xi h|^2 = k_1^2\left[(b - a)^2 + a^2\left(\frac{(\Delta k_\perp)^2}{k_1 k_0}\right)^2 - 2a(b - a)\frac{(\Delta k_\perp)^2}{k_1 k_0}\right]|h|^2.$$

We consider only the ratio

$$\text{“}d\lambda\text{”} = |\mathfrak{L}_\xi h|^2/\xi^\alpha\xi_\alpha$$

without integrating and thereby avoid the question of divergent integrals. We take $a = 1$, $b = 1 + \epsilon$. The minimum of the ratio “$d\lambda$” is then given [for $(\Delta k_\perp)^2(k_1 k_0)^{-1}$ small] when

$$\epsilon = -\frac{(\Delta k_\perp)^2}{k_1 k_0} < 0.$$

Thus the minimizing vector is timelike and tends to a null vector as $(\Delta k_\perp)^2 \to 0$. The value of the ratio is, in this limit,

$$\text{“}d\lambda\text{”} \simeq 2\,[(\Delta k_\perp)^2/k_0]^2\,|h|^2.$$

The eigenvalue tends to zero if the wave vanishes or if it becomes more nearly plane-fronted. The deviation from symmetry goes quadratic in both the amplitude of the wave and the quantity $\Delta k_\perp(\Delta k_\perp/k_0)$.

The quadratic dependence on the amplitude of the wave is reasonable, since we are dealing with a minimum principle and the flat space is symmetric. The other factors can be understood by considering the behavior of a pulse of radiation. If the pulse did not spread and traveled at the speed of light, then there would be a null Killing vector and λ would be zero. The eigenvalue λ depends on $\Delta k_\perp$ as $[\Delta k_\perp(\Delta k_\perp/k_0)]^2$, since the spread is slower for pulses with larger center momentum. Finally, the “center of mass” of such a pulse does not travel at unit velocity but at a somewhat slower group velocity. Hence our minimizing vector is timelike: The vector gives the velocity that keeps an observer centered in a slowly diffusing wave packet. This is obviously the trajectory that keeps the space looking most time-independent to such an observer.

IV. APPLICATIONS TO SHORT GRAVITATIONAL WAVES

In the example of the flat torus, Sec. IIIA, we saw that the ground-state solution was sampling the

average curvature of the disturbance. This is interesting because the averaged curvature gives a sort of effective energy. If we agree to ignore small scale effects, then a solution which is averaged over many wavelengths of a rapidly varying ripple is the same as a slowly varying space which contains matter of the same density as the average energy density in the ripples. A very rapidly varying situation can mimic a slowly varying one if we have only coarse measuring instruments.

These ideas can be made precise, and perhaps the most thorough investigation of them has been by Isaacson,[5] who has considered the high-frequency limit for gravitational radiation. Because of the nonlinear nature of the Einstein equations, gravitational radiation produces a gravitational field with an effective-source stress tensor which is more or less the pseudotensor. (See, for instance, Ref. 18, p. 341.) Isaacson, following Brill and Hartle,[19] has given a method, when the rapidly varying wave can be separated out from the smoothly curving background, for averaging to obtain an effective stress tensor which is invariant over a wide class of gauge transformations, a much more general class than the usual pseudotensor treatment allows.

One problem he has considered is the following: Let $g_{\alpha\beta}$ be a vacuum metric, which admits a coordinate system (Isaacson's *steady coordinates*) such that the metric can be written

$$g_{\alpha\beta} = \gamma_{\alpha\beta} + \epsilon h_{\alpha\beta},$$

where the metric $\gamma_{\alpha\beta}$ is a slowly varying function of position; $h_{\alpha\beta}$ is a rapidly varying function of position and $h_{\alpha\beta}$ satisfies a certain generalized wave equation "$\Box$"$h_{\alpha\beta} = 0$ in the space $\gamma_{\alpha\beta}$. The statement that γ is a slowly changing background while h is rapidly changing means that $\langle h_{\mu\nu}\rangle = O(\lambda\!\!\!^{-}/L)$, where $\lambda\!\!\!^{-}$ is the (short) wavelength of the radiation and L is a typical length in the background. A discussion of the gauge invariance of the wave-background splitting is given in the Appendix; Isaacson has discussed the invariance of the average effective stress tensor [the r.h.s. of Eq. (13), below] and of Eq. (14). We demand that the averaged stress tensor (defined by Isaacson) for $h_{\alpha\beta}$ should give the background $\gamma_{\alpha\beta}$ when inserted as a source into the field equations for $\gamma_{\alpha\beta}$. We symbolically write

$$R^{(0)}_{\alpha\beta}(\gamma) = -\epsilon^2\langle R^{(2)}_{\alpha\beta}(h, \gamma)\rangle, \tag{13}$$

$$\epsilon R^{(1)}_{\alpha\beta}(h, \gamma) \equiv \epsilon\text{"}\Box\text{"}h_{\alpha\beta} = 0, \tag{14}$$

where the number 0, 1, 2, $\cdots$ refer to the powers of ϵ appearing in an expansion of the equation $R_{\alpha\beta}(\gamma + \epsilon h)$.

The equations must be solved simultaneously in a consistent manner, since (14) involves the background metric $\gamma_{\alpha\beta}$ which is obtained from (13) which involves $h_{\alpha\beta}$. Because of this consistency aspect, the derivatives of h must be of order ϵ^{-1}, i.e., $\partial^n h = O(\epsilon^{-n})$. This can easily be seen by the following argument due to Isaacson. Derivatives of the background are $\partial\gamma \sim L^{-1}\gamma$; those of h are $\partial h \sim h\lambda\!\!\!^{-1}$; the "energy density" in the wave is then $\rho \sim c^4G^{-1}\epsilon^2\lambda\!\!\!^{-2}$, and the curvature of the background is $\sim L^{-2}$. Then we have, by the Einstein equations,

$$R_{\alpha\beta} \sim L^{-2} \gtrsim \left(\frac{G}{c^4}\right)\left(\frac{c^4}{G}\right)\left(\frac{\epsilon}{\lambda\!\!\!^{-}}\right)^2 \Rightarrow \epsilon \lesssim \frac{\lambda\!\!\!^{-}}{L}.$$

If there is matter present which is also curving up the space, then the inequality holds; if the curvature is due totally to the wave, we have approximate equality $\epsilon \sim \lambda\!\!\!^{-}$ (we take $L = 1$). This means that $\lambda\!\!\!^{-} \to 0$ and $\epsilon \to 0$ are the same limit for a fixed background, and to emphasize this we write $O(\lambda\!\!\!^{-})$ instead of $O(\epsilon)$.

We write $\pounds_\xi\gamma_{\alpha\beta} = \dot\gamma_{\alpha\beta}$ and assume ξ is only slowly varying, so $\xi = O(1)$ and $\partial\xi = O(1)$. Then we find, recalling $\epsilon\,\partial h = O(1)$,

$$O(1) = 2\xi_{(\alpha;\beta)} = \dot\gamma_{\alpha\beta} + \epsilon h_{\alpha\beta,\sigma}\xi^\sigma + O(\lambda\!\!\!^{-}).$$

And, since $g^{\alpha\beta} = \gamma^{\alpha\beta} + O(\lambda\!\!\!^{-})$, we find

$$4\xi_{(\alpha;\beta)}\xi^{(\alpha;\beta)} = \dot\gamma_{\alpha\beta}\dot\gamma_{\mu\nu}\gamma^{\alpha\nu}\gamma^{\mu\beta} + 2\epsilon h_{\alpha\beta,\sigma}\xi^\sigma\dot\gamma_{\mu\nu}\gamma^{\mu\alpha}\gamma^{\nu\beta} + \epsilon^2 h_{\alpha\beta,\sigma}h_{\mu\nu,\rho}\xi^\sigma\xi^\rho\gamma^{\alpha\mu}\gamma^{\beta\nu} + O(\lambda\!\!\!^{-}). \tag{15}$$

To this order, the denominator in (1) is just $\int\sqrt{-\gamma}\,\xi^\alpha\xi^\beta\gamma_{\alpha\beta}\,d^4x$, and upon integration the first term in (15) yields a number depending only on ξ and $\gamma^{\alpha\beta}$, which we denote by $4\lambda_\gamma[\xi]$. Since we have assumed $\xi = O(1)$, $\partial\xi = O(1)$, and $\partial\gamma^{\alpha\beta} = O(1)$, and since it is the integral of a rapidly oscillating quantity $h_{\alpha\beta,\sigma}$ times slowly varying factors, the integral of the second term in (15), although *a priori* of order unity, is actually much smaller. Thus this term is at most $O(\lambda\!\!\!^{-})$ and we need consider only the last term, which (again noting the product of rapidly and slowly varying terms) we write as

$$\int\epsilon^2\langle\gamma^{\mu\alpha}h_{\alpha\beta,\sigma}\gamma^{\beta\nu}h_{\mu\nu,\rho}\rangle\xi^\sigma\xi^\rho\sqrt{-\gamma}\,d^4x + O(\lambda\!\!\!^{-}).$$

The average is over many wavelengths, but over a region which is much smaller than the scale of the slowly changing background. Let the colon denote the covariant derivative in the background, then noting that $O(\lambda\!\!\!^{-1}) = h_{\alpha\beta,\sigma} = h_{\alpha\beta:\sigma} + O(1)$, we write

[19] D. R. Brill and J. B. Hartle, Phys. Rev. 135, B271 (1967).

this term as

$$\int \epsilon^2 \langle \gamma^{\mu\alpha} h_{\alpha\beta;\sigma} \gamma^{\beta\nu} h_{\mu\nu;\rho} \rangle \xi^\sigma \xi^\rho \sqrt{-\gamma}\, d^4x + O(\lambda\lambda) = 32\pi \int T^{(\mathrm{av})}_{\sigma\rho} \xi^\sigma \xi^\rho \sqrt{-\gamma}\, d^4x + O(\lambda\lambda).$$

Here

$$T^{(\mathrm{av})}_{\sigma\rho} = \frac{\epsilon^2}{32\pi} \langle \gamma^{\mu\alpha} h_{\alpha\beta;\sigma} \gamma^{\beta\nu} h_{\mu\nu;\rho} \rangle$$

is the average stress tensor of gravitational waves as defined by Isaacson. It is this average stress tensor which determines the background $\gamma_{\alpha\beta}$ according to $R_{\alpha\beta}(\gamma) = 8\pi(T^{(\mathrm{av})}_{\alpha\beta} - \frac{1}{2}\gamma_{\alpha\beta} T^{(\mathrm{av})})$. This term is clearly also independent of λ as $\lambda \to 0$, and we find

$$\lambda[\xi] = \lambda_\gamma[\xi] + \lambda_{\mathrm{rad}}[\xi] \tag{16}$$

for high-frequency radiation, where

$$\lambda_{\mathrm{rad}}[\xi] = \frac{\int 8\pi \xi \cdot \overleftrightarrow{T}^{(\mathrm{av})} \cdot \xi \sqrt{-\gamma}\, d^4x}{\int \xi \cdot \xi \sqrt{-\gamma}\, d^4x}.$$

Both terms are independent of λ in the short-wavelength limit. If ξ is timelike, then[20] $\xi \cdot \overleftrightarrow{T}^{(\mathrm{av})} \cdot \xi \geq 0$.

We note, moreover, that the additional term λ_{rad} allows in some sense a distinction to be made between fluid and gravitational-radiation-filled universes; the eigenvalue λ is lower for fluid universes with the same large scale "shape." Since λ_γ can be bounded by a curvature (in the background $\gamma_{\alpha\beta}$) and since $\overleftrightarrow{T}^{(\mathrm{av})}$ is clearly also a curvature in the background metric, we see that $\lambda[\xi]$ is indeed sampling only the large-scale curvature of the space and "smoothing over" the ripples, as might be expected from an integral estimate.[21] This derivation makes explicit the result suggested by the torus calculation in Sec. IIIA.

To carry the discussion of the energy density in gravitational radiation a bit further, let us consider the Robertson–Walker (R–W) metrics, which have the form

$$ds^2 = -d\tau^2 + \Omega^2(\tau)\, d\sigma^2, \tag{17}$$

where $d\sigma^2$ is the line element of the homogeneous- and isotropic-space sections; these sections are flat or have unit (by choice of length scale) positive or negative curvature. Because of the high symmetry, these sections each have six Killing vectors: translation along and rotation around each of the axes for the flat sections; two sets of "rotations" for the curved ones. (For a discussion of the metrics of this type, see, e.g. Ref. 22.)

Because the τ-constant sections in these metrics have Killing vectors, if we form the ratio (for ξ_i a vector in the 3-space, described by the 3-metric ${}^3g_{ij} \equiv g_{ij}$, ${}^3g_{il}\,{}^3g^{ik} = \delta^k_l$)

$${}^{(3)}\lambda[\xi] = \frac{\int \xi^{(i|j)} \xi_{(i|j)} \sqrt{{}^3g}\, d^3x}{\int \xi_i \xi_j\, {}^3g^{ij} \sqrt{{}^3g}\, d^3x},$$

${}^{(3)}{}_0\lambda = 0$, since we can pick ξ Killing. In this case, because of the symmetry we have ${}^{(3)}{}_0\lambda = 0$, even though there is a length scale introduced by $\Omega(\tau)$.

For an example of a space which is on the large scale identical to the R–W types, and in spirit of recent observational discoveries, we suppose that the universe at the present time is given by an R–W form and that its behavior is dominated by the matter in it, but that it contains 3°K blackbody *gravitational* radiation distributed in a uniform and isotropic way through the universe.[23] In the "now" constant-time slice, we compute ${}^{(3)}\lambda[\xi]$ for ξ Killing in the background, and, by arguments like those leading to Eq. (16), we find

$${}^{(3)}\lambda[\xi] = {}^{(3)}\lambda_{\mathrm{rad}}[\xi] = \frac{8\pi \int T^{(\mathrm{av})}_{ij} \xi^i \xi^j \sqrt{{}^3g}\, d^3x}{\int {}^3\gamma_{ij} \xi^i \xi^j \sqrt{{}^3\gamma}\, d^3x},$$

where here γ_{ij} is the background R–W metric. Because of the assumed isotropy and homogeneity of both the background and the radiation, $T^{(\mathrm{av})}_{ij}$ must be proportional to γ_{ij} where the proportionality factor is $\rho/3$, since this is a massless radiation field. Since $\rho = \rho(\tau)$ is constant on space slices, we have ${}^{(3)}\lambda[\xi] = {}^{(3)}\lambda_{\mathrm{rad}}[\xi] = (8\pi/3)\rho$. Here ρ is the energy density due to the gravitational radiation, i.e., the density appropriate to 3°K blackbody radiation.

[In this simple situation, it is clear that a minimum with respect to the background is a minimum in the full metric, since $\lambda = \lambda_\gamma + \text{const}$. In fact, by an argument similar to that for first-order perturbation theory in quantum mechanics, it is easy to see that to first order in $\overleftrightarrow{T}^{(\mathrm{av})}$, the minimum is given by the same vector field in the full metric as in the background. One can

[20] Reference 5, Eq. (4.1) *et seq.*

[21] We must still justify the assumption that $\partial\xi \sim 1$ for the ground state. We note that to change the integrals of the second and third terms in Eq. (15) would require $\partial\xi \sim \lambda^{-1}$, so that oscillations in ξ will be "in phase" with those in the radiation field. In this case, the second integral would be $O(\lambda^{-1})$, while the third would still be $O(1)$. The first integral would then, however, be $\lambda_\gamma[\xi] = O(\lambda^{-2})$. So we conclude that the ground-state eigenvalue can certainly be made smaller by taking a slowly varying vector field, and the ground-state eigenvalue is given by Eq. (16).

[22] S. W. Hawking, Astrophys. J. **140**, 1 (1967).

[23] The observations of 3°K electromagnetic radiation show that it is in fact remarkably isotropic. See R. B. Partridge and D. T. Wilkenson, Phys. Rev. Letters **18**, 557 (1967).

also calculate the first-order correction to the vector fields exactly as is done in nonrelativistic quantum theory:

$$\zeta_{(m)} = \xi_{(m)} + 8\pi \sum_{k \neq m} \left(\int \xi_{(m)} \cdot \overleftrightarrow{T}^{(\mathrm{av})} \cdot \xi_{(k)} \, dV \right) \frac{\xi_{(k)}}{{}_m\lambda - {}_k\lambda},$$

where $\xi_{(i)}$ is the ith eigenvector of $-\mathfrak{D}$ in the background, ${}_i\lambda$ is the corresponding eigenvalue, and $\zeta_{(i)}$ is the ith eigenvector of the perturbed $-\mathfrak{D}$. In this formula, the vectors $\xi_{(i)}$ and $\zeta_{(i)}$ are considered referred to the background space $\gamma_{\alpha\beta}$.]

Since 3°K blackbody radiation has an energy density of $\sim 10^{-34}$ g/cc, we see that this yields a much longer characteristic length than that provided by the large-scale background. (The background scale is necessarily shorter than that given by the amount of matter observed in the galaxies $\sim 10^{-30}$ g/cc, and even shorter—magic density $\sim 10^{-29}$ g/cc—if we assume that there is sufficient deceleration to close the universe with the observed Hubble velocity.)

Actually, the energy density of the radiation contributes to the curving of the background. In the previous example, the radiation was as symmetric in the large as in the background.[24] However, if the background is determined by some other factor, say a distribution of dust or electromagnetic radiation, then the gravitational radiation, if weak enough, will not change the background significantly and such an integral over the different Killing vectors in the background will give six different numbers (for R–W background) characterizing the "stress" in the gravitational radiation.

This result is perhaps the most interesting of this work. We have here apparently an invariant method for specifying some parameters of gravitational radiation.

V. OUTLOOK

The method presented here in Eqs. (1) and (2) is a straightforward generalization of the idea of a Killing field. The differential equation (2), in spaces of Minkowski signature, can be considered a coordinate condition for the time, say. Detailed investigation of this idea may yield very useful results in the future.

We also have left for future investigation the question of using the vector fields defined by these recipes to give new candidates for conserved momentum or energy objects. This may also prove quite a fruitful field of investigation.

Perhaps the most significant and unexpected results of the ideas in this work are their applications to spaces which contain short-wave gravitational radiation, and their uses in specifying some numerical parameters for the radiation. As we have seen, the ground-state eigenvalue measures, for different spaces with the same large-scale shape, the contributions of the gravitational-energy density in the space. This gives an additional length scale $\lambda_{\mathrm{rad}}^{-\frac{1}{2}}[\xi]$ in addition to the sizes obviously present in a situation with large-scale size L and Riemann-tensor variations on a length scale $\lambda\!\!\!^{-}$. With each invariantly defined vector in the background, there is an associated length scale which measures some component of the stress of the gravitational energy.

The question remains: Can a complete specification of the space be done in this way? Restricting consideration to just the ground state ξ_0, this does not seem possible because the eigenvalue ${}_0\lambda$ contains parts due to the background as well as due to the radiation and there seems to be no way to separate them. However, it is plausible that the entire spectrum of the operator $-\mathfrak{D}$ may give a sufficiently powerful specification of the space it is expressed in, that the complete solution, background plus radiation, can be expressed in an expansion in terms of the eigenvectors and eigenvalues of $-\mathfrak{D}$. While this complete specification may be overambitious, it is hoped that a more clear-cut identification of the spaces which satisfy Isaacson's requirement (that they admit a "steady" coordinate system) may be possible by these methods. This is certainly an important application, if it can in fact be done. A metric which appears to contain gravitational radiation can always be analyzed by computing the Riemann-tensor components in a tetrad frame. If there is radiation present which is curving up the space, we have seen that its characteristic Riemann tensor will be typically

$$R.... \sim \epsilon \, \partial^2 h = O(\epsilon^{-1}) = O(\lambda\!\!\!^{-1}),$$

and so will be overwhelming in the short-wavelength limit. The integral method given here, when applied to a space of the Isaacson type, gives a finite result for the integrals involved, as $\lambda\!\!\!^{-} \to 0$, and in fact the limit is of the order of the large-scale background curvature. The Isaacson metrics can be singled out from among the (perhaps) wider class of metrics which have Riemann-tensor variations, in a tetrad frame, of order $\lambda\!\!\!^{-1}$ on a length scale $\lambda\!\!\!^{-}$.

One other interesting application for these ideas may be to define, invariantly, a background metric in any situation, by averaging the metric along the eigenvector fields of $-\mathfrak{D}$.

[24] The gravitational radiation can supply the entire energy density curving up the space. The solution with the R–W symmetries was discussed by D. R. Brill, Nuovo Cimento Suppl. Vol. II, No. 1 (1964). The value of $\lambda[\xi]$ for ξ Killing in the background of this space gives $\rho_g = -\ddot{\Omega}(\tau)/\Omega(\tau)$, the deceleration parameter for the large-scale evolution.

ACKNOWLEDGMENTS

I gratefully acknowledge helpful conversations with Dr. R. E. Burridge, Dr. F. G. Friedlander, and Dr. S. W. Hawking. I am especially grateful to Professor C. W. Misner for the advice and encouragement he has given me during the course of this work. This work was begun while I held an NSF Graduate Fellowship at the University of Maryland.

APPENDIX: GAUGE INVARIANCE OF BACKGROUND-WAVE SPLITTING

We assume, as stated in Sec. IV, that there exists one coordinate system (steady coordinates) such that we can write

$$g_{\mu\nu} = \gamma_{\mu\nu} + \epsilon h_{\mu\nu}, \tag{A1}$$

with

$$\gamma = O(1), \quad h = O(1), \tag{A2}$$

where $\gamma_{\mu\nu}$ is a slowly varying function of position, and there is some large length L (which we set equal to 1) such that

$$L\partial\gamma = O(1), \quad L^2\partial^2\gamma = O(1), \tag{A3}$$

while $h_{\mu\nu}$ is a rapidly varying function of position with the property that (from the field equations, as discussed in Sec. IV)

$$\epsilon^m \partial^m h = O(1). \tag{A4}$$

Since we assume h is a high-frequency radiation, we assume this holds through $m = 2$.

Furthermore, since we demand γ be a background, we assume

$$\langle h \rangle = O(\epsilon), \tag{A5}$$

where this average is over a linear extent l, such that $\epsilon \ll l \ll 1 = L$. Coordinates which fulfill this requirement are acceptable "steady coordinates."

The first problem to discuss is whether an average can be done in a meaningful way. We contend that it can, because we can make a correspondence between the geodesics in g and those in γ.

We do this by looking for a solution to the geodesic equation of the form $x = {}^0x + \epsilon^2\,{}^2x$. That such a solution exists can be seen by writing out the geodesic equation

$$\frac{d^2({}^0x^\mu + \epsilon^2\,{}^2x^\mu)}{ds^2} + ({}^0\Gamma^\mu{}_{\nu\sigma} + {}^h\Gamma^\mu{}_{\nu\sigma})\frac{d\,{}^0x^\mu}{ds}\frac{d\,{}^0x^\sigma}{ds} = O(\epsilon), \tag{A6}$$

where ${}^h\Gamma^\mu{}_{\nu\sigma}$ are the terms containing derivatives of $h_{\alpha\beta}$, and where we assume ${}^2x$ is rapidly varying solution, so $\epsilon^2\partial^2\,{}^2x = O(1)$. Since solutions of the geodesic equation are unique, there will thus be a ${}^2x(s)$ correction of the form we have written. Further, by averaging this Eq. (A6) (in any crude way), we find the background geodesic equation, which identifies the parameter s.

We then suppose that we want to do averages by parallel transport along geodesics in g. We see that some quantity A, when carried back along geodesics in g, differs from the same quantity carried back along a geodesic in γ by a term of order $\epsilon^2(\partial A/\partial x)$. Since all our high frequency quantities have $\partial A/\partial x \sim \epsilon^{-1}A$, we see that this term is small and vanishes in the high frequency limit. So we can do averages in γ, and make small errors; in the limit $\epsilon \to 0$, we obtain the same result as if we carry back along geodesics in g.

We thus postulate the following sort of average:

$$\langle T^{\mu\nu}(x)\rangle = \int_{N_4} \sqrt{\gamma}\, d^4x T^{\mu'\nu'}(x') f(x, x') \gamma^{\mu\nu}_{\mu'\nu'},$$

where f is a smooth and slowly varying weighting function with $\partial f \sim f$ and $\int f = 1$. $\gamma^{\mu'\nu'}_{\mu\nu}$ is a bitensor[25] (with the appropriate rank) of parallel transport along the geodesics in γ. N_4 is a region in γ on which a geodesic coordinate patch centered on x is nonsingular; so $L_N \sim L = 1$.

We want f to fall off smoothly with "Euclidean distance" from x. While we could make these ideas rigorous and give estimates of errors, here we simply take the following viewpoint.

Consider some l such that $\lambda \ll l \ll L_N$. At x, construct a Minkoskian coordinate system. Since $l \ll L_N$, we can neglect the deviation from flatness and treat the Minkowski coordinates x, y, z, and t as in flat space. In this frame,

$$\langle T^{\mu\nu}(x)\rangle = \int_{N_4} d^4y f(x - y) T^{\mu\nu}(y),$$

$$N_4 = \left\{ y \middle| \sum_{i=0}^{3} (x^i - y^i)^2 \leq l^2 \right\},$$

agrees with our previous definition sufficiently well, where now f falls off within a Euclidean distance l from y.

This average is the sense in which we mean $\langle h \rangle = O(\epsilon)$ as $\epsilon \to 0$.

Now consider coordinate ripples. First consider a high-frequency transformation:

$$x \to x + \sigma\xi, \quad g_{\mu\nu} \to g_{\mu\nu} - 2\sigma\xi_{(\mu;\nu)}.$$

We want this to be a small-gauge transformation, so σ is a small parameter and $\xi_{(\mu;\nu)} = O(1)$. However, our high-frequency assumption demands $\xi = O(\sigma)$, since derivatives are large.

[25] J. L. Synge, *Relativity: The General Theory* (North-Holland Publ. Co., Amsterdam, 1964).

With this estimate of the orders, we find

$$\epsilon h \to \epsilon h - 2\sigma\xi_{(\mu,\nu)}$$

as the change in h under the coordinate change, to first order in ϵ and σ.

We also have ${}^0\gamma \to \gamma = {}^0\gamma + \delta\gamma$, the change in γ being given as $\delta\gamma$. We will show that $\delta\gamma$ is small enough so that our separation criteria (A1)–(A5) are still satisfied through $O(\epsilon)$ or $O(\sigma)$.

By a discussion similar to the one above, we can see that the geodesics in g are still given to order ϵ or σ by the geodesics in γ. Thus $\langle h\rangle$ is unchanged at order ϵ, so $\langle h\rangle = O(\epsilon)$ is still true, and so we must consider only the quantity $\delta\gamma = -2\langle\sigma\xi_{(\mu,\nu)}\rangle$, which might contribute a secular shift in γ. We find

$$\langle\xi_{\mu,\nu}\rangle = \int_{N_4} d^4x'(\xi_{\mu'}f(x'-x)\delta^{\gamma'}_{\nu'})_{,\gamma'} - \int_{N_4} d^4x'\xi_{\mu'}f_{,\nu'} = O(\sigma). \quad \text{(A7)}$$

The first term vanishes because it can be taken to a surface integral where f vanishes; the second is of order σ because of our assumptions on ξ.

Thus we find

$$\gamma_{\mu\nu} = {}^0\gamma_{\mu 0} + \sigma^2\rho_{\mu\nu}, \quad \text{(A8)}$$

where $\rho_{\mu\nu}$ is of order one, and so we have all our criteria (A1)–(A5) satisfied by the split

$$\gamma + (\epsilon h - 2\sigma\xi_{(\mu,\nu)})$$

if they were initially satisfied by the wave–background separation $\gamma + \epsilon h$.

We see that

$$\begin{aligned}\gamma &= {}^0\gamma(1 + O(\sigma^2)),\\ \partial\gamma &= \partial\,{}^0\gamma(1 + O(\sigma)),\\ \partial^2\gamma &= \partial^2\,{}^0\gamma(1 + O(1)),\end{aligned} \quad \text{(A9)}$$

which estimates allow that $\rho_{\mu\nu}$ may be rapidly varying, so $\partial^n\rho_{\mu\nu} = O(\sigma^{-n}\rho_{\mu\nu})$.

Though γ still satisfies the criteria (A1)–(A5), we may wonder about $R^{(0)}_{\mu\nu\sigma\tau}(\gamma)$, since we see that second derivatives of γ may change by substantial amounts. To investigate this, we change variables in (A7) to $y' = x' - x$ to find

$$\begin{aligned}\langle\xi_{\mu,\nu}\rangle &= \int_{N_4}\xi_{\mu'}(y'+x)_{,\nu'}f(y')\,d^4y'\\ &= \frac{\partial}{\partial x^\nu}\left(\int_{N_4}\xi_\mu(y'+x)f(y')\,d^4y\right),\end{aligned}$$

so the correction term is $\rho_{\mu\nu} = \rho_{(\mu,\nu)}$.

Counting orders, we see that it is only those terms in $R^{(0)}_{\mu\nu\sigma\tau}$ which involve two derivatives on $\rho_{\mu\nu}$ which are of order one. But because of the symmetries of the Riemann tensor, these terms exactly cancel one another. This is familiar from linear gravity theory, and may be checked explicitly. Thus

$$R^{(0)}_{\mu\nu\sigma\tau}(\gamma) = R^{(0)}_{\mu\nu\sigma\tau}({}^0\gamma)[1 + O(\sigma)].$$

If we consider long-wavelength coordinate transformations, we have $\xi = O(1)$, $\partial\xi = O(1)$, and then

$$g \to \gamma_{\mu\nu} - \sigma\mathcal{L}_\xi\gamma_{\mu\nu} + \epsilon\sigma h_{\mu\nu,\lambda}\xi^\lambda + \epsilon h_{\mu\nu}.$$

The first gauge term consists of a slow coordinate change in $\gamma_{\mu\nu}$ and so is easily handled, and, for instance, allows precisely the same local Minkowskian space for averages as was discussed above. Since this is so, we still have $\langle h\rangle$ unchanged and $\langle h\rangle = O(\epsilon)$, and we must consider only

$$\langle h_{\mu\nu,\lambda}\xi^\lambda\rangle = \int(\xi^\lambda f h_{\mu\nu})_{,\lambda}\,d^4x - \int h_{\mu\nu}(\xi^\lambda f)_{,\lambda}\,d^4x = O(\epsilon),$$

so the new term $\epsilon\sigma\langle h_{\mu\nu,\lambda}\xi^\lambda\rangle$ is of third order in smallness, and again γ is invariant, as is $R^{(0)}_{\alpha\beta\sigma\tau}(\gamma)$.

We note that this invariance of $R^{(0)}_{\alpha\beta\sigma\tau}$ corroborates Isaacson's finding[5] that $T^{(\mathrm{av})}_{\mu\nu}$ is invariant to this order, where

$$\begin{aligned}R^0_{\mu\nu}(\gamma) - \tfrac{1}{2}\gamma_{\mu\nu}R^{(0)} &= 8\pi T^{(\mathrm{av})}_{\mu\nu}\\ &\equiv -\epsilon^2\langle R^{(2)}_{\mu\nu} - \tfrac{1}{2}\gamma_{\mu\nu}R^{(2)}\rangle.\end{aligned} \quad \text{(A10)}$$

Isaacson[5] showed the invariance of the r.h.s. of (A10) under small-gauge transformations; we have just shown the invariance of the left-hand side.

Isaacson also discussed[5] the invariance of the field equation obeyed to lowest order by h:

$$R^{(1)}_{\cdots\cdot}(\gamma + \epsilon h) \equiv \text{“}\Box\text{”}h = 0.$$

This follows for the high-frequency coordinate transformations because of the symmetry $\xi_{(\mu,\nu)}$ of the gauge term and the symmetries of $R^{(1)}_{\mu\nu\sigma\tau}$, so that the highest-order terms exactly cancel one another. In the case of low-frequency coordinate transformations, the dominant terms in “$\Box$”h are $\partial^2 h_{\mu\nu}$ (ordinary derivatives), so the dominant contribution from the gauge term is $\epsilon\sigma\xi^\lambda\partial^2 h_{\alpha\beta,\lambda}$. This vanishes because “$\Box$”$h = 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Multiplicity Problem for Compact Subgroups of Noncompact Groups

WILLIAM H. KLINK*
Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa, and Argonne National Laboratory, Argonne, Illinois

(Received 6 November 1966)

A method is shown for calculating the multiplicities occurring in the reduction of a reducible representation of a compact group. The reducible representation is obtained by constructing induced irreducible representations of a noncompact group and then restricting the elements of this representation to elements of the compact subgroup.

I. INTRODUCTION

One of the ways in which noncompact groups are being used in elementary particle physics is the formation of supermultiplets which have an infinite number of "particles" in them. One assumes that particles are grouped into finite-dimensional multiplets; then an infinite number of these finite-dimensional multiplets are needed to "fill up" the infinite-dimensional supermultiplet which is generated by the noncompact group.

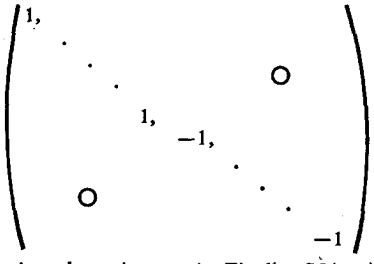
For example, the nonrelativistic hydrogen atom has an infinite number of bound states, classified by the "n" quantum number. The finite-dimensional multiplets of the nonrelativistic hydrogen atom are the angular-momentum multiplets, coming from the group $SO(3)$,[1] which are contained in the "hidden degeneracy" finite-dimensional multiplets arising from the group $SO(4)$. These multiplets are then brought together into an infinite-dimensional supermultiplet generated by the group $SO(4, 1)$.[2]

Other schemes, such as that of Cook, Goebel, and Sakita,[3] take finite-dimensional meson multiplets to form an infinite-dimensional supermultiplet, while Rühl and Fronsdal[4] use as their starting point the fact that $SU(6)$, which couples ordinary spin and the internal $SU(3)$ symmetry, is contained in the noncompact group $SL(6, C)$. Gell-Mann, Ne'eman, and Dothan[5] discuss the possible uses of other groups, like $U(6, 6)$ and $SL(3, R)$, in forming infinite chains of multiplets. Finally, Barut has shown[6] how one can use noncompact groups to classify the multiplets coming from the n-dimensional harmonic oscillator and the rigid rotator.

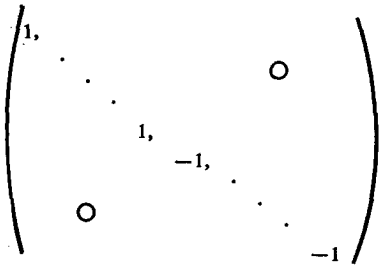
Now the mathematical problem which is common to all these examples is to find an irreducible unitary representation of a noncompact group which gives the correct energy levels and the degeneracy of these levels for a given physical system. The physical system generally has finite-dimensional multiplets coming from a compact group—which in turn is an expression of the symmetry (or approximate symmetry) of the system. One assumes that there are an infinite number of these "symmetry multiplets" and they

* Permanent address: Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa.

[1] The notation for groups is as follows: $SL(n, C)$, $SL(n, R)$ denotes the group of $n \times n$ unimodular matrices over the complex and real fields, respectively. $SU(n, m)$ is the set of complex unimodular $(n + m) \times (n + m)$ matrices leaving invariant a form



with n plus-one's and m minus-one's. Finally, $SO(p, q)$ is the set of real unimodular $(p + q) \times (p + q)$ matrices leaving invariant the form



with p plus-one's and q minus-one's.

[2] H. Bacry, Nuovo Cimento **41**, A, 222 (1966); E. C. G. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, Phys. Letters **19**, 322 (1965); M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966); A. O. Barut, P. Budini, and C. Fronsdal, Proc. Roy. Soc. (London) **A291**, 106 (1966).

[3] T. Cook, C. J. Goebel, and B. Sakita, Phys. Rev. Letters **15**, 35 (1965); C. J. Goebel, T. Cook, and B. Sakita, "Proceedings of the Third Coral Gables Conference on Symmetry Principles at High Energy (1966);" T. Cook and B. Sakita, "Induced Representations of Strong Coupling Groups," preprint, Argonne National Laboratory, May (1966).

[4] P. Budini and C. Fronsdal, Phys. Rev. Letters **14**, 968 (1965); W. Rühl, Nuovo Cimento **44**, 572 (1966). See also the articles by C. Fronsdal and W. Rühl, *High Energy Physics and Elementary Particles* (IAEA, Vienna, 1965).

[5] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Rev. Letters **17**, 148 (1965).

[6] A. O. Barut, Phys. Rev. **139**, B1107, B1433 (1965); P. Budini, Nuovo Cimento **44**, 363, 418 (1966); A. Böhm, "Dynamical Groups of Simple Non-relativistic Models," preprint, International Center for Theoretical Physics, IC/65.82; N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965); B. Vitale, three lectures given at the Institute for Mathematical Sciences, Madras, India (March, 1966).

are to be properly arranged with respect to the unitary irreducible representation of some noncompact group. Thus, if G is some noncompact group and K is a compact subgroup of G, the mathematical question to be asked is, How often does an irreducible representation of K appear in an irreducible representation of G?

Let $\mathfrak{G}(g)$ be an irreducible unitary representation of G, while $\mathfrak{K}(k)$ is an irreducible unitary representation of K.[7] Then $\mathfrak{G}(k)$ is a reducible representation of K and can be decomposed into a direct sum of irreducible representations of K:

$$\mathfrak{G}(k) = \sum_{\mathfrak{K}} n(\mathfrak{G}, \mathfrak{K})\mathfrak{K}(k), \tag{1}$$

where $n(\mathfrak{G}, \mathfrak{K})$ is the multiplicity; that is, the number of times the representation $\mathfrak{K}(k)$ appears in the decomposition of the reducible representation $\mathfrak{G}(k)$. In applications of noncompact groups to particle physics, an important problem is to get the correct multiplicity. This paper will demonstrate a simple technique for finding $n(\mathfrak{G}, \mathfrak{K})$ for a class of representations of G, namely, the induced representations. Such representations have been extensively studied by Mackey[8] and Gel'fand and his collaborators,[9] and generate a large number of the known irreducible representations of the noncompact groups. In what follows, no prior knowledge of induced representation theory is assumed, so that Sec. II will present an outline of the relevant theory needed for the later sections, which deal with the multiplicity problem for semisimple groups (Secs. III and IV) and semidirect product groups (Sec. V).

II. INDUCED REPRESENTATION THEORY

The main results of Mackey's work needed in this paper are what he calls the subgroup theorem and the Frobenius reciprocity theorem. To understand these results it is necessary to explain what is meant by an induced representation.[10]

Let G be a given noncompact group and H_1 a subgroup of G (H_1 may, for example, be the compact subgroup K mentioned above). Let $\mathfrak{H}_1$ be an irreducible representation of H_1 acting on the vector space $\mathfrak{V}(\mathfrak{H}_1)$. Consider functions f which map elements g of G into $\mathfrak{V}(\mathfrak{H}_1)$:

$$g \xrightarrow{f} \mathfrak{V}(\mathfrak{H}_1), \quad f(g) \in \mathfrak{V}(\mathfrak{H}_1). \tag{2}$$

[Often $\mathfrak{V}(\mathfrak{H}_1)$ will be a one-dimensional space—then f carries the elements g into the complex numbers.] Now consider the set of functions $f(g)$ satisfying the condition $f(h_1 g) = \mathfrak{H}_1(h_1)f(g)$ for all h_1 in H_1 and g in G. This set forms a vector space

$$\hat{\mathfrak{V}}(\mathfrak{H}_1) = \{f \mid f(g) \in \mathfrak{V}(\mathfrak{H}_1), \quad f(h_1 g) = \mathfrak{H}_1(h_1)f(g), \quad \forall h_1 \in H_1, g \in G\}. \tag{3}$$

The representation defined by $\mathfrak{U}^{\mathfrak{H}_1}(g')f(g) = f(gg')$ on the vector space $\hat{\mathfrak{V}}(\mathfrak{H}_1)$ is called the induced representation of G. The induced representation $\mathfrak{U}^{\mathfrak{H}_1}(g')$ will be a unitary representation of G if it is possible to make $\hat{\mathfrak{V}}(\mathfrak{H}_1)$ into a Hilbert space possessing the correct measure with respect to the group G. That is, it is necessary to find a measure $\mu(g_1, g_2)$ such that

$$(f(g_1), f'(g_2)) = \int f^*(g_1)f'(g_2)\, d\mu(g_1, g_2) \tag{4}$$

(where g_1, g_2 are elements of G and * means complex conjugation). Then unitarity means

$$\begin{aligned}(\mathfrak{U}^{\mathfrak{H}_1}(g')f(g_1), \mathfrak{U}^{\mathfrak{H}_1}(g')f'(g_2)) &= \int [\mathfrak{U}^{\mathfrak{H}_1}(g')f(g_1)]^*[\mathfrak{U}^{\mathfrak{H}_1}(g')f'(g_2)]\, d\mu(g_1, g_2) \\ &= \int f^*(g_1 g')f'(g_2 g')\, d\mu(g_1, g_2) \\ &= \int f^*(g_1)f'(g_2)\, d\mu(g_1 g'^{-1}, g_2 g'^{-1}).\end{aligned} \tag{5}$$

If it is possible to find a measure satisfying

$$\int d\mu(g_1, g_2) = \int d\mu(g_1 g'^{-1}, g_2 g'^{-1}), \tag{6}$$

then

$$(\mathfrak{U}^{\mathfrak{H}_1}(g')f(g_1), \mathfrak{U}^{\mathfrak{H}_1}(g')f'(g_2)) = (f(g_1), f'(g_2)), \tag{7}$$

and the representation is unitary. References will be given in Secs. IV and VI which show how measures can be constructed for certain classes of induced representations of noncompact groups. Notice that it is not necessary that $\mathfrak{H}_1$ be a unitary representation of H in order that $\mathfrak{U}^{\mathfrak{H}_1}(g)$ be a unitary representation of G. However, it is a necessary (but not sufficient) condition that $\mathfrak{H}_1$ be irreducible in order that $\mathfrak{U}^{\mathfrak{H}_1}(g)$ be irreducible.

With the above definition of an induced representation, it is possible to paraphrase Mackey's subgroup theorem in the following way: Let H_1, H_2 be subgroups of the noncompact group G and let $\mathfrak{H}_1$ be an

[7] The notation to be used is that capital letters will denote the various groups while small letters will denote elements of the groups. Script letters will be used to denote the representations of a group. A superscript on a Script letter indicates a particular representation or class of representations.

[8] G. W. Mackey, "The Theory of Group Representations," Dept. of Mathematics, The University of Chicago (1955).

[9] I. M. Gel'fand and M. A. Neumark, *Unitäre Darstellungen der Klassischen Gruppen* (Akademie Verlag, Berlin, 1957).

[10] See Ref. 8, pp. 135 and 146 for the subgroup theorem and the Frobenius reciprocity theorem, respectively. Only a restricted definition of induced representation will be used in this paper; for the more general definition, see Ref. 8, p. 119.

irreducible representation of H_1. Assume further that it is possible to write $G = H_1H_2$. Then $\mathfrak{U}^{\mathcal{H}_1}(g)$ restricted to elements of H_2 is equivalent to $\mathfrak{U}^{\mathcal{J}}(h_2)$, where $\mathcal{J}$ is the restriction of the representation $\mathcal{H}_1$ to the subgroup $H_1 \cap H_2 \equiv J$; that is, if $\mathcal{H}_1(j) \equiv \mathcal{J}(j)$, then

$$\mathfrak{U}^{\mathcal{H}_1}(h_2) \cong \mathfrak{U}^{\mathcal{J}}(h_2), \tag{8}$$

where $\cong$ means "is equivalent to." Thus, the subgroup theorem allows one to write the reducible representation $\mathfrak{U}^{\mathcal{H}_1}(h_2)$ of H_2 as a reducible representation induced by a subgroup of H_2.

To carry the decomposition further it is necessary to introduce the Frobenius reciprocity theorem. Let J be a subgroup of the compact group K, and let $\mathcal{J}$, $\mathcal{K}$ be irreducible representations of J and K, respectively. Then the Frobenius reciprocity theorem says that $\mathfrak{U}^{\mathcal{J}}(k)$ contains $\mathcal{K}(k)$ the same number of times as $\mathcal{K}(j)$ contains $\mathcal{J}(j)$:

$$\mathfrak{U}^{\mathcal{J}}(k) \cong \sum_{\mathcal{K}} n(\mathcal{K}, \mathcal{J})\mathcal{K}(k),$$
$$\mathcal{K}(j) \cong \sum_{\mathcal{J}} n(\mathcal{K}, \mathcal{J})\mathcal{J}(j). \tag{9}$$

Now let $K = H_2$. Then Eqs. (8) and (9) can be combined to give

$$\mathfrak{U}^{\mathcal{H}_1}(k) \cong \mathfrak{U}^{\mathcal{J}}(k) \cong \sum_{\mathcal{K}} n(\mathcal{K}, \mathcal{J})\mathcal{K}(k). \tag{10}$$

But the multiplicity $n(\mathcal{G}, \mathcal{K})$ of Eq. (1) is what we are interested in finding and, according to Eq. (10), this is equal to $n(\mathcal{K}, \mathcal{J})$. Further, $n(\mathcal{K}, \mathcal{J})$ can be obtained solely from the compact groups J, K. Thus, the problem of finding the multiplicity $n(\mathcal{G}, \mathcal{K})$ has been reduced to the problem of finding the multiplicity $n(\mathcal{K}, \mathcal{J})$ and this problem can be solved in several different ways.

For example, since J and K are compact groups, one can use the characters of the irreducible representations of J and K to compute $n(\mathcal{K}, \mathcal{J})$. Let $\chi^{\mathcal{K}}(k)$ be the character of the irreducible representation $\mathcal{K}(k)$, while $\varphi^{\mathcal{J}}(j)$ is the character of the irreducible representation $\mathcal{J}(j)$. Then taking the trace of the lower equation in (9) gives

$$\chi^{\mathcal{K}}(j) = \sum_{\mathcal{J}} n(\mathcal{K}, \mathcal{J})\varphi^{\mathcal{J}}(j). \tag{11}$$

The characters satisfy orthogonality relations, so that

$$\int \varphi^{\mathcal{J}}(j)\varphi^{\mathcal{J}'*}(j)\, d\mu(j) = \delta_{\mathcal{J},\mathcal{J}'}, \tag{12}$$

where $d\mu(j)$ is the Haar measure on the group J. Multiplying Eq. (11) by $\varphi^{\mathcal{J}'*}(j)$ and integrating with respect to $d\mu(j)$ gives

$$\int \chi^{\mathcal{K}}(j)\varphi^{\mathcal{J}'*}(j)\, d\mu(j) = \int \sum_{\mathcal{J}} n(\mathcal{K}, \mathcal{J})\varphi^{\mathcal{J}}(j)\varphi^{\mathcal{J}'*}(j)\, d\mu(j)$$
$$= \sum_{\mathcal{J}} n(\mathcal{K}, \mathcal{J})\delta_{\mathcal{J},\mathcal{J}'}$$
$$= n(\mathcal{K}, \mathcal{J}') \tag{13}$$

or

$$n(\mathcal{K}, \mathcal{J}) = \int \chi^{\mathcal{K}}(j)\varphi^{\mathcal{J}*}(j)\, d\mu(j). \tag{14}$$

Thus, a knowledge of the characters is sufficient to find the multiplicities. For a discussion of character theory and also how to find the characters of the classical compact groups, see Ref. 11. For other techniques which can be used in finding the $n(\mathcal{K}, \mathcal{J})$ (such as using weight diagrams) see reference 12.

Notice that if K is the largest or maximal compact subgroup contained in G, it is possible to put an upper bound on $n(\mathcal{K}, \mathcal{J})$. For J must be at least the identity subgroup, which has the trivial representation $\mathcal{J}(j) = 1$. Then, according to the Frobenius reciprocity theorem, $\mathfrak{U}^1(k)$ (called the regular representation of K) contains $\mathcal{K}(k)$ as many times as $\mathcal{K}(e)$ contains 1 (e is the identity element). But $\mathcal{K}(e)$ is the identity matrix so that $n(\mathcal{K}, 1) = \dim \mathcal{K}$. Therefore, for K the maximal compact subgroup $n(\mathcal{K}, 1) \leq \dim \mathcal{K}$.

The question now arises, How often is it possible to write $G = H_1H_2$ and How many of the unitary irreducible representations of G can be obtained from induced representations? The answers to these questions will be taken up in Secs. III and V for G, a semisimple group, while in Sec. V semidirect product groups will be considered.

III. G, A SEMISIMPLE GROUP

When G is a semisimple group, it is possible to use the Iwasawa decomposition[13]

$$G = ANK_m, \tag{15}$$

where K_m is the maximal compact subgroup of G, A is an Abelian subgroup of G, N is a nilpotent subgroup of G and a normal subgroup of AN.

Now it is known that, from the Iwasawa decomposition, it is possible to generate a class of irreducible

[11] D. E. Littlewood, *The Theory of Group Characters* (The Clarendon Press, Oxford, England, 1940); J. P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226 1560 (1964); D. Speiser, *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon and Breach, Science Publishers, New York, 1962).

[12] R. E. Rehrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962); A. Salam, *Seminar on Theoretical Physics* (IAEA, Vienna, 1963).

[13] R. Hermann, *Lie Groups for Physicists* (W. A. Benjamin, Inc. New York, 1966).

unitary representations of G called the principal nondegenerate series.[14] Choose H_1 of the subgroup theorem so that

$$H_1 = C(A)AN, \tag{16}$$

where $C(A)$ is the centralizer of A.[15] Then

$$\mathcal{K}_1(h_1) = \mathcal{C}(c)\mathcal{A}(a), \tag{17}$$

where $\mathcal{C}(c)$ is an irreducible unitary representation of $C(A)$ and $\mathcal{A}(a)$ is an irreducible unitary representation of A. Often $C(A)$ will be Abelian so that $\mathcal{C}$ and $\mathcal{A}$ are both one-dimensional. In any event $\mathfrak{U}^{\mathcal{K}_1}(g)$, with $\mathcal{K}_1$ defined in Eq. (17), will generate a class of unitary irreducible representations on a Hilbert space defined by the inner product

$$(f, f') = \int_{G/H_1} f^*(g) f'(g)\, d\mu(g), \tag{18}$$

where $d\mu(g)$ is the measure on the coset space G/H_1 and the f, f' are elements of the vector space $\hat{\mathfrak{V}}(\mathcal{K}_1)$ defined in Eq. (3).

Choose $H_2 = K_m$. Then $G = H_1H_2$ and the principal nondegenerate series comes from the induced representations $\mathfrak{U}^{\mathcal{K}_1}(g)$ [where H_1 is defined in Eq. (17)] so that the subgroup theorem is satisfied. Further, $J = H_1 \cap H_2 = C(A)AN \cap K_m$ so that

$$\begin{aligned}\mathfrak{U}^{\mathcal{K}_1}(k_m) &\cong \mathfrak{U}^{\mathfrak{J}}(k_m) \\ &\cong \sum_{\mathcal{K}} n(\mathcal{K}_m, \mathfrak{J})\mathcal{K}_m(k_m).\end{aligned} \tag{19}$$

If K is not the maximal compact subgroup K_m, it follows that $K \subset K_m \subset G$, since K is assumed to be compact. Therefore

$$\begin{aligned}\mathfrak{U}^{\mathcal{K}_1}(k) &\cong \mathfrak{U}^{\mathfrak{J}}(k) \\ &\cong \sum_{\mathcal{K}_m} n(\mathcal{K}_m, \mathfrak{J})\mathcal{K}_m(k) \\ &\cong \sum_{\mathcal{K}_m, \mathcal{K}} n(\mathcal{K}_m, \mathfrak{J}) n(\mathcal{K}_m, \mathcal{K})\mathcal{K}(k)\end{aligned} \tag{20}$$

and the multiplicity, $n(\mathfrak{S}, \mathcal{K})$ of Eq. (1), is

$$n(\mathfrak{S}, \mathcal{K}) = \sum_{\mathcal{K}_m} n(\mathcal{K}_m, \mathfrak{J}) n(\mathcal{K}_m, \mathcal{K}); \tag{21}$$

that is, the calculation of the multiplicity $n(\mathfrak{S}, \mathcal{K})$ still depends only on the multiplicities of subgroups of K_m.

Using the decomposition $H_1 = C(A)AN$, it is also possible to find the multiplicities arising from the so-called supplementary nondegenerate series. In this case

$$\mathcal{K}_1(h_1) = \mathcal{C}(c)\mathcal{A}(a), \tag{17}$$

where now $\mathcal{A}(a)$ is a *nonunitary* irreducible representation of A and the inner product is defined by a measure much more complicated than that of Eq. (18). But once the proper nonunitary irreducible representations of A have been chosen, the reduction and calculation of multiplicities proceeds exactly as in the case of the principal series.

A second class of unitary irreducible representations of G arising from induced representations is the degenerate series. To get this class of representations, consider elements a of A which have the property that, considered as matrices, some of their matrix elements are degenerate. Pick a class $\bar{A}$ of these elements and consider

$$C(\bar{A}) = \{k \mid k \in K_m, ka = ak, \forall a \in \bar{A}\} \tag{22}$$

and, as before, let $H_1 = C(\bar{A})AN$. Then, depending on the group under consideration, a certain class of one-dimensional representations of H_1 will induce the principal and supplementary degenerate series of G. (For more details see Ref. 14 and the references given therein.)

Other remaining representations are the so-called discrete series of unitary irreducible representations and certain exceptional representations. These representations will not be treated in this paper, although the discrete series has been used in physical applications, as, for example, in the n-dimensional harmonic oscillator where the relevant group is $SU(n, 1)$.[6] There is some hope that the discrete and exceptional representations can be written as induced representations, the main problem being to construct an appropriate Hilbert space.[16] If the appropriate Hilbert space could be constructed, it would be possible to calculate the multiplicities along the lines sketched in this section.

IV. SOME EXAMPLES OF MULTIPLICITIES FOR THE SEMISIMPLE GROUPS

Consider first the group $G = SL(n, C)$. Its Iwasawa decomposition is given in Ref. 13 and can be written as

$$A = \begin{pmatrix} |g_{11}| & & & 0 \\ & |g_{22}| & & \\ & & \ddots & \\ 0 & & & |g_{nn}| \end{pmatrix},$$

[14] E. M. Stein, *High Energy Physics and Elementary Particles*, A. Salam, Ed. (IAEA, Vienna 1965). This is a main reference for this paper. It contains a discussion of the various classes of irreducible representations of the noncompact groups and also gives many mathematical references, including references to other work done on the multiplicity problem.

[15] The centralizer $C(A)$ is the set of elements of K_m which commute with A; thus $C(A) = \{k \mid k \in K_m, ak = ka\ \forall\ a \in A\}$.

[16] E. Thieleker, Argonne National Laboratory (private communication).

$$N = \begin{pmatrix} 1 & g_{12} & \cdots & g_{1n} \\ & 1 & \cdots & g_{2n} \\ & & 1 & \cdot \\ & & & \cdot \\ & \mathbf{O} & & \cdot \\ & & & 1 \end{pmatrix},$$

$$K_m = SU(n), \tag{23}$$

where

$$|g_{11}| = \frac{1}{|g_{22}| \cdots |g_{nn}|}$$

in order that the determinant of A be one. It is not hard to check that $C(A)$, the centralizer of A, is the diagonal matrix

$$\frac{g_{11}}{|g_{11}|}, \frac{g_{22}}{|g_{22}|}, \cdots, \frac{g_{nn}}{|g_{nn}|}.$$

The unitary irreducible representations of $H_1 = C(A)AN$ are

$$h_1 \to \mathcal{K}_1(g_{jj}) \equiv \mathcal{K}_1(g_{22}, \cdots, g_{nn}).$$

$$\mathcal{K}_1(g_{jj}) = |g_{22}|^{i\rho_2}\left(\frac{g_{22}}{|g_{22}|}\right)^{m_2} \cdots |g_{nn}|^{i\rho_n}\left(\frac{g_{nn}}{|g_{nn}|}\right)^{m_n}, \tag{24}$$

where the $\mathbf{m} \equiv m_j$ are integers and the $\boldsymbol{\rho} \equiv \rho_j$ real numbers. Thus, $\mathfrak{U}^{\mathcal{K}_1}(g)$ will be labeled in the principal nondegenerate series as $\mathfrak{U}^{\mathbf{m},\boldsymbol{\rho}}(g)$. To see how $\mathfrak{U}^{\mathbf{m},\boldsymbol{\rho}}(k_m)$ decomposes into irreducible representations of $K_m = SU(n)$, it is necessary to calculate

$$J = H_1 \cap K_m = C(A)AN \cap SU(n)$$

$$= \begin{pmatrix} \frac{g_{11}}{|g_{11}|} & & & \mathbf{O} \\ & \frac{g_{22}}{|g_{22}|} & & \\ & & \cdot & \\ & & & \cdot \\ & & & \cdot \\ \mathbf{O} & & & \frac{g_{nn}}{|g_{nn}|} \end{pmatrix}$$

$$= \begin{pmatrix} e^{i\alpha_1} & & & \\ & e^{i\alpha_2} & & \\ & & \cdot & \\ & & & \cdot \\ & & & \cdot \\ & & & e^{i\alpha_n} \end{pmatrix}, \tag{25}$$

where

$$e^{i\alpha_j} = \frac{g_{jj}}{|g_{jj}|}, \quad j = 1, \cdots, n.$$

Then the relevant representations of J are, according to Eq. (8),

$$\mathfrak{J}(\alpha_2, \cdots, \alpha_n) = e^{i(m_2\alpha_2 + m_3\alpha_3 + \cdots + m_n\alpha_n)} = e^{i(\mathbf{m}\cdot\boldsymbol{\alpha})}, \tag{26}$$

where $\boldsymbol{\alpha} = (\alpha_2, \cdots, \alpha_n)$ and

$$\mathfrak{U}^{\mathbf{m},\boldsymbol{\rho}}(k_m) \cong \mathfrak{U}^{\mathfrak{J}}(k_m) \equiv \mathfrak{U}^{\mathbf{m}}(k_m) \cong \sum_{[\boldsymbol{\lambda}]} n([\boldsymbol{\lambda}], \mathbf{m})\mathfrak{K}_m^{[\boldsymbol{\lambda}]}(k_m), \tag{27}$$

where the $[\boldsymbol{\lambda}]$ are the dominant weights which label the irreducible representations of $SU(n)$. Now by the Frobenius reciprocity theorem n $([\boldsymbol{\lambda}], \mathbf{m})$ is given by the number of times the representation $\mathbf{m}$ of J occurs in the representation $[\boldsymbol{\lambda}]$ of $SU(n)$, when the elements of $SU(n)$ are restricted to elements of J. But these elements are precisely the diagonal elements of $SU(n)$ and are easily obtained from the weights associated with the $[\boldsymbol{\lambda}]$ representation.

Thus, consider the set of weights $|[\boldsymbol{\lambda}], \mathbf{w}_d\rangle$ associated with the irreducible representation $[\boldsymbol{\lambda}]$, $\mathbf{w}_d$ being one of the $d = 1, \cdots, D$ weights of the representation $[\boldsymbol{\lambda}]$ (D is the dimension of the representation). Once the weights are known (ways of finding weights for a representation $[\boldsymbol{\lambda}]$ are found in Ref. 12), the diagonal group elements of $SU(n)$ have $[\boldsymbol{\lambda}]$ representation matrices which can be written as

$$e^{i\mathbf{w}_d\cdot\boldsymbol{\alpha}} = \begin{pmatrix} e^{i\mathbf{w}_1\cdot\boldsymbol{\alpha}} & & & \mathbf{O} \\ & e^{i\mathbf{w}_2\cdot\boldsymbol{\alpha}} & & \\ & & \cdot & \\ & & & \cdot \\ & & & \cdot \\ \mathbf{O} & & & e^{i\mathbf{w}_D\cdot\boldsymbol{\alpha}} \end{pmatrix}. \tag{28}$$

The trace of this matrix is

$$\chi^{[\boldsymbol{\lambda}]}(\boldsymbol{\alpha}) = e^{i\mathbf{w}_1\cdot\boldsymbol{\alpha}} + e^{i\mathbf{w}_2\cdot\boldsymbol{\alpha}} + \cdots + e^{i\mathbf{w}_D\cdot\boldsymbol{\alpha}}, \tag{29}$$

so that the multiplicity, according to Eq. (14), is

$$n([\boldsymbol{\lambda}], \mathbf{m}) = \frac{1}{(2\pi)^{n-1}}\int_0^{2\pi} \chi^{[\boldsymbol{\lambda}]}(\boldsymbol{\alpha})e^{-i\mathbf{m}\cdot\boldsymbol{\alpha}}\, d\boldsymbol{\alpha}$$

$$= \frac{1}{(2\pi)^{n-1}}\int_0^{2\pi} (e^{i\mathbf{w}_1\cdot\boldsymbol{\alpha}} + e^{i\mathbf{w}_2\cdot\boldsymbol{\alpha}} + \cdots + e^{i\mathbf{w}_D\cdot\boldsymbol{\alpha}}) \times e^{-i\mathbf{m}\cdot\boldsymbol{\alpha}}\, d\boldsymbol{\alpha}, \tag{30}$$

which is equal to zero unless $\mathbf{m}$ equals some weight vector $\mathbf{w}_d$. Thus, the multiplicity $n([\boldsymbol{\lambda}], \mathbf{m})$ can immediately be read off of the weight diagram.

For example, in $SL(2, C)$, $K_m = SU(2)$ and the

weight diagrams are one-dimensional:

$$\begin{array}{ccccccc} -2j & -2j+1 & & & -2j-1 & 2j & \\ \text{——x——} & \text{x——} & \cdots & & \text{——x——} & \text{x——} & , \end{array} \tag{31}$$

where j is the ordinary "total spin" quantum number. Then

$$\begin{aligned} n(j, m) &= 1, \quad |m| \leq 2j, \\ &= 0, \quad |m| > 2j. \end{aligned} \tag{32}$$

To see how a multiplicity greater than one can occur, consider $SL(3, C)$ with $K_m = SU(3)$. Then the 8-dimensional representation of $SU(3)$ labeled by $[\boldsymbol{\lambda}] = [1, 1]$ has two weights $(0, 0)$ which differ by "total isotopic spin." Nevertheless, for $\mathbf{m} = (0, 0)$, $n([1, 1], (0, 0)) = 2$.

To get the multiplicity for the supplementary nondegenerate series, it is sufficient to note that the unitary irreducible representations of H_1 in Eq. (24) labelled by $\mathbf{m}$ and $\boldsymbol{\rho}$ go into nonunitary irreducible representations with $\mathbf{m} = \mathbf{0}$ and $\boldsymbol{\rho}$ becoming a set of pure imaginary numbers over a certain range. Once this is known, the multiplicity is obtained in exactly the same way as was done for the principal series, which means that the multiplicity for the supplementary series is given in Eq. (30) with $\mathbf{m} = \mathbf{0}$.

For details concerning the nondegenerate and degenerate series of representations of $SL(n, C)$, see Refs. 9 and 14. In Ref. 9 the same results for the multiplicities of $SL(n, C)$ as those given here are obtained, but different techniques are used.

To get the multiplicities arising from the degenerate series of representations of $SL(n, C)$, it is necessary to specify the degenerate elements $\bar{A}$ of Eq. (22). Pick a set of $r \leq n$ integers $n_1, \cdots, n_r$ such that $n_1 + n_2 + \cdots + n_r = n$. The degeneracy of an element a of $\bar{A}$ is specified by these r integers in the sense that for a fixed $n_1, \cdots, n_r$, an element a can be written as

$$a = \begin{pmatrix} |g_{11}| & & & & & & & & \\ & |g_{11}| & & & & & & & \\ & & \ddots & & & & & & \\ & & & |g_{11}| & & & & & \\ & & & & |g_{22}| & & & & \\ & & & & & \ddots & & & \\ & & & & & & |g_{22}| & & \\ & & & & & & & \ddots & \\ & & & & & & & & |g_{rr}| \end{pmatrix} \begin{pmatrix} \circ & & & & & & & & \\ & \ddots & & & & & & & \\ & & \circ & & & & & & \\ & & & \ddots & & & & & \\ & & & & \ddots & & & & \\ & & & & & \ddots & & & \\ & & & & & & \ddots & & \\ & & & & & & & \ddots & \\ & & & & & & & & |g_{rr}| \end{pmatrix} \tag{33}$$

with each g_{jj} occurring n_j times. Once the $n_1, \cdots, n_r \equiv \mathbf{n}$ are given, it is possible to write $C(\bar{A})$ as

$$\begin{aligned} C(\bar{A}) &\equiv C(\mathbf{n}) \\ &= \begin{pmatrix} U(n_1) & & & \circ \\ & U(n_2) & & \\ & & \ddots & \\ \circ & & & U(n_r) \end{pmatrix} \\ &= \begin{pmatrix} \frac{g_{11}}{|g_{11}|} SU(n_1) & & & \circ \\ & & \ddots & \\ \circ & & & \frac{g_{rr}}{|g_{rr}|} SU(n_r) \end{pmatrix} \\ &= \begin{pmatrix} e^{i\theta_1} SU(n_1) & & & \circ \\ & & \ddots & \\ \circ & & & e^{i\theta_r} SU(n_r) \end{pmatrix}. \end{aligned} \tag{34}$$

Then representations of $H_1 = C(\mathbf{n})AN$ are chosen to be

$$h_1 \to \mathcal{H}_1(g_{22}, \cdots, g_{rr}) = |g_{22}|^{i\rho_2} e^{im_2\theta_2} \cdots |g_{rr}|^{i\rho_r} e^{im_r\theta_r} \quad (35)$$

and the principal degenerate series of irreducible representations is written as $\mathfrak{U}^{m_2,\cdots,m_r,\rho_2,\cdots,\rho_r}(g) \equiv \mathfrak{U}^{\mathbf{m},\boldsymbol{\rho}}(g)$.

It is to be noted that picking the one-dimensional representation of H_1 given in Eq. (35) amounts to picking the trivial one-dimensional representations for all the $SU(n_j)$ groups.

The reduction of the reducible representation $\mathfrak{U}^{\mathbf{m},\boldsymbol{\rho}}(k_m)$ is readily obtained, since $J = H_1 \cap K_m = C(\mathbf{n})$ and $\mathfrak{z} = e^{i\mathbf{m}\cdot\boldsymbol{\theta}} = e^{i(m_2\theta_2+\cdots+m_r\theta_r)}$. To get the multiplicity $n([\boldsymbol{\lambda}], \mathbf{m}, \boldsymbol{\rho})$ it is necessary to see how many times the representation $e^{i\mathbf{m}\cdot\boldsymbol{\theta}}$ of $C(\mathbf{n})$ is contained in the representation $[\boldsymbol{\lambda}]$ of $SU(n)$, when the elements of $SU(n)$ are restricted to $C(\mathbf{n})$. Now since $C(\mathbf{n})$ consists, in part, of direct sums of $SU(n_j)$ groups, the Dynkin diagram corresponding to $SU(n)$ can be broken up into subdiagrams in the following way:

$$\underbrace{0\text{——}0}_{SU(n_1)}\text{——}0\text{——}\underbrace{0\text{——}0\text{——}0}_{SU(n_2)}\text{——}0\text{——}\underbrace{0}_{SU(n_3)}\text{——}0\cdots 0\text{——}\underbrace{0\text{——}0\text{——}0}_{SU(n_r)}, \quad (36)$$

Since none of the $SU(n_j)$ groups are connected to each other, the weight diagrams of the $[\boldsymbol{\lambda}]$ representation of $SU(n)$ will break up into a series of orthogonal subspaces for each of the $SU(n_j)$ groups. Further, since the only representation of any $SU(n_j)$ being considered is the identity representation, the weight $\mathbf{w}_d$, $d = 1, \cdots, D$ of the D-dimensional $[\boldsymbol{\lambda}]$ representation of $SU(n)$ must be of the form

$$\mathbf{w}_d = (\underbrace{0, \cdots, 0}_{n_1}, u_2, \underbrace{0, \cdots, 0}_{n_2}, u_3, \underbrace{0, \cdots, 0}_{n_3}, \cdots u_r, \underbrace{0, \cdots, 0}_{n_r}) \quad (37)$$

in order to contain the representation $e^{i\mathbf{m}\cdot\boldsymbol{\theta}}$ of $C(\mathbf{n})$. Then the multiplicity $n([\boldsymbol{\lambda}], \mathbf{m})$ of the representation $e^{i\mathbf{m}\cdot\boldsymbol{\theta}}$ will be given by the number of times the weight $\mathbf{w}_d$ of Eq. (37) occurs in the $[\boldsymbol{\lambda}]$ representation of $SU(n)$, and, by the Frobenius reciprocity theorem, this gives the multiplicity of representations occurring in the decomposition of $U^{\mathbf{m},\boldsymbol{\rho}}(k_m)$.

As an example, consider the group $SL(3, C)$ and its principal degenerate series representations labeled by $(n_1, n_2) = (2, 1)$. Then

$$C(\mathbf{n}) = C(2, 1) = \begin{pmatrix} U(2) & 0 \\ 0 & U(1) \end{pmatrix} = \begin{pmatrix} e^{i\theta_1}SU(2) & 0 \\ 0 & e^{i\theta_2} \end{pmatrix} \quad (38)$$

with $\theta_1 = -\theta_2$ to make the determinant of the matrix equal to one. The representations of $SL(3, C)$ are labelled by m_2, ρ_2 and the reducible representation $\mathfrak{U}^{m_2,\rho_2}(k_m)$ has a multiplicity $n([\boldsymbol{\lambda}], m_2, \rho_2)$, which is calculated using the Dynkin diagram

$$\underbrace{0\text{——}0}_{SU(2)} \quad (39)$$

and looking for weights $\mathbf{w}_d = (0, u_2)$ of the $[\boldsymbol{\lambda}]$ representation of $SU(3)$. Thus, for $[\boldsymbol{\lambda}] = [1, 1]$ there is a weight $\mathbf{w}_d = (0, 0)$ occurring once when the $SU(2)$ representation is the trivial one-dimensional representation. Therefore, $n([1, 1], 0, \rho_2) = 1$.

As a second class of groups, consider $G = SL(n, R)$.[17] The Iwasawa decomposition will be similar to that of $SL(n, C)$, except that K_m will be $SO(n)$ rather than $SU(n)$:

$$A = \begin{pmatrix} g_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g_{nn} \end{pmatrix}, \qquad N = \begin{pmatrix} 1, & g_{12}, \cdots, g_{1n} \\ & 1, & \cdots, g_{2n} \\ & & \ddots \\ & & \ddots \\ \mathbf{0} & & 1 \end{pmatrix} \quad (40)$$

with the coefficients g_{ij} of A, N, and K_m being real. The centralizer $C(A)$ will contain no elements other than those already in A so that $H_1 = AN$. Because H_1 now has only real coefficients, its irreducible representations will no longer be labelled by $\mathbf{m}$ and $\boldsymbol{\rho}$. Rather, the $\mathbf{m}$ label is replaced by a "parity" label so that the representations of H_1 are written as

$$h_1 \to \mathcal{H}_1(g_{22}, \cdots, g_{nn}) = |g_{22}|^{i\rho_2} P_2(\text{sgn } g_{22}) \cdots |g_{nn}|^{i\rho_n} P_n(\text{sgn } g_{nn}) \quad (41)$$

in the principal nondegenerate series. In Eq. (41) sgn g_{ii} is the sign of g_{ii} and $P_i(\text{sgn } g_{ii})$ takes on the values ± 1, depending on whether the parity is plus or minus. The principal nondegenerate series of

[17] B. D. Romm, American Mathematical Society Translations, Ser. 2, **58**, 155 (1966); E. Thieleker, "On Some Infinite Dimensional Representations of Lie Groups", preprint, Applied Mathematics Division, Argonne National Laboratory, Argonne, Illinois (1966).

representations are then written as

$$\mathfrak{U}^{\mathfrak{K}_1}(g) = \mathfrak{U}^{P_2,\cdots,P_n,\rho_2,\cdots,\rho_n}(g) \equiv \mathfrak{U}^{\mathbf{P},\rho}(g).$$

To get the multiplicity occurring in the reducible representation $\mathfrak{U}^{\mathbf{P},\rho}(k_m)$, it is necessary to calculate $J = AN \cap SO(n)$. Now $SO(n)$ can be written as

$$SO(n) = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & & & \mathbf{O} \\ -\sin\theta_1 & \cos\theta_1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \times \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 & & \mathbf{O} \\ 0 & 1 & 0 & & \\ -\sin\theta_2 & 0 & \cos\theta_2 & & \\ & & & 1 & \\ & & & & \ddots \\ \mathbf{O} & & & & 1 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \cos\theta_{(n^2-n)/2} & \sin\theta_{(n^2-n)/2} \\ \mathbf{O} & & & -\sin\theta_{(n^2-n)/2} & \cos\theta_{(n^2-n)/2} \end{pmatrix} \tag{42}$$

and since AN has only zeros for its subdiagonal elements, it is clear that $\sin\theta_i$ must equal zero, which implies that $\cos\theta_i = \pm 1$ so that J is the finite group of determinant one-diagonal matrices having only ± 1 as possible entries:

$$J = \left\{ \begin{pmatrix} \pm 1 & & & \mathbf{O} \\ & \pm 1 & & \\ & & \ddots & \\ \mathbf{O} & & & \pm 1 \end{pmatrix} \right\}. \tag{43}$$

J is Abelian of order 2^{n-1} so that $\mathfrak{J}(j)$ will be labeled by the "parity" representations of Eq. (41).

Finally, using the subgroup theorem and the Frobenius reciprocity theorem, we get

$$\mathfrak{U}^{\mathbf{P},\rho}(k_m) \cong \mathfrak{U}^{\mathfrak{J}}(k_m) \equiv \mathfrak{U}^{\mathbf{P}}(k_m) \cong \sum_{[\lambda]} n([\boldsymbol{\lambda}], \mathbf{P}) \mathfrak{K}_m^{[\lambda]}(k_m), \tag{44}$$

where $n([\boldsymbol{\lambda}], \mathbf{P})$ is the number of times the reducible representation $\mathfrak{K}^{[\lambda]}(j)$ contains the irreducible representation $\mathfrak{J}^{\mathbf{P}}(j)$ of J (j is an element of J).

Since J is a finite group and its elements are obtained by having $\cos\theta_i = \pm 1$ in Eq. (42), it is possible to get the character of the irreducible representations $[\boldsymbol{\lambda}]$ of $SO(n)$ and use Eq. (14) to calculate $n([\boldsymbol{\lambda}], \mathbf{P})$. For the fundamental representation of $SO(n)$ the multiplicity is then immediately obtained; for the other irreducible representations of $SO(n)$ there seems to be no such easy technique.

To see what happens in the case of the fundamental representation $[\boldsymbol{\lambda}] = [0, \cdots, 0, 1]$, consider $n = 3$, $K_m = SO(3)$. From Eq. (42), $SO(3)$ can be written as

$$SO(3) = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_3 & \sin\theta_3 \\ 0 & -\sin\theta_3 & \cos\theta_3 \end{pmatrix}. \tag{42}$$

J will consist of $2^{n-1} = 2^2$ elements:

$$J = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right\}. \tag{43}$$

Here each element of J is written as an element of $SO(3)$ with the $\cos\theta_i$ taking on the value 0 or π radians. To get the multiplicity of the $\mathfrak{J}$ representation with parity label $\mathbf{P} = (+, -)$, for example, it is merely necessary to take the trace of the elements in Eq. (43) and use Eq. (14) to get

$$\begin{aligned} n([1], (+, -)) &= \frac{1}{2^2} \sum_J \mathfrak{J}^{(+,-)}(j) \chi^{[1]}(j) \\ &= \tfrac{1}{4} \sum [3 + (+1)(-1) + (-1)(-1) + (-1)(-1)] \\ &= 1. \end{aligned} \tag{44}$$

Repeating this simple calculation for the other

possible parity representations gives

$$\begin{aligned} n([1],(+,+)) &= 0 \\ n([1],(+,-)) &= n([1],(-,+)) \\ &= n([1],(-,-)) \\ &= 1. \end{aligned} \tag{45}$$

It is clear that this procedure can easily be generalized to arbitrary n.

To get the multiplicity for arbitrary representations of $SO(n)$ seems rather long and complicated. A possibility is to build irreducible tensors from the defining vector

$$\mathbf{T}' = k_m \mathbf{T}. \tag{46}$$

That is, corresponding to an irreducible representation $[\boldsymbol{\lambda}]$ of $SO(n)$, there is a tensor $T_{[\lambda]}$ which can be written as a D-dimensional column vector, where D is the dimension of the $[\boldsymbol{\lambda}]$ representation:

$$T'_{[\lambda]} = \mathfrak{D}^{[\lambda]}(k_m) T_{[\lambda]}. \tag{47}$$

Since we are concerned only with the diagonal elements of the J subgroup of $SO(n)$, $\mathfrak{D}^{[\lambda]}(k_m)$ will be a matrix with only plus and minus ones on the diagonal. Once this matrix is known, it is easy to get its trace and carry out a computation similar to Eq. (44). However, the problem of getting the tensor $T_{[\lambda]}$ is in general nontrivial; Ref. 18 presents a method which is generally applicable.

As a last example, consider the group $SO(p,q)$. $SO(p,q)$ can conveniently be parameterized as

$$\begin{pmatrix} \cos\beta_1 & \sin\beta_1 & \\ -\sin\beta_1 & \cos\beta_1 & \\ & & I_{p+q-2} \end{pmatrix} \begin{pmatrix} \cos\beta_2 & 0 & \sin\beta_2 & \\ 0 & 1 & 0 & \\ -\sin\beta_2 & 0 & \cos\beta_2 & \\ & & & I_{p+q-3} \end{pmatrix} \cdots$$

$$\begin{pmatrix} \cosh\beta_{p+1} & 0 \cdots & \sinh\beta_{p+1} & \\ 0 & & 0 & \\ \vdots & & \vdots & \\ \sinh\beta_{p+1} & 0 \cdots & \cosh\beta_{p+1} & \\ & & & I_{q-1} \end{pmatrix} \cdots \begin{pmatrix} I_{p+q-2} & & \\ & \cos\beta_{(p+q)^2-(p+q)/2} & \sin\beta_{(p+q)^2-(p+q)/2} \\ & -\sin\beta_{(p+q)^2-(p+q)/2} & \cos\beta_{(p+q)^2-(p+q)/2} \end{pmatrix}, \tag{48}$$

where there are $(p^2-p)/2 + (q^2-q)/2$ matrices which involve cosine and sine terms and

$$\frac{(p+q)^2-(p+q)}{2} - \frac{(p^2-p)}{2} + \frac{q^2-q}{2} = pq$$

matrices which involve cosh and sinh terms.

In the Iwasawa decomposition, $K_m = SO(p) \otimes SO(q)$, while if the convention $p \geq q$ is chosen, A can be written as

$$A = \begin{vmatrix} \cosh\beta_1 & & & \cdots & & \sinh\beta_1 & & & \cdots \\ & \cosh\beta_2 & & & & & \sinh\beta_2 & & \\ & & \ddots & & & & & \ddots & \\ & & & \cosh\beta_q & & & & & \sinh\beta_q \\ & & & & I_{p-q} & & & & \\ \sinh\beta_1 & & & & & \cosh\beta_1 & & & \\ & \sinh\beta_2 & & & & & \cosh\beta_2 & & \cdots \\ & & \ddots & & & & & \ddots & \\ & & & \sinh\beta_q & \cdots & & & & \cosh\beta_q \end{vmatrix} . \tag{49}$$

[18] M. Hamermesh, *Group Theory* (Addison Wesley Publishing Co., Reading, Massachusetts, 1962).

Then the centralizer $C(A)$ is seen to be

$$C(A) = \begin{pmatrix} I_q & 0 & 0 \\ 0 & SO(p-q) & 0 \\ 0 & 0 & I_q \end{pmatrix}, \tag{50}$$

where I_q is the q-dimensional identity matrix.

The principal nondegenerate series of representations of $SO(p,q)$ are induced by representations of $H_1 = C(A)AN$:

$$h_1 \to \mathcal{H}_1(h_1) = \mathcal{R}^{[\lambda]}(SO(p-q)) \times (\rho_1)^{i\rho_1}(\beta_2)^{i\rho_2}\cdots(\beta_q)^{i\rho_q}, \tag{51}$$

where $\mathcal{R}^{[\lambda]}(SO(p-q))$ is the $[\lambda]$ irreducible representation matrix of $SO(p-q)$. Then $U^{\mathcal{H}_1}(SO(p,q)) = U^{[\lambda],\rho_1\cdots\rho_q}(SO(p,q))$ is an irreducible representation of $SO(p,q)$.

The multiplicity of $U^{[\lambda],\rho_1\cdots\rho_q}(SO(p)\otimes SO(q))$ is obtained from $U^{\mathfrak{J}}(SO(p)\times SO(q))$ where $J = C(A) = SO(p-q)$ has the representation

$$\mathfrak{J} = \mathcal{R}^{[\lambda]}(SO(p-q)).$$

By the Frobenius reciprocity theorem, $n([\lambda], [\lambda_p, \lambda_q])$ {where $[\lambda_p, \lambda_q]$ is the irreducible representation label for $SO(p)\otimes SO(q)$} is given by the number of times the irreducible representation $\mathcal{R}^{[\lambda_p]}(SO(p))\otimes \mathcal{R}^{[\lambda_q]}(SO(q))$ of $SO(p)\otimes SO(q)$ contains the representation $\mathcal{R}^{[\lambda]}(SO(p-q))\otimes 1$ of $SO(p-q)\otimes(e)$ [(e) is the identity subgroup]; thus,

$$n([\lambda], [\lambda_p, \lambda_q]) = n([\lambda], [\lambda_p]) \times \text{dimension } [\lambda_q]. \tag{52}$$

$n([\lambda], [\lambda_p])$ is the number of times the $[\lambda]$ representation of $SO(p-q)$ is contained in the $[\lambda_p]$ representation of $SO(p)$. It can be obtained either by using weight-diagram techniques or calculating the irreducible characters using Eq. (14). In any case, there are procedures for getting $n([\lambda], [\lambda_p])$, which, since they are in general complicated, will not be given here.

V. G, A SEMIDIRECT PRODUCT GROUP

G a semidirect product group means that G can be written as

$$G = TL, \qquad T \cap L = (e), \tag{53}$$

where L is a subgroup and T is an invariant Abelian subgroup of G. In contrast to the semisimple groups discussed in Sec. III, a complete representation theory for semidirect product groups exists[8] in the sense that all irreducible unitary representations can be obtained as induced representations of the so-called little groups.

Let $\mathcal{T}(t)$ be an irreducible representation of T. The set of elements of G satisfying

$$\{g \mid \mathcal{T}(gtg^{-1}) = \mathcal{T}(t) \forall t \in T\} \tag{54}$$

forms a subgroup of G called the little group. Choose $H_1(\mathcal{T})$ to be the little group and let $\mathcal{H}_1$ be an irreducible representation of $H_1(\mathcal{T})$. Then $\mathfrak{U}^{\mathcal{H}_1}(g)$ is a representation of G and further, a certain prescribed class of irreducible representations of H_1 generates all the irreducible representations of G.

However, we are interested in finding the multiplicity resulting from the decomposition of the reducible representation $\mathfrak{U}^{\mathcal{H}_1}(h_2)$. If H_2 is chosen to be L, then $G = H_1H_2$, so that it is possible to write

$$\mathfrak{U}^{\mathcal{H}_1}(l) \cong \mathfrak{U}^{\mathfrak{J}}(l) \cong \sum_{\mathfrak{L}} n(\mathfrak{L}, \mathfrak{J})\mathfrak{L}(l), \tag{55}$$

where $\mathfrak{J}$ is the representation of the subgroup $J = L \cap H_1$ as defined in Eq. (8) and $\mathfrak{L}(l)$ is an irreducible unitary representation of L. It is also possible to find the multiplicity occurring in the decomposition of $\mathfrak{U}^{\mathcal{H}_1}(k)$, where K is some subgroup of L, by using Eq. (21).

As a simple example of a multiplicity calculation for a semidirect product group, consider $SO(3)$ as a subgroup of the three-dimensional Euclidean group, which can be written in matrix form as

$$\begin{pmatrix} SO(3) & \begin{matrix} T_x \\ T_y \\ T_z \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{pmatrix} \equiv \begin{pmatrix} SO(3) & \mathbf{T} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I_3 & \mathbf{T} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} SO(3) & \mathbf{0} \\ 0 & 1 \end{pmatrix}. \tag{56}$$

The representations of the invariant Abelian subgroup T are $e^{i\mathbf{p}\cdot\mathbf{T}}$, so that the little group $H_1(\mathbf{p})$ is

$$H_1(\mathbf{p}) = \begin{pmatrix} \cos\theta & \sin\theta & 0 & T_x \\ -\sin\theta & \cos\theta & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{57}$$

for $\mathbf{p} \neq \mathbf{0}$; its representations are

$$\mathcal{H}_1(\mathbf{p}) = e^{i\mathbf{p}\cdot\mathbf{T}}e^{im\theta}, \tag{58}$$

where $m = 0, \pm1, \pm2, \cdots$ labels the irreducible

representations of

$$SO(2) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\begin{aligned} \mathfrak{U}^{\mathfrak{K}_1}(SO(3)) &\equiv \mathfrak{U}^{\mathfrak{p}\cdot\mathfrak{D},m}(SO(3)) \\ &\cong \mathfrak{U}^{m}(SO(3)) \\ &\cong \sum_j n(j,m)\mathfrak{K}^j(SO(3)), \end{aligned} \tag{59}$$

where the integer j labels the irreducible representations of $SO(3)$.

Since $J = L \cap H_1 = SO(2)$, the multiplicity $n(j, m)$ is found from the Frobenius reciprocity theorem to be

$$\begin{aligned} n(j,m) &= 1, \quad |m| \leq j, \\ n(j,m) &= 0, \quad |m| > j. \end{aligned} \tag{60}$$

There are, of course, more irreducible unitary representations than those given above; all of the representations of the three-dimensional Euclidean group are given in Ref. 19.

As a last example of multiplicity calculations, consider the case when $H_2 = L$ is not a compact group. Choose G to be the Poincaré group and L to be the Lorentz group. The unitary irreducible representations of the Poincaré group are well known[20]; in this example we will consider only the positive mass representations labeled by M. For these representations the little group which induces the representations of the Poincaré group is

$$H_1(M) = \begin{pmatrix} 1 & 0 & T_0 \\ 0 & SO(3) & \mathbf{T} \\ 0 & 0 & 1 \end{pmatrix}, \tag{61}$$

while $J = H_1 \cap L = SO(3)$. Therefore,

$$\mathfrak{U}^{M,j}(l) \cong \mathfrak{U}^{j}(l) \cong \sum_{\mathfrak{L}} n(\mathfrak{L}, j)\mathfrak{L}(l), \tag{62}$$

where the integer j again labels the irreducible representations of $SO(3)$.

But the irreducible representations of the covering group of L, namely $SL(2, C)$, were discussed in Sec. IV and were labeled by m, ρ. Since $SL(2, C)$ is the covering group of L, all of the unitary, irreducible representations of L will be contained in those of $SL(2, C)$. Thus, to get the multiplicity $n(\mathfrak{L}, j)$ it is merely necessary to note that, by the Frobenius reciprocity theorem, $\mathfrak{U}^{m,\rho}(l)$ is contained in $\mathfrak{U}^{j}(l)$ as many times as $\mathfrak{U}^{m,\rho}(SO(3))$ contains $\mathfrak{K}^{j}(SO(3))$. But this multiplicity was calculated in Sec IV and shown to be

$$\begin{aligned} n(j,m,\rho) &= 1, \quad |m| \leq j, \\ n(j,m,\rho) &= 0, \quad |m| > j. \end{aligned} \tag{32}$$

Therefore, the multiplicity $n(M, j, m, \rho)$ is given by

$$\begin{aligned} n(M,j,m,\rho) &= 1, \quad |m| \leq j, \\ n(M,j,m,\rho) &= 0, \quad |m| > j, \end{aligned} \tag{63}$$

a result already obtained by Joos[21] in his calculation of the Clebsch–Gordan coefficients resulting from the reducible representation $u^{M,j}(L)$.

VI. CONCLUSION

The calculation of the multiplicity occurring in the reduction of a reducible representation of the Lorentz group [Eq. (62) has shown that it is not necessary that H_2 be a compact group]. In most physical applications H_2 has been a compact group, but, aside from this physical motivation, the reduction theory needed when H_2 is any noncompact subgroup of any noncompact group G is much more complicated than that given in Sec. II. Nevertheless, Mackey[8] has shown that the subgroup theorem is generally applicable to H_2, a noncompact group, and also shown that Frobenius reciprocity theorem generalizes as expected. Thus, with the general subgroup theorem and the general Frobenius reciprocity theorem, it should be possible to get multiplicities for most subgroups of a given group, as long as the irreducible representations of the group can be written as induced representations.

ACKNOWLEDGMENTS

The author wishes to thank Dr. K. C. Wali of the High Energy Division of Argonne National Laboratory for the hospitality shown him while this work was being completed. He also wishes to thank Dr. E. Thieleker of the Applied Mathematics Division of Argonne for many helpful discussions on the mathematics of group theory.

[19] J. S. Lomont, *Applications of Finite Groups* (Academic Press Inc., New York, 1959), p.331.
[20] See Ref. 19, page 328.
[21] H. Joos, Fortschr. Physik **10**, 3 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Exact Occupation Kinetics for One-Dimensional Arrays of Dumbbells

R. B. McQuistan and D. Lichtman
Physics Department and Surface Studies Laboratory, University of Wisconsin at Milwaukee, Milwaukee, Wisconsin

(Received 12 June 1967)

By utilizing a branch-dependent counting technique, exact relationships are developed which describe the ensemble average of the kinetics of occupation by dumbbells of finite one-dimensional arrays of compartments. It is shown that, as the number of compartments per array tends to infinity, $\langle\theta(t)\rangle$, the ensemble average of the fraction of an array which is occupied, is given by

$$\langle\theta(t)\rangle = 1 - \exp\{-2[1 - \exp(-\nu t)]\},$$

where ν is the striking frequency and t is time. By contrasting these results with the statistics of one-dimensional arrays of dumbbells, it is demonstrated that it is inappropriate to employ classical statistical-mechanical methods (e.g., the Bethe approximation) to treat the kinetic aspects of occupation where configurational correlation exists. (Here we define configurational correlation to be the situation in which the occupation of a compartment precludes the occupation of one of its nearest neighbors.)

I. INTRODUCTION

It is well known[1] that the methods of statistical mechanics can be used to treat the kinetic aspects of physical phenomena only when certain assumptions are made regarding the details of the mechanisms involved. Thus, with appropriate assumptions such essentially kinetic processes as carrier recombination, adsorption, crystallization, etc., can usually be treated by recourse to statistical concepts. Even in situations where the dimensionality of the phenomenon is restricted and where occupational correlation of the Fermi–Dirac type exists (e.g., adsorption processes), statistical methods are still applicable.[2,3] However, as the present paper will show, when considering particles which can occupy more than one compartment or when the occupation of one compartment precludes the occupation of one of the nearest-neighbor compartments, a correlation (which we will call configurational correlation) renders the classical statistical-mechanical approach inappropriate. The neglect of such configurational correlation has led to serious difficulties in the literature. For example, it has resulted in the failure of the Bethe approximation[4–6] and similar statistical-mechanical techniques to predict nonunity values for such quantities as (1) the saturation coverage[7] and (2) the maximum degree of crystallinity.[8]

The Bethe approximation, when applied[7] to the kinetics of the occupation by dumbbells of an array in which each compartment has two nearest neighbors, predicts that $\langle S\rangle$, the average probability of success when attempting to place a dumbbell on such an array on which q dumbbells reside, is

$$\langle S\rangle = \frac{2(1-\theta)^2}{2-\theta},$$

where $\theta \equiv 2q/N$ is the coverage or the fraction of compartments occupied. This equation indicates that θ must be unity before the probability of success is zero. Clearly this conclusion is incorrect because of the creation of isolated, vacant compartments (see Fig. 1) which cannot accommodate additional dumbbells. Consequently, after a long period of time a saturation situation arises in which the probability of success becomes zero even though the coverage has not attained the value of one.

Numerous attempts of an analytic[9,10] and Monte Carlo[11–13] nature have been made to obviate this difficulty, with particular interest centered on the value of the saturation coverage. Similar difficulties were encountered by Gornick and Jackson[8] when dealing, in a statistical manner, with a more general problem of the crystallization of linear polymer chains, i.e., they calculated a maximum degree of crystallinity of unity. This result, as they point out, is incorrect because the length of some crystallizable sequences may exceed the length available for them in the polymer chain, so that they are "wasted" as far as further crystallization is concerned.

[1] R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1952), Chap. XII.
[2] R. H. Fowler, Proc. Cambridge Phil. Soc. **31**, 260 (1935).
[3] K. J. Laidler, *Catalysis*, P. H. Emmett, Ed. (Reinhold Publishing Corporation, New York, 1954), Chap. 3.
[4] H. A. Bethe, Proc. Roy. Soc. (London) A **150**, 552 (1935).
[5] R. Peierls, Proc. Cambridge Phil. Soc. **32**, 471 (1936).
[6] G. S. Rushbrook, Proc. Cambridge Phil. Soc. **34**, 424 (1938).
[7] J. K. Roberts and A. R. Miller, Proc. Cambridge Phil. Soc. **35**, 293 (1939).
[8] F. Gornick and J. L. Jackson, J. Chem. Phys. **38**, 1150 (1963).
[9] J. K. Roberts, Proc. Roy. Soc. (London) A **161**, 141 (1937).
[10] J. K. Roberts, Proc. Cambridge Phil. Soc. **34**, 577 (1938).
[11] J. K. Roberts, Proc. Cambridge Phil. Soc. **34**, 399 (1938).
[12] P. A. Redhead, Trans. Faraday Soc. **57**, 641 (1961).
[13] D. R. Rossington and R. Borst, Surface Sci. **3**, 202 (1965).

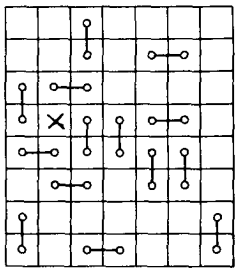


FIG. 1. The creation of isolated vacancies when dumbbells are placed on a two-dimensional array.

An analogous difficulty is also encountered[14] in the problem of determining the average number of cars which can be accommodated along a curb which is divided into spaces which are half the length of the cars.

It has been suggested[8] that these difficulties result from the nature of the approximations inherent in the statistical methods (e.g., the Bethe approximation) which are inappropriate for values of the coverage, crystallinity, etc., near unity. However, in a previous article[15] we have developed relationships which exactly describe the occupation statistics for linear arrays of dumbbells. We have shown that in the limit, as the number of compartments per array tends to infinity, these relationships reduce to those calculated using the Bethe approximation for the case when the number of nearest neighbors is two. We conclude, therefore, *that the saturation value of unity predicted by the Bethe method is not the result of any approximations employed.* Rather, the difficulty arises because *the counting procedure used in all classical statistical calculations is not applicable to dynamic processes, such as adsorption and crystallization, when configurational correlations exist.* The reason for this is that the statistical-mechanical approach to this problem deals essentially with the question of the number of ways of arranging q dumbbells on a linear array of N compartments without regard to the dynamic aspects of the process by which the dumbbells are deposited. When dealing with particles which occupy single compartments, this approach is appropriate because the occupation of a compartment is the only factor which precludes further occupation (Fermi–Dirac correlation). Thus one can assume that the occupation rate is proportional to the fraction of compartments which are empty. For dumbbells which occupy two adjacent compartments, however, occupation of a particular compartment may be precluded because its nearest neighbors are occupied, i.e., isolated vacancies may exist. Therefore one cannot determine the rate of occupation by dumbbells of a linear array of compartments by asking the question, "How many ways can q indistinguishable dumbbells be arranged on a linear array of N compartments?" Rather one must ask the question, "After m spatially random attempts to place indistinguishable dumbbells on a linear array of N compartments, what is the average number of dumbbells on the array?" When dealing with particles which occupy a single compartment, the answers to both these questions lead to identical expressions which can be utilized to describe the kinetics of occupation. However, when considering dumbbells, these two questions have different answers because not only do the number of dumbbells on the array affect the answer but the configuration of these dumbbells can also influence the number of additional dumbbells which the array can accommodate.

II. CALCULATION

To answer the second of these questions we consider an ensemble composed of a large number of systems, each of which is a linear array of N compartments. These systems are struck by dumbbells at the same rate in a spatially random manner. If a double vacancy exists where the dumbbell strikes, the dumbbell will stick; if not, it will be rejected. For example, (see Fig. 2) if $N = 5$ after the first attempt to place a

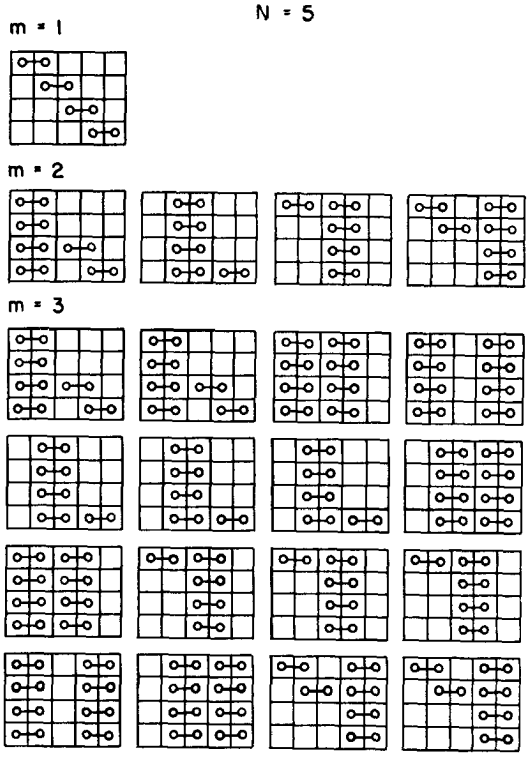


FIG. 2. The arrangements possible after m attempts to place dumbbells on a linear array of 5 compartments in a spatially random manner.

[14] P. E. Ney, Ph.D. thesis, Columbia University, N.Y. (1960).

[15] D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).

dumbbell on each system of 5 compartments in a random geometric manner, $\frac{1}{4}$ of the systems will have the dumbbell occupying the first and second compartment; $\frac{1}{4}$ of the systems will have the dumbbell occupying the second and third compartment, etc. After each system has been struck once, each of the four arrangements shown under $m = 1$ in Fig. 2 has an equal probability of occurrence. For those systems in which the first dumbbell occupies the first and second compartments, the second dumbbell cannot occupy the first and second compartment or the second and third compartment, but it can occupy the third and fourth, as well as the fourth and fifth compartments. Each of the arrangements shown in the first box under $m = 2$ is thus equally probable. If the first dumbbell had occupied compartments two and three, then the second dumbbell could only have occupied the fourth and fifth compartments as shown in the second box under $m = 2$, etc. Thus, for $N = 5$ after each array has been struck twice by dumbbells, the average coverage (averaged over the entire ensemble) is $\langle\theta(2, 5)\rangle = 0.55$. This process is then continued for any m.

The previously described counting procedure may be generalized in the following manner. After m attempts to place dumbbells on systems of linear arrays of compartments, there will be $q(m, N)$ dumbbells on $N(N-1)^m$ sites; thus, after the mth attempt, we may write

$$2q(m, N) + N_v(m, N) = N(N-1)^m, \tag{1}$$

where $N_v(m, N)$ is the number of vacant compartments after the mth attempt. Dividing Eq. (1) by $N(N-1)^m$ and defining $\langle\theta\rangle \equiv 2q(m, N)/N(N-1)^m$, we may write

$$\langle\theta(m, N)\rangle = 1 - \frac{N_v(m, N)}{N(N-1)^m} = 1 - \frac{1}{N(N-1)^m}\sum_{p=1}^{N-2} pN_p(m, N), \tag{2}$$

where $N_p(m, N)$ is the number of p-tuple contiguous vacant compartments; the sum extends to the largest value of p, i.e., $N-2$.

To determine $N_p(m, N)$ we observe that after m attempts each $(N-s)$-tuple contiguous vacancy will lead in the next generation to the following:

s $(N-s)$-tuple contiguous vacant sites (cvs),
0 $(N-s-1)$-tuple (cvs),
2 $(N-s-2)$-tuple (cvs),
. .
. .
. .
2 2-tuple (cvs),
2 1-tuple (cvs) (isolated sites).

This leads to a recursion relation

$$N_p[m, N] = (N-p-2)N_p(m-1, N) + 2\sum_{h=p}^{N-2} N_h(m-1, N) - 2N_{p+1}(m-1, N)\sum_{h=1}^{N-3}\delta_{ph}, \quad p \le N-2, \tag{3}$$

where δ_{ph} is the Kronecker delta and where $N_p(1, N) = 2$. If this expression [Eq. (3)] is utilized in Eq. 2 to calculate $N_v(m, N)$, we obtain (see Appendix A)

$$N_v(m, N) = \sum_{p=1}^{N-2} pN_p(m, N) = (N-1)\sum_{p=1}^{N-2} pN_p(m-1, N) - 2\sum_{p=2}^{N-2}(p-1)N_p(m-1, N). \tag{4}$$

If Eq. (3) is used again in Eq. (4), it yields

$$\sum_{p=1}^{N-2} pN_p(m, N) = (N-1)^2\sum_{p=1}^{N-2} pN_p(m-2, N) - 2[(N-1)+(N-2)]\sum_{p=2}^{N-2}(p-1)N_p(m-2, N) + 2^2\sum_{p=3}^{N-2}(p-2)N_p(m-2, N). \tag{5}$$

The repeated use of Eq. (3) in resulting equations r times yields

$$\begin{aligned}\sum_{p=1}^{N-2} pN_p(m, N) = {} & (N-1)^r\sum_{p=1}^{N-2} pN_p(m-r, N) \\ & - 2[(N-1)^{r-1} + (N-1)^{r-2}(N-2) + (N-1)^{r-3}(N-2)^2 + \cdots \\ & + (N-1)(N-2)^{r-2} + (N-2)^{r-1}]\sum_{p=2}^{N-2}(p-1)N_p(m-r, N) \\ & + 2^2[(N-1)^{r-2} + (N-1)^{r-3}(N-2)(N-3)^0 + (N-1)^{r-3}(N-2)^0(N-3) + \cdots \\ & + (N-1)^1(N-2)^{r-3}(N-3)^0 + (N-2)^{r-2} + \cdots \\ & + (N-1)^1(N-2)^0(N-3)^{r-3} + (N-1)^0(N-2)^1(N-3)^{r-3} + (N-3)^{r-2}]\end{aligned}$$

$$\times \sum_{p=3}^{N-2}(p-2)N_p(m-r,N)+\cdots$$
$$=(N-1)^r\Bigg\{\sum_{p=1}^{N-2}pN_p(m-r,N)$$
$$+\left(\frac{-2}{N-1}\right)^1\left[\sum_{n_1=0}^{r-1}\left(\frac{N-2}{N-1}\right)^{n_1}\right]\sum_{p=2}^{N-2}(p-1)N_p(m-r,N)$$
$$+\left(\frac{-2}{N-1}\right)^2\left[\sum_{n_1=0}^{r-2}\left(\frac{N-2}{N-1}\right)^{n_1}\sum_{n_2=0}^{r-2-n_1}\left(\frac{N-3}{N-1}\right)^{n_2}\right]\sum_{p=3}^{N-2}(p-2)N_p(m-r,N)+\cdots$$
$$+\left(\frac{-2}{N-1}\right)^k\left[\sum_{n_1=0}^{r-k}\left(\frac{N-2}{N-1}\right)^{n_1}\sum_{n_2=0}^{r-k-n_1}\left(\frac{N-3}{N-1}\right)^{n_2}\cdots\sum_{n_k=0}^{r-k-n_1-n_2-\cdots-n_{k-1}}\left(\frac{N-k-1}{N-1}\right)^{n_k}\right]$$
$$\times\sum_{p=k+1}^{N-2}(p-k)N_p(m-r,N)+\cdots\Bigg\}. \tag{6}$$

Examination of the coefficient of

$$\sum_{p=k+1}^{N-2}(p-k)N_p(m-r,N)$$

reveals it to be (see Appendix B)

$$\frac{(-2)^k}{k!}\sum_{j=0}^{k}(-1)^j{}_kC_j\left(\frac{N-1-j}{N-1}\right)^r.$$

Thus Eq. (6) may be written

$$N_v(m,N)=(N-1)^r\Bigg\{\sum_{k=0}^{N-3}\frac{(-2)^k}{k!}\sum_{j=0}^{k}{}_kC_j\left(\frac{N-1-j}{N-1}\right)^r \times\sum_{p=k+1}^{N-2}(p-k)N_p(m-r,N)\Bigg\}. \tag{7}$$

If r takes on the value $m-1$, Eq. (7) becomes

$$N_v(m,N)=(N-1)^{m+1}\Bigg\{\sum_{k=0}^{N-3}\frac{(-2)^k}{k!}\left(\frac{N-1-k}{N-1}\right)\left(\frac{N-2-k}{N-1}\right) \times\sum_{j=0}^{k}(-1)^j{}_kC_j\left(\frac{N-1-j}{N-1}\right)^{m-1}\Bigg\}. \tag{8}$$

Therefore, the average coverage $\langle\theta(m,N)\rangle$, after m attempts to place a dumbbell on a linear array of N compartments, is given by [see Eq. (2)]

$$\langle\theta(m,N)\rangle=1-\left(\frac{N-1}{N}\right)\sum_{k=0}^{N-3}\frac{(-2)^k}{k!}\left(\frac{N-1-k}{N-1}\right)\left(\frac{N-2-k}{N-1}\right) \times\sum_{j=0}^{k}(-1)^j{}_kC_j\left(\frac{N-1-j}{N-1}\right)^{m-1}. \tag{9}$$

In order to treat θ for large N as a function of time, we note that m, the number of times that an array of N compartments is struck by dumbbells, is given by

$$m=Nvt,$$

where v is the striking frequency per unit compartment per unit time and t is time. Thus the last factor in Eq. (9) becomes

$$\left(\frac{N-1-j}{N-1}\right)^{m-1}=\left(\frac{N-1-j}{N-1}\right)^{-1}\left(\frac{N-1-j}{N-1}\right)^{m}$$
$$=\left(\frac{N-1-j}{N-1}\right)^{-1}\left(\frac{N-1-j}{N-1}\right)^{Nvt}$$
$$=\left(1-\frac{j}{N-1}\right)^{-1}\left\{\left(1-\frac{j}{N-1}\right)^{-N/j}\right\}^{-jvt}. \tag{10}$$

As N tends to infinity, this becomes

$$\exp[-jvt].$$

The last sum in Eq. 9 therefore yields

$$\sum_{j=0}^{k}(-1)^j{}_kC_j\exp[-jvt]=\{1-\exp[-vt]\}^k, \tag{11}$$

so that Eq. (9) becomes

$$\theta(t)=1-\sum_{k=0}^{\infty}\frac{(-2)^k}{k!}\{1-\exp[-vt]\}^k$$
$$=1-\exp\{-2[1-\exp(-vt)]\}. \tag{12}$$

Figure 3 shows θ as a function of time as given by Eq. (12.) For comparison $\langle\theta(m,10)\rangle$, according to Eq. (9), is also shown in this figure. At $t=\infty$, $0=1-e^{-2}$; thus, if dumbbells are placed in a random manner on an infinite array of compartments, after a long time 13.5% of the compartments will be isolated and vacant.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to T. N. Casselman and F. N. Simon for many helpful discussions, and to L. P. Levine for the necessary computer calculations.

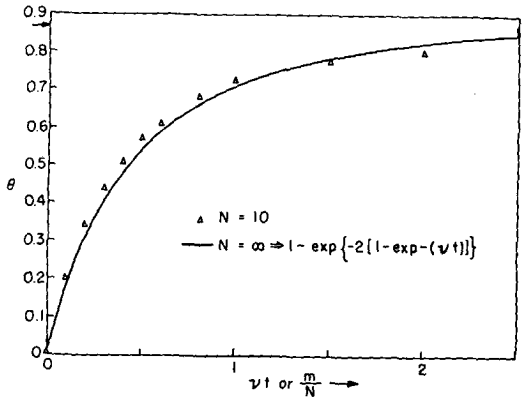


FIG. 3. The fraction of the compartments filled for $N = 10$ and $N \to \infty$ as a function of m/N and exposure.

APPENDIX A

$$N_v(m, N) = \sum_{p=1}^{N-2} pN_p(m, N),$$

where

$$N_p(m, N) = (N - p - 2)N_p(m - 1, N) + 2\sum_{h=p}^{N-2} N_h(m - 1, N) - 2N_{p+1}(m - 1, N)\sum_{h=1}^{N-3} \delta_{ph}.$$

Thus

$$\begin{aligned} N_v(m, N) &= \sum_{p=1}^{N-2} p(N - p - 2)N_p(m - 1, N) \\ &\quad + 2\sum_{p=1}^{N-2} p\sum_{h=p}^{N-2} N_h(m - 1, N) \\ &\quad - 2\sum_{p=1}^{N-2} pN_{p+1}(m - 1, N)\sum_{h=1}^{N-3} \delta_{ph} \\ &= \sum_{p=1}^{N-2} p(N - p - 2)N_p(m - 1, N) \\ &\quad + 2\sum_{p=1}^{N-2} p(p + 1)N_p(m - 1, N) \\ &\quad - 2\sum_{p=2}^{N-2} (p - 1)N_p(m - 1, N) \\ &= (N - 1)\sum_{p=1}^{N-2} pN_p(m - 1, N) \\ &\quad - 2\sum_{p=2}^{N-2} (p - 1)N_p(m - 1, N). \end{aligned}$$

APPENDIX B

For example, the coefficient of

$$\sum_{p=3}^{N-2} (p - 2)N_p(m - r, N)$$

is

$$\begin{aligned} &\left(\frac{-2}{N-1}\right)^2\left[\sum_{n_1=0}^{r-2}\left(\frac{N-2}{N-1}\right)^{n_1}\sum_{n_2=0}^{r-2-n_1}\left(\frac{N-3}{N-1}\right)^{n_2}\right] \\ &= \left(\frac{-2}{N-1}\right)^2\left\{\sum_{n_1=0}^{r-2}\left(\frac{N-2}{N-1}\right)^{n_1}\left[\frac{1-\left(\frac{N-3}{N-1}\right)^{r-1-n_1}}{1-\left(\frac{N-3}{N-1}\right)}\right]\right\} \\ &= \left(\frac{-2}{N-1}\right)^2\left(\frac{N-1}{2}\right)\sum_{n_1=0}^{r-2}\left(\frac{N-2}{N-1}\right)^{n_1} - \left(\frac{-2}{N-1}\right)^2\left(\frac{N-1}{2}\right)\left(\frac{N-3}{N-1}\right)^{r-1}\sum_{n_1=0}^{r-2}\left(\frac{N-2}{N-3}\right)^{n_1} \\ &= \left(\frac{-2}{N-1}\right)^2\frac{(N-1)}{2}\left[\frac{1-\left(\frac{N-2}{N-1}\right)^{r-1}}{1-\left(\frac{N-2}{N-1}\right)}\right] - \left(\frac{-2}{N-1}\right)^2\left(\frac{N-1}{2}\right)\left(\frac{N-3}{N-1}\right)^{r-1}\left[\frac{\left(\frac{N-2}{N-3}\right)^{r-1}-1}{\left(\frac{N-2}{N-3}\right)-1}\right] \\ &= \left(\frac{-2}{N-1}\right)^2\frac{(N-1)^2}{2!}\left[1-\left(\frac{N-2}{N-1}\right)^{r-1}\right] - \left(\frac{-2}{N-1}\right)^2\left(\frac{N-1}{2}\right)\left(\frac{N-3}{1}\right)^{r-1}\left[\left(\frac{N-2}{N-3}\right)^{r-1}-1\right] \\ &= \left(\frac{-2}{N-1}\right)^2\frac{(N-1)^2}{2!}\left[1-\left(\frac{N-2}{N-1}\right)^{r-1}-\left(\frac{N-3}{N-1}\right)\left(\frac{N-2}{N-3}\right)^{r-1}+\left(\frac{N-3}{N-1}\right)^{r}\right] \\ &= \frac{(-2)^2}{2!}\left\{1-\left(\frac{N-2}{N-1}\right)^{r}\left[1+\left(\frac{N-3}{N-1}\right)\right]+\left(\frac{N-3}{N-1}\right)^{r}\right\} \\ &= \frac{(-2)^2}{2!}\left\{1-2\left(\frac{N-2}{N-1}\right)^{r}+\left(\frac{N-3}{N-1}\right)^{r}\right\} \\ &= \frac{(-2)^2}{2!}\sum_{j=0}^{2}(-1)^j{}_2C_j\left(\frac{N-1-j}{N-1}\right)^{r}. \end{aligned}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Crossing and Unitarity in a Multichannel Static Model. I

P. O. G. EHRHARDT AND D. B. FAIRLIE

Mathematics Department, University of Durham, Durham City, Great Britain

(Received 28 November 1967)

The infinite multichannel-scattering process of the two-particle scattering of mesons of arbitrary isospin off an isospin-$\frac{1}{2}$ target, inelastic two-particle channels being admitted, is considered in the static limit. The restrictions on the S matrix arising from crossing under $SU(2)$ and two-particle unitarity are expressed in terms of three equations alone in the limit of degenerate meson masses. These equations are solved by a perturbative technique, the expansion parameter being a measure of the strength of the inelastic scattering. The principal result of this paper is that the inelastic amplitudes are all related, so that if one is zero all must be zero. Comparisons are made with strong-coupling theory, which also requires an infinite number of channels.

1. INTRODUCTION

One of the most challenging problems in strong-interaction physics is the interplay between the restrictions imposed upon scattering amplitudes by invariance requirements under an internal-symmetry group and the constraints of unitarity. The principal difficulty in constructing even a simple dynamical model incorporating both constraints is that the admission of inelastic-scattering processes forces us to treat an infinite number of scattering channels. This difficulty is circumvented in the work of Cushing,[1] who has solved a model which is an abstraction of the pion–nucleon system with two-body unitarity, with $\pi\pi$, NN, and πN elastic-scattering amplitudes, as the inelasticity here arises in the annihilation channels of the πN elastic-scattering process. However, when inelastic-scattering processes are permitted, the problem of dealing with an infinite-dimensional S matrix arises immediately. In this paper we consider such a model, namely, that of the scattering of pseudoscalar mesons of isospin λ off spinless "nucleons" in a p-wave static approximation, taking into account two-particle unitarity. The two $I = \lambda \pm \frac{1}{2}$ elastic amplitudes for the scattering $\lambda + \frac{1}{2} \to \lambda + \frac{1}{2}$ are related by crossing under $SU(2)$, while the amplitude in the $I = \lambda + \frac{1}{2}$ channel is linked to the $I = \lambda + \frac{1}{2}$ amplitude for the elastic scattering of a meson of isospin $\lambda + 1$ off a "nucleon" through the single inelastic amplitude

$$\lambda + \tfrac{1}{2} \to (\lambda + 1) + \tfrac{1}{2}$$

by the requirements of unitarity. Since there is no *a priori* reason for setting any of the inelastic amplitudes equal to zero, there is no terminal value of λ and the S matrix is infinite-dimensional. The equations become uncoupled only when all the inelastic amplitudes vanish, and the solutions of this two-channel problem are known and extensively discussed in the literature.[2–4]

In other words, this model shows that the usual approximation of truncation of multichannel dynamics to a finite number of coupled channels is not mathematically legitimate: if one inelastic amplitude is permitted to be nonzero, an infinitude of inelastic channels are also nonvanishing. In some respects this result is analogous to those of Aks[5] and Cheung and Toll[6] on the impossibility of vanishing production amplitudes in a nontrivial collision process.

The model also bears some resemblance to the results of Geobel's strong-coupling theory,[7] which also demands an infinite set of particles with scattering amplitudes determined in an effective-range approximation. In so far as we treat two-particle unitarity exactly, our model is more sophisticated than Goebel's, although we are compelled to simplify it in other directions—for example, by ignoring spin. It is also interesting to compare Cushing's work[1] with the extensions of strong-coupling theory proposed by Kuriyan and Sudarshan.[8] Both have solutions for a finite number of channels, and the common ingredient is the inclusion of t-channel processes.

2. MULTICHANNEL UNITARITY AND CROSSING

As we mentioned in the Introduction, the choice of an appropriate notation is crucial for the solution of our model. We define the S-matrix elements for the

[1] J. T. Cushing, Phys. Rev. **148**, 1558 (1966).

[2] A. W. Martin and W. D. McGlinn, Phys. Rev. **136**, B1515 (1964).

[3] J. Rothleitner, Zeits.Phys. **177**, 287 (1964).

[4] V. A. Meshcheryakov, Phys. Letters **24B**, 63 (1967).

[5] S. O. Aks, J. Math. Phys. **6**, 516 (1965).

[6] F. F. K. Cheung and J. S. Toll, Phys. Rev. **160**, 1072 (1967).

[7] C. J. Goebel, *Proceedings of Dubna Conference 1962* (unpublished); C. J. Goebel, T. Cook, and B. Sakita, Phys. Rev. Letters **15**, 35 (1965).

[8] J. Kuriyan and E. C. G. Sudarshan, Phys. Letters **21**, 106 (1966); J. Kuriyan and E. C. G. Sudarshan, *ibid.* **16**, 825 (1966).

elastic process

$$\lambda + \tfrac{1}{2} \to \lambda + \tfrac{1}{2} \tag{1a}$$

describing the elastic scattering of a meson, of isospin λ, off a nucleon, by $S_{\mp}(\lambda(\lambda+1), \omega)$, where $\mp$ refer to the channels with total isospin $\lambda_{\mp}$, respectively, ω is the s-channel energy in the static limit (fixed target), and we have chosen to label S according to the Casimir operator of $SU(2)$, rather than λ itself, for reasons which will become apparent later. We describe the inelastic S-matrix element for the process

$$\lambda + \tfrac{1}{2} \to (\lambda+1) + \tfrac{1}{2} \tag{1b}$$

by $S((\lambda+1), \omega)$. This process occurs entirely in the $I = (\lambda + \frac{1}{2})$ channel. The matrix elements $S_{\mp}(\omega)$, $S(\omega)$ satisfy the following analyticity requirements:

(a) $S_{\mp}(\omega)$, $S(\omega)$ are meromorphic functions of ω in the complex ω-plane cut along the real axis from $-\infty < \omega \leq -1$ and $1 \leq \omega < \infty$;

(b) Hermitian analyticity, $S^{x}_{\mp}(\omega) = S_{\mp}(\omega^{x})$:

$$S^{x}(\omega) = S(\omega^{x});$$

(c) Crossing symmetry, $S(-\omega) = S(\omega)$:

$$S_{\alpha}(\lambda(\lambda+1), -\omega) = C_{\alpha\beta} S_{\beta}(\lambda(\lambda+1), \omega) \qquad (\alpha, \beta = -, +),$$

with

$$C_{\alpha\beta} = \frac{1}{2\lambda+1}\begin{pmatrix} -1 & 2(\lambda+1) \\ 2\lambda & 1 \end{pmatrix};$$

(d) Two-particle unitarity:

$$|S_{+}(\lambda(\lambda+1), \omega)|^2 + |S(\lambda+1, \omega)|^2 = 1, \qquad \omega \text{ real} \geq 1,$$

$$S_{+}(\lambda(\lambda+1), \omega) S^{x}(\lambda+1, \omega) + S^{*}_{-}((\lambda+1)(\lambda+2), \omega) S(\lambda+1, \omega) = 0 \qquad \omega \text{ real} \geq 1;$$

(e) Pole structure: $S_{\mp}(\lambda(\lambda+1), \omega)$ possesses pole terms arising from the "nucleon" isobars $N_{\lambda \pm \frac{1}{2}}$; $S(\lambda+1, \omega)$ does not possess a pole in the limit of a degenerate isobaric spectrum.

As may be seen from the equations pertaining to requirements (c) and (d), crossing relates different isospin channels belonging to the same elastic process, while unitarity relates different processes with a common isospin channel. Thus the set of equations does not terminate, as there is no *a priori* reason for taking any of the $S(\lambda, \omega)$ as zero. Our task in disentangling the equations is facilitated by Wanders's trick[9] of decomposing the elastic-matrix elements into symmetric and antisymmetric parts. Let

$$S_{-}(\lambda(\lambda+1), \omega) = A(\lambda(\lambda+1), \omega)[B(\lambda(\lambda+1), \omega) - \lambda - 1],$$

$$S_{+}(\lambda(\lambda+1), \omega) = A(\lambda(\lambda+1), \omega)[B(\lambda(\lambda+1), \omega) + \lambda], \tag{2}$$

where A and B are antisymmetric functions of ω; then the crossing relation given in (c) is automatically satisfied. We note in passing that, under the transformation $\lambda \to -(\lambda+1)$, S_{-} and S_{+} are interchanged, while $C_{\alpha\beta} \to C_{\beta\alpha}$; thus crossing is maintained. This is the reason for the choice of labeling according to $\lambda(\lambda+1)$ rather than λ itself. The unitarity equations then become

$$|A(\lambda(\lambda+1), \omega)|^2 |B(\lambda(\lambda+1), \omega) + \lambda|^2 + |S(\lambda+1, \omega)|^2 = 1, \tag{3}$$

$$A(\lambda(\lambda+1), \omega)[B(\lambda(\lambda+1), \omega) + \lambda] S^{x}(\lambda+1, \omega) + A^{x}((\lambda+1)(\lambda+2), \omega) \times [B((\lambda+1)(\lambda+2, \omega) - \lambda - 2]^{x} S(\lambda+1, \omega) = 0. \tag{4}$$

To ensure compatibility of (3) and (4) when λ is replaced by $-\lambda - 2$, we require that

$$S(-(\lambda+1), \omega) = \pm S(\lambda+1, \omega). \tag{5}$$

The problem of handling the incorporation of inelastic effects becomes one of notation; we succeed in reducing the coupled equations for the infinite-dimensional S matrix to three at the price of the rather drastic imposition of complete degeneracy of the meson spectrum; i.e., all the mesons in the theory, whatever their isospin, have the same mass, taken here as unity, so that all the two-particle unitarity thresholds coincide. Starting from the known solutions,[2,3] where there is no inelasticity, a solution of these equations, incorporating inelasticity, is found as a perturbation series in a parameter ϵ which determines the strength of a typical inelastic coupling. The chief consequence of this model is that the inelasticities cannot be arbitrarily assigned; and this implies that all the meson–nucleon isobar-coupling constants in the model bear some definite relationship to one another. Equations (3)–(5), for ω real > 1, together with the symmetry requirements on A and B (and S) as functions of ω, are the basic equations of the theory. An exact solution is known when $S = 0$[2,3]; in its basic form, it is determined by the equation

$$\operatorname{Re} B_0(\lambda(\lambda+1), \omega) = \tfrac{1}{2}, \tag{6}$$

[9] G. Wanders, Nuovo Cimento **23**, 817 (1962).

which gives

$$B_0(\lambda(\lambda+1),\omega) = \tfrac{1}{2} + \frac{i}{\pi}\log(\omega + [\omega^2-1]^{\frac{1}{2}}) + i(\omega^2-1)^{\frac{1}{2}}q(\omega), \quad (7)$$

where $q(\omega)$ is an arbitrary meromorphic antisymmetric function of ω, and the subscript zero is used to indicate the solution with $S=0$. Note that in this case B_0 has no explicit dependence on λ. The corresponding function A_0 is determined from the recurrence relation

$$A_0(\lambda(\lambda+1),\omega)A_0((\lambda+1)(\lambda+2),\omega) = \frac{\pm 1}{b^2-(\lambda+1)^2}, \quad (8)$$

with the initial condition $A_0(0,\omega) = 1/b$, where we have written b in place of B_0. In this form, it is not evident that $A_0(\lambda(\lambda+1),\omega)$ depends on λ only through the combination $\lambda(\lambda+1)$; however, a power-series solution indicates that A_0 has the formal expansion

$$A_0(\lambda(\lambda+1),\omega) = (\pm 1)^{(|2\lambda+1|+1)/2}\left(\frac{1}{b} + \frac{1}{2}\frac{\lambda(\lambda+1)}{b^3} + \frac{3\lambda^2(\lambda+1)^2 - 2\lambda(\lambda+1)}{8b^5} + \cdots\right) \quad (|\lambda| > 0), \quad (9)$$

where now the dependence is explicit. The choice of sign in (8) is arbitrary: this is reflected in the power series (9), though the sign factor is only valid for integral λ with the second choice for (8). When S is nonzero but small, we suppose it proportional to an inelasticity parameter ϵ, which is pure real or pure imaginary according to the choice (∓ 1) for (8). To first order in ϵ, Eqs. (4), (5), and the symmetry requirements on S give

$$S(\lambda,\omega) = \frac{\epsilon f(\lambda^2)}{b^2-\lambda^2}, \quad (10)$$

where $f(\lambda^2)$ is an undetermined real function of λ. Writing Eq. (3) for λ replaced by $-\lambda-1$ and eliminating the function $A(\lambda(\lambda+1),\omega)$, we have

$$\frac{|B(\lambda(\lambda+1),\omega)+\lambda|^2}{|B(\lambda(\lambda+1),\omega)-\lambda-1|^2} = \frac{1-|S((\lambda+1),\omega)|^2}{1-|S(\lambda,\omega)|^2}. \quad (11)$$

Setting

$$B(\lambda(\lambda+1),\omega) = B_0(\omega) + |\epsilon|^2 B_1(\lambda(\lambda+1),\omega) \quad (12)$$

and using (10), we find that the address of B_1 as a function of ω, together with the interchangeability of λ with $-\lambda-1$, may be satisfied with the choice of $f(\lambda^2)$ as $(\lambda^2-\frac{1}{4})^{\frac{1}{2}}$, giving

$$B_1(\lambda(\lambda+1),\omega) = \frac{-b[b^2-\lambda(\lambda+1)-\frac{1}{2}]}{(b^2-\lambda^2)[b^2-(\lambda+1)^2]}. \quad (13)$$

We do not know whether this choice for $f(\lambda^2)$ is unique; we do not expect so, but this choice does lead to a particularly simple expression for B_1. Finally we use (3) again to determine A_1, the correction to order $|\epsilon|^2$ to A_0; this leads to the equation

$$A_1(\lambda(\lambda+1),b) + A_1(\lambda(\lambda+1),(1-b)) = -\frac{B_1(b)}{b+\lambda} - \frac{B_1(1-b)}{1-b+\lambda} - \frac{(\lambda+1)^2-\frac{1}{4}}{[b^2-(\lambda+1)^2][(1-b)^2-(\lambda+1)^2]}, \quad (14)$$

where

$$A(\lambda(\lambda+1),b) = A_0(\lambda(\lambda+1),b)[1+|\epsilon|^2 A_1(\lambda(\lambda+1),b)]. \quad (15)$$

The solution of Eq. (14) may be expressed as follows:

$$A_1(\lambda(\lambda+1),b) = S(b) + 2\int_0^\infty dx \int_{-\infty}^{\infty} dp\, a(p)\sin(p-b)x\tan\frac{x}{2}, \quad (16)$$

where $S(b)$ and $a(b)$ are even and odd parts of

$$-\frac{B_1(b)}{b+\lambda} + \frac{2b+1}{4[b^2-(\lambda+1)^2]}.$$

This solution is derived and evaluated explicitly in the Appendix.

This perturbative type of solution could, in principle, be carried further, but one feature of the solution is already clear, namely, that the relative magnitudes of the permitted inelastic processes cannot be arbitrarily assigned, but are determined by (15). This feature is ignored in the standard finite coupled-channel calculations, where the potentially infinite set of two-body coupled-scattering processes is arbitrarily truncated. This conclusion could be criticized on the grounds that it was derived with the simplifying assumption of coincident unitarity thresholds: for if indeed the meson spectrum may be arbitrarily assigned, then we should be able to recover the finite model by sending all the meson masses except a finite number to infinity. However, if Goebel's theory[7] is any guide, not only are the meson isobar-coupling ratios determined by the theory, but we can expect the meson spectrum to be determined too. Since we have only a perturbative solution, we cannot complete our solution by fitting the pole terms in the

S-matrix amplitudes [analyticity requirement (d)] using the arbitrariness in the basic solutions (15) and (16).[2,3,9–11] This arbitrariness takes the form of two meromorphic antisymmetric sets of functions

$$q(\lambda(\lambda+1),\omega) \quad \text{and} \quad D(\lambda(\lambda+1),\omega),$$

the first set appearing in the general solution of Eq. (1), the second as an arbitrary multiplicative factor in S and S^*; i.e., given a solution $S(\lambda,\omega)$ and $S_\pm(\lambda(\lambda+1),\omega)$, then $D((\lambda-1)\lambda,\omega)D(\lambda(\lambda+1),\omega)S(\lambda,\omega)$ and $D^2(\lambda(\lambda+1),\omega)S_\mp(\lambda(\lambda+1),\omega)$ are also solutions where each D is meromorphic, antisymmetric, and unitary on the real axis. This procedure would yield a connection between the meson isobar couplings more subtle than that imposed by Goebel's theory. Unfortunately, at this stage in the development of multichannel scattering with exact two-particle scattering, this program does not appear to be technically feasible.

Note added in proof: In a recent paper by B. Sakita ("Strong Coupling of Multipartial Wave Meson Isotriplet," unpublished University of Wisconsin report), an exact solution of the Chew Low equation, in the strong-coupling limit with an arbitrary crossing matrix, is exhibited. This solution is an entirely even function of the energy ω and is obtained by mapping the problem onto the ω^2 plane.

ACKNOWLEDGMENT

P. O. G. Ehrhardt thanks Lancashire County Council for the award of a Research Grant.

APPENDIX

We solve the equation

$$s(b)+s(1-b)=f(b)+f(1-b), \tag{A1}$$

where $s(b)$ is an even function of b and $f(b)$ is arbitrary. It is easy to see that it is sufficient to consider the case where $f(b)$ is odd, as a general $f(b)$ may be split into its symmetric and antisymmetric parts, and the symmetric part taken immediately to the left-hand side. If $S(k)$ and $F(k)$ are the Fourier transforms of s and f, respectively, Eq. (A1), upon transformation, becomes

$$S(k)(1+e^{-ikb})=F(k)(1-e^{-ikb}), \tag{A2}$$

where the parity of s and f is reflected in the coefficient of the exponential in the above equation. This equation may be immediately solved for $S(k)$ and inverted to give

$$s(b)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i(b-p)k}\tan\frac{kp}{2}f(p)\,dp\,dk. \tag{A3}$$

The oddness of $f(p)$ may be exploited to show explicitly that the solution is symmetric in b by the transformation to integrals over positive values of p and k:

$$s(b)=\frac{2}{\pi}\int_0^{\infty}\int_0^{\infty}\cos bk\sin pk\tan\frac{kp}{2}f(p)\,dp\,dk. \tag{A4}$$

The general solution of the homogeneous equation for $s(b)$ must be added to this solution: any S symmetric about both the origin and antisymmetric about the point $b=\frac{1}{2}$ will do; i.e.,

$$s(b)=\sum_{n=0}^{\infty}a_n\cos(2n+1)\pi b, \tag{A5}$$

with a_n arbitrary coefficients, satisfies the homogeneous equation. The solution of Eq. (14) can then be evaluated explicitly. Since we want a solution which is a function of $\lambda(\lambda+1)$, we take the sum of (14) with the equivalent equation with λ replaced by $-(\lambda+1)$ [the difference is automatically satisfied by our solution of Eq. (11)]. For $f(b)$ in Eq. (A1) this gives

$$\begin{aligned}f(b)=&\frac{1}{8}\left[\frac{1}{(b+\lambda)^2}-\frac{1}{(b-\lambda)^2}\right]\\&+\frac{1}{8}\left[\frac{1}{(b-\lambda-1)^2}-\frac{1}{(b+\lambda+1)^2}\right]\\&+\frac{\lambda}{4(2\lambda+1)}\left\{\frac{1}{b+\lambda}+\frac{1}{b-\lambda}\right\}+\frac{(\lambda+1)}{4(2\lambda+1)}\\&\times\left[\frac{1}{(b+\lambda+1)}+\frac{1}{(b-\lambda-1)}\right].\end{aligned} \tag{A6}$$

By using the identity

$$\begin{aligned}i\tan(kp/2)=&\,1-2e^{-ikp}+2e^{-2ikp}\cdots(-1)^ne^{-nikp}\\&+(-1)^ne^{-nikp}i\tan(kp/2),\end{aligned} \tag{A7}$$

the integrals in (A3) may be performed for the first two brackets in (A6) to give the finite sum (λ being taken as integral)

$$\begin{aligned}&-\frac{1}{8}\left(\frac{1}{(b+\lambda)^2}+\frac{1}{(b-\lambda)^2}\right)+\frac{1}{2}\sum_{-\lambda}^{\lambda}(-1)^n\frac{1}{(b+n)^2}\\&-\frac{1}{8}\left(\frac{1}{(b+\lambda+1)^2}+\frac{1}{(b-\lambda-1)^2}\right).\end{aligned} \tag{A8}$$

A similar device may be used to compute the integral for the last two brackets as an infinite series of poles, which may be rearranged into the derivative of the

[10] W. D. McGlinn and C. H. Albright, Nuovo Cimento **27**, 834 (1962).

[11] A. A. Cunningham, Nuovo Cimento **50**, 535 (1967); A. A. Cunningham, J. Math. Phys. **8**, 716 (1967).

logarithm of a product of Gamma functions. Alternatively, we can exploit the identity

$$A_0(\lambda(\lambda+1), b)A_0(\lambda(\lambda+1), (1-b)) \\ (b+\lambda)(1-b+\lambda) = 1. \quad \text{(A9)}$$

Differentiating with respect to λ, we have

$$\frac{d}{d\lambda}\log A_0(b) + \frac{d}{d\lambda}\log A_0(1-b) \\ + \frac{1}{b+\lambda} + \frac{1}{1-b+\lambda} = 0. \quad \text{(A10)}$$

Now $(d/d\lambda)\log A_0(b)$ is an even function of b; hence $(d/d\lambda)\log A_0(b) + \lambda/(b^2-\lambda^2)$ is the solution of (A1) with

$$f(b) = -\frac{b}{b^2-\lambda^2} = -\frac{1}{2}\left[\frac{1}{b-\lambda} + \frac{1}{b+\lambda}\right]. \quad \text{(A11)}$$

Thus, the solution for the second two terms in (A6) may be written succintly as

$$\frac{1}{4}\left(\frac{-\lambda}{(2\lambda+1)}\frac{d}{d\lambda} + \frac{\lambda+1}{2\lambda+1}\frac{d}{d(\lambda+1)}\right)\log A_0(b) \\ + \frac{1}{4(2\lambda+1)}\left(\frac{\lambda^2}{b^2-\lambda^2} - \frac{(\lambda+1)^2}{b^2-(\lambda+1)^2}\right). \quad \text{(A12)}$$

Collecting together (A8), (A12), and the symmetric parts of the right-hand side of (14), for the solution of (14) we finally obtain

$$A_1(\lambda(\lambda+1), b) \\ = -\frac{1}{4(2\lambda+1)}\left(\lambda\frac{\partial}{\partial\lambda} - (\lambda+1)\frac{\partial}{\partial(\lambda+1)}\right) \\ \times \log A_0(\lambda(\lambda+1), b) \\ + \frac{1}{2(2\lambda+1)}\left(\frac{1}{b^2-\lambda^2} - \frac{1}{b^2-(\lambda+1)^2}\right) \\ - \frac{1}{8}\left(\frac{1}{(b+\lambda)^2} + \frac{1}{(b-\lambda)^2}\right. \\ \left. + \frac{1}{(b+\lambda+1)^2} + \frac{1}{(b-\lambda-1)^2}\right) \\ + \frac{1}{2}\sum_{-\lambda}^{\lambda}(-1)^n\frac{1}{(b+n)^2}. \quad \text{(A13)}$$

This completes the solution of Eqs. (4), (11), and (14) in the main text. The arbitrariness in solution of this type of equation [(A5)] is just that of the arbitrary phase factor $D(\omega)$, as we are working to order ϵ^2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Poincaré Group and the Invariant Relativistic Equations for Massive Particles of Any Spin*

JOSÉ A. DE AZCÁRRAGA AND LUIS J. BOYA
Facultad de Ciencias, Instituto de Física Teórica, Barcelona, Spain

(Received 29 June 1967)

The aim of this paper is to clarify some aspects of the connection between the Poincaré group and the invariant equations for nonzero-mass particles of any spin (Bargmann–Wigner equations). With this purpose we first make some general considerations about the representations of the Poincaré group and analyze the equivalence between the realizations corresponding to a given class and a selected "canonical" one which was given by Wigner. For the spin-$\frac{1}{2}$ case the equivalence is given by the Chakrabarti transformation, and for higher spins we introduce a generalization of it; we also consider specifically the case $S = 1$. We give the form of the equations which provide the corresponding canonical realization; some comments about the equivalence of the theories provided by the different realizations are also made.

1. INTRODUCTION

The main point on the realization of the Hilbert space for the manifold of states of a relativistic particle through covariant equations is the apparent incompatibility between irreducibility and manifest covariance. We do not intend to solve this problem, but rather to establish a clearer connection between representations and covariant equations.

* This research has been sponsored in part by the U.S. Air Force through the European Office of Aerospace Research, EOAR, U.S. Air Force, under grant AF-EOAR-64-17.

We start with the following theorem[1]:

Theorem: Every continuous representation of $\bar{\mathfrak{I}}_+^\uparrow$ (universal covering group of the restricted Poincaré group $\mathfrak{I}_+^\uparrow$) corresponding to the case $p^2 = p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2 > 0$ is equivalent to a representation over the Hilbert space of the functions on the mass hyperboloid Ω_m of the form

$$(U(a, A)\psi)(p) = \exp ia\cdot p\; Q(p, A)\psi(\Lambda(A^{-1})p), \quad (1.1)$$

[1] A. S. Wightman, in *Les Houches Lectures* (Hermann & Cie., Paris, 1960). We follow in this section the notations used here.

logarithm of a product of Gamma functions. Alternatively, we can exploit the identity

$$A_0(\lambda(\lambda+1), b)A_0(\lambda(\lambda+1), (1-b)) \\ (b+\lambda)(1-b+\lambda) = 1. \quad \text{(A9)}$$

Differentiating with respect to λ, we have

$$\frac{d}{d\lambda}\log A_0(b) + \frac{d}{d\lambda}\log A_0(1-b) \\ + \frac{1}{b+\lambda} + \frac{1}{1-b+\lambda} = 0. \quad \text{(A10)}$$

Now $(d/d\lambda)\log A_0(b)$ is an even function of b; hence $(d/d\lambda)\log A_0(b) + \lambda/(b^2-\lambda^2)$ is the solution of (A1) with

$$f(b) = -\frac{b}{b^2-\lambda^2} = -\frac{1}{2}\left[\frac{1}{b-\lambda} + \frac{1}{b+\lambda}\right]. \quad \text{(A11)}$$

Thus, the solution for the second two terms in (A6) may be written succintly as

$$\frac{1}{4}\left(\frac{-\lambda}{(2\lambda+1)}\frac{d}{d\lambda} + \frac{\lambda+1}{2\lambda+1}\frac{d}{d(\lambda+1)}\right)\log A_0(b) \\ + \frac{1}{4(2\lambda+1)}\left(\frac{\lambda^2}{b^2-\lambda^2} - \frac{(\lambda+1)^2}{b^2-(\lambda+1)^2}\right). \quad \text{(A12)}$$

Collecting together (A8), (A12), and the symmetric parts of the right-hand side of (14), for the solution of (14) we finally obtain

$$A_1(\lambda(\lambda+1), b) \\ = -\frac{1}{4(2\lambda+1)}\left(\lambda\frac{\partial}{\partial\lambda} - (\lambda+1)\frac{\partial}{\partial(\lambda+1)}\right) \\ \times \log A_0(\lambda(\lambda+1), b) \\ + \frac{1}{2(2\lambda+1)}\left(\frac{1}{b^2-\lambda^2} - \frac{1}{b^2-(\lambda+1)^2}\right) \\ - \frac{1}{8}\left(\frac{1}{(b+\lambda)^2} + \frac{1}{(b-\lambda)^2}\right. \\ \left. + \frac{1}{(b+\lambda+1)^2} + \frac{1}{(b-\lambda-1)^2}\right) \\ + \frac{1}{2}\sum_{-\lambda}^{\lambda}(-1)^n\frac{1}{(b+n)^2}. \quad \text{(A13)}$$

This completes the solution of Eqs. (4), (11), and (14) in the main text. The arbitrariness in solution of this type of equation [(A5)] is just that of the arbitrary phase factor $D(\omega)$, as we are working to order ϵ^2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Poincaré Group and the Invariant Relativistic Equations for Massive Particles of Any Spin*

JOSÉ A. DE AZCÁRRAGA AND LUIS J. BOYA
Facultad de Ciencias, Instituto de Física Teórica, Barcelona, Spain

(Received 29 June 1967)

The aim of this paper is to clarify some aspects of the connection between the Poincaré group and the invariant equations for nonzero-mass particles of any spin (Bargmann–Wigner equations). With this purpose we first make some general considerations about the representations of the Poincaré group and analyze the equivalence between the realizations corresponding to a given class and a selected "canonical" one which was given by Wigner. For the spin-$\frac{1}{2}$ case the equivalence is given by the Chakrabarti transformation, and for higher spins we introduce a generalization of it; we also consider specifically the case $S = 1$. We give the form of the equations which provide the corresponding canonical realization; some comments about the equivalence of the theories provided by the different realizations are also made.

1. INTRODUCTION

The main point on the realization of the Hilbert space for the manifold of states of a relativistic particle through covariant equations is the apparent incompatibility between irreducibility and manifest covariance. We do not intend to solve this problem, but rather to establish a clearer connection between representations and covariant equations.

* This research has been sponsored in part by the U.S. Air Force through the European Office of Aerospace Research, EOAR, U.S. Air Force, under grant AF-EOAR-64-17.

We start with the following theorem[1]:

Theorem: Every continuous representation of $\bar{\mathfrak{I}}_+^\uparrow$ (universal covering group of the restricted Poincaré group $\mathfrak{I}_+^\uparrow$) corresponding to the case $p^2 = p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2 > 0$ is equivalent to a representation over the Hilbert space of the functions on the mass hyperboloid Ω_m of the form

$$(U(a, A)\psi)(p) = \exp ia\cdot p\; Q(p, A)\psi(\Lambda(A^{-1})p), \quad (1.1)$$

[1] A. S. Wightman, in *Les Houches Lectures* (Hermann & Cie., Paris, 1960). We follow in this section the notations used here.

where a is a translation, A belongs to the unimodular group in two dimensions $SL(2, C)$, and $\Lambda(A)$ is the restricted Lorentz transformation associated to it[2] (the connection between the occurring groups can be seen in the Appendix).

The Q factors, which can depend on the particular point p, satisfy the conditions

$$Q(p, A_1)Q(\Lambda(A_1^{-1})p, A_2) = Q(p, A_1A_2), \quad (1.2a)$$

$$Q(p, I) = I, \quad (1.2b)$$

which show that the $Q(p, A)$ which operate in the "internal" variable of ψ constitute a sort of representation of $SL(2, C)$.

As usual, two representations are equivalent if

$$U_2(a, A) = VU_1(a, A)V^{-1}. \quad (1.3)$$

Without loss of generality,[1] we can take $U_2(a, I) = U_1(a, I)$; then, if

$$(V\psi)(p) = V(p)\psi(p) \quad \text{(local operator)}, \quad (1.4)$$

it is easily seen that

$$Q_2(p, A) = V(p)Q_1(p, A)V(\Lambda(A^{-1})p)^{-1}. \quad (1.5)$$

If A belongs to the little group L_k of k [$A \in L_k$ if $\Lambda(A)k = k$,] Eq. (1.5) is the expression of the equivalence of two representations of it; namely,

$$Q_0'(k, A) = V(k)Q_0(k, A)V(k)^{-1}; \quad (1.6)$$

Q_0 is then a representation of the unitary unimodular group $SU(2, C)$, which is the little group of a p such that $p^2 > 0$.

It is known[1] that there is a $V(p)$ which verifies (1.5) for every p of an orbit and every $A \in SL(2, C)$ if and only if there is a $V(k)$ which verifies (1.6) for only a k and any $A \in L_k$. In fact, given a representation Q_0 of the little group of a vector k, $Q(p, A)$ is determined by the relation

$$Q(p, A) = Q(k, A_{p\leftarrow k}^{-1})^{-1}Q_0(k, A_{p\leftarrow k}^{-1}AA_{p'\leftarrow k})Q(k, A_{p'\leftarrow k}^{-1}), \quad (1.7)$$

where $A_{p\leftarrow k}$ is a standard transformation which verifies

$$\Lambda(A_{p\leftarrow k})k = p,$$

and where

$$A_{p\leftarrow k}^{-1}AA_{p'\leftarrow k} \equiv A_k \in L_k \quad \text{and} \quad \Lambda(A^{-1})p = p'.$$

(1.7) can be written, taking into account (1.2) in the form

$$Q(p, A) = Q(p, A_{p\leftarrow k})Q_0(k, A_k)Q(p', A_{p'\leftarrow k})^{-1}. \quad (1.8)$$

[2] The homomorphism $A \to \Lambda(A)$ is between $SL(2, C)$ and the restricted Lorentz group $L_+^\uparrow$; see the Appendix. See also Ref. 1 or A. J. Macfarlane, J. Math. Phys. **3**, 1116 (1962).

Distinct $Q(p, A)$ are obtained from distinct $Q(p, A_{p\leftarrow k})$, and it is natural to define the canonical representation[3] by taking $Q(p, A_{p\leftarrow k}) = I$ for it. Then,

$$Q(p, A) = Q_0(k, A_k),$$

and in that case

$$(U_0(a, A)\varphi)(p) = \exp ia\cdot p\, Q_0(k, A_k)\varphi(\Lambda(A^{-1})p), \quad (1.9)$$

and (1.8) shows the equivalence between the two representations with

$$V(p) = Q(p, A_{p\leftarrow k}).$$

We note that the irreducibility of the representation is not necessary for the preceding reasonings.

2. EQUIVALENCE BETWEEN DIFFERENT THEORIES AND SCALAR PRODUCT

Let φ be the generic vector of the Hilbert-space support of the canonical realization and ψ that corresponding to another realization of the same class. Then, restricting ourselves to the elements of $SL(2, C)$, we have

$$(U(A)\varphi)(p) = Q_0\varphi(p'), \quad (2.1a)$$

$$(U(A)\psi)(p) = V(p)Q_0V(p')^{-1}\psi(p'), \quad (2.1b)$$

where

$$Q(p, A) = V(p)Q_0V(p')^{-1}.$$

We define the ordinary scalar product in φ space:

$$\langle\varphi \mid \varphi\rangle = \int \varphi(p)^\dagger\varphi(p)\, d\Omega_m, \quad (2.2)$$

where $d\Omega_m$ is the well-known invariant measure $dp_1\, dp_2\, dp_3/|p_0|$ on the mass hyperboloid Ω_m. This scalar product is positive-definite and invariant, as is evident.

We want to show now that it is possible to establish the equivalence between the theories provided by φ and ψ by means of an *isometric* and *covariant* application, which is $V(p)$, and the corresponding invariant scalar product in ψ space.

To see this, we make the correspondence

$$\psi(p) = V(p)\varphi(p), \quad (2.3)$$

which is covariant, as is shown by the following commutative diagram:

$$\begin{array}{ccc} \varphi(p) = V(p)^{-1}\psi(p) & \xrightarrow{U_0(A)} & Q_0\varphi(p') = Q_0V(p')^{-1}\psi(p') \\ \downarrow V(p) & & \downarrow V(p) \\ \psi(p) & \xrightarrow{U(A)} & V(p)Q_0V(p')^{-1}\psi(p'). \end{array}$$

[3] E. P. Wigner, Ann. Math. **40**, 149 (1939). Another realization which considers the tensor product of the representation $(m, 0)$ and the $D(s, 0)$ representation of the homogeneous Lorentz group has been given by R. Shaw [Nuovo Cimento **33**, 1074 (1964)].

[In fact, it may be seen directly because

$$U(A) = Q(p, A)P(A) = V(p)Q_0V(p')^{-1}P(A) = V(p)Q_0P(A)V(p)^{-1},$$

the action of $P(A)$ being reduced to change p by $p' = \Lambda(A^{-1})p$.]

We take now the scalar product in ψ space in the way

$$\langle \psi \mid \psi \rangle = \int \psi(p)^\dagger \rho(p)\psi(p)\, d\Omega_m, \tag{2.4a}$$

$$\rho(p) = V(p)^{-1\dagger}V(p)^{-1}. \tag{2.4b}$$

This scalar product is again clearly covariant and positive-definite [see (2.1b); $VV^\dagger$ and $V^{-1\dagger}V^{-1}$ are nonnegative-definite operators]. Then, the ψ space being endowed with the above metric, we have shown the equivalence between the theories constructed over these spaces.

3. THE DIRAC EQUATION

We now apply the preceding considerations to the representation of $\bar{\mathfrak{P}}^\uparrow_+$ provided by the Dirac equation in order to obtain the relation existent between it and the corresponding canonical realization.

As it is well known, the four-spinors solution of the Dirac equation

$$(\gamma(p) - m)\psi(p) = 0$$

take values on the two sheets of Ω_m, corresponding to the two signs of the energy; this is, in fact, a general feature of the solutions of the manifestly local covariant equations. The ψ's form a support space of a representation of $\bar{\mathfrak{P}}^\uparrow_+$, which is defined in the following way:

$$(U(a, A)\psi)(p) = \exp ia \cdot p\, S(A)\psi(\Lambda(A^{-1})p), \qquad p \in \Omega_m, \tag{3.1}$$

the $S(A)$ being four-squared matrices satisfying

$$S(A)\gamma(p)S(A)^{-1} = \gamma(\Lambda(A)p). \tag{3.2}$$

We take the γ's in a representation in which γ^0 is diagonal, $\gamma^{0\dagger} = \gamma^0$ and $\gamma^{i\dagger} = -\gamma^i$, $i = 1, 2, 3$; $\gamma(p) = \gamma^0p^0 - \sum \gamma^ip^i$.

Let us reduce the representation (3.1) to the canonical form. We have to use the little group of $k = (\epsilon m, 0, 0, 0)$, ϵ being ± 1 for $p \in \Omega^\pm_m$. As we know, every A can be written in the form

$$A = A_{p \leftarrow k}(A^{-1}_{p \leftarrow k}AA_{p' \leftarrow k})A^{-1}_{p' \leftarrow k},$$

the bracket belonging to L_k and hence to $SU(2, C)$ [universal covering group of $SO(3, R)$]. Then the following relation holds:

$$S(A) = S(A_{p \leftarrow k})S(A_k)S(A^{-1}_{p' \leftarrow k}). \tag{3.3}$$

$S(A_k)$ is a representation of $SU(2, C)$; then, according to (1.5), the equivalence between the two realizations is brought about by taking

$$S(A_{p \leftarrow k}) = V(p).$$

It is not difficult to find the corresponding form of $V(p)$ and $V(p)^{-1}$, which can be written as

$$V(p) = \frac{\gamma^0\gamma(p)^\dagger + \epsilon m}{[2\epsilon m(\epsilon m + p^0)]^{\frac{1}{2}}},$$

$$V(p)^{-1} = \frac{\gamma^0\gamma(p) + m}{[2\epsilon m(\epsilon m + p^0)]^{\frac{1}{2}}}. \tag{3.4}$$

Since the points $(\epsilon m, 0, 0, 0)$ enable us to reach only points p belonging to Ω^+_m under the action of $SL(2, C)$, the ϵ is to be taken $+$ or $-$ according to $p \in \Omega^\pm_m$ or $p \in \Omega^-_m$.

Let us apply the obtained transformation to the Dirac spinors. We have

$$\varphi(p) = V(p)^{-1}\psi(p) \tag{3.5a}$$

and $(p \in \Omega_m)$

$$V(p)^{-1}(\gamma(p) - m)V(p) = (\epsilon m\gamma^0 - m). \tag{3.5b}$$

Taking into account the diagonal form of γ^0, Eqs. (3.5) show that $\varphi(p)$ has only the two upper components different from zero if $p \in \Omega^+_m$, and the two lower ones if $p \in \Omega^-_m$. Then the $\varphi(p)$, with $p \in \Omega_m$, decompose in the sum of two $\varphi_i(p)$, $i = 1, 2$, each of them having support in a different sheet of Ω_m and corresponding to a definite sign of the energy. In fact, we have shown explicitly the equivalence of the representation $\bar{\mathfrak{P}}^\uparrow_+$ provided by the Dirac spinors with the direct sum of the two irreducible representations $[m, \frac{1}{2}, +]$ and $[m, \frac{1}{2}, -]$. Of course, the representation of $SU(2, C)$; acts only on $\varphi(p)$ it is not difficult to show that $S(A_k)$ decomposes in the direct sum $D^{\frac{1}{2}} + D^{\frac{1}{2}}$ of representations of $SU(2, C)$. The realization over the φ space has then the canonical form (1.9).

The equation which gives rise to the canonical realization can be obtained applying to the arbitrary spinor $\psi(p)$ the transformation

$$V_C(p) = \frac{\gamma^0\gamma(p)^\dagger + \epsilon K}{[2\epsilon K(\epsilon K + p^0)]^{\frac{1}{2}}}, \quad \epsilon = \pm 1, \quad p \in \Omega^\pm_m, \tag{3.6}$$

which is obtained by considering the transformation which brings the arbitrary vector p to the vector $(\epsilon K, 0, 0, 0)$, K being $+(p^2)^{\frac{1}{2}}$ and therefore a number for the corresponding mass shell. In this way, we obtain

$$(\epsilon K\gamma^0 - m)\varphi(p) = 0. \tag{3.7}$$

This equation decomposes, giving two *analogous*

equations, each of them corresponding to a definite sign of the energy, which can be written as

$$(K - m)\varphi_i(p) = 0.$$

In fact, Eq. (3.7) is the Chakrabarti form of the Dirac equation[4] and (3.6) is the Chakrabarti transformation.

The correspondence between the two realizations is made by $V(p)$; the new scalar product is

$$\langle \varphi \mid \varphi \rangle = \int \varphi(p)^\dagger \varphi(p)\, d\Omega_m$$

and is equal to

$$\int \psi(p)^\dagger V(p)^{-1\dagger} V(p)^{-1} \psi(p)\, d\Omega_m = \int \psi(p) \frac{\gamma^0 \gamma(p)}{\epsilon m} \psi(p)\, d\Omega_m = \epsilon \int \psi(p)^\dagger \gamma^0 \psi(p)\, d\Omega_m,$$

which is the usual scalar product for the Dirac equation.

4. THE PROCA EQUATIONS

As it is well known, the Proca equations in momentum space read

$$(p^2 - m^2)\psi(p) = 0, \tag{4.1a}$$

$$(p, \psi(p)) = 0, \tag{4.1b}$$

the $\psi(p)$ being a vector-valued function on Ω_m (4.1a) and tangent to it (4.1b); $(\ ,\ )$ is the usual Lorentz scalar product. The Proca equations describe massive particles of spin $S = 1$, the corresponding representation of $\bar{\mathcal{T}}^\uparrow_+$ being defined by

$$(U(a, A)\psi)(p) = \exp ia \cdot p\, \Lambda(A)\psi(\Lambda(A^{-1})p). \tag{4.2}$$

$\Lambda(A)$ is the natural representation of $L^\uparrow_+$ in the Minkowski space.

Let us reduce (4.2) to the canonical form. We have to consider the standard element $\Lambda_{k\leftarrow p}$, which brings the generic vector p to the point $(\epsilon m, 0, 0, 0)$. In this representation, $\Lambda_{k\leftarrow p}$ has the explicit form

$$\Lambda_{k\leftarrow p} = \begin{pmatrix} \frac{p_0}{\epsilon m} & -\frac{p_1}{\epsilon m} & -\frac{p_2}{\epsilon m} & -\frac{p_3}{\epsilon m} \\ -\frac{p_1}{\epsilon m} & 1 + \frac{p_1^2}{[\ ]} & \frac{p_1 p_2}{[\ ]} & \frac{p_1 p_3}{[\ ]} \\ -\frac{p_2}{\epsilon m} & \frac{p_2 p_1}{[\ ]} & 1 + \frac{p_2^2}{[\ ]} & \frac{p_2 p_3}{[\ ]} \\ -\frac{p_3}{\epsilon m} & \frac{p_3 p_1}{[\ ]} & \frac{p_3 p_2}{[\ ]} & 1 + \frac{p_3^2}{[\ ]} \end{pmatrix}, \tag{4.3}$$

where $[\ ] = \epsilon m(\epsilon m + p_0)$.

[4] A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963).

The new functions are taken as

$$\varphi(p) = \frac{(K + m)}{2m}(\Lambda_{k\leftarrow p})\psi(p);$$

and taking into account that

$$0 = (p, \psi(p)) = (\Lambda_{k\leftarrow p} p, \Lambda_{k\leftarrow p}\psi(p)),$$

it appears evident that they have only three nonzero components. The canonical representation is then defined by

$$(U(a, A)\varphi)(p) = \exp ia \cdot p\, D^1(\Lambda(A_k))\varphi(\Lambda(A^{-1})p), \tag{4.4}$$

$\Lambda(A_k)$ being a rotation which leaves invariant the vector $(\epsilon m, 0, 0, 0)$ and which operates in the space tangent to Ω_m at k . Λ

Operating in a similar form as we did in the Dirac case, the canonical form of the Proca equations is found to be

$$(K - m)\varphi(p) = 0, \tag{4.5a}$$

$$(k, \varphi(p)) = 0, \tag{4.5b}$$

and again we have that the positive definite product

$$\langle \psi \mid \psi \rangle = -\int (\psi(p),^\dagger \psi(p))\, d\Omega_m$$

is equal to

$$\langle \varphi \mid \varphi \rangle = \int \varphi(p)^\dagger \varphi(p)\, d\Omega_m,$$

since

$$-(\varphi(p)^\dagger, \varphi(p)) = \varphi(p)^\dagger \varphi(p).$$

5. BARGMANN–WIGNER EQUATIONS

We consider now the Bargmann–Wigner (B–W) equations for massive particles of any spin.[5] As is known, the B–W equations for nonzero mass particles can be written as a set of $n = 2s$ (s = spin) equations of the form

$$(\gamma(p)^{(r)} - m)\psi(p, \xi_1 \cdots \xi_r \cdots \xi_n) = 0, \tag{5.1}$$

ψ being a symmetric function of the n four-valued spin variables ξ_r, having in all $2\ (2s + 1)$ independent components because they describe particles with a unique spin, but they include the two signs of the energy. $\gamma(p)^{(r)}$ acts on the variable ξ_r; if $n = 1$, we have the Dirac equation.

If we consider ψ as an element of the 4^{2s}-dimensional space obtained by taking the tensor product of the $2s$ four-dimensional spaces, $\gamma(p)^{(r)}$ can be written in the

[5] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **34**, 211 (1948).

form

$$\gamma(p)^{(r)} = 1 \otimes \cdots \otimes \gamma(p) \otimes \cdots \otimes 1,$$

the factor $\gamma(p)$ being placed in the rth position.

Following a procedure analogous to that used in the Dirac case, we define, with $V(p)$ given by (3.4),

$$V(p)^{(r)} = 1 \otimes \cdots \otimes V(p) \otimes \cdots \otimes 1, \qquad (5.2)$$

and we obtain

$$(V(p)^{(r)})^{-1}\gamma(p)^{(r)}V(p)^{(r)} = \epsilon m^{\gamma O(r)}.$$

Moreover, taking in account the following property of the Kronecker product of matrices,

$$(D_1 \otimes D_2 \cdots \otimes D_n)(D_1' \otimes D_2' \cdots \otimes D_n') = (D_1D_1' \otimes D_2D_2' \otimes \cdots \otimes D_nD_n'),$$

it is obvious that $V(p)^{(r)}$ commutes with $V(p)^{(r')}$. Then, if

$$B(p) = \prod_{r=1}^{n} V(p)^{(r)},$$

it is evident that

$$B(p)^{-1}(\gamma(p)^{(r)} - m)B(p) = \epsilon m\gamma^{0(r)} - m.$$

Then the functions $\varphi(p, \xi_1 \cdots \xi_n)$, which come from the $\psi(p, \xi_1 \cdots \xi_n)$ solution of (5.1), take values on Ω_m^+ for $\xi_r = 1, 2$ and on Ω_m^- for $\xi_r = 3, 4$, being zero for the other values in each case. There are

$$\binom{2s + 2 - 1}{2s} = 2s + 1$$

independent components for each sign of the energy, and the corresponding spaces are support of the representation D^s of $SU(2, C)$, as corresponds to the canonical realization.

The form of the equations (5.1) analogous to the form (3.7) for the Dirac equation can be obtained in a similar way. We put[6]

$$A(p) = \prod_{r=1}^{2s} V_C(p)^{(r)}, \qquad (5.3)$$

with $V_C(p)$ given by (3.6); applying this transformation, we obtain the B–W equations in the form ($\varphi = A^{-1}\psi$):

$$(\epsilon K\gamma^{0(r)} - m)\varphi(p, \xi_1 \cdots \xi_n) = 0. \qquad (5.4)$$

Equations (5.4) give rise to two analogous sets of equations, each of them for a sign of the energy whose solutions are support of the D^s representation of $SU(2, C)$ as stated.

We consider now the equivalence between the theories provided by (5.1) and (5.4). In φ space the scalar product is

$$\langle \varphi \mid \varphi \rangle = \int \varphi^\dagger(p, \xi_1 \cdots \xi_n)\varphi(p, \xi_1 \cdots \xi_n)\, d\Omega_m;$$

it is positive-definite and equal to

$$\langle \psi \mid \psi \rangle = \int \psi(p, \xi_1 \cdots \xi_n)^\dagger B(p)^{-1\dagger}B(p)^{-1}\psi(p, \xi_1 \cdots {}_1\, \xi_n)\, d\Omega_m$$

$$= \int \psi^\dagger \underbrace{\epsilon\gamma^0 \otimes \cdots \otimes \epsilon\gamma^0}_{\text{n factors}}\psi\, d\Omega_m.$$

Finally, it is interesting to note that the existence of the ϵ factor is only necessary for half-integer spin (in that case, n is odd); for integer spin it is irrelevant. This is on the basis of the spin-statistics theorem of Pauli.

ACKNOWLEDGMENTS

One of us (J. A. A.) is indebted to the Fundación Juan March, Madrid, for financial support during the completion of this work.

APPENDIX

The relations between the groups $\mathfrak{T}_+^\uparrow$, $\bar{\mathfrak{T}}_+^\uparrow$, $SL(2, C)$, and $L_+^\uparrow$ can be easily seen in the following diagram of exact sequences:

$$\begin{array}{ccccccccc} & & & & 1 & & 1 & & \\ & & & & \downarrow & & \downarrow & & \\ & & & & Z_2 & = & Z_2 & & \\ & & & & \downarrow & & \downarrow & & \\ 1 & \to & \mathrm{Tr} & \to & \bar{\mathfrak{T}}_+^\uparrow & \rightleftarrows & SL(2, C) & \rightleftarrows & 1 \\ & & & & \downarrow & & \downarrow & & \\ 1 & \to & \mathrm{Tr} & \to & \mathfrak{T}_+^\uparrow & \rightleftarrows & L_+^\uparrow & \rightleftarrows & 1 \\ & & & & \downarrow & & \downarrow & & \\ & & & & 1 & & 1 & & \end{array}$$

Z_2 is the group of two elements, Tr is the group of translations. The diagram shows that $\bar{\mathfrak{T}}_+^\uparrow(\mathfrak{T}_+^\uparrow)$ is the semidirect product of $SL(2, C)(L_+^\uparrow)$ and Tr.

[6] A similar procedure to that given here has been used by Pursey to generalize the Foldy–Wouthuysen transformation [D. L. Pursey, Nucl. Phys. **53**, 174 (1964).

Retarded and Advanced Electromagnetic Fields in Friedmann Universes*

AMNON KATZ†
Department of Physics, University of Washington, Seattle, Washington

(Received 9 February 1968)

Maxwell's equations of classical electrodynamics are solved in the framework of the curved space-time of the Friedmann universes. Explicit formulas are given for the retarded and advanced fields in terms of the four-current.

1. INTRODUCTION

This paper treats Maxwell's equations of classical electrodynamics[1] in the framework of the curved space-time of the Friedmann universes.[2] Retarded and advanced solutions are found and stated explicitly. The metric and Maxwell's equations are introduced in Sec. 2. The solutions are given in Sec. 9.

We write the Friedmann metric in a form that makes it the conformal transform of a static homogeneous metric. Since Maxwell's equations are conformal-invariant, it is then enough to solve them in the static homogeneous case. We treat the spherical (closed), flat, and hyperbolic (open) cases simultaneously. (The static homogeneous universe in the flat case is, of course, an ordinary Minkowski space. We carry it along for completeness and comparison.)

In the spherical (closed) case, the total charge in the universe must vanish and no solution exists for just a single charge. In Sec. 10 we show that the solution we have obtained does in fact correspond in this case to a pair of oppositely charged particles. The "countercharge" appears at the diametrically opposite point of space at the advanced or retarded time and is surrounded by the opposite kind of field (in the sense of retarded or advanced). In Sec. 11 we show how it is possible to superimpose solutions of this kind to obtain the field of a pair of oppositely charged particles following arbitrarily prescribed world lines and with independent choice of retarded or advanced field for either particle.

We use Greek indices that range from 0 to 3, and Latin indices that range from 1 to 3. The summation convention is used. The symbol $|\mathbf{x}|$ stands for $((x^1)^2 + (x^2)^2 + (x^3)^2)^{\frac{1}{2}}$. This quantity is also denoted by r. This r is quite distinct from r that appears as a component index (as in x^r).

* Supported in part by the U.S. Atomic Energy Commission under contract A.T.(45-1)1388, program B.

† Present address: The Weizmann Institute, Rehovoth, Israel.

[1] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Co., Reading, Mass., 1959), p. 272.

[2] A. Friedmann, Z. Physik **10**, 377 (1922); **21**, 326 (1924). See also Ref. 1, p. 336.

2. MAXWELL'S EQUATIONS IN A FRIEDMANN UNIVERSE

We propose to solve Maxwell's equations

$$(-g)^{-\frac{1}{2}}\{(-g)^{\frac{1}{2}}g^{\mu\rho}g^{\nu\sigma}F_{\rho\sigma}\}_{,\nu} = \frac{4\pi}{c^2}J^{\mu}, \tag{2.1}$$

$$F_{\mu\nu,\rho} + F_{\nu\rho,\mu} + F_{\rho\mu,\nu} = 0 \tag{2.2}$$

in a Friedmann universe. The commas in the last equations denote ordinary derivatives. We write the metric of the Friedmann universe in the form

$$ds^2 = [f(t)]^2\left\{c^2\,dt^2 - \frac{dx^2 + dy^2 + dz^2}{[S(r)]^2}\right\}, \tag{2.3}$$

where $r \equiv (x^2 + y^2 + z^2)^{\frac{1}{2}}$ and

$$S(r) \equiv 1 + \kappa(r^2/4a^2), \tag{2.4}$$

with κ being $+1$, -1, or 0. The function $f(t)$ is given by

$$f(t) = \begin{cases} 1 - \cos(ct/a), & \text{when } \kappa = 1, \\ (t/t_0)^2, & \text{when } \kappa = 0, \\ \cosh(ct/a) - 1, & \text{when } \kappa = -1. \end{cases} \tag{2.5}$$

The Friedmann line element is often stated in terms of a different time coordinate τ related to t by

$$d\tau = f(t)\,dt. \tag{2.6}$$

The line element then has the form $ds^2 = c^2\,d\tau^2 - (f/S)^2(dx^2 + dy^2 + dz^2)$. A standard clock comoving with the space coordinates measures τ. However, the use of t has the advantage that the line element of Eq. (2.3) appears as a conformally transformed line element of a static homogeneous universe. Since the Maxwell equations are conformal-invariant, it suffices to solve them in the static homogeneous case.

From now on and in the following sections we disregard Eq. (2.5) and put

$$f(t) = 1 \tag{2.7}$$

in Eq. (2.3).

The static homogeneous universe is of cylindrical nature: it is the direct product of a one-dimensional space (time) by a three-dimensional homogeneous space (space). It therefore makes sense to separate space

indices from time indices and use the tensor properties of the various quantities under transformations of the space coordinates only.

When Eqs. (2.3) and (2.7) are substituted into Eqs. (2.1) and (2.2) and the latter are separated into space components (denoted by Latin indices) and time components (denoted by the index 0), we find

$$-(F_{0s}/cS)_{,s} = 4\pi J^0/cS^3, \tag{2.8}$$

$$-(F_{r0}/cS)_{,0} + (cSF_{rs})_{,s} = 4\pi J^r/cS^3, \tag{2.9}$$

$$F_{rs,0} + F_{0r,s} - F_{0s,r} = 0, \tag{2.10}$$

$$F_{rs,t} + F_{st,r} + F_{tr,s} = 0. \tag{2.11}$$

These are the equations we set out to solve.

3. RETARDED AND ADVANCED SOLUTIONS

For the case $\kappa = 0$, the static homogeneous universe is an ordinary Minkowski universe, and the well-known retarded and advanced solutions of Liénard and Wiechert[3] apply. Guided by analogy with this case, we search in the general case for retarded and advanced solutions of the form

$$F_{\mu\nu}(\mathbf{x}, t) = \int d^3\mathbf{y}\left\{G_{\mu\nu\alpha}(\mathbf{x}, \mathbf{y})J^\alpha\left(\mathbf{y}, t \mp \frac{d_{\mathbf{xy}}}{c}\right) + H_{\mu\nu\rho}(\mathbf{x}, \mathbf{y})J^\rho{}_{,0}\left(\mathbf{y}, t \mp \frac{d_{\mathbf{xy}}}{c}\right)\right\}. \tag{3.1}$$

In the last term $J^0{}_{,0}$ was omitted as it can be expressed in terms of $J^r{}_{,r}$ and $J^r{}_{,r0}$ through the continuity equation. The $d_{\mathbf{xy}}$ is the metric distance between $\mathbf{x}$ and $\mathbf{y}$ in the three-dimensional homogeneous space.[4]

In the Minkowski case ($\kappa = 0$), because of translation invariance, the Green's functions G and H depend only on $\mathbf{y} - \mathbf{x}$. In the other cases this is no longer true. Still, G and H do not depend on $\mathbf{x}$ and $\mathbf{y}$ separately. Our first task is to find just how they depend on $\mathbf{x}$ and $\mathbf{y}$.

In the curved homogeneous spaces, the role of translations is played by the coordinate transformation

$$\mathbf{z}_\mathbf{b} = \frac{(\mathbf{z} - \mathbf{b})(1 + \kappa\mathbf{z}\cdot\mathbf{b}/4a^2) + (\mathbf{z}^2 - \mathbf{z}\cdot\mathbf{b})(\kappa\mathbf{b}/4a^2) + (\mathbf{b}^2 - \mathbf{z}\cdot\mathbf{b})(\kappa\mathbf{z}/4a^2)}{1 + \kappa\mathbf{z}\cdot\mathbf{b}/2a^2 + \kappa^2(\mathbf{z}^2/4a^2)(\mathbf{b}^2/4a^2)}, \tag{3.2}$$

which leaves the form of the metric invariant, but turns the point $\mathbf{b}$ into the origin of space coordinates. [Dot products in Eq. (3.2) are Euclidean: $\mathbf{z}\cdot\mathbf{b} = z_x b_x + z_y b_y + z_z b_z$.] Rather than evaluate $F_{\mu\nu}$ at $\mathbf{x}$, we may translate $\mathbf{x}$ to the origin, compute $F_{\mu\nu}$ at the origin in the new coordinates, and transform back. This yields[5] (the field and current in the new coordinate system are denoted $\tilde{F}, \tilde{J}$)

$$\begin{aligned} F_{\mu\nu}(\mathbf{x}, t) &= (\partial x^\alpha_\mathbf{x}/\partial x^\mu)(\partial x^\beta_\mathbf{x}/\partial x^\nu)\tilde{F}_{\alpha\beta}(\mathbf{0}, t) \\ &= (\partial x^\alpha_\mathbf{x}/\partial x^\mu)(\partial x^\beta_\mathbf{x}/\partial x^\nu) \\ &\quad\times \int d^3\mathbf{y}_\mathbf{x}\left\{G_{\alpha\beta\gamma}(\mathbf{0}, \mathbf{y}_\mathbf{x})\tilde{J}^\gamma\left(\mathbf{y}_\mathbf{x}, t \mp \frac{d_{\mathbf{xy}}}{c}\right) + H_{\alpha\beta\gamma}(\mathbf{0}, \mathbf{y}_\mathbf{x})\tilde{J}^\gamma\left(\mathbf{y}_\mathbf{x}, t \mp \frac{d_{\mathbf{xy}}}{c}\right)\right\} \\ &= (\partial x^\alpha_\mathbf{x}/\partial x^\mu)(\partial x^\beta_\mathbf{x}/\partial x^\nu)\int d^3\mathbf{y}\left|\det\frac{\partial y_\mathbf{x}}{\partial y}\right| \\ &\quad\times\left\{G_{\alpha\beta\rho}(\mathbf{0}, \mathbf{y}_\mathbf{x})(\partial y^\rho_\mathbf{x}/\partial y^\sigma)J^\sigma\left(\mathbf{y}, t \mp \frac{d_{\mathbf{xy}}}{c}\right) + H_{\alpha\beta\rho}(\mathbf{0}, \mathbf{y}_\mathbf{x})(\partial y^\rho_\mathbf{x}/\partial y^\sigma)j^\sigma\left(\mathbf{y}, t \mp \frac{d_{\mathbf{xy}}}{c}\right)\right\}. \end{aligned} \tag{3.3}$$

In the last equation we have used

$$d_{\mathbf{xy}} = d_{0\mathbf{y}_\mathbf{x}}. \tag{3.4}$$

The four-dimensional $x_\mathbf{b}$ in Eq. (3.3) stands for $(\mathbf{x}_\mathbf{b}, t)$ and we have

$$\partial x^0_\mathbf{b}/\partial x^r = \partial x^r_\mathbf{b}/\partial x^0 = \delta_{r0}. \tag{3.5}$$

The expression $\partial x^\alpha_\mathbf{x}/\partial x^\mu$ is defined by

$$\partial x^\alpha_\mathbf{x}/\partial x^\mu \equiv [\partial x^\alpha_\mathbf{b}/\partial x^\mu]_{\mathbf{b}=\mathbf{x}}, \tag{3.6}$$

and it turns out that

$$\partial x^r_\mathbf{x}/\partial x^s = \delta_{rs}/S(|\mathbf{x}|), \tag{3.7}$$

where $|\mathbf{x}|$ is the Euclidean length of $\mathbf{x}$. The determinant of the transformation is

$$\det\frac{\partial x_\mathbf{b}}{\partial \mathbf{x}} = \left|\frac{S(|\mathbf{x}_\mathbf{b}|)}{S(|\mathbf{x}|)}\right|^3. \tag{3.8}$$

Thus, upon comparing Eq. (3.3) with Eq. (3.1) [and substituting Eqs. (3.5), (3.7), and (3.8)], we find

$$G_{0r\alpha}(\mathbf{x}, \mathbf{y}) = \frac{1}{S(|\mathbf{x}|)}\bar{G}_{0r\beta}(\mathbf{y}_\mathbf{x})\frac{\partial y^\beta_\mathbf{x}}{\partial y^\alpha}\frac{1}{[S(|\mathbf{y}|)]^3}, \tag{3.9}$$

$$G_{rs\alpha}(\mathbf{x}, \mathbf{y}) = \frac{1}{[S(|\mathbf{x}|)]^2}\bar{G}_{rs\beta}(\mathbf{y}_\mathbf{x})\frac{\partial y^\beta_\mathbf{x}}{\partial y^\alpha}\frac{1}{[S(|\mathbf{y}|)]^3}, \tag{3.10}$$

[3] A. Lienard, L'eclairage electrique **16**, 5, 53, 106 (1898); E. Wiechert, Archives Neerland (2) **5**, 549 (1900); Ann. Phys. (N.Y.) **4**, 676 (1901). See also Ref. 1, p. 174.

[4] We always take $d_{\mathbf{xy}}$ as the shortest distance between $\mathbf{x}$ and $\mathbf{y}$. In the closed case $\mathbf{x}$ and $\mathbf{y}$ are connected by an infinite number of geodesics that wrap around the spherical space many times (as well as by two that go around less than once). The length of any of these geodesics could be used to construct retarded or advanced fields. We restrict ourselves to the shortest. Solutions based on the others may be recovered from ours by superposition (see Sec. 11).

[5] This transformation affects only space indices. Time components are scalar under this transformation and so is the operator of differentiation with respect to the time.

$$H_{0rp}(\mathbf{x}, \mathbf{y}) = \frac{1}{S(|\mathbf{x}|)} \bar{H}_{0rq}(\mathbf{y}_\mathbf{x}) \frac{\partial y^q_\mathbf{x}}{\partial y^p} \frac{1}{[S(|\mathbf{y}|)]^3}, \tag{3.11}$$

$$H_{rsp}(\mathbf{x}, \mathbf{y}) = \frac{1}{[S(|\mathbf{x}|)]^2} \bar{H}_{rsq}(\mathbf{y}_\mathbf{x}) \frac{\partial y^q_\mathbf{x}}{\partial y^p} \frac{1}{[S(|\mathbf{y}|)]^3}. \tag{3.12}$$

For $\kappa = 0$, these equations reduce to dependence on the difference of $\mathbf{x}$ and $\mathbf{y}$.

4. CHARGED PARTICLE AS THE SOURCE

We now consider J^μ to be the current corresponding to a particle carrying one unit of charge and following a world line given parametrically as

$$y^\alpha = \eta^\alpha(\tau). \tag{4.1}$$

We choose the parameter τ to be the proper time of the particle (in the static homogeneous universe), so that

$$g_{\alpha\beta}\dot{\eta}^\alpha\dot{\eta}^\beta = c^2. \tag{4.2}$$

Dots denote differentiations with respect to τ. The current is

$$J^\mu(\mathbf{y}, t) = \int d\tau\, \dot{\eta}^\mu(\tau)\delta^3(\mathbf{y} - \boldsymbol{\eta}(\tau))\delta(t - \eta^0(\tau))[S(|\boldsymbol{\eta}|)]^3. \tag{4.3}$$

We next substitute this current into the integrals that appear in the expressions for the field $F_{\mu\nu}$. We find

$$F_{0r}(\mathbf{x}, t) = \frac{1}{S(|\mathbf{x}|)} \times \int d\tau \left\{ \bar{G}_{0r\beta}(\boldsymbol{\eta}_\mathbf{x}) \frac{\partial \eta^\beta_\mathbf{x}}{\partial \eta^\alpha} \dot{\eta}^\alpha \delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \bar{H}_{0rq}(\boldsymbol{\eta}_\mathbf{x}) \frac{\partial \eta^q_\mathbf{x}}{\partial \eta^p} \dot{\eta}^p \delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) \right\}, \tag{4.4}$$

$$F_{rs}(\mathbf{x}, t) = \frac{1}{[S(|\mathbf{x}|)]^2} \times \int d\tau \left\{ \bar{G}_{rs\beta}(\boldsymbol{\eta}_\mathbf{x}) \frac{\partial \eta^\beta_\mathbf{x}}{\partial \eta^\alpha} \dot{\eta}^\alpha \delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + H_{rsq}(\boldsymbol{\eta}_\mathbf{x}) \frac{\partial \eta^q_\mathbf{x}}{\partial \eta^p} \dot{\eta}^p \delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) \right\}. \tag{4.5}$$

In the last equations, $\partial\eta^\beta_\mathbf{x}/\partial\eta^\alpha$ stands for $\partial y^\beta_\mathbf{x}/\partial y^\alpha$ taken for $\mathbf{y} = \boldsymbol{\eta}$.

5. SOURCE IN THE VICINITY OF THE ORIGIN

We now assume [for a particular point of space–time $(\mathbf{x}, t)$] that at the time $t \mp d_{\mathbf{x}\eta}/c$ our particle was (or will be) at the origin. That is, we assume that, for that value of τ for which $\eta^0 = t \mp d_{\mathbf{x}\eta}/c$, we have $\boldsymbol{\eta} = \mathbf{0}$. This assumption entails no sacrifice of generality, since it can always be made true by an appropriate choice of the origin of coordinates.

When $\boldsymbol{\eta} = \mathbf{0}$,

$$\boldsymbol{\eta}_\mathbf{x}|_{\eta=0} = -\mathbf{x}. \tag{5.1}$$

This value may now be substituted for $\boldsymbol{\eta}_\mathbf{x}$, except where a derivative of $\boldsymbol{\eta}$ need be taken; in that case we must keep $\boldsymbol{\eta}_\mathbf{x}$ to first order in $\boldsymbol{\eta}$:

$$\boldsymbol{\eta}_\mathbf{x} = \boldsymbol{\eta} S(|\mathbf{x}|) - \mathbf{x} + O(\eta^2). \tag{5.2}$$

When the fields of Eqs. (4.4) and (4.5) are substituted into Maxwell's equations [(2.8) through (2.11)], they are subject to derivatives with respect to $\mathbf{x}$ and t, not τ. However, those terms that contain $J^0{}_{,0}$ should be transformed into terms in $J^r{}_{,r}$ and $J^r{}_{,r0}$, and the transformation involves taking a derivative with respect to τ. Let us now perform this transformation explicitly:

$$\begin{aligned}
&\int d^3\mathbf{y}\, \bar{G}_{\mu\nu0}(\mathbf{y}_\mathbf{x}) J^0{}_{,0}\left(\mathbf{y}, t \mp \frac{d_{\mathbf{x}\eta}}{c}\right) \\
&\quad = \int d\tau\, \bar{G}_{\mu\nu0}(\boldsymbol{\eta} S(|\mathbf{x}|) - \mathbf{x}) \dot{\eta}^0 \delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) \\
&\quad = \int d\tau\, \delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right)\left(\dot{\eta}^0 \pm \frac{1}{c}\frac{\partial d_{\mathbf{x}\eta}}{\partial \boldsymbol{\eta}} \cdot \boldsymbol{\eta}\right) \\
&\qquad \times \bar{G}_{\mu\nu0}(\boldsymbol{\eta} S(|\mathbf{x}|) - \mathbf{x}) \mp \int d\tau \frac{1}{c}\frac{\partial d_{\mathbf{x}\eta}}{\partial \boldsymbol{\eta}} \cdot \boldsymbol{\eta} \\
&\qquad \times \bar{G}_{\mu\nu0}(-\mathbf{x}) \delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right).
\end{aligned} \tag{5.3}$$

We have already put $\boldsymbol{\eta} = \mathbf{0}$ in G in the last term. We may also substitute

$$\partial d_{\mathbf{y}\mathbf{x}}/\partial \mathbf{y}|_{\mathbf{y}=\mathbf{0}} = -\mathbf{x}/|\mathbf{x}| \tag{5.4}$$

in the same term. In the first term an integration by parts is called for, after which $\boldsymbol{\eta} = \mathbf{0}$ may be substituted there too. When all this is done, we find

$$\begin{aligned}
&\int d^3\mathbf{y}\, \bar{G}_{\mu\nu0}(\mathbf{y}_\mathbf{x}) J^0{}_{,0}\left(\mathbf{y}, t - \frac{d_{\mathbf{x}\mathbf{y}}}{c}\right) \\
&= S(|\mathbf{x}|)\frac{\partial}{\partial \mathbf{x}} \cdot \int d\tau\, \bar{G}_{\mu\nu0}(-\mathbf{x}) \delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) \\
&\quad \pm \int d\tau \frac{1}{c}\frac{\mathbf{x}}{|\mathbf{x}|} \cdot \dot{\boldsymbol{\eta}}\, \bar{G}_{\mu\nu0}(-\mathbf{x}) \delta'\left(t - \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right).
\end{aligned} \tag{5.5}$$

We have used the fact that $\partial\eta^\alpha_\mathbf{x}/\partial\eta^0 = \delta_{\alpha0}$ identically. After $\boldsymbol{\eta} = \mathbf{0}$ is substituted, we also find $\partial\eta^r_\mathbf{x}/\partial\eta^s = \delta_{rs}$.

In conclusion to this section we therefore find that, for our special choice of origin, the fields are (at $\mathbf{x}$

and t)

$$F_{0r}(\mathbf{x}, t) = \frac{1}{S(|\mathbf{x}|)}\left\{\bar{G}_{0r\alpha}(-\mathbf{x})\int d\tau\, \dot{\eta}^{\alpha}\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \bar{H}_{0rp}(-\mathbf{x})\int d\tau\, \dot{\eta}^{p}\delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right)\right\}, \tag{5.6}$$

$$F_{rs}(\mathbf{x}, t) = \frac{1}{[S(|\mathbf{x}|)]^2}\left\{\bar{G}_{rs\alpha}(-\mathbf{x})\int d\tau\, \dot{\eta}^{\alpha}\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \bar{H}_{rsp}(-\mathbf{x})\int d\tau\, \dot{\eta}^{p}\delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right)\right\}. \tag{5.7}$$

When these fields are substituted in Maxwell's equations [(2.8) through (2.11)], derivatives with respect to $\mathbf{x}$ may be taken directly in Eqs. (5.6) and (5.7); derivatives with respect to t are taken by putting a prime on the δ function in Eqs. (5.6) and (5.7) with one exception: in terms involving $\dot{\eta}^0\delta(t - d_{\mathbf{x}\cdot\eta}/c - \eta^0)$ we do not put a prime on the δ function to take the time derivative of $F_{\mu\nu}$; rather, we use Eq. (5.5).

6. FURTHER REDUCTION OF $\bar{G}$ AND $\bar{H}$

The tensorial character of the various quantities under space rotation and the antisymmetry of $F_{\mu\nu}$ may be used to reduce the number of unknown functions in Eqs. (5.6) and (5.7). Using these considerations, we recast these equations as follows:

$$F_{0r}(\mathbf{x}, t) = \frac{1}{S(r)}\left\{d\tau\left[A(r)\dot{\eta}^r + C(r)\frac{x^r}{r}\dot{\eta}^0 + M(r)\frac{x^r x^s}{r^2}\dot{\eta}^s\right]\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \int d\tau\left[B(r)\dot{\eta}^r + N(r)\frac{x^r x^s}{r^2}\dot{\eta}^s\right] \times \delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right)\right\}, \tag{6.1}$$

$$F_{rs}(\mathbf{x}, t) = \frac{1}{[S(r)]^2}\int d\tau\left(\frac{x^s}{r}\dot{\eta}^r - \frac{x^r}{r}\dot{\eta}^s\right) \times \left[U(r)\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + V(r)\delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right)\right]. \tag{6.2}$$

In the last equations r stands for $|\mathbf{x}|$. In order to solve our problem we have to find A, B, C, M, N, U, and V as functions of r. In the following we suppress the dependence of the various functions on r and we denote differentiation with respect to r by a prime.

We are now ready to substitute the $F_{\mu\nu}$ into Maxwell's equations. Remember that whenever a time derivative is indicated for the term containing C in Eq. (6.1), the result is to be found in Eq. (5.5). We indicate it symbolically as

$$C\frac{x^r}{r}\dot{\eta}^0\delta'\left(t \mp \frac{d_{\mathbf{x}\cdot\eta}}{c} - \eta^0\right) \to -S\left\{\frac{C}{r}\delta_{rs} + \left(C' - \frac{C}{r}\right)\frac{x^r x^s}{r^2}\right\}\dot{\eta}^s\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) \pm S\frac{C}{c}\frac{x^r x^s}{r^2}\dot{\eta}^s\delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right). \tag{6.3}$$

7. THE EQUATIONS

When Eqs. (6.1) and (6.2) are substituted into Maxwell's equations and the rule (6.3) used, we find the following. The left-hand side of Eq. (2.8) becomes

$$-\left(\frac{F_{0s}}{cS}\right)_{,s} = -\frac{1}{c}\left[\left(\frac{A}{S^2}\right)' + \left(\frac{M}{S^2}\right)' + \frac{2M}{rS^2} \pm \frac{C'}{cS^2}\right] \times \int d\tau\frac{\mathbf{x}}{r}\cdot\dot{\boldsymbol{\eta}}\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \left[S\left(\frac{B}{S^2}\right)' + S\left(\frac{N}{S^2}\right)' + \frac{2N}{rS} - \frac{C/c \pm A \pm M}{cS^2}\right] \times \int d\tau\frac{\mathbf{x}}{r}\cdot\dot{\boldsymbol{\eta}}\delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) \mp \frac{B+N}{cS^2} \times \int d\tau\frac{\mathbf{x}}{r}\cdot\dot{\boldsymbol{\eta}}\delta''\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \left[S\left(\frac{C}{S^2}\right)' + \frac{2C}{rS}\right]\int d\tau\,\eta^0\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right). \tag{7.1}$$

The left-hand side of Eq. (2.9) becomes

$$-\left(\frac{F_{r0}}{cS}\right)_{,0} + (cSF_{rs})_{,s} = c\left\{\left[\left(\frac{U}{S}\right)' + \frac{U}{rS} - \frac{C}{rc^2S}\right]\delta_{rs} - \left[\left(\frac{U}{S}\right)' - \frac{U}{rS} + \frac{C'}{c^2S} - \frac{C}{rc^2S}\right]\frac{x^r x^s}{r^2}\right\} \times \int d\tau\,\dot{\eta}^s\delta\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + c\left\{\left[\left(\frac{V}{S}\right)' + \frac{V}{rS} \mp \frac{U}{cS^2} + \frac{A}{c^2S^2}\right]\delta_{rs} - \left[\left(\frac{V}{S}\right)' - \frac{V}{rS} \mp \frac{U}{cS^2} \mp \frac{C}{c^3S^2} - \frac{M}{c^2S^2}\right]\frac{x^r x^s}{r^2}\right. \times \int d\tau\,\dot{\eta}^s\delta'\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right) + \frac{1}{S^2}\left\{\left[\frac{B}{c} \mp V\right]\delta_{rs} + \left[\frac{N}{c} \pm V\right]\frac{x^r x^s}{r^2}\right\} \times \int d\tau\,\dot{\eta}^s\delta''\left(t \mp \frac{d_{\mathbf{x}\eta}}{c} - \eta^0\right). \tag{7.2}$$

The left-hand side of Eq. (2.10) becomes

$$
\begin{aligned}
&F_{rs,0} + F_{s0,r} - F_{r0,s} \\
&\quad = \left[\left(\frac{A}{S}\right)' - \frac{M}{rS}\right] \\
&\qquad \times \int d\tau \left(\frac{x^s}{r}\dot{\eta}^r - \frac{x^r}{r}\dot{\eta}^s\right)\delta\left(t \mp \frac{d_{x\eta}}{c} - \eta^0\right) \\
&\qquad + \left[\frac{U}{S^2} + \left(\frac{B}{S}\right)' \mp \frac{A}{cS^2} - \frac{N}{rS}\right] \\
&\qquad \times \int d\tau \left(\frac{x^s}{r}\dot{\eta}^r - \frac{x^r}{r}\dot{\eta}^s\right)\delta'\left(t \mp \frac{d_{x\eta}}{c} - \eta^0\right) \\
&\qquad + \frac{1}{S^2}\left[V \mp \frac{B}{c}\right] \\
&\qquad \times \int d\tau \left(\frac{x^s}{r}\dot{\eta}^r - \frac{x^r}{r}\dot{\eta}^s\right)\delta''\left(t \mp \frac{d_{x\eta}}{c} - \eta^0\right). \qquad (7.3)
\end{aligned}
$$

The left-hand side of Eq. (2.11) vanishes identically.

Whatever the motion of the particle, if its position at time $t - d_{x0}/c$ is the origin, it cannot be in position $\mathbf{x}$ at the time t. Thus the current at $(\mathbf{x}, t)$ vanishes and expressions (7.1), (7.2), and (7.3) must all be set equal to zero identically in $\dot{\eta}^\alpha$. This yields

$$\left(\frac{A}{S^2}\right)' + \left(\frac{M}{S^2}\right)' + \frac{2M}{rS^2} \pm \frac{C'}{cS^2} = 0, \qquad (7.4)$$

$$\left(\frac{B}{S^2}\right)' + \left(\frac{N}{S^2}\right)' + \frac{2N}{rS^2} - \frac{C/c \pm A \pm M}{cS^3} = 0, \qquad (7.5)$$

$$B + N = 0, \qquad (7.6)$$

$$\left(\frac{C}{S^2}\right)' + \frac{2C}{rS^2} = 0, \qquad (7.7)$$

$$\left(\frac{U}{S}\right)' + \frac{U}{rS} - \frac{C}{rc^2S} = 0, \qquad (7.8)$$

$$\left(\frac{U}{S}\right)' - \frac{U}{rS} + \frac{C'}{c^2S} - \frac{C}{rc^2S} = 0, \qquad (7.9)$$

$$\left(\frac{V}{S}\right)' + \frac{V}{rS} \mp \frac{U}{cS^2} + \frac{A}{c^2S^2} = 0, \qquad (7.10)$$

$$\left(\frac{V}{S}\right)' - \frac{V}{rS} \mp \frac{U}{cS^2} \mp \frac{C}{c^3S^2} - \frac{M}{c^2S^2} = 0, \qquad (7.11)$$

$$B \mp cV = 0, \qquad (7.12)$$

$$N \pm cV = 0, \qquad (7.13)$$

$$\left(\frac{A}{S}\right)' - \frac{M}{rS} = 0, \qquad (7.14)$$

$$\left(\frac{B}{S}\right)' \mp \frac{A}{cS^2} - \frac{N}{rS} + \frac{U}{S^2} = 0. \qquad (7.15)$$

Equations (7.4)–(7.15) are twelve equations (some differential, some algebraic) in the seven unknown functions A, B, C, M, N, U, V.

8. SOLUTION

It turns out that Eqs. (7.4)–(7.15) are not all independent; yet there are enough of them to determine the solution up to an arbitrary multiplicative constant. The solution so determined is

$$A = \pm cU = \mp \frac{1}{cr^2} S(2 - S) = \mp \frac{1}{cr^2}\left(1 - \kappa^2 \frac{r^4}{16a^4}\right), \qquad (8.1)$$

$$B = -N = \pm cV = -\frac{S}{c^2 r} = -\frac{1}{c^2 r} - \frac{\kappa r}{c^2 4a^2}, \qquad (8.2)$$

$$C = \left(\frac{S}{r}\right)^2 = \frac{1}{r^2} + \kappa \frac{1}{2a^2} + \kappa^2 \frac{r^2}{16a^4}, \qquad (8.3)$$

$$M = \pm \frac{2S}{cr^2} = \pm \frac{2}{cr^2} \pm \kappa \frac{2}{c4a^2}. \qquad (8.4)$$

The arbitrary multiplicative constant has been chosen so that our solution coincides with the ordinary Liénard–Wiechert solution when $\kappa = 0$ and for $\kappa \neq 0$ in the limit $r \ll a$. For a source at the origin (where r is in fact infinitely smaller than a), this ensures the correctness of our solution also at the source.

All that remains to be done is to reconstruct $\bar{G}$ and $\bar{H}$ from the functions in Eqs. (8.1)–(8.4), substitute these $\bar{G}$ and $\bar{H}$ in Eqs. (3.9)–(3.12) to find G and H, and use these G and H in Eq. (3.1).

9. RESULTS

We are now in a position to write down explicitly the retarded and advanced solutions of Maxwell's equations in a Friedmann universe. They are

$$
\begin{aligned}
F_{0r}(\mathbf{x}, t) &= \frac{1}{cS(|\mathbf{x}|)} \int \frac{d^3\mathbf{y}}{[S(|\mathbf{y}|)]^3} \\
&\quad \times \left\{ \frac{[S(|\mathbf{y}_\mathbf{x}|)]^2}{|\mathbf{y}_\mathbf{x}|^3} J^0\left(\mathbf{y}, t \mp \frac{d_{xy}}{c}\right)(-y^r_\mathbf{x}) \right. \\
&\quad - \left[\frac{S(|\mathbf{y}_\mathbf{x}|)(2 - S(|\mathbf{y}_\mathbf{x}|))}{c\,|\mathbf{y}_\mathbf{x}|^2}\delta_{rp} - \frac{2S(|\mathbf{y}_\mathbf{x}|)}{c\,|\mathbf{y}_\mathbf{x}|^4} y^r_\mathbf{x} y^p_\mathbf{x}\right] \\
&\quad \times \frac{\partial y^p_\mathbf{x}}{\partial y^s} J^s\left(\mathbf{y}, t \mp \frac{d_{xy}}{c}\right) - \frac{S(|\mathbf{y}_\mathbf{x}|)}{c^2\,|\mathbf{y}_\mathbf{x}|} \\
&\quad \left. \times [\delta_{rp} - (y^r_\mathbf{x} y^p_\mathbf{x}/y^2_\mathbf{x})] \frac{\partial y^p_\mathbf{x}}{\partial y^s} J^s{}_{,0}\left(\mathbf{y}, t \pm \frac{d_{xy}}{c}\right)\right\}, \qquad (9.1)
\end{aligned}
$$

$$F_{rs}(\mathbf{x}, t) = -\frac{1}{c[S(|\mathbf{x}|)]^2}\int\frac{d^3\mathbf{y}}{[S(|\mathbf{y}|)]^3}$$
$$\times\left\{\frac{S(|\mathbf{y}_\mathbf{x}|)(2-S(|\mathbf{y}_\mathbf{x}|))}{c^2|\mathbf{y}_\mathbf{x}|^3}\left[y^s_\mathbf{x}\frac{\partial y^r_\mathbf{x}}{\partial y^p}-y^r_\mathbf{x}\frac{\partial y^s_\mathbf{x}}{\partial y^p}\right]\right.$$
$$\times J^p\left(\mathbf{y}, t\mp\frac{d_{\mathbf{xy}}}{c}\right)+\frac{S(|\mathbf{y}_\mathbf{x}|)}{c^3|\mathbf{y}_\mathbf{x}|^2}$$
$$\left.\times\left[y^s_\mathbf{x}\frac{\partial y^r_\mathbf{x}}{\partial y^p}-y^r_\mathbf{x}\frac{\partial y^s_\mathbf{x}}{\partial y^p}\right]J^p{}_{,0}\left(\mathbf{y}, t-\frac{d_{\mathbf{xy}}}{c}\right)\right\}. \tag{9.2}$$

The plus or minus signs refer to the retarded and advanced solutions. The function S is defined in Eq. (2.4); the rule to find $\mathbf{y}_\mathbf{x}$ in terms of $\mathbf{y}$ and $\mathbf{x}$ is given in Eq. (3.2).

For the retarded or advanced field of a charged particle following the world line $y^\alpha = \eta^\alpha(\tau)$ (τ being the proper time), use J^μ of Eq. (4.3). If the particle happens to be at the origin at time $t \mp d_{\mathbf{x}\cdot\mathbf{0}}/c$ (the case we used to solve for the unknown functions), things simplify considerably. In that case we have

$$F_{0r}(\mathbf{x}, t) = \frac{S(r)x^r}{r^3}\int d\tau\,\dot\eta^0\delta\left(t\mp\frac{d_{\mathbf{x}\eta}}{c}-\eta^0\right)$$
$$\mp\left[\frac{2-S(r)}{cr^2}\delta_{rs}-\frac{2x^rx^s}{cr^4}\right]$$
$$\times\int d\tau\,\dot\eta^s\delta\left(t\mp\frac{d_{\mathbf{x}\eta}}{c}-\eta^0\right)$$
$$-\frac{1}{c^2r}\left[\delta_{rs}-\frac{x^rx^s}{r^2}\right]$$
$$\times\int d\tau\,\dot\eta^s\delta'\left(t\mp\frac{d_{\mathbf{x}\eta}}{c}-\eta^0\right), \tag{9.3}$$

$$F_{rs}(\mathbf{x}, t) = \frac{2-S(r)}{c^2S(r)r^3}\int d\tau\,(x^s\dot\eta^r-x^r\dot\eta^s)$$
$$\times\delta\left(t-\frac{d_{\mathbf{x}\eta}}{c}-\eta^0\right)\mp\frac{1}{c^3S(r)r^2}$$
$$\times\int d\tau\,(x^s\dot\eta^r-x^r\dot\eta^s)\delta'\left(t-\frac{d_{\mathbf{x}\eta}}{c}-\eta^0\right). \tag{9.4}$$

The dot indicates derivation with respect to τ, and r is $|\mathbf{x}|$. When a derivative of $F_{\mu\nu}$ at $(\mathbf{x}, t)$ is desired, one can apply the space- or time-derivative operator to Eq. (9.3) and (9.4) and then effect the substitution (6.3).

10. COUNTER SOURCES IN A CLOSED UNIVERSE

In this section and the next we concentrate on the spherical (closed) universe ($\kappa = +1$) and study a feature peculiar to this case: charged particles in a spherical universe must exist in pairs, and retarded and advanced fields must appear symmetrically.

Consider the retarded solution in Eqs. (9.3) and (9.4). It is singular at the origin where the charged particle is; it is an empty-space solution for all finite $\mathbf{x}$. However, for $r = \infty$ it again develops a singularity. In the closed spherical case, $r = \infty$ is just another point of the three-dimensional space, the one diametrically opposite to the origin $r = 0$. The fact that the coordinates of this point are infinite reflects a singularity of the system of coordinates, not of the space.

In order to study the neighborhood of $r = \infty$, we introduce a transformation of coordinates that interchanges the roles of $r = 0$ and $r = \infty$ and moves each point to the diametrically opposite point. It is

$$\boldsymbol{\xi} = -\mathbf{x}4a^2/r^2, \tag{10.1}$$

and may be inverted into

$$\mathbf{x} = -\boldsymbol{\xi}4a^2/\rho^2, \tag{10.2}$$

where ρ stands for $|\boldsymbol{\xi}|$. We have

$$\partial x^r/\partial\xi^s = -(4a^2/\rho^2)(\delta_{rs}-2\xi^r\xi^s/\rho^2). \tag{10.3}$$

Also,

$$\mathbf{x}/r = -\boldsymbol{\xi}/\rho \tag{10.4}$$

and

$$S(r) = (4a^2/\rho^2)S(\rho). \tag{10.5}$$

The distance $d_{\mathbf{x}0}$, reexpressed in the new coordinates, becomes

$$d_{\mathbf{x}0} = d_{\xi\infty} = \pi a - d_{\xi 0}. \tag{10.6}$$

All these relations may be used to transform the metric and the retarded electromagnetic field. When this is done, we find that the metric is unchanged—it is the same form in $\boldsymbol{\xi}$ as Eq. (2.3) is in $\mathbf{x}$. As for the electromagnetic field, the coefficients in Eqs. (9.3) and (9.4) become the same forms in $\boldsymbol{\xi}$ as they originally were of $\mathbf{x}$, except that a number of signs are changed so that the *new coefficients are those of the advanced solution* with an over-all minus sign. The argument in the δ functions becomes

$$t-\frac{d_{\xi\eta}}{c}-\eta^0 = t-\pi\frac{a}{c}+\frac{d_{\xi\eta}}{c}-\eta^0. \tag{10.7}$$

We therefore find that the retarded field of particles following the world line $\mathbf{x} = \boldsymbol{\eta}^0(\tau)$ $t = \eta^0(\tau)$ (which passes through the origin at time $t - d_{\mathbf{x}0}/c$) and carrying a positive unit of charge, *is at the same time* the advanced field of a particle carrying a negative unit of charge and following the world line $\boldsymbol{\xi} = \boldsymbol{\eta}(\tau)$, $t = \eta^0(\tau) + \pi a/c$ (which passes through the diametrically opposite point at the time $t + d_{\mathbf{x}\infty}/c$).

Thus the solution we are considering does not correspond to just a single particle but to a pair of particles of opposite charge carrying out related motions. For just one charged particle in the whole spherical space, no solution exists. This could be expected on intuitive grounds, for the "lines of force" starting at a charge cannot in this case go out to infinity and must therefore end in an opposite charge. Thus it turns out that the field of a system of many particles cannot be constructed as a superposition of solutions of each particle alone. One must deal at least in pairs. The fields of Eqs. (9.3) and (9.4) correspond to a pair of particles carrying out related motions. Fields of pairs of particles carrying out arbitrary motions are constructed in the next section.

11. AN ARBITRARY PAIR IN A CLOSED UNIVERSE

In this section we consider a pair of oppositely charged particles moving along arbitrary world lines in the spherical universe and we construct the corresponding electromagnetic field. It should be remembered that the whole expansion and recontraction of the universe takes place over a period of $2\pi a/c$ of our coordinate time t [see Eq. (2.5)]. We therefore only prescribe the world lines of the pair of particles for such a period, say for $0 < t < 2\pi a/c$. But in the static spherical universe these world lines must be continued. Because of the continuity equation, they cannot end (except possibly where they meet). It is the freedom as to the particular continuation that we use to accommodate the prescribed world lines of the pair of particles.

For the purpose of the present section, let us specify the world line of the positive particle as $\mathbf{x} = \mathbf{y}(t)$ and that of the negative particle as $\mathbf{x} = \mathbf{z}(t)$. Both $\mathbf{y}(t)$ and $\mathbf{z}(t)$ are specified for $0 < t < 2\pi a/c$. Let us now continue $\mathbf{y}(t)$ for all times in such a way that

$$\mathbf{y}(t - 2\pi na/c) = \mathbf{z}(t), \tag{11.1}$$

where n is an integer and t again ranges from 0 to $2\pi a/c$. Now consider our retarded solution for the positive particle following the world line $\mathbf{x} = \mathbf{y}(t)$. This solution entails a negative particle on the world line

$$\mathbf{x} = -\mathbf{y}(t - \pi a/c)\frac{4a^2}{[\mathbf{y}(t - \pi a/c)]^2} \tag{11.2}$$

surrounded by its advanced potential. On the field we have so far, superimpose the retarded field of a positive particle following the world line (11.2). The particles on the world line (11.2) now cancel out, but a negative particle surrounded by an advanced field now appears on the world line

$$\mathbf{x} = \mathbf{y}(t - 2\pi a/c). \tag{11.3}$$

Now superimpose the retarded field of a positive particle following the world line (11.3) to cancel out the negative particle and repeat the procedure. After $2n$ steps we find a negative particle on the world line

$$\mathbf{x} = \mathbf{y}(t - 2n\pi a/c). \tag{11.4}$$

But, by Eq. (11.1) (for $0 < t < 2\pi a/c$), this is the world line $\mathbf{x} = \mathbf{z}(t)$ prescribed in advance for the negative particle.

In the field we have constructed, the positive particle is surrounded by its retarded field and the negative particle by its advanced field. A similar construction could lead to the advanced field surrounding the positive particle and the retarded field surrounding the negative particle. If retarded (or advanced) fields are desired for both particles, the following device may be used.[6] Consider an arbitrary world line $\mathbf{x} = \mathbf{u}(t)$. Construct the field of a positive particle following $\mathbf{x} = \mathbf{y}(t)$ surrounded by its retarded field and a negative particle surrounded by its advanced field following $\mathbf{x} = \mathbf{u}(t)$. Likewise construct the solution involving a negative particle surrounded by its retarded field following $\mathbf{x} = \mathbf{z}(t)$ and a positive particle surrounded by its advanced field following $\mathbf{x} = \mathbf{u}(t)$. When these two fields are superimposed, the particles on the world line $\mathbf{x} = \mathbf{u}(t)$ cancel out. We are left with a pair of particles following the originally prescribed world lines, both surrounded by their retarded fields.

The constructions discussed above involve a large amount of arbitrariness. More arbitrariness could have been introduced by further variations on the construction. This arbitrariness reflects the fact that the solution of Maxwell's equations for a given system of charges is determined only up to the addition of an arbitrary empty-space solution.

Note added in proof: In a recent paper by B. Sakita ("Strong Coupling of Multipartial Wave Meson Isotriplet," unpublished University of Wisconsin report), an exact solution of the Chew Low equation, in the strong-coupling limit with an arbitrary crossing matrix, is exhibited. This solution is an entirely even function of the energy ω and is obtained by mapping the problem onto the ω^2 plane.

ACKNOWLEDGMENT

The author is indebted to P. C. Peters for discussions that were particularly fruitful in clarifying the peculiarities of the closed case.

[6] I owe this device to P. C. Peters.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Electro-Optical Effects. II

M. M. CARROLL
University of California, Berkeley, California
AND
R. S. RIVLIN
Lehigh University, Bethlehem, Pennsylvania

(Received 5 July 1967)

The constitutive equations governing the propagation of electromagnetic waves of small amplitude in a centro-symmetric isotropic material to which a static electric field is applied involve, in general, six scalar functions of the static field strength and wave frequency. It is shown that if the material is nondissipative, four of these functions are real, while the remaining two are complex conjugates. Conditions are also derived such that the material shall be nonabsorptive with respect to plane-electromagnetic waves and it is seen that a nonabsorptive material is not necessarily nondissipative, but is necessarily stable. It is shown that for a stable, nondissipative material, there are four real velocities corresponding to any direction of propagation. If it is assumed that these are neither zero nor infinite for any direction of propagation, then for each direction of propagation two of these velocities are positive and two are negative.

1. INTRODUCTION

In previous papers,[1,2] a theory has been developed for the propagation of electromagnetic waves of small amplitude in an isotropic, centro-symmetric material to which a static electric field is applied. The constitutive equations of the theory express the linear dependence of sinusoidally varying magnetic intensity and electric displacement fields on sinusoidally varying electric and magnetic induction fields, and they are expressed in terms of the complex fields. The material properties enter into the constitutive equations through six scalar functions of the magnitude of the applied static electric field and of $\iota\omega$, where $\iota = \sqrt{-1}$ and ω is the angular frequency. In the present paper, we explore certain restrictions which, in certain cases, may be imposed on the six constitutive functions. These constitutive functions are denoted by α_1, β_1, α_3, β_2, α_7, and β_7; they enter into the constitutive equations through Eqs. (2.4) and (2.3).

Firstly, we define a nondissipative material at a given angular frequency and given applied static electric field as one for which the energy dissipation per cycle is zero for arbitrary applied sinusoidal fields. We then show in Secs. 3 and 4 that for such a material four of the constitutive functions (α_1, β_1, α_7, and β_7) are real and the remaining two (α_3 and β_2) are complex conjugates.

We define a nonabsorptive material at the angular frequency ω and static field $\mathfrak{E}$ as one for which electromagnetic waves of this frequency and propagation direction arbitrarily oriented with respect to the static field undergo no absorption. In Sec. 5 we obtain restrictions on the constitutive functions for a nonabsorptive material. We see that certain of these restrictions are implied by those for a nondissipative material. The remaining ones, however, are not, unless we assume also that the nondissipative material in stable, i.e., that a wave of angular frequency ω cannot build up in amplitude as it propagates.

Finally, we show in Sec. 7 that for a nondissipative stable material there are two positive and two negative velocities of propagation, provided that we assume that propagation with infinite or zero velocity is not possible in any direction.

2. THE CONSTITUTIVE EQUATIONS

We consider a static electric field $\mathfrak{E}$ to be applied to a material which is isotropic and possesses a center of symmetry. We now consider that small electric and magnetic induction fields which vary sinusoidally with angular frequency ω are superposed on this static field. Let $\mathbf{E}$ and $\mathbf{B}$ be the complex electric and magnetic induction fields, respectively. Then we may write

$$(\mathbf{E}, \mathbf{B}) = (\mathbf{e}, \mathbf{b})e^{\iota\omega t}, \tag{2.1}$$

where $\mathbf{e}$ and $\mathbf{b}$ are complex vectors independent of time. We shall assume that the corresponding complex electric displacement field $\mathbf{D}$ and the complex magnetic intensity field $\mathbf{H}$ are related to $\mathbf{E}$ and $\mathbf{B}$ by linear relations. They may then be written in the forms

$$(\mathbf{D}, \mathbf{H}) = (\mathbf{d}, \mathbf{h})e^{\iota\omega t}, \tag{2.2}$$

where $\mathbf{d}$ and $\mathbf{h}$ are complex vectors independent of time. It follows from previous papers[1,2] that the constitutive equations relating $\mathbf{d}$ and $\mathbf{h}$ with $\mathbf{e}$ and

where $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Lambda}$ are 3×3 matrices defined in $\mathbf{b}$ are

$$\mathbf{d} = \boldsymbol{\Phi}\cdot\mathbf{e} + \boldsymbol{\Psi}\cdot\mathbf{b}, \quad \mathbf{h} = \boldsymbol{\Omega}\cdot\mathbf{e} + \boldsymbol{\Lambda}\cdot\mathbf{b}, \tag{2.3}$$

a rectangular Cartesian coordinate system, in which $\mathfrak{E} = (\mathfrak{E}_1, \mathfrak{E}_2, \mathfrak{E}_3)$, by

$$\begin{aligned} \Phi_{ij} &= \alpha_1\delta_{ij} + \alpha_7\mathfrak{E}_i\mathfrak{E}_j, \quad \Psi_{ij} = -\alpha_3\varepsilon_{ijk}\mathfrak{E}_k, \\ \Lambda_{ij} &= \beta_1\delta_{ij} + \beta_7\mathfrak{E}_i\mathfrak{E}_j, \quad \Omega_{ij} = -\beta_2\varepsilon_{ijk}\mathfrak{E}_k. \end{aligned} \tag{2.4}$$

[1] R. A. Toupin and R. S. Rivlin, Arch. Ratl. Mech. Anal. 7, 434 (1961).
[2] M. M. Carrol and R. S. Rivlin, J. Math. Phys. 8, 2088 (1967).

In Eqs. (2.4), α_1, β_1, α_3, β_2, α_7, and β_7 are functions of $\iota\omega$ and of $\mathfrak{E}\cdot\mathfrak{E}$.

From (2.3) it is easy to derive relations expressing $\mathbf{e}$ and $\mathbf{h}$ in terms of $\mathbf{d}$ and $\mathbf{b}$. We obtain

$$\mathbf{e} = \mathbf{P}\cdot\mathbf{d} + \mathbf{Q}\cdot\mathbf{b}, \quad \mathbf{h} = \mathbf{R}\cdot\mathbf{d} + \mathbf{S}\cdot\mathbf{b}, \tag{2.5}$$

where $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{S}$ are 3×3 matrices defined in terms of $\mathbf{\Phi}$, $\mathbf{\Psi}$, $\mathbf{\Omega}$, and $\mathbf{\Lambda}$ by

$$\begin{aligned} \mathbf{P} &= \mathbf{\Phi}^{-1}, \quad & \mathbf{Q} &= -\mathbf{\Phi}^{-1}\mathbf{\Psi}, \\ \mathbf{R} &= \mathbf{\Omega}\mathbf{\Phi}^{-1}, \quad & \mathbf{S} &= \mathbf{\Lambda} - \mathbf{\Omega}\mathbf{\Phi}^{-1}\mathbf{\Psi}. \end{aligned} \tag{2.6}$$

It follows from (2.1), (2.2), and (2.5) that

$$\mathbf{E} = \mathbf{P}\cdot\mathbf{D} + \mathbf{Q}\cdot\mathbf{B}, \quad \mathbf{H} = \mathbf{R}\cdot\mathbf{D} + \mathbf{S}\cdot\mathbf{B}. \tag{2.7}$$

3. NONDISSIPATIVE MATERIALS

We now derive the conditions on the matrices $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{S}$ in order that the energy dissipated per cycle shall be zero, for all $\mathbf{B}$ and $\mathbf{D}$ sufficiently small so that the linear constitutive equations shall be valid. Let T be the energy dissipated per cycle. Then

$$T = \int_0^{2\pi/\omega} \left\{ (\mathfrak{E} + \mathbf{E}^+)\cdot\frac{d\mathbf{D}^+}{dt} + \mathbf{H}^+\cdot\frac{d\mathbf{B}^+}{dt} \right\} dt. \tag{3.1}$$

From (2.1) and (2.2), we have

$$\begin{aligned} &(\mathbf{D}^+, \mathbf{B}^+, \mathbf{E}^+, \mathbf{H}^+) \\ &\quad = (\mathbf{d}^+, \mathbf{b}^+, \mathbf{e}^+, \mathbf{h}^+)\cos\omega t - (\mathbf{d}^-, \mathbf{b}^-, \mathbf{e}^-, \mathbf{h}^-)\sin\omega t, \\ &(\mathbf{D}^-, \mathbf{B}^-, \mathbf{E}^-, \mathbf{H}^-) \\ &\quad = (\mathbf{d}^+, \mathbf{b}^+, \mathbf{e}^+, \mathbf{h}^+)\sin\omega t + (\mathbf{d}^-, \mathbf{b}^-, \mathbf{e}^-, \mathbf{h}^-)\cos\omega t. \end{aligned} \tag{3.2}$$

Introducing (3.2) into (3.1) and carrying out the integration, we obtain

$$T = \pi(\mathbf{e}^-\cdot\mathbf{d}^+ - \mathbf{e}^+\cdot\mathbf{d}^- + \mathbf{h}^-\cdot\mathbf{b}^+ - \mathbf{h}^+\cdot\mathbf{b}^-). \tag{3.3}$$

Now, from (2.5), we have

$$\begin{aligned} \mathbf{e}^+ &= \mathbf{P}^+\cdot\mathbf{d}^+ - \mathbf{P}^-\cdot\mathbf{d}^- + \mathbf{Q}^+\cdot\mathbf{b}^+ - \mathbf{Q}^-\cdot\mathbf{b}^-, \\ \mathbf{h}^+ &= \mathbf{R}^+\cdot\mathbf{d}^+ - \mathbf{R}^-\cdot\mathbf{d}^- + \mathbf{S}^+\cdot\mathbf{b}^+ - \mathbf{S}^-\cdot\mathbf{b}^-, \\ \mathbf{e}^- &= \mathbf{P}^-\cdot\mathbf{d}^+ + \mathbf{P}^+\cdot\mathbf{d}^- + \mathbf{Q}^-\cdot\mathbf{b}^+ + \mathbf{Q}^+\cdot\mathbf{b}^-, \\ \mathbf{h}^- &= \mathbf{R}^-\cdot\mathbf{d}^+ + \mathbf{R}^+\cdot\mathbf{d}^- + \mathbf{S}^-\cdot\mathbf{b}^+ + \mathbf{S}^+\cdot\mathbf{b}^-. \end{aligned} \tag{3.4}$$

Introducing (3.4) into (3.3), we obtain

$$\begin{aligned} T = \pi[&(\mathbf{d}^+\cdot\mathbf{P}^-\cdot\mathbf{d}^+ + \mathbf{d}^-\cdot\mathbf{P}^-\cdot\mathbf{d}^-) \\ &+ (\mathbf{b}^+\cdot\mathbf{S}^-\cdot\mathbf{b}^+ + \mathbf{b}^-\cdot\mathbf{S}^-\cdot\mathbf{b}^-) \\ &+ \mathbf{d}^+\cdot(\mathbf{P}^+ - \tilde{\mathbf{P}}^+)\cdot\mathbf{d}^- + \mathbf{b}^+\cdot(\mathbf{S}^+ - \tilde{\mathbf{S}}^+)\cdot\mathbf{b}^- \\ &+ \mathbf{d}^+\cdot(\mathbf{Q}^- + \tilde{\mathbf{R}}^-)\cdot\mathbf{b}^+ + \mathbf{d}^-\cdot(\mathbf{Q}^- + \tilde{\mathbf{R}}^-)\cdot\mathbf{b}^- \\ &+ \mathbf{d}^+\cdot(\mathbf{Q}^+ - \tilde{\mathbf{R}}^+)\cdot\mathbf{b}^- + \mathbf{b}^+\cdot(\mathbf{R}^+ - \tilde{\mathbf{Q}}^+)\cdot\mathbf{d}^-], \end{aligned} \tag{3.5}$$

where $\tilde{\mathbf{P}}^+$, $\tilde{\mathbf{Q}}^-$, $\cdots$, etc., denote the transposes of the matrices $\mathbf{P}^+$, $\mathbf{Q}^-$, $\cdots$, etc.

We shall suppose that $T = 0$ for all $\mathbf{b}$ and $\mathbf{d}$. It follows, by differentiation of the relation (3.5), that

$$\begin{aligned} &\mathbf{P}^- = \mathbf{S}^- = 0, \quad \mathbf{P}^+ = \tilde{\mathbf{P}}^+, \quad \mathbf{S}^+ = \tilde{\mathbf{S}}^+, \\ &\mathbf{Q}^- + \tilde{\mathbf{R}}^- = 0, \qquad \mathbf{Q}^+ - \tilde{\mathbf{R}}^+ = 0. \end{aligned} \tag{3.6}$$

Thus, if the energy dissipated per cycle is zero, we may conclude the following:

(i) $\mathbf{P}$ and $\mathbf{S}$ are real, symmetric matrices;

(ii) $\mathbf{Q}^+$ is the transpose of $\mathbf{R}^+$ and $\mathbf{Q}^-$ is the negative of the transpose of $\mathbf{R}^-$, i.e., $\mathbf{Q}$ is the transjugate matrix of $\mathbf{R}$.

From the relations (2.6) it follows that

$$\begin{aligned} &\mathbf{\Phi} = \mathbf{P}^{-1}, \quad \mathbf{\Psi} = -\mathbf{P}^{-1}\mathbf{Q}, \\ &\mathbf{\Omega} = \mathbf{R}\mathbf{P}^{-1}, \quad \mathbf{\Lambda} = \mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q}. \end{aligned} \tag{3.7}$$

From (3.7) and the results (i) and (ii), we readily obtain the following results:

(a) $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ are real, symmetric matrices;

(b) $\mathbf{\Psi}$ is the transjugate matrix of $-\mathbf{\Omega}$.

Conversely, it is evident that if the conditions (a) and (b) are satisfied, so are the conditions (i) and (ii); in either case the energy dissipated in a cycle of angular frequency ω is zero. Thus, (i) and (ii), or (a) and (b), are necessary and sufficient conditions that the energy dissipated in a cycle of angular frequency ω be zero. A material which satisfies these conditions for a given angular frequency ω and static field $\mathfrak{E}$ shall be called *nondissipative at angular frequency ω and electric field $\mathfrak{E}$*.

4. RESTRICTIONS ON CONSTITUTIVE FUNCTIONS FOR NONDISSIPATIVE MATERIALS

We shall now discuss the restrictions on the constitutive functions α_1, β_1, α_3, β_2, α_7, and β_7 which result from the assumption that the material considered is nondissipative at a specified angular frequency ω and a field $\mathfrak{E}$ of given magnitude but arbitrary direction.

The expressions for Φ_{ij}, Ψ_{ij}, Ω_{ij}, and Λ_{ij} in terms of the α's and β's are given by (2.4). We note that the α's and β's are independent of the direction of $\mathfrak{E}$. It therefore follows from the first and third parts of (2.4) that if the condition (a) in Sec. 3 is satisfied for $\mathfrak{E}$ of given magnitude and arbitrary direction, then α_1, α_7, β_1, and β_7 are all real; i.e.,

$$\alpha_1^- = \alpha_7^- = \beta_1^- = \beta_7^- = 0. \tag{4.1}$$

Again, if the condition (b) in Sec. 3 is satisfied, it follows from the second and fourth parts of (2.4) that

$$\alpha_3 = \beta_2^*, \tag{4.2}$$

where the star is used to denote the complex conjugate. From (4.2) it follows that

$$\alpha_3^+ = \beta_2^+ \quad \text{and} \quad \alpha_3^- + \beta_2^- = 0. \tag{4.3}$$

It is evident that if the relations (4.1) and (4.3) are satisfied, then the conditions (a) and (b) of Sec. 3 are satisfied for all $\mathfrak{E}$ of the specified magnitude.

5. NONABSORPTIVE MATERIALS

It has been shown[2] that if a plane-electromagnetic wave of angular frequency ω and wavenumber k is

propagated in the direction of the unit vector $\mathbf{n}$ in the presence of a static electric field $\mathfrak{E}$, in a material to which the constitutive equations (2.3) apply, there are four possible complex velocities for the wave. Denoting the inverses of these (the complex slownesses) by η_1, η_2, η_3, η_4, we have

$$\eta_1, \eta_2 = \frac{1}{2(\beta_1 + \beta_7\mathfrak{E}_1^2)} \times [-(\alpha_3 + \beta_2)\mathfrak{E}_3 \pm \{(\alpha_3 + \beta_2)^2\mathfrak{E}_3^2 + 4\alpha_1(\beta_1 + \beta_7\mathfrak{E}_1^2)\}^{\frac{1}{2}}], \tag{5.1}$$

$$\eta_3, \eta_4 = \frac{1}{2A}[-(\alpha_3 + \beta_2)\mathfrak{E}_3 \pm \{(\alpha_3 + \beta_2)^2\mathfrak{E}_3^2 + 4\alpha_1 A\}^{\frac{1}{2}}], \tag{5.2}$$

where

$$A = \frac{\alpha_1\beta_1 + \alpha_7\beta_1\mathfrak{E}_3^2 + \alpha_3\beta_2\mathfrak{E}_1^2}{\alpha_1 + \alpha_7(\mathfrak{E}_1^2 + \mathfrak{E}_3^2)}, \tag{5.3}$$

and the rectangular Cartesian reference system x is chosen, as it may be without loss of generality, so that $\mathbf{n}$ is the direction of the x_3 axis and $\mathfrak{E}$ lies in the x_1x_3 plane, i.e., $\mathfrak{E} = (\mathfrak{E}_1, 0, \mathfrak{E}_3)$.

We consider here the restrictions which are placed on the six complex constitutive functions α_1, β_1, α_3, β_2, α_7, and β_7, if it is assumed that none of the waves suffers absorption for any direction of the applied static field. This implies that η_1, η_2, η_3, and η_4 given by (5.1) and (5.2) must be real for all values of $\mathfrak{E}_1$ and $\mathfrak{E}_3$ such that $\mathfrak{E}_1^2 + \mathfrak{E}_3^2 = \mathfrak{E}^2$(const). The necessary and sufficient conditions for this to be the case are:

(i) $(\alpha_3 + \beta_2)/(\beta_1 + \beta_7\mathfrak{E}_1^2)$ must be real,

(ii) $(\alpha_3 + \beta_2)/A$ must be real,

(iii) $\alpha_1/(\beta_1 + \beta_7\mathfrak{E}_1^2)$ must be real,

(iv) α_1/A must be real,

(v) $\dfrac{(\alpha_3 + \beta_2)^2\mathfrak{E}_3^2}{4(\beta_1 + \beta_7\mathfrak{E}_1^2)^2} + \dfrac{\alpha_1}{\beta_1 + \beta_7\mathfrak{E}_1^2}$ must be nonnegative,

(vi) $\dfrac{(\alpha_3 + \beta_2)^2\mathfrak{E}_3^2}{4A^2} + \dfrac{\alpha_1}{A}$ must be nonnegative.

A is given by Eq. (5.3).

The above six conditions must, of course, be satisfied for all values of $\mathfrak{E}_1$ and $\mathfrak{E}_3$ such that $\mathfrak{E}_1^2 + \mathfrak{E}_3^2 = \mathfrak{E}^2$, where $\mathfrak{E}$ is the magnitude of the applied static field.

Taking $\mathfrak{E}_1 = 0$ in condition (i), we have

$$[(\alpha_3 + \beta_2)/\beta_1]^- = 0, \tag{5.4}$$

and hence from (i),

$$[\beta_7/\beta_1]^- = 0. \tag{5.5}$$

Condition (iii) then yields

$$[\beta_1/\alpha_1]^- = [\beta_7/\alpha_1]^- = 0. \tag{5.6}$$

Conditions (ii) and (iv) may be expressed as

$$[(\alpha_3 + \beta_2)/A]^- = [\alpha_1/A]^- = 0. \tag{5.7}$$

It is immediately evident that the results (5.4)–(5.7) are equivalent to

$$[(\alpha_3 + \beta_2)/\alpha_1]^- = [\beta_1/\alpha_1]^- = [A/\alpha_1]^- = [\beta_7/\alpha_1]^- = 0, \tag{5.8}$$

and that if (5.8) is satisfied, the conditions (i)–(iv) are satisfied. Thus, (5.8) are necessary and sufficient conditions that (i)–(iv) be satisfied. We shall examine further the implications of the relation $(A/\alpha_1)^- = 0$.

From (5.3) we have

$$\frac{A}{\alpha_1} = \frac{\beta_1}{\alpha_1} + \frac{(\alpha_3\beta_2 - \alpha_7\beta_1)\mathfrak{E}_1^2}{\alpha_1(\alpha_1 + \alpha_7\mathfrak{E}^2)}. \tag{5.9}$$

Thus the conditions (5.8) may be replaced by

$$\left(\frac{\alpha_3 + \beta_2}{\alpha_1}\right)^- = \left(\frac{\beta_1}{\alpha_1}\right)^- = \left[\frac{\alpha_3\beta_2 - \alpha_7\beta_1}{\alpha_1(\alpha_1 + \alpha_7\mathfrak{E}^2)}\right]^- = \left(\frac{\beta_7}{\alpha_1}\right)^- = 0. \tag{5.10}$$

It is evident that the relations (5.10) imply the conditions (i)–(iv) of this section. Thus, (5.10) together with conditions (v) and (vi) are necessary and sufficient conditions that the material be nonabsorptive at angular frequency ω and static electric field strength $\mathfrak{E}$.

These conditions do not imply the conditions (4.1) and (4.3) that the material be nondissipative at angular frequency ω and static electric field strength $\mathfrak{E}$; i.e., a nonabsorptive material is not necessarily nondissipative. It is easily seen that the relations (4.1) and (4.3) imply the relations (5.10) but not necessarily the conditions (v) and (vi).

We shall now investigate the implications of the conditions (v) and (vi) for a nonabsorptive material. Since these relations must be valid for all directions of the static electric field, taking $\mathfrak{E}_3 = 0$, $\mathfrak{E}_1 = \mathfrak{E}$ and using (5.3), we obtain,

$$\frac{\alpha_1}{\beta_1 + \beta_7\mathfrak{E}^2} \geq 0 \quad \text{and} \quad \frac{\alpha_1(\alpha_1 + \alpha_7\mathfrak{E}^2)}{\alpha_1\beta_1 + \alpha_3\beta_2\mathfrak{E}^2} \geq 0. \tag{5.11}$$

Again, taking $\mathfrak{E}_1 = 0$, $\mathfrak{E}_3 = \mathfrak{E}$ and again using (5.3), we obtain

$$\frac{1}{\beta_1^2}[(\alpha_3 + \beta_2)^2\mathfrak{E}^2 + 4\alpha_1\beta_1] \geq 0. \tag{5.12}$$

6. STABLE MATERIALS

We consider a plane-sinusoidal wave of angular frequency ω and complex slowness η, travelling parallel to the x_3 axis. The complex electromagnetic field vectors $\mathbf{E}$, $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{H}$ at time t are given by

$$(\mathbf{E}, \mathbf{B}, \mathbf{D}, \mathbf{H}) = (\mathbf{e}, \mathbf{b}, \mathbf{d}, \mathbf{h})e^{\iota\omega(\eta x_3 - t)} = (\mathbf{e}, \mathbf{b}, \mathbf{d}, \mathbf{h})e^{-\omega\eta^- x_3}e^{\iota\omega(\eta^+ x_3 - t)} \tag{6.1}$$

We note that the wave propagates in the positive or

negative direction of the x_3 axis accordingly as η^+ is positive or negative. In the former case ($\eta^+ > 0$) the wave increases in amplitude as it travels along the axis if $\eta^- < 0$ and decreases in amplitude if $\eta^- > 0$. Again, if $\eta^+ < 0$, the wave decreases in amplitude along its direction of propagation if $\eta^- < 0$ and increases if $\eta^- > 0$. Thus, the condition that the wave does not increase in amplitude along its direction of propagation is

$$\eta^+\eta^- \geq 0. \tag{6.2}$$

We describe a material as *stable* if all plane waves which can propagate in it in any direction do so with a complex slowness which satisfies the relation (6.2). We note that the nonabsorptive material discussed in Sec. 5 is automatically stable.

We now consider the conditions imposed on the constitutive functions α_1, β_1, α_3, β_2, α_7, and β_7, for a given field strength, if we add to the assumption that the material is nondissipative the assumption that it is stable.

From Sec. 4 we note that if the material is nondissipative, α_1, β_1, $\alpha_3 + \beta_2$, α_7, β_7, and A are all real. Thus, the complex slownesses η_1, η_2 given by (5.1) will be real unless the condition (v) of Sec. 5 is violated. If it is, η_1^- and η_2^- will have opposite signs while $\eta_1^+ = \eta_2^+$. Consequently, either η_1 or η_2 will violate the condition (6.2) and the material will be unstable. Similarly, it follows from (5.2) that, unless the condition (vi) of Sec. 5 is satisfied, the material will be unstable. We thus obtain the result that a nondissipative material will be stable if and only if the conditions (v) and (vi) of Sec. 5 are satisfied for all directions of the static electric field $\mathcal{E}$.

For a stable nondissipative material it follows that the values of η_1, η_2, η_3, and η_4, corresponding to any direction of propagation, are all real.

7. THE SIGNS OF THE SLOWNESSES

In this section we shall consider the material to be nondissipative and stable and shall add the further assumption that all four slownesses of the wave (which we have shown to be real in Sec. 6) are neither zero nor infinite for any direction of propagation relative to the applied static field. We will show that for any direction of propagation two of the slownesses are positive and two are negative.

In order to do this we write the secular equation for the slowness η in a fixed rectangular Cartesian system $\bar{x}$ in which the applied static field has components $\bar{\mathcal{E}}_i$ and the direction of propagation is that of the unit vector $\mathbf{n}(= \bar{n}_i)$.

In the coordinate system x, the slownesses η are given by the quadratic equations

$$(\beta_1 + \beta_7\mathcal{E}^2 - \beta_7\mathcal{E}_3^2)\eta^2 + (\alpha_3 + \beta_2)\mathcal{E}_3\eta - \alpha_1 = 0 \tag{7.1}$$

and

$$[(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2) + (\alpha_7\beta_1 - \alpha_3\beta_2)\mathcal{E}_3^2]\eta^2 + (\alpha_1 + \alpha_7\mathcal{E}^2)[(\alpha_3 + \beta_2)\mathcal{E}_3\eta - \alpha_1] = 0.$$

Now, suppose the coordinate systems $\bar{x}$ and x are related by

$$x_i = a_{ij}\bar{x}_j. \tag{7.2}$$

Since in the systems x and $\bar{x}$ the unit vector $\mathbf{n}$ has components δ_{i3} and $\bar{n}_i$, respectively, we obtain

$$a_{3i} = \bar{n}_i. \tag{7.3}$$

Also, since the components of the static field $\mathcal{E}$ in the systems x and $\bar{x}$ are $(\mathcal{E}_1, 0, \mathcal{E}_3)$ and $\bar{\mathcal{E}}_i$, respectively, we have

$$\mathcal{E}_3 = a_{3j}\mathcal{E}_j = \bar{n}_j\bar{\mathcal{E}}_j. \tag{7.4}$$

Substituting from (7.4) in (7.1), we have

$$(\beta_1 + \beta_7\mathcal{E}^2 - \beta_7\bar{n}_j\bar{n}_k\bar{\mathcal{E}}_j\bar{\mathcal{E}}_k)\eta^2 + (\alpha_3 + \beta_2)\bar{n}_j\mathcal{E}_j\eta - \alpha_1 = 0 \tag{7.5}$$

and

$$[(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2) + (\alpha_7\beta_1 - \alpha_3\beta_2)\bar{n}_j\bar{n}_k\bar{\mathcal{E}}_j\bar{\mathcal{E}}_k]\eta^2 + (\alpha_1 + \alpha_7\mathcal{E}^2)[(\alpha_3 + \beta_2)\bar{n}_j\bar{\mathcal{E}}_j\eta - \alpha_1] = 0.$$

We shall introduce the notation

$$\bar{\eta}_i = \eta\bar{n}_i. \tag{7.6}$$

$\bar{\eta}_i$ are the components in the system $\bar{x}$ of the vector slowness of a wave propagating in the direction $\mathbf{n}$ with slowness η. From (7.6) and (7.5) we obtain

$$[(\beta_1 + \beta_7\mathcal{E}^2)\delta_{jk} - \beta_7\bar{\mathcal{E}}_j\bar{\mathcal{E}}_k]\bar{\eta}_j\bar{\eta}_k + (\alpha_3 + \beta_2)\bar{\mathcal{E}}_j\bar{\eta}_j - \alpha_1 = 0 \tag{7.7}$$

and

$$[(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2)\delta_{jk} + (\alpha_7\beta_1 - \alpha_3\beta_2)\mathcal{E}_j\bar{\mathcal{E}}_k]\bar{\eta}_j\bar{\eta}_k + (\alpha_1 + \alpha_7\mathcal{E}^2)[(\alpha_3 + \beta_2)\bar{\mathcal{E}}_j\bar{\eta}_j - \alpha_1] = 0.$$

Each of the equations (7.7) represents a quadric in the $\bar{\eta}_1$, $\bar{\eta}_2$, $\bar{\eta}_3$ space. The four slownesses corresponding to any direction of propagation are given by the intersections of a line drawn in the direction of propagation with the quadrics. Since for any direction of propagation there are four (not necessarily distinct) real values for the slowness, none of which is zero or infinite, each of the quadrics is an ellipsoid which contains the origin as an internal point. Thus, for any direction of propagation, each of the equations (7.1) gives two values for the slowness, one positive and one negative.

ACKNOWLEDGMENT

As far as one of the authors (R. S. R) was concerned, this work was supported by a grant from the National Science Foundation to Brown University.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Modified Admittance, Perpendicular Susceptibilities, and Transformation of Correlations of the Ising Model

SHOON KYUNG KIM*
Department of Chemistry, University of Louisville, Louisville, Kentucky

(Received 16 January 1968)

A modified admittance is introduced to give Kubo's admittance at nonzero frequencies and to give the isothermal static admittance at zero frequency within the scope of the Kubo linear-response theory. The method is demonstrated by exact calculations of the frequency-dependent perpendicular susceptibility at zero field and its modified susceptibility of the regular Ising model. The results appear as linear combinations of the equilibrium spin–spin correlation functions of the lattice. The results are valid for all dimensions and all frequencies and temperatures. A $(q+1)\times(q+1)$ matrix $\mathbf{a}^{(q)}$ describes the linear combinations explicitly, where q is the coordination number of the lattice. The properties of this matrix are extensively discussed as a special case of a matrix $\mathbf{A}^{(q)}(\xi)$, which satisfies a simple quadratic equation of the form $[\mathbf{A}^{(q)}(\xi)]^2=(1+\xi)^q$ for arbitrary values of ξ. Fisher's algebraic transformation of the spin–spin correlation functions for the regular Ising lattice is derived from the linear relation which holds between the perpendicular susceptibility and the corresponding modified susceptibility. By means of the product rules of the matrix $\mathbf{A}^{(q)}(\xi)$, the higher-order spin–spin correlation functions are expressed in terms of the lower-order ones in complete generality.

1. INTRODUCTION

Since Onsager has published his celebrated derivation of the partition function of the two-dimensional Ising model,[1] there have been numerous works on the model.[2] These are mainly concerned with its equilibrium properties. Recently Allan and Betts[3] performed exact calculations of the frequency-dependent initial perpendicular susceptibilities of the honeycomb and square Ising lattices using the Kubo linear-response theory.[4]

They have observed that the zero-frequency limit of the susceptibility for the case of the square lattice is different from the corresponding isothermal static susceptibility calculated by Fisher[5] in terms of the equilibrium theory. The difference is due to the basic assumption of the Kubo theory that the external force is turned in adiabatically into the system, which is represented initially by the canonical ensemble. Actually, Kubo's statement in regard to this point is more specific.[4] Let $\chi_{BA}(\omega)$ be the admittance with respect to two physical quantities A and B, expressed as a function of the frequency ω of the external force. Then, the zero-frequency limit of the admittance $\chi_{BA}(0)$ gives the adiabatic static admittance, which is, in general, different from the isothermal static admittance χ^T_{BA}. Within the scope of the dynamical theory, however, one must be able to calculate the isothermal static admittance if one manages to recover the difference $\chi^T_{BA}-\chi_{BA}(0)$ caused by the adiabatic condition by means of a suitable mathematical transformation.

One of the purposes of the present communication is to introduce a modified admittance $\chi^T_{BA}(\omega)$ which gives the Kubo admittance $\chi_{BA}(\omega)$ at nonzero frequencies, but gives the isothermal static admittance χ^T_{BA} at zero frequency. That is, we require the modified admittance to satisfy the following equation:

$$\chi^T_{BA}(\omega)=\begin{cases}\chi_{BA}(\omega), & \text{when } \omega\neq 0,\\ \chi^T_{BA}, & \text{when } \omega=0.\end{cases} \tag{1.1}$$

The method is based on the fluctuation–dissipation theorem. First we observe that the difference between two admittances $\chi^T_{BA}-\chi_{BA}(0)$ is linearly related to the zero-frequency Fourier component of the Kubo correlation function[4] [see Eq. (2.9)]. This zero-frequency component is absent in the response function, of which the half-interval Fourier transform gives the Kubo admittance. This absence seems rather essential in the Kubo theory since the response function should vanish (in the Abel limit) when the response time interval becomes infinite. Now, if we consider the Fourier transforms of the Kubo correlation and response functions, there exists a linear relation between them which has a highly singular term arising from the zero-frequency component of the Kubo correlation function. By means of the

* This work was supported in part by the National Aeronautics and Space Administration through Sustaining University Program Grant NGR18-002-005 with the University of Louisville.

[1] L. Onsager, Phys. Rev. **65**, 117 (1944).

[2] For reviews see L. P. Kadanoff *et al.*, Rev. Mod. Phys. **39**, 395 (1967); M. E. Fisher and R. J. Burford, Phys. Rev. **156**, 583 (1967); E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. **4**, 308 (1963).

[3] G. A. T. Allan and D. D. Betts, *The Frequency Dependent Initial Perpendicular Susceptibility of the Ising Model* (Department of Physics, University of Alberta, Edmonton, Canada, June 14, 1967); Can. J. Phys. **46**, 15 (1968).

[4] R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957). See also R. Kubo, *Lectures in Theoretical Physics* (Interscience Publ., Inc., New York, 1959), Vol. 1, p. 120.

[5] M. E. Fisher, J. Math. Phys **4**, 124 (1963).

Hilbert transform[6] of the linear relation, we shall show that this singular term gives rise to the required difference. We may note here that, in the Kubo treatment of the fluctuation–dissipation theorem, the singular term is hidden, since Kubo considered quantities of which the invariant parts are subtracted out from the start.

We demonstrate the present method by the exact calculations of the perpendicular magnetic susceptibility $\chi_{xx}(\omega)$ and the corresponding modified susceptibility $\chi^T_{xx}(\omega)$ for the regular Ising model. The results appear as linear combinations of the equilibrium spin–spin correlation functions and are valid for all dimensions. A $(q+1)\times(q+1)$ matrix $\mathsf{a}^{(q)}$ describes the linear combinations explicitly. We shall give a detailed discussion of the matrix $\mathsf{a}^{(q)}$ as a special case of a matrix $\mathsf{A}^{(q)}(\xi)$, where ξ is an arbitrary parameter. This matrix satisfies a simple quadratic equation of the form $(\mathsf{A}^{(q)}(\xi))^2 = (1+\xi)^q$. It is essential to understand the properties of this matrix in the discussion of Fisher's algebraic transformation[7] of the spin–spin correlation functions for the regular Ising lattice. We show, in fact, that Fisher's transformation is a natural consequence of the linear relation which exists between $\chi_{xx}(\omega)$ and $\chi^T_{xx}(\omega)$. By means of the product rule for the matrix $\mathsf{A}^{(q)}(\xi)$, which is derived in the Appendix, we express the higher-order spin–spin correlation functions in terms of the lower-order ones in complete generality.

The isothermal perpendicular static susceptibility χ^T_{xx} obtained as the modified susceptibility $\chi^T_{xx}(\omega)$ at $\omega = 0$ proves to be in complete agreement with that of Fisher's static calculation, although the present results are more explicit than Fisher's expressions.[5] The present results for the Kubo susceptibilities $\chi_{xx}(\omega)$, when specialized for the honeycomb and the square Ising lattices, are also in agreement with the results of Allan and Betts which we have mentioned in the beginning. We have observed that the zero-frequency limit of the Kubo susceptibility $\chi_{xx}(0)$ gives the corresponding χ^T_{xx} only when q is odd.

We shall show also that, in the high-frequency limit, the susceptibility is simply proportional to E/ω^2, where E is the lattice energy. The limiting properties with respect to temperatures seem to reveal further physical insights into the present problem.

2. MODIFIED ADMITTANCE

To begin with, we define two kinds of correlation functions in time with respect to two physical quantities A and B by

$$\phi^{\pm}_{BA}(t) = \mathrm{Tr}\,\rho(AB(t) \pm B(t)A),$$
$$\rho = \exp(-\beta\mathcal{H}_0)/\mathrm{Tr}\exp(-\beta\mathcal{H}_0), \quad (2.1)$$

where ρ is the canonical density matrix defined by the Hamiltonian $\mathcal{H}_0$ which describes the natural motion of the system, $B(t)$ is the Heisenberg operator defined by $\mathcal{H}_0$, and $\beta = 1/(kT)$. Kubo's correlation and response functions[4] are given by $\phi^{+}_{BA}(t)/2$ and $\phi^{-}_{BA}(t)/(i\hbar)$, respectively. The present notations are chosen differently from Kubo's in order to ensure the symmetry in the present work.

We assume the existence of the Fourier transforms of the two correlation functions

$$\Phi^{\pm}_{BA}(\omega) = \frac{1}{2\hbar}\int_{-\infty}^{+\infty}\phi^{\pm}_{BA}(t)e^{-i\omega t}\,dt, \quad (2.2)$$

where the limits of the integrals are taken to be the Abel limits. Then we have

$$\Phi^{-}_{BA}(\omega) = \tanh(\beta\hbar\omega/2)\Phi^{+}_{BA}(\omega) - \pi\beta\,\mathrm{Tr}\,\rho A^0B^0\omega\delta(\omega), \quad (2.3)$$

where $\delta(\omega)$ is the Dirac delta function and A^0 and B^0 are the diagonal parts of A and B with respect to the Hamiltonian $\mathcal{H}_0$. We can prove Eq. (2.3) using the analytic property of the function $\mathrm{Tr}\,\rho AB(t)$ in the domain $0 \le \mathrm{Im}\,t \le \hbar\beta$ or writing $\phi^{\pm}_{BA}(t)$ in the matrix forms in the representation which diagonalizes the Hamiltonian $\mathcal{H}_0$ and using the property of the delta function. The second term in the right-hand side of Eq. (2.3) arises since $\phi^{+}_{BA}(t)$ has a zero-frequency Fourier component equal to $2\,\mathrm{Tr}\,\rho A^0B^0$, whereas $\phi^{-}_{BA}(t)$ does not have it for $\mathrm{Tr}\,\rho A^0B^0 = \mathrm{Tr}\,\rho B^0A^0$ if the trace exists. The second term of Eq. (2.3) is highly singular, but its Hilbert transform exists [see Eq. (2.7)].

According to the Kubo linear-response theory and the fluctuation–dissipation theorem,[4] the admittance $\chi_{BA}(\omega)$ with respect to two physical quantities A and B satisfies the following relations:

$$\chi_{BA}(\omega) = \frac{1}{i\hbar}\int_0^{\infty}\phi^{-}_{BA}(t)e^{-i\omega t}\,dt$$
$$= -i\Phi^{-}_{BA}(\omega) + \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\Phi^{-}_{BA}(\omega')}{(\omega'-\omega)}\,d\omega'. \quad (2.4)$$

If we substitute Eq. (2.3) into this, we may arrive at a function defined by

$$\chi^T_{BA}(\omega) = -i\Phi^{-}_{BA}(\omega) + \frac{1}{\pi}\int_{-\infty}^{+\infty}\Phi^{+}_{BA}(\omega')\frac{\tanh(\beta\hbar\omega'/2)}{\omega'-\omega}\,d\omega', \quad (2.5)$$

which satisfies

$$\chi^T_{BA}(\omega) = \chi_{BA}(\omega) + \beta\,\mathrm{Tr}\,\rho A^0B^0u(\omega), \quad (2.6)$$

[6] *Tables of Integral Transforms*, A. Erdelyi, Ed. (McGraw-Hill Book Company, New York, 1954), Vol. 2, p. 243.
[7] M. E. Fisher, Phys. Rev. **113**, 969 (1959).

where $u(\omega)$ is defined by

$$u(\omega) = \int_{-\infty}^{+\infty} \frac{\omega'\delta(\omega')}{\omega' - \omega}\, d\omega' = \lim_{\epsilon \to 0} \frac{\epsilon^2}{\epsilon^2 + \omega^2}$$
$$= \begin{cases} 0 & \text{for} \quad \omega \neq 0, \\ 1 & \text{for} \quad \omega = 0. \end{cases} \tag{2.7}$$

Thus, the function $\chi_{BA}^T(\omega)$, which we call the modified admittance, has a discontinuity at $\omega = 0$ and is equal to the Kubo admittance $\chi_{BA}(\omega)$ when $\omega \neq 0$. Since Tr $\rho A^0 B^0 u(\omega)$ is real, we have for the imaginary and real parts of Eq. (2.6)

$$\text{Im}\, \chi_{BA}^T(\omega) = \text{Im}\, \chi_{BA}(\omega),$$
$$\text{Re}\, \chi_{BA}^T(\omega) = \text{Re}\, \chi_{BA}(\omega) + \beta\, \text{Tr}\, \rho A^0 B^0 u(\omega). \tag{2.8}$$

In the following we shall show that the modified admittance actually gives the isothermal static admittance χ_{BA}^T when $\omega = 0$.

The Kubo result for the difference between the isothermal static admittance χ_{BA}^T and the Kubo adiabatic admittance $\chi_{BA}(0)$ is given by[4]

$$\chi_{BA}^T - \chi_{BA}(0) = \beta\, \text{Tr}\, \rho A^0 B^0 - \beta \bar{A}\bar{B}, \tag{2.9}$$

where $\bar{A}$ and $\bar{B}$ are the equilibrium expectations of A and B. From Eqs. (2.6) and (2.9) we have immediately

$$\chi_{BA}^T(0) = \chi_{BA}^T, \tag{2.10}$$

provided that

$$\bar{A}\bar{B} = \text{Tr}\, \rho A \times \text{Tr}\, \rho B = 0. \tag{2.11}$$

This condition is always achieved if we replace A by $A - \bar{A}$ or B by $B - \bar{B}$, or both. This replacement does not alter all the equations derived above.

For later use we shall write down the Kubo admittance and the modified admittance for the case where $A = B$. In this case the Fourier transforms $\Phi_{AA}^+(\omega)$ and $\Phi_{AA}^-(\omega)$ are real, so that from Eqs. (2.4) and (2.8) we have for the imaginary parts of the admittances

$$\text{Im}\, \chi_{AA}^T(\omega) = \text{Im}\, \chi_{AA}(\omega) = -\Phi_{AA}^-(\omega) \tag{2.12}$$

and for the real parts

$$\text{Re}\, \chi_{AA}^T(\omega) = \text{Re}\, \chi_{AA}(\omega) + \beta\, \text{Tr}\, \rho A^0 A^0 u(\omega), \tag{2.13}$$

where

$$\text{Re}\, \chi_{AA}^T(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \Phi_{AA}^+(\omega') \frac{\tanh(\beta\hbar\omega'/2)}{\omega' - \omega}\, d\omega', \tag{2.14}$$

$$\text{Re}\, \chi_{AA}(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\Phi_{AA}^-(\omega')}{(\omega' - \omega)}\, d\omega'. \tag{2.15}$$

The last equation is simply the Kramers–Kronig relation.[4]

3. PERPENDICULAR SUSCEPTIBILITY OF THE ISING MODEL

We apply the method developed in the preceding section and calculate the zero-field perpendicular susceptibility and the corresponding modified susceptibility of the Ising model. We take the z axis as the axis of anisotropy of the Ising model and apply a periodic magnetic field in the x direction. Then the total Hamiltonian of a lattice of N spins may be written as

$$\mathcal{H} = \mathcal{H}_0 - M_x H_x \cos(\omega t). \tag{3.1}$$

In terms of the Pauli spin matrices σ_j^x, σ_j^y, and σ_j^z, the unperturbed Hamiltonian $\mathcal{H}_0$ and the magnetization M_x are given by

$$\mathcal{H}_0 = -J \sum_{(i,j)} \sigma_i^z \sigma_j^z, \tag{3.2}$$

$$M_x = m \sum_{j=1}^{N} \sigma_j^x, \tag{3.3}$$

where the first sum is over all nearest-neighbor pairs and J is the coupling constant.

For the calculation of the zero-field perpendicular susceptibility $\chi_{xx}(\omega)$ and its modified susceptibility $\chi_{xx}^T(\omega)$, it is necessary to know the explicit time dependence of the correlation functions defined by Eq. (2.1), which, in the present case, take the forms

$$\phi_{xx}^{\pm}(t) = \text{Tr}\, \rho(M_x M_x(t) \pm M_x(t) M_x). \tag{3.4}$$

To obtain the Heisenberg operator $M_x(t)$ described by the unperturbed Hamiltonian, we first solve the Heisenberg equation of motion for the x component of a spin σ_j^x, located at the jth site of the lattice. If we assume the cyclic boundary condition for the lattice, the solution does not depend on the specific location of the spin. A straightforward calculation gives

$$\sigma_j^x(t) = \cos^q(\omega_0 t) \Pi_j(t) \sigma_j^x, \quad \hbar\omega_0 = 2J, \tag{3.5}$$

where the operator $\Pi_j(t)$ is defined by

$$\Pi_j(t) = \prod_{k=1}^{q} (1 - i \tan(\omega_0 t) \sigma_{j+k}^z \sigma_j^z), \tag{3.6}$$

and σ_{j+1}^z, σ_{j+2}^z, $\cdots$, σ_{j+q}^z are the spins located at the nearest-neighbor sites to the jth site. We note here that, owing to the assumed Hamiltonian $\mathcal{H}_0$, there appear only the nearest-neighbor spins and the spin itself in the description of the motion of a spin.

Substituting Eqs. (3.3) and (3.5) into Eq. (3.4), followed by the expansion of $\Pi_j(t)$ in the powers of $\tan(\omega_0 t)$, and using the property of the trace, we have the explicit time dependence of the functions $\phi_{xx}^{\pm}(t)$. The results appear as linear combinations of the sums

of the spin–spin correlation functions defined by

$$\Theta_{2n}^{(q)} = \mathrm{Tr}\,\rho \sum_{1\le k_1<k_2<\cdots<k_{2n}\le q} \sigma_{k_1}^z\sigma_{k_2}^z\cdots\sigma_{k_{2n}}^z,$$

$$\Theta_0^{(q)} = 1,$$

$$\Theta_{2n+1}^{(q)} = \mathrm{Tr}\,\rho \sum_{1\le k_1<k_2<\cdots<k_{2n+1}\le q} \sigma_0^z\sigma_{k_1}^z\sigma_{k_2}^z\cdots\sigma_{k_{2n+1}}^z, \tag{3.7}$$

where $\sigma_1^z, \sigma_2^z, \cdots, \sigma_q^z$ are the nearest neighbors to σ_0^z. There are $\binom{q}{2n}$ terms in $\Theta_{2n}^{(q)}$ and $\binom{q}{2n+1}$ terms in $\Theta_{2n+1}^{(q)}$. We note here that the correlation functions of only even number spins occur in the sums. The correlation functions of odd number spins are zero, since the Hamiltonian $\mathcal{H}_0$ is quadratic in σ^z. The odd number suffix in $\Theta_{2n+1}^{(q)}$ means only the number of the nearest-neighbor spins involved in one term in the sum. The present notation is chosen for later convenience in the discussion of the algebraic relations between $\Theta_n^{(q)}(n = 0, 1, \cdots, q)$. Hereafter we shall call $\Theta_n^{(q)}$ the spin-correlation sum.

In terms of these spin-correlation sums, the functions $\phi_{xx}^{\pm}(t)$ take the forms

$$\phi_{xx}^{+}(t) = 2Nm^2\cos^q(\omega_0 t)\sum_{n=0}^{[q/2]}(i\tan(\omega_0 t))^{2n}\Theta_{2n}^{(q)},$$

$$\phi_{xx}^{-}(t) = 2Nm^2\cos^q(\omega_0 t)\sum_{n=0}^{[(q-1)/2]}(i\tan(\omega_0 t))^{2n+1}\Theta_{2n+1}^{(q)}, \tag{3.8}$$

where the notation $[\alpha]$ denotes the greatest integer contained in α. Previously, Allan and Betts obtained the second equation of Eqs. (3.8) to calculate the perpendicular susceptibility of the honeycomb and square Ising lattices.

Our next step is to calculate the Fourier transforms of $\phi_{xx}^{\pm}(t)$ defined by Eq. (2.2). We shall first decompose the trigonometric functions of Eq. (3.8) into the Fourier components to obtain

$$\phi_{xx}^{+}(t) = 2^{-q+1}Nm^2\sum_{n=0}^{[q/2]}\sum_{\mu=0}^{q}\Theta_{2n}^{(q)}a_{2n,\mu}^{(q)}e^{i(q-2\mu)\omega_0 t}, \tag{3.9}$$

$$\phi_{xx}^{-}(t) = 2^{-q+1}Nm^2\sum_{n=0}^{[(q-1)/2]}\sum_{\mu=0}^{q}\Theta_{2n+1}^{(q)}a_{2n+1,\mu}^{(q)}e^{i(q-2\mu)\omega_0 t}, \tag{3.10}$$

where the matrix elements $a_{\nu\mu}^{(q)}$ is defined by the generating function

$$(1-x)^{\nu}(1+x)^{q-\nu} = \sum_{\mu=0}^{q}a_{\nu\mu}^{(q)}x^{\mu}; \quad \nu = 0, 1, \cdots, q. \tag{3.11}$$

It is this $(q+1)\times(q+1)$ matrix $\mathrm{a}^{(q)}$ which plays the central role in the algebraic procedures involved in the present problem. It satisfies a simple quadratic equation of the form $(\mathrm{a}^{(q)})^2 = 2^q$ and has many other useful properties. The detailed discussion of the matrix relevant to the present work is given in the Appendix. Hereafter, we shall omit the superscripts q on $\Theta_n^{(q)}$ and $a_{\nu\mu}^{(q)}$ whenever convenient.

According to Eqs. (3.9) and (3.10), the resonance frequencies of the system are given by $(q-2\mu)\omega_0$ $(\mu = 0, 1, 2, \cdots, q)$, each of which corresponds to the energy associated with flipping a spin which is surrounded by q neighboring spins of which $(q-\mu)$ spins are up and μ spins are down.

When q is odd, there exists no zero-frequency Fourier component in both expansions, Eqs. (3.9) and (3.10). The unexpected absence of the component in $\phi_{xx}^{+}(t)$ can be easily understood from the physical ground that it is impossible to flip a central spin surrounded by an odd number of spins without compensation of energy, because the resultant spin of the neighboring spins does not vanish.

Next, when q is even, from Eq. (A22) in the Appendix we have

$$a_{2n+1,q/2} = 0, \tag{3.12}$$

$$a_{2n,q/2} = (-1)^n\binom{q}{q/2}\binom{q/2}{n}\Big/\binom{q}{2n}. \tag{3.13}$$

Accordingly, in agreement with the general discussion given in Sec. 2, $\phi_{xx}^{-}(t)$ does not have the zero-frequency Fourier component and $\phi_{xx}^{+}(t)$ does have the zero-frequency component equal to $2\Delta_q(0)$, where $\Delta_q(0)$ is given by

$$\Delta_q(0) = \mathrm{Tr}\,\rho M_x^0 M_x^0 = 2^{-q}Nm^2\sum_{n=0}^{[q/2]}\Theta_{2n}a_{2n,q/2}. \tag{3.14}$$

Here M_x^0 is the diagonal part of M_x with respect to the Hamiltonian $\mathcal{H}_0$. This term does not vanish in general unless the temperature is absolute zero [see Eq. (4.3)].

Substitution of Eq. (3.9) and (3.10) into Eq. (2.2) yields, for the Fourier transforms of the correlation functions,

$$\Phi_{xx}^{+}(\omega) = \frac{\pi Nm^2}{2^{q-1}\hbar}\sum_{n=0}^{[q/2]}\sum_{\mu=0}^{q}\Theta_{2n}a_{2n,\mu}\delta(\omega-(q-2\mu)\omega_0), \tag{3.15}$$

$$\Phi_{xx}^{-}(\omega) = \frac{\pi Nm^2}{2^{q-1}\hbar}\sum_{n=0}^{[(q-1)/2]}\sum_{\mu=0}^{q}\Theta_{2n+1}a_{2n+1,\mu}\delta(\omega-(q-2\mu)\omega_0). \tag{3.16}$$

According to Eq. (2.12), the imaginary parts of $\chi_{xx}^{T}(\omega)$ and $\chi_{xx}(\omega)$ are given by

$$\mathrm{Im}\,\chi_{xx}^{T}(\omega) = \mathrm{Im}\,\chi_{xx}(\omega) = -\Phi_{xx}^{-}(\omega). \tag{3.17}$$

Since the imaginary part of the susceptibility is a measure for the energy absorption, we see that the absorption of energy occurs at the frequency precisely equal to $(q-2\mu)\omega_0$, $\mu = 0, 1, 2, \cdots, q$.

The real parts of $\chi_{xx}^T(\omega)$ and $\chi_{xx}(\omega)$ are calculated by means of Eqs. (2.14) and (2.15). The results are

$$\operatorname{Re}\chi_{xx}^T(\omega) = \frac{\Delta_q(\omega)}{kT} + \frac{Nm^2}{2^{q-1}\hbar\omega_0}\sum_{n=0}^{[q/2]}{\sum_{\mu=0}^{q}}'\Theta_{2n}a_{2n,\mu}\frac{\tanh[K(q-2\mu)]}{q-2\mu-(\omega/\omega_0)}, \tag{3.18}$$

$$\operatorname{Re}\chi_{xx}(\omega) = \frac{Nm^2}{2^{q-1}\hbar\omega_0}\sum_{n=0}^{[(q-1)/2]}{\sum_{\mu=0}^{q}}'\Theta_{2n+1}a_{2n+1,\mu}\frac{1}{q-2\mu-(\omega/\omega_0)}, \tag{3.19}$$

where the primes denote that the terms involving $q = 2\mu$ are omitted and

$$K = \hbar\omega_0/(2kT) = J/(kT), \tag{3.20}$$

$$\Delta_q(\omega) = 2^{-q}Nm^2\sum_{n=0}^{[q/2]}\Theta_{2n}a_{2n,q/2}u(\omega),$$

$$a_{2n,q/2} = 0 \quad \text{for} \quad q = \text{odd}. \tag{3.21}$$

Here the function $u(\omega)$, as defined by Eq. (2.7), gives a finite discontinuity to the modified susceptibility at $\omega = 0$ when q is even. The above two susceptibilities are linearly related by the general equation (2.13), which, in the present case, becomes

$$\operatorname{Re}\chi_{xx}^T(\omega) = \operatorname{Re}\chi_{xx}(\omega) + \beta\Delta_q(\omega). \tag{3.22}$$

We show later that Fisher's algebraic transformation[7] for the regular Ising lattice arises from this equation.

The calculations of $\chi_{xx}(\omega)$ by Allan and Betts for the honeycomb and square Ising lattices proves to be in complete agreement with the present results specialized for $q = 3$ and 4. Whenever one needs to write down the whole matrix $\mathbf{a}^{(q)}$ for a particular value of q, it is simpler to use the generating function given by Eq. (3.11) directly than to use the general expressions given by Eqs. (A18) and (A19). We give here the matrix $\mathbf{a}^{(q)}$ for the cases $q = 2$, 3, and 4 as examples:

$$\mathbf{a}^{(2)} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix},$$

$$\mathbf{a}^{(3)} = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -3 & 3 & -1 \end{bmatrix}, \tag{3.23}$$

$$\mathbf{a}^{(4)} = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 1 & 2 & 0 & -2 & -1 \\ 1 & 0 & -2 & 0 & -1 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}.$$

4. LIMITING PROPERTIES OF THE SUSCEPTIBILITIES

In this section we shall discuss the limiting properties of susceptibilities $\operatorname{Re}\chi_{xx}^T(\omega)$ and $\operatorname{Re}\chi_{xx}(\omega)$ with respect to the temperatures and frequencies. The limiting properties seem to reveal some physical insights into the present problem. Because of the linear relation Eq. (3.22), one may discuss either one of them—whichever is more convenient—then the other follows from the difference term.

A. High-Temperature Limit

It is known that the spin–spin correlation functions fall off as $K(=J/kT)$ or faster at high temperatures.[5] Accordingly, from Eq. (3.18), when $K \ll 1$, we have

$$\operatorname{Re}\chi_{xx}^T(\omega) = \frac{Nm^2}{kT}\left\{1 + 2^{-q}\sum_{\mu=0}^{q}\binom{q}{\mu}\frac{(\omega/\omega_0)}{(q-2\mu)-(\omega/\omega_0)} + O(K)\right\}, \tag{4.1}$$

where use has been made of $\Theta_0 = 1$ and $a_{0\mu} = \binom{q}{\mu}$ and a convention that the second sum is zero when $\omega = 0$. The first term is due to the thermal agitation. The subsequent sum describes the resonance of a spin surrounded by q nearest-neighbor spins which are randomly oriented. From Eqs. (3.21) and (3.22), for the difference between two susceptibilities at high temperatures when q is even, we have

$$\beta\Delta_q(\omega) = \chi_{xx}^T(\omega) - \chi_{xx}(\omega) = 2^{-q}\binom{q}{q/2}Nm^2\beta u(\omega) + O(K^2). \tag{4.2}$$

B. Low-Temperature Limit

It is well known[5] that, as $T \to 0$,

$$\Theta_n \to \binom{q}{n} + O(\exp(-\alpha/T)), \tag{4.3}$$

where α is a positive quantity. Accordingly, from Eq. (3.18) as $T \to 0$, we have

$$\operatorname{Re}\chi_{xx}^T(\omega) \to 2qNm^2(\hbar\omega_0)^{-1}\tanh(qK)[q^2 - (\omega/\omega_0)^2]^{-1}, \tag{4.4}$$

where use has been made of $a_{0,n} = \binom{q}{n}$ and Eq. (A16). The only effective resonance frequence is $q\omega_0$, which corresponds to the situation where all the nearest spins are either up or down. In this limiting expression the contribution from the term $\beta\Delta_q(0)$ given by Eq. (3.21) proves to be zero regardless of q. This makes sense, because at low temperatures the process is essentially adiabatic, and there should be no difference between adiabatic and isothermal process. If we

consider the next higher term in Eq. (4.4), then the term $\beta\Delta_q(\omega)$ contributes a term proportional to $\exp(-\alpha/T)/T$ when q is even.

C. High-Frequency Limit

In this case there should be no difference between two susceptibilities, since $\Delta_q(\omega) = 0$ when $\omega \neq 0$. It is most convenient to use Eq. (3.19), for there is no additional temperature dependence in the coefficients of spin-correlation sums. We expand the frequency-dependent factor in the powers of (ω_0/ω) and use the sum rules given by Eq. (A20). Then, we have for $\omega \to \infty$

$$\operatorname{Re}\chi_{xx}(\omega) = -\frac{Nm^2}{J}\left(\frac{\omega_0}{\omega}\right)^2\left\{\Theta_1 + (6\Theta_3 + (3q-2)\Theta_1)\left(\frac{\omega_0}{\omega}\right)^2 + O\left(\left(\frac{\omega_0}{\omega}\right)^4\right)\right\}, \quad (4.5)$$

where the leading term is proportional to the energy of the Ising lattice, for Θ_1 is proportional to the energy. If there exists a Ising model which is measureable, then the measurements of the susceptibility at the high-frequency limit should give the energy.

D. Static Susceptibilities

When the frequency ω is equal to zero, Eq. (3.18) gives the isothermal static susceptibility, because $\operatorname{Tr}\rho M_x = 0$ satisfies the condition Eq. (2.11):

$$\chi^T_{xx}(0) = Nm^2 2^{-q}/(kT)\sum_{n=0}^{[q/2]}\sum_{\mu=0}^{q}\Theta_{2n}a_{2n,\mu}\tau(K(q-2\mu)), \quad (4.6)$$

where

$$\tau(\alpha) = (1/\alpha)\tanh\alpha, \quad \tau(0) = 1. \quad (4.7)$$

Fischer has obtained also a general expression for the static susceptibility χ^T_{xx}. Based upon his expression, Fisher[5] has given the detailed expressions for the cases of $q = 2$, 3, and 4. The present result supplies the detailed expressions for general cases.

From Eq. (3.19), we obtain the Kubo adiabatic susceptibility:

$$\chi_{xx}(0) = \frac{Nm^2}{2^{q-2}\hbar\omega_0}\sum_{n=0}^{[(q-1)/2]}\sum_{\mu=0}^{[(q-1)/2]}\Theta_{2n+1}a_{2n+1,\mu}/(q-2\mu). \quad (4.8)$$

When q is odd, this equation is further simplified to the following form:

$$\chi_{xx}(0) = -\frac{Nm^2}{\hbar\omega_0}\sum_{n=0}^{[(q-1)/2]}\Theta_{2n+1}n!\Big/\left(-\frac{q}{2}\right)_{n+1}, \quad (4.9)$$

where use has been made of Eq. (A21), and $(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1)$, $(\alpha)_0 = 1$. This equation gives the isothermal static susceptibility as well, since $\Delta_q(0) = 0$ for $q =$ odd.

5. FISHER'S ALGEBRAIC TRANSFORMATION OF SPIN–SPIN CORRELATION FUNCTIONS

Based upon the linear relation between the Kubo and the modified susceptibilities given by Eq. (3.22), we shall derive Fisher's algebraic transformation of spin–spin correlation functions.[7] We substitute Eqs. (3.18) and (3.19) into Eq. (3.22) and compare the both sides of the equation; then we obtain a set of linearly independent equations:

$$\sum_{n=0}^{[(q-1)/2]}\Theta_{2n+1}a_{2n+1,\mu} = \tanh(K(q-2\mu))\sum_{n=0}^{[q/2]}\Theta_{2n}a_{2n,\mu}, \quad \mu = 0, 1, 2, \cdots, [(q-1)/2]. \quad (5.1)$$

The equations for μ lying outside the range specified here are redundant. The reason is that the set of equations is invariant under the transformation $\mu \to q - \mu$, because of the symmetry property of the matrix $a_{\nu\mu}$ given in Eqs. (A15):

$$a_{\nu\mu} = (-1)^\nu a_{\nu,q-\mu}. \quad (5.2)$$

The equations in the set of Eq. (5.1) are linearly independent, since the determinant of the $[(q+1)/2] \times [(q+1)/2]$ matrix $(a_{2n+1,\mu})$ which describes the left-hand side of Eq. (5.1) can be shown to be nonzero:

$$\det|a_{2n+1,\mu}| = (-4)^{\frac{1}{2}[\frac{1}{2}(q-1)][\frac{1}{2}(q+1)]}. \quad (5.3)$$

By means of the orthogonality relation of the matrix a given by Eq. (A16), we can solve Eq. (5.1) for Θ_{2n+1} to obtain

$$\Theta_{2n+1} = 2^{-q+1}\sum_{m=0}^{[q/2]}\sum_{\mu=0}^{[(q-1)/2]}\Theta_{2m}a_{2m,\mu} \times \tanh(K(q-2\mu))a_{\mu,2n+1}, \quad n = 0, 1, 2, \cdots, [(q-1)/2]. \quad (5.4)$$

If we recall the definitions of spin-correlation sums Θ_n given by Eq. (3.7), we see that Eq. (5.4) is Fisher's algebraic transformations of spin–spin correlation functions for the regular Ising lattice of which the Hamiltonian $\mathcal{H}_0$ is defined by Eq. (3.2). Actually Fischer introduced a transformation which is valid for more general Ising lattices.[7] Equation (5.4) is valid only for the regular Ising lattice. Within this limitation, however, this equation seems to give the most general and explicit results of this kind.

In the analogous manner we may express Θ_{2n} in terms of Θ_{2n+1}. Using the orthogonality relations

Eq. (A16), we obtain

$$\Theta_{2n} = a_{q/2,2n} + 2^{-q+1}\sum_{m=0}^{[(q-1)/2]}\sum_{\mu=0}^{[(q-1)/2]}\Theta_{2m+1}a_{2m+1,\mu} \times \coth(K(q-2\mu))(a_{\mu,2n} - a_{q/2,2n}), \quad n = 0, 1, 2, \cdots, [q/2], \quad (5.5)$$

where the following obvious convention is used:

$$a_{q/2,2n} = 0 \quad \text{for} \quad q = \text{odd}. \quad (5.6)$$

It is interesting to note that, when q = odd and $n = 0$, Eq. (5.5) gives an equation which holds among Θ_{2n+1}, for $\Theta_0 = 1$. When q = even and $n = 0$, Eq. (5.5) is simply $1 = 1$, for $a_{v0} = 1$.

One application of Eq. (5.5) may be to express the modified susceptibility in terms of Θ_{2n+1}. We only need the following equation:

$$\Delta_q(0)/(2^qNm^2) = \sum_{n=0}^{[q/2]}\Theta_{2n}a_{2n,q/2} = 2^q - \sum_{n=0}^{[(q-1)/2]}\sum_{\mu=0}^{q}{}'\Theta_{2n+1}a_{2n+1,\mu}\coth(K(q-2\mu)). \quad (5.7)$$

Substituting Eqs. (5.7) and (3.19) into Eq. (3.22), we obtain the required result:

$$\operatorname{Re}\chi_{xx}^T(\omega) = \frac{Nm^2}{kT} + \frac{Nm^2}{2^{q-1}\hbar\omega_0}\sum_{n=0}^{[(q-1)/2]}\sum_{\mu=0}^{q}{}'\Theta_{2n+1}a_{2n+1,\mu} \times \left(\frac{1}{q-2\mu-(\omega/\omega_0)} - \frac{K}{\tanh(K(q-2\mu))}\right). \quad (5.8)$$

One of the most useful solutions of Eq. (5.1) may be to express the higher-order spin-correlation functions Θ_q, $\Theta_{q-1}, \cdots, \Theta_{[q/2]+1}$ by the lower-order spin-correlation functions Θ_0, $\Theta_1, \cdots, \Theta_{[q/2]}$. We first rewrite Eq. (5.1) in the form

$$\sum_{\mu=0}^{q}(e^{(-q+2s)K} - (-1)^\mu e^{(q-2s)K})a_{s\mu}\bar{\Theta}_\mu = 0, \quad s = 0, 1, 2, \cdots, q, \quad (5.9)$$

where use has been made of the symmetry property of the matrix a given by the first equation of Eqs. (A15) and $\bar{\Theta}_\mu$ are defined by

$$\bar{\Theta}_\mu = \Theta_\mu \Big/ \binom{q}{\mu}. \quad (5.10)$$

Using the product rule of the matrix given by Eq. (A30), we can transform Eq. (5.9) into a form

$$\sum_{\mu=0}^{q}A_{2n+1,\mu}^{(q)}(-t^2)t^{-\mu}\bar{\Theta}_\mu = 0, \quad t = \tanh K, \quad n = 0, 1, 2, \cdots, [(q-1)/2], \quad (5.11)$$

where the matrix $\mathsf{A}^{(q)}(\xi)$ is defined by the generating equation

$$(1-x)^v(1+\xi x)^{q-v} = \sum_{\mu=0}^{q}A_{v\mu}^{(q)}(\xi)x^\mu, \quad v = 0, 1, 2, \cdots, q. \quad (5.12)$$

This matrix is a simple generalization of the matrix a. It satisfies the orthogonality and symmetry relations given by Eqs. (A3), (A6), and (A7). If we use the product rules given by Eq. (A34), we can transform Eq. (5.11) to obtain a set of linear equations, of which only $[(q+1)/2]$ are linear independent:

$$\sum_{\mu=0}^{q}(B_{v\mu}(c_2) - B_{q-v,\mu}(c_2))s_2^\mu\bar{\Theta}_\mu = 0, \quad (5.13)$$

where

$$c_2 = \cosh(2K), \quad s_2 = \sinh(2K), \quad (5.14)$$

and the matrix $\mathsf{B}(\xi)$ is defined by

$$(\xi - x)^v = \sum_{\mu=0}^{q}B_{v\mu}(\xi)x^\mu; \quad B_{v\mu}(\xi) = (-1)^\mu\binom{v}{\mu}\xi^{v-\mu}, \quad v, \mu = 0, 1, 2, \cdots, q. \quad (5.15)$$

By definition, the matrix $\mathsf{B}(\xi)$ is triangular, i.e., $B_{v\mu}(\xi) = 0$ for $v < \mu$. By means of the orthogonality relation of this matrix given by Eq. (A.24), we have the required solution:

$$\bar{\Theta}_\lambda = -\sum_{v=[q/2]+1}^{q}\sum_{\mu=0}^{[q/2]}B_{\lambda v}(c_2)[B_{v\mu}(c_2) - B_{q-v,\mu}(c_2)]s_2^{\mu-\lambda}\bar{\Theta}_\mu, \quad \lambda = q, q-1, \cdots, \left[\frac{q}{2}\right] + 1, \quad (5.16)$$

where the spin–spin correlation sums $\bar{\Theta}_\mu$ ($\mu = 0, 1, 2, \cdots, q$) are defined by Eqs. (3.7) and (5.10).

In the application of this general expression to a specific value of q, it is more convenient to introduce a matrix defined by

$$\Gamma_{v\mu}(c_2) = B_{v\mu}(c_2) - B_{q-v,\mu}(c_2), \quad (5.17)$$

which satisfies a simple symmetry relation $\Gamma_{\mu\nu} = -\Gamma_{q-v,\mu}$. Then, Eq. (5.16) becomes

$$\bar{\Theta}_\lambda = -\sum_{v=[q/2]+1}^{q}\sum_{\mu=0}^{[q/2]}\Gamma_{\lambda v}(c_2)\Gamma_{v\mu}(c_2)s_2^{\mu-\lambda}\bar{\Theta}_\mu, \quad (5.18)$$

where λ is the same as in Eq. (5.16). We write down the detailed expressions for the case $q = 6$ as an example:

$$\bar{\Theta}_4 = -x^2 + 2x(x^2+1)\bar{\Theta}_1 - (5x^2+1)\bar{\Theta}_2 + 4x\bar{\Theta}_3,$$
$$\bar{\Theta}_5 = -x(3x^2+1) + (6x^4+8x^2+1)\bar{\Theta}_1 - 5x(3x^2+1)\bar{\Theta}_2 + 10x^2\bar{\Theta}_3,$$
$$\bar{\Theta}_6 = -(6x^4+3x^2+1) + 6x(2x^4+3x^2+1)\bar{\Theta}_1 - 15x^2(2x^2+1)\bar{\Theta}_2 + 20x^3\bar{\Theta}_3,$$

where $x = \coth(2K)$.

[*Note added in proof:* After the completion of this work the author noticed that G. A. T. Allan and D. D. Betts, in the published version of their work,[3] resolved the discrepancy between χ_{xx}^{T} and $\chi_{xx}(0)$ of the Ising lattice by inserting a constant magnetic field H_z in the Hamiltonian $\mathcal{H}_0$ and taking the limit $\omega \to 0$ before the limit $H_z \to 0$ in the end. Professor Michael E. Fisher at Cornell University (private communication) kindly informed the author that the discrepancy is resolved by altering one of the interactions J at each site of the Ising lattice.]

ACKNOWLEDGMENTS

The author wishes to thank Dr. J. Dillon, Dean of the Graduate School, University of Louisville, who arranged the financial support for this work. The author is also grateful to Dr. Sung S. Yun at Ohio University and Dr. J. Brown, Dr. R. Wilson, and K. Lim at Louisville for stimulating discussions.

APPENDIX: MATHEMATICAL PROPERTIES OF THE MATRICES $\mathbf{A}^{(q)}(\xi)$, $\mathbf{a}^{(q)}$, AND $\mathbf{B}^{(q)}(\xi)$

The Matrix $\mathbf{A}^{(q)}(\xi)$

For a given integer q, a $(q+1) \times (q+1)$ matrix $\mathsf{A}^{(q)}(\xi)$, with an arbitrary parameter ξ, is defined by the generating function

$$(1-x)^{\nu}(1+\xi x)^{q-\nu} = \sum_{\mu=0}^{q} A_{\nu\mu}^{(q)}(\xi) x^{\mu}, \quad \nu = 0, 1, \cdots, q. \tag{A1}$$

If we introduce a transformation of the variable $x \to y$,

$$y = (1-x)/(1+\xi x), \tag{A2}$$

of which the inverse is the same as itself, provided that $\xi \neq -1$, we can show that the matrix elements satisfy the orthogonality relation

$$\sum_{s=0}^{q} A_{\nu s}^{(q)}(\xi) A_{s\mu}^{(q)}(\xi) = (1+\xi)^{q} \delta_{\nu\mu}, \quad \nu, \mu = 0, 1, \cdots, q, \tag{A3}$$

where $\delta_{\nu\mu}$ is Kronecker's delta. A direct calculation shows that Eq. (A3) holds even when $\xi = -1$. Hereafter, whenever there exist no confusions, we shall suppress the superscript q of the matrix $\mathsf{A}^{(q)}$. We assume also that the subscripts of the matrix run from 0 to q unless specified otherwise.

Since the matrix satisfies a simple quadratic equation $(\mathsf{A}(\xi))^2 = (1+\xi)^q$, its eigenwerts are $\pm(1+\xi)^{q/2}$. The determinant and the trace of the matrix prove to be

$$\det |\mathsf{A}(\xi)| = (-1-\xi)^{q(q+1)/2},$$
$$\operatorname{Tr} \mathsf{A}(\xi) = \begin{cases} 0, & q = \text{odd}, \\ (1+\xi)^{q/2}, & q = \text{even}. \end{cases} \tag{A4}$$

Accordingly, the degeneracy of the positive eigenwert (when $\xi > -1$) is larger by one than, or equal to that of, the negative eigenwert.

From the symmetry of the bilinear form of x and y defined by

$$\sum_{\nu,\mu=0}^{q} \binom{q}{\nu} \xi^{\nu} A_{\nu\mu}(\xi) y^{\nu} x^{\mu} = (1 + (x + y - xy)\xi)^{q}, \tag{A5}$$

we obtain a symmetry property of the matrix:

$$\binom{q}{\nu} \xi^{\nu} A_{\nu\mu}(\xi) = \binom{q}{\mu} \xi^{\mu} A_{\mu\nu}(\xi). \tag{A6}$$

Directly from the symmetry of the generating function involved in the transformation $\nu \to q - \nu$, we have

$$A_{\nu\mu}(\xi) = (-\xi)^{\mu} A_{q-\nu,\mu}(\xi^{-1}) = (-1)^{\nu} \xi^{q-\nu} A_{\nu,q-\mu}(\xi^{-1}). \tag{A7}$$

It is a simple matter to obtain the explicit form of the matrix by means of the binomial theorem:

$$A_{\nu\mu}(\xi) = \sum_{s=l_1}^{l_2} (-1)^{\mu} \binom{\nu}{\mu - s} \binom{q-\nu}{s} (-\xi)^{s}, \tag{A8}$$

where the lower limit of the summation l_1 is the larger of 0 and $\mu - \nu$ and the upper limit l_2 is the smaller of $q - \nu$ and μ. For the values of s lying outside of this range, the product of the binomial coefficients simply vanishes. From Eq. (A9) it is possible to represent the matrix elements in terms of the hypergeometric function[8]:

$$A_{\nu\mu}^{(q)}(\xi) = (-1)^{\mu} \binom{\nu}{\mu} {}_2F_1(\nu - q, -\mu; \nu - \mu + 1; -\xi). \tag{A9}$$

This equation holds even when q is not an integer, provided that $|\xi| < 1$. Alternate representations are also possible if we use the symmetry properties given by Eqs. (A7) and (A6). From Eq. (A9) we can show that the matrix elements are related to the Jacobi polynomials[9] $P_{\mu}^{\alpha\beta}$ as follows:

$$A_{\nu\mu}^{(q)}(\xi) = (-1)^{\mu}(1+\xi)^{\mu} P_{\mu}^{\alpha\beta}\left(\frac{1-\xi}{1+\xi}\right),$$
$$P_{\mu}^{\alpha\beta}(\xi) = (-2)^{-\mu}(1+\xi)^{\mu} A_{\nu\mu}^{(q)}\left(\frac{1-\xi}{1+\xi}\right), \tag{A10}$$

where

$$\alpha = \nu - \mu > -1, \quad \beta = q - \nu - \mu > -1.$$

[8] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, Eds. (U.S. Government Printing Office, National Bureau of Standards, Washington, D.C., 1964), Appl. Math. Ser. 55, p. 556.

[9] See Ref. 8, p. 773.

We give the explicit form of the matrix $\mathsf{A}^{(2)}(\xi)$ as an example:

$$\mathsf{A}^{(2)}(\xi) = \begin{pmatrix} 1 & 2\xi & \xi^2 \\ 1 & \xi - 1 & -\xi \\ 1 & -2 & 1 \end{pmatrix}.$$

The Matrix $\mathsf{a}^{(q)}$

It is a special case of the matrix $\mathsf{A}^{(q)}(\xi)$:

$$\mathsf{a}^{(q)} = \mathsf{A}^{(q)}(1). \tag{A11}$$

In the main text, this matrix has risen from the Fourier expansion of the trigonometric function

$$\sin^{\nu}\theta \cos^{q-\nu}\theta = i^{-\nu}2^{-q}\sum_{\mu=0}^{q} a_{\nu\mu}e^{i(q-2\mu)\theta}, \tag{A12}$$

where we have omitted the superscript q for convenience. The generating function is written as

$$(1-x)^{\nu}(1+x)^{q-\nu} = \sum_{\mu=0}^{q} a_{\nu\mu}x^{\mu} \tag{A13}$$

and the orthogonality and symmetry properties are, from Eqs. (A3), (A6), and (A7),

$$\sum_{s=0}^{q} a_{\nu s}a_{s\mu} = 2^{q}\delta_{\nu\mu}, \tag{A14}$$

$$\binom{q}{\nu}a_{\nu\mu} = \binom{q}{\mu}a_{\mu\nu}, \quad a_{\nu\mu} = (-1)^{\mu}a_{q-\nu,\mu} = (-1)^{\nu}a_{\nu,q-\mu}. \tag{A15}$$

From these we obtain the orthogonality relations in "half" dimension:

$$\sum_{n=0}^{[q/2]} a_{\nu,2n}a_{2n,\mu} = 2^{q-1}(\delta_{\nu\mu} + \delta_{q-\nu,\mu}),$$
$$\sum_{n=0}^{[(q-1)/2]} a_{\nu,2n+1}a_{2n+1,\mu} = 2^{q-1}(\delta_{\nu\mu} - \delta_{q-\nu,\mu}). \tag{A16}$$

If we write the generating function of a in the form

$$(1-x^2)^{\nu}(1+x)^{q-2\nu} = \sum_{\mu=0}^{q} a_{\nu\mu}x^{\mu}, \tag{A17}$$

we obtain an expression, which is sometime more convenient than $A(\xi = 1)$, given by Eq. (A8):

$$a_{\nu\mu} = \sum_{s}(-1)^{s}\binom{\nu}{s}\binom{q-2\nu}{\mu-2s}, \quad 0 \le \nu, \quad \mu \le q, \tag{A18}$$

where the summation over s is limited to the range where the whole number s gives nonvanishing binomial coefficients. From this we obtain the first few elements of $a_{\nu\mu}$ in terms of $k_{\nu}(= q - 2\nu)$:

$$a_{\nu 0} = 1, \quad a_{\nu 1} = k_{\nu}, \quad 2!\,a_{\nu 2} = k_{\nu}^2 - q,$$
$$3!\,a_{\nu 3} = k_{\nu}^3 - (3q-2)k_{\nu},$$
$$4!\,a_{\nu 4} = k_{\nu}^4 - (8-6q)k_{\nu}^2 + 3q(q-2). \tag{A19}$$

By means of these we obtain the sum rules for the matrix, which we need in the present problem:

$$2^{-q}\sum_{\mu} a_{\nu\mu} = \delta_{\nu 0},$$
$$2^{-q}\sum_{\mu} a_{\nu\mu}(q-2\mu) = \delta_{\nu,1},$$
$$2^{-q}\sum_{\mu} a_{\nu\mu}(q-2\mu)^2 = q\delta_{\nu 0} + 2!\,\delta_{\nu 2},$$
$$2^{-q}\sum_{\mu} a_{\nu\mu}(q-2\mu)^3 = (3q-2)\delta_{\nu 1} + 3!\,\delta_{\nu 3},$$
$$2^{-q}\sum_{\mu} a_{\nu\mu}(q-2\mu)^4 = q(3q-2)\delta_{\nu 0} + 4(3q-4)\delta_{\nu 2} + 4!\,\delta_{\nu 4}, \tag{A20}$$

where the summations over μ are from 0 to q. These equations could be obtained by differentiation of Eq. (A12) by θ. By integrating Eq. (A12) over θ from 0 to π, we obtain

$$\sum_{\mu=0}^{q} a_{2n+1,\mu}/(q-2\mu) = -2^{q-1}n!/(-q/2)_{n+1}, \quad q = \text{odd}, \tag{A21}$$

where

$$(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1),$$
$$(\alpha)_0 = 1.$$

We may need the following special elements of the matrix a:

$$a_{\nu 0} = 1, \quad a_{0\nu} = \binom{q}{\nu},$$

and for $q = 2p$ (p = an integer)

$$a_{p,2n+1} = 0, \quad a_{p,2n} = (-1)^n\binom{p}{n},$$
$$a_{2n+1,p} = 0, \quad a_{2n,p} = (-1)^n\binom{2p}{p}\binom{p}{n}\Big/\binom{2p}{2n} = 4^p(\tfrac{1}{2}-n)_p/p!. \tag{A22}$$

The specific examples of the matrix for the cases $q = 2$, 3, and 4 are given by Eq. (3.23).

The Matrix $\mathsf{B}^{(q)}(\xi)$

This matrix is generated by the equation

$$(\xi - x)^{\nu} = \sum_{\mu=0}^{q} B_{\nu\mu}^{(q)}(\xi)x^{\mu}, \quad \nu = 0, 1, \cdots, q. \tag{A23}$$

We can show the following orthogonal and symmetry relations in a manner analogous to the case of the matrix $\mathsf{A}^{(q)}(\xi)$:

$$\sum_{s=0}^{q} B_{\nu s}(\xi)B_{s\mu}(\xi) = \delta_{\nu\mu}, \tag{A24}$$

$$\binom{q}{\nu}B_{\nu\mu}(\xi) = \binom{q}{\mu}B_{q-\mu,q-\nu}(\xi), \quad 0 \le \nu, \quad \mu \le q, \tag{A25}$$

where we have omitted the superscript q for convenience. The explicit form is

$$B_{\nu\mu}(\xi) = (-1)^{\mu}\binom{\nu}{\mu}\xi^{\nu-\mu}. \tag{A26}$$

It is a triangular matrix, i.e., $B_{\nu\mu}(\xi) = 0$ for $\mu > \nu$. In particular, when $\xi = 0$ or $\xi = 1$, we have

$$B_{\nu\mu}(0) = (-1)^{\nu}\delta_{\nu\mu}, \quad B_{\nu\mu}(1) = A_{\nu\mu}(0). \tag{A27}$$

We give $\mathsf{B}^{(2)}(\xi)$ as an example:

$$\mathsf{B}^{(2)}(\xi) = \begin{pmatrix} 1 & 0 & 0 \\ \xi & -1 & 0 \\ \xi^2 & -2\xi & 1 \end{pmatrix}.$$

Product Rule

One of the most important properties of the matrix $\mathsf{A}^{(q)}(\xi)$ arises from the following bilinear forms of x and y:

$$\sum_{s=0}^{q}\sum_{\mu=0}^{q} A_{\nu s}^{(q)}(\xi_1)y^s A_{s\mu}(\xi)x^{\mu} = (1 - y + (\xi + y)x)^{\nu}(1 + \xi_1 y + (\xi - \xi_1 y)x)^{q-\nu}, \quad \nu = 0, 1, \cdots, q. \tag{A28}$$

Expansion of the right-hand side of this equation in the powers of x yields the product rule:

$$\sum_{s=0}^{q} A_{\nu s}^{(q)}(\xi_1)y^s A_{s\mu}^{(q)}(\xi) = (-1)^{\mu}(1 + \xi_1 y)^{q-\nu}(\xi + y)^{\mu}(1 - y)^{\nu-\mu} \times A_{\nu\mu}^{(q)}\left(-\frac{(\xi - \xi_1 y)(1 - y)}{(1 + \xi_1 y)(\xi + y)}\right), \quad \nu, \mu = 0, 1, \cdots, q. \tag{A29}$$

When the argument of the matrix in the right-hand side of this equation becomes zero or infinite, we may use the corresponding limiting values of the right-hand side as a whole or we may rederive the equation directly from Eq. (A28). We consider only the following special cases which we need in the present problem:

(i) When $\xi = \xi_1$ and $y = 1$,

$$\sum_{s=0}^{q} A_{\nu s}(\xi)A_{s\mu}(\xi) = (1 + \xi)^{q}\delta_{\nu\mu},$$

which is the orthogonal relation already given by Eq. (A3). In this proof, however, we did not use the unnecessary condition, $\xi \neq -1$.

(ii) When $\xi = \xi_1 = 1$,

$$\sum_{s=0}^{q} a_{\nu s}y^s a_{s\mu} = (-1)^{\mu}(1 + y)^{q}\left(\frac{1 - y}{1 + y}\right)^{\nu-\mu} A_{\nu\mu}\left(-\left(\frac{1 - y}{1 + y}\right)^2\right). \tag{A30}$$

(iii) When $y = \xi$ and $\xi_1 = 1$,

$$\sum_{s=0}^{q} a_{\nu s}\xi^s A_{s\mu}(\xi) = (-2\xi)^{\mu}(1 + \xi)^{q-\mu}B_{\nu\mu}\left(\frac{1 - \xi}{1 + \xi}\right). \tag{A31}$$

(iv) When $y = -\xi$ and $y = 1$,

$$\sum_{s=0}^{q} a_{\nu s}(-\xi)^s A_{s\mu}(\xi) = (-2\xi)^{\mu}(1 + \xi)^{q-\mu}B_{q-\nu,\mu}\left(\frac{1 - \xi}{1 + \xi}\right). \tag{A32}$$

From Eq. (A31) and Eq. (A32) we obtain the following equations as a generalization of Eqs. (A16):

$$2\sum_{n=0}^{[q/2]} a_{\nu,2n}\xi^{2n}A_{2n,\mu}(\xi) = (-2\xi)^{\mu}(1 + \xi)^{q-\mu}\left(B_{\nu\mu}\left(\frac{1 - \xi}{1 + \xi}\right) + B_{q-\nu,\mu}\left(\frac{1 - \xi}{1 + \xi}\right)\right), \tag{A33}$$

$$2\sum_{n=0}^{[(q-1)/2]} a_{\nu,2n+1}\xi^{2n+1}A_{2n+1,\mu}(\xi) = (-2\xi)^{\mu}(1 + \xi)^{q-\mu}\left(B_{\nu\mu}\left(\frac{1 - \xi}{1 + \xi}\right) - B_{q-\nu,\mu}\left(\frac{1 - \xi}{1 + \xi}\right)\right). \tag{A34}$$

When $\xi = 1$, these equations in fact reduce to Eqs. (A16), owing to the first equation of Eqs. (A27).

The author has obtained a matrix $\mathsf{Q}^{(q)}(\alpha, \beta, \gamma)$ with three arbitrary parameters, which includes the matrices $\mathsf{A}^{(q)}(\xi)$ and $\mathsf{B}^{(q)}(\xi)$ as its special cases. It satisfies a simple quadratic equation and symmetry relations analogous to those discussed in the present work. We do not discuss it here since it is not necessary within the scope of the present paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Stationary Electromagnetic Fields in General Relativity

B. K. DATTA
Department of Mathematics, Surendranath College, Calcutta
AND
A. K. RAYCHAUDHURI
Physics Department, Presidency College, Calcutta

(Received 10 October 1967)

The paper presents some solutions of the Einstein–Maxwell equations corresponding to a superposed electric and magnetic field at right angles to each other. The solutions are stationary and include the case of null field also. There are, however, some singularities in the metric tensor.

1. INTRODUCTION

A very common type of electromagnetic field that one meets with in the laboratory is the case of static superposed electric and magnetic fields at right angles to each other. In this case the Poynting vector $\mathbf{E} \times \mathbf{H}$ does not vanish, although there is no energy flux.

One may wonder how such a field can be accommodated in the formalism of general relativity. The essentially static nature would lead one to expect that the metric-tensor components would be independent of time (in the language of groups of motion, the space–time must admit a motion with a timelike generator). However, a completely static metric would require all components such as R_{i4} to vanish ($i = 1, 2, 3$; $x^4 = t$), whereas the crossed fields would require at least some of these components to survive. The only way to accommodate such fields thus appears to be to introduce not a static but a stationary metric of the type formally similar to what one meets in case of rotation.

In this paper we propose to investigate some such source-free fields and, indeed, some fields which may be called null fields in the sense that

$$F^{\mu\nu}F_{\mu\nu} = 0, \quad {}^*F^{\mu\nu}F_{\mu\nu} = 0,$$

i.e.,

$$E^2 = H^2, \quad \mathbf{E}\cdot\mathbf{H} = 0,$$

but these are nevertheless stationary nonradiation fields.

2. FIELD EQUATIONS

We start with the metric

$$ds^2 = f\,dt^2 - e^{2\psi}(dx^2 + dy^2) - l\,dz^2 + 2m\,dz\,dt \tag{1}$$

and assume further that f, ψ, l, and m are functions of x alone. It is obvious that, if z is an angle coordinate, the above metric is cylindrically symmetric. Numbering x, y, z, t as 1, 2, 3, 4, we assume further that the only nonvanishing contravariant components of the electromagnetic-field tensor are $F^{13}(= -F^{31})$ and $F^{14}(= -F^{41})$. For such a field one has

$$T^4{}_4 = -T^3{}_3 \tag{2}$$

so that the Einstein equations

$$R^\mu{}_\nu - \tfrac{1}{2}R\delta^\mu{}_\nu = -8\pi T^\mu{}_\nu \tag{3}$$

give

$$R^3{}_3 + R^4{}_4 = 0. \tag{4}$$

In view of (4), it is possible to introduce Weyl-like canonical coordinates such that[1]

$$fl + m^2 = x^2. \tag{5}$$

Further, the fields being source-free, one has from the Maxwell equations

$$F^{31} = A/(-g)^{\frac{1}{2}} \tag{6}$$

and

$$F^{41} = B/(-g)^{\frac{1}{2}}, \tag{7}$$

where A and B are arbitrary constants and g is the determinant of the metric tensor.

For the line element (1) with Eq. (5) satisfied, the Ricci tensor components are

$$(-g)^{\frac{1}{2}}R^1{}_1 = -\left[x\psi_{11} - \psi_1 - \frac{1}{2x}(l_1 f_1 + m_1^2)\right], \tag{8}$$

$$(-g)^{\frac{1}{2}}R^2{}_2 = -(x\psi_{11} + \psi_1), \tag{9}$$

$$(-g)^{\frac{1}{2}}R^3{}_3 = -\frac{1}{2}\frac{d}{dx}\left(\frac{fl_1 + mm_1}{x}\right), \tag{10}$$

$$(-g)^{\frac{1}{2}}R^4{}_4 = -\frac{1}{2}\frac{d}{dx}\left(\frac{lf_1 + mm_1}{x}\right), \tag{11}$$

$$(-g)^{\frac{1}{2}}R^3{}_4 = \frac{1}{2}\frac{d}{dx}\left(\frac{fm_1 - mf_1}{x}\right), \tag{12}$$

$$(-g)^{\frac{1}{2}}R^4{}_3 = \frac{1}{2}\frac{d}{dx}\left(\frac{ml_1 - lm_1}{x}\right), \tag{13}$$

[1] W. J. Van Stockum, Proc. Roy. Soc. (Edinburgh) **57**, 135 (1937).

where the subscript 1 indicates differentiation with respect to x.

Now, with the usual expression for the electromagnetic stress-energy tensor,

$$T^{\mu}{}_{\nu} = -\frac{1}{4\pi}(F^{\mu\alpha}F_{\nu\alpha} - \tfrac{1}{4}\delta^{\mu}{}_{\nu}F_{\alpha\beta}F^{\alpha\beta}), \tag{14}$$

the field equations give

$$-\frac{1}{x}(fB^2 - lA^2 + 2mAB) = -\left[x\psi_{11} - \psi_1 - \frac{1}{2x}(l_1f_1 + m_1^2)\right], \tag{15}$$

$$\frac{1}{x}(fB^2 - lA^2 + 2mAB) = -(x\psi_{11} + \psi_1), \tag{16}$$

$$\frac{1}{x}(lA^2 + fB^2) = -\frac{1}{2}\frac{d}{dx}\left(\frac{fl_1 + mm_1}{x}\right), \tag{17}$$

$$-\frac{1}{x}(lA^2 + fB^2 = -\frac{1}{2}\frac{d}{dx}\left(\frac{lf_1 + mm_1}{x}\right), \tag{18}$$

$$-\frac{2}{x}(mA^2 + fAB) = \frac{1}{2}\frac{d}{dx}\left(\frac{fm_1 - mf_1}{x}\right), \tag{19}$$

$$-\frac{2}{x}(mB^2 - lAB) = \frac{1}{2}\frac{d}{dx}\left(\frac{ml_1 - lm_1}{x}\right). \tag{20}$$

Among the four equations (17)–(20), because of the symmetry of $R_{\mu\nu}$ and $T_{\mu\nu}$, only three can be independent. Furthermore, Eqs. (17) and (18) are identical in view of (5). Hence there are effectively only two independent equations in the set (17)–(20) to determine f, l, and m, of which again only two are independent by virture of Eq. (5). From Eqs. (17) and (18) we get

$$\frac{4}{x}(lA^2 + fB^2) = -\frac{d}{dx}\left(\frac{fl_1 - lf_1}{x}\right). \tag{21}$$

We take Eqs. (19), (20), and (21) as the basic equations to determine f, l, and m.

We have not been able to get the general solution of these equations. However, we have been able to obtain a special class of solutions in the following manner.

Writing

$$u = f/l, \qquad v = m/l, \tag{22}$$

Eq. (5) gives

$$u + v^2 = x^2/l^2. \tag{23}$$

Equations (19)–(21) become

$$\frac{4}{x}(mA^2 + fAB) = -\frac{d}{dx}\left[\frac{u^2x}{u + v^2}\frac{d}{dx}\left(\frac{v}{u}\right)\right], \tag{24}$$

$$\frac{4}{x}(mB^2 - lAB) = \frac{d}{dx}\left[\frac{x}{u + v^2}\frac{dv}{dx}\right], \tag{25}$$

$$\frac{4}{x}(lA^2 + fB^2) = \frac{d}{dx}\left[\frac{x}{u + v^2}\frac{du}{dx}\right]. \tag{26}$$

If now we assume that there is a linear relation between u and v,

$$v = au + b, \tag{27}$$

we find that Eqs. (24)–(26) require

$$aA^2 + BA - bB^2 = 0. \tag{28}$$

Regarded as a quadratic in A, Eq. (28) gives

$$A = -(B/2a)(1 + \mu), \tag{29}$$

where

$$\mu = \pm(1 + 4ab)^{\frac{1}{2}}. \tag{30}$$

Hence, for the fields to be real, we have

$$1 + 4ab \geq 0. \tag{31}$$

The solutions divide naturally into two classes, accordingly as the sign of equality or inequality holds in (31), and we consider them separately.

3. $1 + 4ab = 0$ NULL FIELD

Of the tensors ${}^*F^{\mu\nu}$ dual to $F^{\mu\nu}$, only ${}^*F^{24}$, $(= -{}^*F^{42})$, and ${}^*F^{23}$ $(= -{}^*F^{32})$ exist; hence ${}^*F^{\alpha\beta}F_{\alpha\beta}$ vanishes for the fields under consideration. If $1 + 4ab = 0$, then, from Eqs. (29) and (39),

$$A = -(B/2a) \tag{32}$$

and a little calculation shows that $F^{\alpha\beta}F_{\alpha\beta}$ also vanishes. The field in this case is thus a null field, although it is not a radiation field, as is clear from the stationary character of the metric.

With (29) satisfied, Eqs. (24)–(26) reduce to one single equation. Using (22) and (23) to eliminate f, l, and m, we have, from Eqs. (24)–(26), after one integration

$$\frac{x}{(au + 1/4a)^2}\frac{du}{dx} = \frac{4B^2}{a}x + \frac{C}{a}, \tag{33}$$

where we have used (31) with the equality sign and Eq. (32). C is an arbitrary constant of integration.

On further integration we get

$$\frac{1}{au + 1/4a} = -[4B^2x + C\log x + D] = -\xi, \tag{34}$$

say, or

$$au = -1/\xi - 1/4a, \tag{35}$$

where D is again a constant of integration.

From (22), (23), (27), and (35) we now get

$$l = -x\xi, \tag{36}$$

$$f = (x/4a^2)(4a + \xi), \tag{37}$$

and

$$m = (x/2a)(2a + \xi). \tag{38}$$

Substituting these in Eqs. (15) and (16), one gets

$$\psi = -\tfrac{1}{4}\log(x/x_0), \tag{39}$$

where x_0 is an integration constant.

The solution given by (36)–(39) has singularities at $x = 0$ and as $x \to \infty$. Further, for large enough values of x, ξ is positive and hence z is timelike. In particular, if z is an angle coordinate (as in the cylindrically symmetric case), this would give rise to closed timelike lines.

4. NONNULL FIELDS

If the inequality holds in (31),

$$1 + 4ab > 0, \tag{31'}$$

the field is nonnull. Writing

$$\theta = \int \frac{du}{u + v^2}, \tag{40}$$

we have, using (27) and (30),

$$k^2 \exp(\mu\theta) = \frac{au + b + (1/2a) - (\mu/2a)}{au + b + (1/2a) + (\mu/2a)} \tag{41}$$

or

$$u = \frac{\mu}{2a^2}\frac{1 + \xi^2}{1 - \xi^2} - \frac{\mu^2}{4a^2} - \frac{1}{4a^2}, \tag{42}$$

where

$$\xi = k\exp(\mu\theta/2), \tag{43}$$

k being an arbitrary constant of integration, and

$$v = \frac{\mu}{2a}\frac{1 + \xi^2}{1 - \xi^2} - \frac{1}{2a}. \tag{44}$$

Thus, from (22), (23), (42), and (44), we have

$$l = \frac{a}{\mu}x\frac{(1 - \xi^2)}{\xi}, \tag{45}$$

$$f = \frac{1}{2a}x\frac{(1 + \xi^2)}{\xi} - \frac{\mu x}{4a}\frac{(1 - \xi^2)}{\xi} - \frac{1}{4a\mu}\frac{x(1 - \xi^2)}{\xi}, \tag{46}$$

and

$$m = \tfrac{1}{2}x\frac{(1 + \xi^2)}{\xi} - \frac{1}{2\mu}\frac{x(1 - \xi^2)}{\xi}. \tag{47}$$

Substituting Eq. (40) in (26), we get

$$\frac{4}{x}(lA^2 + fB^2) = \frac{d}{dx}\left(x\frac{d\theta}{dx}\right). \tag{48}$$

Using now (29), (45), and (46) in (48), we get

$$\frac{d^2\theta}{d\rho^2} = \frac{4B^2}{ak}\exp\left(\rho - \frac{\mu\theta}{2}\right), \tag{49}$$

where

$$\rho = \log x. \tag{50}$$

The first integral of (49) is

$$(d\theta/d\rho)^2 = C^2 - (4\beta^2\mu/ak)e^{\nu}, \tag{51}$$

where

$$\nu = \rho - (\mu\theta/2) \tag{52}$$

and C is an integration constant. Further integration yields

$$e^{-\nu/2} = \frac{1}{2}\left(\frac{x}{x_0}\right)^{\pm C/2} + \frac{2B^2\mu}{akC^2}\left(\frac{x}{x_0}\right)^{\mp C/2}. \tag{53}$$

Thus, using (51) and (43), we get

$$\xi = \frac{kx}{4}\left[\left(\frac{x}{x_0}\right)^{C/2} + \frac{4B^2\mu}{akC^2}\left(\frac{x}{x_0}\right)^{-C/2}\right]^2. \tag{54}$$

Equivalently we may write

$$\xi = \frac{4B^2x}{C^2}\frac{\mu}{a}\cosh^2\left(\frac{C}{2}\log\frac{x}{x_0}\right), \tag{55}$$

where x_0 is a constant different from the x_0 of Eq. (54).

The only component of the metric tensor that is still to be determined is $e^{2\psi}$. Substituting for f, l, m, and A from (45), (46), (47), and (29) in (16), we get

$$x\psi_{11} + \psi_1 = B^2\mu/a\xi \tag{56}$$

so that, substituting from (55), we have the first integral

$$\psi_1 = \frac{C_1}{x} + \frac{C}{2x}\tanh\left(\frac{C}{2}\log\frac{x}{x_0}\right). \tag{57}$$

Hence finally

$$e^{\psi} = C_2x^{C_1}\cosh\left(\frac{C}{2}\log\frac{x}{x_0}\right), \tag{58}$$

where C_1, C_2 are arbitrary constants of integration. Substituting from (58) in (15), we get

$$C_1 = C^2/4. \tag{59}$$

Thus the solution is finally given by (45), (46), (47), (58), and (59). It has apparently singularities at $x = 0$ and as $x \to \infty$.

The condition that the metric tensor can be diagonalized is that m/f must be constant. From (22), (27), (29), and (30) this would require that

$$b = 0 \tag{60}$$

and either

$$\mu = +1, \qquad A = -B/a \tag{61}$$

or

$$\mu = -1, \qquad A = 0. \tag{62}$$

Equations (46) and (47) give in either case

$$f = m/a, \tag{63}$$

so that, with the transformation $t' = t + aZ$, the metric tensor becomes diagonal. Also, with (61), $(F^{41})'$ vanishes and

$$(F^{31})' = -\frac{B}{a(-g)^{\frac{1}{2}}}, \tag{64}$$

while, with (62),

$$(F^{31})' = F^{31} = 0$$

and

$$(F^{41})' = B/(-g)^{\frac{1}{2}}. \tag{65}$$

Thus the two cases correspond, respectively, to an axial magnetic field and radial electric field in the cylindrically symmetric case. These are, of course, equivalent under a duality rotation to an axial electric field and radial magnetic field, respectively.

The metric in the two cases is

$$l = x/\xi, \qquad f = \xi, \tag{66}$$

and

$$f = x/\xi, \qquad l = x\xi, \tag{67}$$

where we have absorbed "a" by a suitable transformation and ψ and ξ are given by (58) and (55), respectively. These solutions already occur in the literature.[2]

[2] W. B. Bonnor, Proc. Phys. Soc. **A66**, 145 (1953).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Local Operator Algebras in the Presence of Superselection Rules

PAUL FEDERBUSH*
Department of Mathematics, University of Michigan, Ann Arbor, Michigan

(Received 8 August 1967)

An introductory study is made of the relation between the local algebra of observables associated to a region of space–time and the algebra of all operators (disregarding superselection rules) associated to the same region of space–time. With a reasonable axiomatization of the problem and assuming the local algebra of observables is a factor of type I, II_∞, or III, the relation between the two algebras is specified up to unitary equivalence in terms of algebraic invariants.

The objects of study in this paper are some of the operator algebras arising in quantum field theory. The general problem to be considered is the relation between the local algebra of observables associated to a region in space–time and the local algebra of all operators constructed from the fields (disregarding superselection rules) associated to the same region in space–time. The situation is abstracted to a mathematical system with several axioms, axioms that seem reasonable from a physical point of view. Some insight is gained into the factor types of the local algebras—assuming indeed that the algebras are factors.

We proceed to describe the mathematical system of interest. First we presume a discrete group, the superselection group, called G. We denote elements of G as g_i, $i \in I$; I is an indexing set, g_0 the identity. To each g_i we associate a Hilbert space H_i; H_0 is associated to g_0. The total Hilbert space H is the orthogonal direct sum of the H_i:

$$H = \oplus H_i. \tag{1}$$

O is a * algebra in $B(H)$, the algebra of bounded operators. O_i, $i \in I$, are vector subspaces of O considered as a vector space over the complex numbers:

$$O - \sum_i O_i. \tag{2}$$

This is taken to mean that O is the algebra generated by finite sums of elements from the subspaces O_i. O_0 is the algebra of local observables.

Using the notation

$$g_i g_j = g_{(i,j)},$$

we assume

$$O_i O_j \subset O_{(i,j)}, \tag{3}$$

$$O_i H_j \subset H_{(i,j)}. \tag{4}$$

* Sloan Foundation Fellow.

and either

$$\mu = +1, \qquad A = -B/a \tag{61}$$

or

$$\mu = -1, \qquad A = 0. \tag{62}$$

Equations (46) and (47) give in either case

$$f = m/a, \tag{63}$$

so that, with the transformation $t' = t + aZ$, the metric tensor becomes diagonal. Also, with (61), $(F^{41})'$ vanishes and

$$(F^{31})' = -\frac{B}{a(-g)^{\frac{1}{2}}}, \tag{64}$$

while, with (62),

$$(F^{31})' = F^{31} = 0$$

and

$$(F^{41})' = B/(-g)^{\frac{1}{2}}. \tag{65}$$

Thus the two cases correspond, respectively, to an axial magnetic field and radial electric field in the cylindrically symmetric case. These are, of course, equivalent under a duality rotation to an axial electric field and radial magnetic field, respectively.

The metric in the two cases is

$$l = x/\xi, \qquad f = \xi, \tag{66}$$

and

$$f = x/\xi, \qquad l = x\xi, \tag{67}$$

where we have absorbed "a" by a suitable transformation and ψ and ξ are given by (58) and (55), respectively. These solutions already occur in the literature.[2]

[2] W. B. Bonnor, Proc. Phys. Soc. **A66**, 145 (1953).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Local Operator Algebras in the Presence of Superselection Rules

PAUL FEDERBUSH*
Department of Mathematics, University of Michigan, Ann Arbor, Michigan

(Received 8 August 1967)

An introductory study is made of the relation between the local algebra of observables associated to a region of space–time and the algebra of all operators (disregarding superselection rules) associated to the same region of space–time. With a reasonable axiomatization of the problem and assuming the local algebra of observables is a factor of type I, II_∞, or III, the relation between the two algebras is specified up to unitary equivalence in terms of algebraic invariants.

The objects of study in this paper are some of the operator algebras arising in quantum field theory. The general problem to be considered is the relation between the local algebra of observables associated to a region in space–time and the local algebra of all operators constructed from the fields (disregarding superselection rules) associated to the same region in space–time. The situation is abstracted to a mathematical system with several axioms, axioms that seem reasonable from a physical point of view. Some insight is gained into the factor types of the local algebras—assuming indeed that the algebras are factors.

We proceed to describe the mathematical system of interest. First we presume a discrete group, the superselection group, called G. We denote elements of G as g_i, $i \in I$; I is an indexing set, g_0 the identity. To each g_i we associate a Hilbert space H_i; H_0 is associated to g_0. The total Hilbert space H is the orthogonal direct sum of the H_i:

$$H = \oplus H_i. \tag{1}$$

O is a * algebra in $B(H)$, the algebra of bounded operators. O_i, $i \in I$, are vector subspaces of O considered as a vector space over the complex numbers:

$$O - \sum_i O_i. \tag{2}$$

This is taken to mean that O is the algebra generated by finite sums of elements from the subspaces O_i. O_0 is the algebra of local observables.

Using the notation

$$g_i g_j = g_{(i,j)},$$

we assume

$$O_i O_j \subset O_{(i,j)}, \tag{3}$$

$$O_i H_j \subset H_{(i,j)}. \tag{4}$$

* Sloan Foundation Fellow.

Observation 1: $O_i^* = O_{i^{-1}}$, with the notation, $(g_i)^{-1} = g_{i^{-1}}$.

Observation 2: The weak closures of the O_i still satisfy (3), (4), and the relation in Observation 1.

We assume from now on that O_i and O are weakly closed, reinterpreting (2) to mean that O is the weakly closed algebra generated by finite sums of the type indicated.

We further require

$$\dim H_i = N, \quad i \in I, \tag{5}$$

that is, the cardinality of bases of all the H_i are the same. The final essential property we require is the existence of a vector ψ contained in H_0 that is a separator and generator for O. In the physical system this is the vacuum vector.

Observation 3: O_0 is a sub-w^* algebra of O.

Observation 4: ψ is a separator and generator for O_0 restricted to H_0.

A system satisfying all the above conditions will be called a superselection paired local algebra system, an SPLA. We wish to study the relation between O_0 and O in an SPLA. In general, the structure is not yet rigid enough to be able to say anything simple. We will restrict ourselves to the case when O_0 is a factor, in which case a structure theorem is accessible. First we give two examples to show that it is possible for either O_0 or O to be a factor without the other being a factor.

Example 1: G contains only two elements, g_0 and g_1. H_0 and H_1 are two-dimensional with bases (a_1, a_2) and (b_1, b_2), respectively. In the basis (a_1, a_2, b_1, b_2) for H we choose

$$\psi = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \tag{6}$$

O_0 consists of diagonal matrices of the form

$$\left(\begin{array}{cc|cc} A & & O & \\ & B & & \\ \hline & & A & \\ O & & & B \end{array}\right), \tag{7}$$

and O_1 the matrices of the form

$$\left(\begin{array}{cc|cc} & O & & A \\ & & B & \\ \hline B & & & \\ & A & O & \end{array}\right). \tag{8}$$

This provides an SPLA with O a factor and O_0 not.

Example 2: Let G be a group with N elements and each H_i a Hilbert space of dimension M^2. We view H as the tensor product $H_A \times H_B$, H_A and H_B of dimensions M^2 and N, respectively. We associate each element of some basis $\{h_i\}$ of H_B with an element of G, $h_i \leftrightarrow g_i$, so that each g_i acts as a unitary operator in H_B, $g_i h_j = h_{(i,j)}$. We let R be a factor, algebraically isomorphic with the full $M \times M$ matrix algebra, in H_A with ϕ as separator and generator. Finally, we identify

$$\begin{aligned} O_i &\sim R \times g_i, \\ H_i &\sim H_A \times h_i, \\ \psi &\sim \phi \times h_0. \end{aligned} \tag{9}$$

This is an SPLA with O_0 a factor and O not a factor, unless G has only one element. Note that this construction generalizes in an obvious way to the case when N and M may be infinite, and R may be any factor with a separator and generator. We will see later that this is a canonical situation and will be referred to as a trivial SPLA.

The following result provides the basic device employed in this paper to study an SPLA.

Lemma: Let $a_1, a_2, \cdots, a_s$ be in $B(H)$, F a factor of infinite type in $B(H)$. Suppose

$$a_i^* a_j \in F \quad \text{all} \quad i \text{ and } j. \tag{10}$$

Then there exists a partial isometry U, with initial projector onto the span of the ranges of the a_i, such that

$$Ua_i \in F \quad \text{all } i. \tag{11}$$

Proof: By induction, assume the lemma true for $s - 1$ (the case $s = 0$ is trivial), so that

$$Va_i \in F, \quad 1 \leq i \leq s - 1, \tag{12}$$

with V a partial isometry onto the span of the ranges of a_i, $1 \leq i \leq s - 1$. We claim $Va_s \in F$, as

$$a_i^* V^* V a_s = a_i^* a_s \in F. \tag{13}$$

Therefore

$$\frac{1}{\epsilon + \sum_{i=1}^{s-1} (Va_i a_i^* V^*)} \sum_{1}^{s-1} (Va_i a_i^* V^*) Va_s \in F. \tag{14}$$

In the limit $\epsilon \xrightarrow[st.]{} 0^+$ we see that $Va_s \in F$, since F is strongly closed, and the initial projector of V is onto the ranges of the a_i, $i \leq s-1$:

$$\begin{aligned} a_s &= V^*Va_s + (1 - V^*V)a_s, \\ Va_s &= V(V^*V)a_s, \\ a_s^*a_s &= a_s^*V^*Va_s + a_s^*(1 - V^*V)a_s, \end{aligned} \tag{15}$$

and therefore

$$a_s^*(1 - V^*V)a_s \in F. \tag{16}$$

We write the polar decomposition of $(1 - V^*V)a_s$:

$$(1 - V^*V)a_s = V'^*h. \tag{17}$$

We note that $h \in F$. Therefore

$$V'(1 - V^*V)a_s \in F. \tag{18}$$

Using the fact that F is a factor of infinite type, we can move the final projectors of V and V' to be orthogonal and construct a new U, so that $Ua_i \in F$ for all i and its intial projector is onto the span of the ranges of the a_i, $1 \leq i \leq s$. ($U = W_1V + W_2V'$ for suitable isometries W_1 and W_2 in F.)

We see that the above lemma is also true for a countably infinite number of a_i instead of a finite number. A variation of the above argument will prove the result for factors of type I_N with a separator and generator. We do not consider factors of type II_1.

From each O_i select a sequence of elements a_j^i, $j = 1, 2, \cdots, N$, N the dimension of the H_i, such that $a_j^i\psi$ is a basis for H_i. We assume all our Hilbert spaces are separable at this point. Denote O_0 restricted to H_0 by F, O_i restricted to H_0 by F_i, and a_j^i restricted to H_0 by b_j^i. We note that

$$b_j^{i*}b_k^i \in F, \quad \text{all} \quad j \text{ and } k, \tag{19}$$

as

$$O_{(i)^{-1}}O_i \subset O_0. \tag{20}$$

Therefore we are in the situation of the lemma. Find a partial isometry V_i with

$$V_ib_j^i \in F. \tag{21}$$

If one is in the I_N factor situation, it is trivial that V_i is unitary. We proceed to show that if the factor is of infinite type, V_i may be picked unitary.

The final projector of V_i is a projection in F. If this projection is equivalent to 1, then it is clear that a new V_i may be picked to be unitary. In general, let P_i be this projection in F. Using V_i, we identify H_i with P_iH_0, and in this new basis for each H_i:

$$b_j^i \in P_iF, \quad \text{all} \quad i \text{ and } j. \tag{22}$$

(We carelessly identify operators in F_i and F that have identical matrix form in the present basis system for H_0 and H_i.) Also

$$F_i \subset P_iF.$$

Now $O_iO_0 \subset O_i$ implies $F_iF \subset F_i$, and therefore F; is a right ideal of F. Since the b_j^i span P_iH_0 for fixed i, the weak closure of F_i, F_i^{cL}, equals P_iF. Calling C_i, O_0 restricted to H_i, it follows from $O_iO_{(i)^{-1}} \subset O_0$ that C_i contains P_iFP_i, and from $O_{(i)^{-1}}O_0O_i \subset O_0$ that $C_i \subset F$ and therefore $C_i = P_iFP_i$. (C_i is weakly closed as it is $E'O_0E'$ with E' a projection commuting with O_0.) F is represented by C_i, uniformly continuously, and thus either F is isomorphic (as a C^* algebra) to P_iFP_i or else F modulo a nontrivial uniformly closed two-sided ideal is isomorphic to P_iFP_i. The second situation is impossible and the first situation implies $P_i \sim 1$. Thus we may pick V_i unitary and work in a basis with $F_i \subset F$, $F_i^{cL} = F$.

Using the axiom $O_iO_j \subset O_{(i,j)}$, we easily show that, in the present bases, O_i restricted to H_i (and weakly closed, for every i and j) is equal to F.

If we now let X be the restriction of an operator X in O_0 to H_0, then the image of X in H_i is $R_iXR_i^{-1}$ with R_i unitary, $R_iFR_i^{-1} = F$, and R_i is independent of X. To establish this we observe that since F has a separator and generator, all isomorphisms are spatial; and the image of O_0 in each H_i is faithful, since $F/I = F$ with I a nontrivial uniformly closed two-sided ideal is impossible. F and the R_i completely describe O_0.

It is important now to establish that each F_i is weakly closed. Consider an increasing sequence of projectors in F_i, weakly approaching 1. Since $O_0O_iO_0 \subset O_i$, such a sequence is easy to come by. We claim the corresponding sequence of elements in O_i (call them T_1^i, T_2^i, $\cdots$) converges weakly, and therefore converges to an element 1_i in O_i. From the fact that $O_iO_{(i)^{-1}} \subset O_0$ and $O_{(i)^{-1}}O_i \subset O_0$, it follows that each of the elements T_k^i restricted to each H_i is a partial isometry. Further, from $O_iO_0 \subset O_i$ it follows that the partial isometries in the sequences are increasing; that is, T_r^i restricted to the range of the initial projector of T_{r-1}^i restricted to H_k, agrees with T_{r-1}^i restricted to H_k. It easily follows that T_r^i converges to an element 1_i and F_i is weakly closed. We may now study O_i.

The essential observation here is that knowledge of the single element 1_i in O_i that restricts to the identity in H_0 is sufficient to characterize O_i, as all of O_i is generated by multiplying 1_i by the algebra O_0. It must be recalled that since $\psi \in H_0$ separates O_i, the restriction of an operator in O_i to H_0 determines the operator.

A simple calculation is required to determine the conditions on S_j^i, the restriction of 1_i to H_j.

The first relation we obtain is

$$1_i X^{\sim} = (R_i^{-1} X R_i)^{\sim} 1_i, \tag{23}$$

where $(R_i^{-1}XR_i)^{\sim}$ is the element of O_0 that restricts to $R_i^{-1}XR_i$ in H_0. This relation holds because these two elements have the same restriction to H_0. From (23) upon examining the restriction of each side of the equation to H_k it follows that

$$X = (R_k)^{-1}(S_k^i)^{-1} R_{(i,k)} (R_i)^{-1} X R_i (R_{(i,k)})^{-1} S_k^i R_k. \tag{24}$$

Since this holds for all X in F, we arrive at

$$S_k^i \cong R_{(i,k)} (R_i)^{-1} (R_k)^{-1}, \tag{25}$$

where the equivalence means that the two sides of the equation, both unitary operators, induce the same automorphism of F. There is only one other set of conditions on the S_k^i to guarantee that the F, R_i, and S_k^i determine an SPLA. This set of relations is obtained by requiring that

$$1_i 1_j \in O_0 1_{(i,j)}. \tag{26}$$

This equation becomes

$$R_{(k,l)} (R_k)^{-1} (S_j^i)^{-1} R_k (R_{(k,l)})^{-1} S_{(j,l)}^i S_l^j = S_l^k. \tag{27}$$

Equations (25) and (27) are the algebraic relations to study for constructing an SPLA from F and the group G. Equation (25) merely states that the map $g_i \to (R_i)^{-1}$ induces a homomorphism from G into the group of outer automorphisms of F. Equation (27) gives the conditions on the S_j^i in terms of the homomorphism just mentioned. Equation (25) determines S_k^i up to a complex number of modulus one. Thus (27) may be looked upon as conditions on these complex factors, since this equation automatically holds up to a complex factor. [It is easily checked that the two sides of (27) induce the same automorphism of F.] Unfortunately, we have not been able to find whether every homomophism of G into Out (F) has at least one solution to (27). However, if there is one solution to (27), and calling this solution S_j^i, we look for all other solutions by writing $(S_j^i)' = \lambda_j^i S_j^i$ with λ_j^i a complex number of modulus one, $(S_j^i)'$ any other solution. The λ_j^i must satisfy the algebraic relations given by (27). Interestingly enough, these state that λ_j^i is a two cocycle of the usual cochain complex of the group G with coefficients in the unimodular complex numbers, the circle group. By changing the basis for the H_i by a complex factor (multiplying each basis element by a complex factor, the factors being constant in each H_i), λ_j^i is changed by a coboundary. Therefore the solutions of (27)—if they exist at all—are in one-to-one correspondence with $H^2(G, T^1)$. Unfortunately, this correspondence is not canonical as we have derived it.

Theorem: In an SPLA with $O_0 = F$, F a factor of type I, II_∞, or III, and group G, O is determined up to unitary equivalence (under basis changes in each H_i separately) by a homomorphism of G into Out (F) and an element of $H^2(G, T^1)$.

It is attractive to conjecture that if, instead of assuming F to be a factor, it is assumed that the restrictions of O_0 to all the H_i are isomorphic and faithful, the same result follows with $H^2(G, T^1)$ replaced by $H^2(G, Z)$, Z the center of F.

Corollary: If F is of type I and G is Abelian with one generator, then the SPLA is trivial; and if G contains more than one element, O is therefore not a factor.

The corollary follows from the fact that, under the stated conditions, Out (F) and $H^2(G, T^1)$ are both trivial.

The corollary may have some implications for the factor-type problem in quantum field theory.

We conclude by remarking that the results in this paper are clearly far from definitive. The obvious examples of several free boson and fermion fields with various choices of superselection rules should be computed to relate the O_0 and O in these examples. The free boson field is studied in detail elsewhere.[1] The result here should also be related to the deeper results of Borchers[2] that also relate the rings of different regions of space to each other.

[1] H. Araki, J. Math. Phys. **5**, 1 (1964); **4**, 1343 (1963).
[2] H. J. Borchers, Commun. Math. Phys. **1**, 281 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Generating Functions for the Exact Solution of the Transport Equation. I*

I. K. ABU-SHUMAYS AND E. H. BAREISS
Argonne National Laboratory, Argonne, Illinois

(Received 19 October 1967)

Generating functions are used as analytic tools for the transformation of the integro-differential transport equations to partial differential equations. The transformations are constructed so that their inverses are known. Therefore, the problem of solving the integro-differential equations is reduced to the more familiar problem of solving partial differential equations. Part I of this series introduces the general method in detail for slab geometry. The general solution is given in analytic form, and the one-to-one correspondence with Case's method of elementary solutions is demonstrated. The results are extended to slab geometry without axial symmetry. Part II extends the method to the general time-dependent, anisotropic case for slab geometry. Parts III and IV treat the stationary and time-dependent problems in one and three space dimensions, introduce in addition the method of characteristics, and include numerical results.

1. INTRODUCTION

To obtain approximate solutions to the Boltzmann equation, various methods of spherical harmonics and discrete ordinates have been used extensively. Each time an approximation seems not sufficiently accurate, the entire problem has to be done over again with a higher order of approximation. In addition, good error estimates are very tedious to obtain. Since the existence of the general solution of the Boltzmann transport equation in closed form has already been proved,[1–3] it is natural and desirable to ask for the solutions of the spherical-harmonics representation. If we can determine a formula for the exact Fourier coefficients in the expansion in spherical harmonics of the solution of the Boltzmann equation, the errors in the corresponding coefficient functions of the usual "truncated" spherical-harmonics methods (as explicitly described by Davison,[4] Weinberg and Wigner,[5] and others) may be obtained by comparison.

In this paper, appropriate generating functions for the exact expansion coefficients in the spherical-harmonics method will be obtained. Then, instead of the infinite system of differential equations for the expansion coefficients of the spherical harmonics, only one partial differential equation for the generating function is needed for the solution of the problem.

Paper I of this series is restricted to the stationary isotropic transport equation in general slab geometry. We solve the equations for the generating functions by the method of separation of variables, and in doing so, obtain formulas for the spectral representation of the generating functions. In particular, in the case of slab geometry with axial symmetry, the spectra are shown to be identical to those of Refs. 1 and 3. In the important case of slab geometry but *without* axial symmetry of the angular distribution, a very simple representation of the general solution results.

When taken at the origin of the parameter, the generating function furnishes the *exact* scalar flux, corresponding to the first expansion term in the spherical harmonics method (which is only an approximation in the usual methods where the system of differential equations was truncated). Similar statements involving derivatives hold for the current and for higher moments.

In the subsequent parts of the series, the time-dependent and anisotropic transport equations for different geometries will be treated. The method of characteristics will be applied, the formulation of well-posed transport problems will be discussed, and numerical results will be presented.

The transport equation under consideration takes the form

$$\frac{\partial \psi}{\partial t} + v\boldsymbol{\Omega} \cdot \nabla \psi + v\sigma\psi = cv\sigma \int_{\{\Omega'\}} f(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}')\psi(r, \boldsymbol{\Omega}', t) d_2\Omega', \quad (1.1)$$

where

$\psi = \psi(r, \boldsymbol{\Omega}, t)$ is the directional flux or angular distribution;
t is the time;
r is the position vector;
v is a constant velocity in the direction of $\boldsymbol{\Omega}$;

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] E. H. Bareiss, J. Math. Anal. Appl. **13**, 53 (1966).

[2] E. H. Bareiss and I. K. Abu-Shumays, "On the Structure of Isotropic Transport Operators in Space," in *Proceedings of Symposia in Applied Mathematics, Vol. 20* (Am. Math. Soc., Providence, R.I., to be published).

[3] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).

[4] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).

[5] A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chain Reactors* (The University of Chicago Press, Chicago, Ill., 1958).

c is the net number of neutrons produced per collision;
f is the scattering function;
σ is the total macroscopic cross section at velocity v;
$\mathbf{\Omega}$ is a unit vector with θ, ϕ its spherical coordinates (θ colatitude, ϕ longitude);
$d_2\Omega \equiv d\mu\, d\phi \equiv \sin\theta\, d\theta\, d\rho$ is the surface element on the unit sphere;
$\nabla \equiv$ grad operates with respect to r only.

2. STATIONARY ISOTROPIC TRANSPORT IN SLAB GEOMETRY WITH AXIAL SYMMETRY

The purpose of this section is to illustrate the technique of applying generating functions to the transport equation. The results of Case[3] and Bareiss[1] are obtained. The expressions containing the Dirac δ function and the principal values follow from well-established relations in analysis.

The stationary transport equation in its simplest form is

$$\mu \frac{\partial}{\partial z}\psi(z,\mu) + \psi(z,\mu) = \frac{c}{2}\int_{-1}^{+1}\psi(z,\mu)\,d\mu. \quad (2.1)$$

The angular distribution $\psi(z,\mu)$ is expanded in Legendre polynomials as follows:

$$\psi(z,\mu) = \sum_{n=0}^{\infty}\frac{2n+1}{2}f_n(z)P_n(\mu), \quad (2.2)$$

where

$$\left.\begin{aligned} f_n(z) &= \int_{-1}^{1}\psi(z,\mu)P_n(\mu)\,d\mu \\ f_n &= 0, \quad \text{for} \quad n<0 \end{aligned}\right\}. \quad (2.3)$$

Direct application of the relation

$$(2n+1)\mu P_n(\mu) = (n+1)P_{n+1}(\mu) + nP_{n-1}(\mu) \quad (2.4)$$

to Eqs. (2.1)–(2.3) yields the recurrence relation

$$(n+1)f'_{n+1}(z) + nf'_{n-1}(z) + (2n+1)f_n(z) - c\delta_{n0}f_0(z) = 0. \quad (2.5)$$

A generating function for the expansion coefficients $f_n(z)$ is now defined as

$$\chi(z,\zeta) = \sum_{n=0}^{\infty}\zeta^n f_n(z) \quad (\zeta = \xi + i\eta). \quad (2.6)$$

Applying (2.6) to (2.3) gives the following transformation of the angular distribution $\psi(z,\mu)$ to the generating function $\chi(z,\zeta)$:

$$\chi(z,\zeta) = \int_{-1}^{+1}\psi(z,\mu)\sum_{n=0}^{\infty}\zeta^n P_n(\mu)\,d\mu = \int_{-1}^{+1}\frac{\psi(z,\mu)\,d\mu}{(1-2\mu\zeta+\zeta^2)^{\frac{1}{2}}}. \quad (2.7)$$

The condition $|\zeta| < 1$ is necessary and sufficient for the convergence of the series

$$\sum_{n=0}^{\infty}\zeta^n P_n(\mu)\,d\mu = (1-2\mu\zeta+\zeta^2)^{-\frac{1}{2}} \quad (\mu\in[-1,1]) \quad (2.8)$$

and therefore the convergence of $\chi(z,\zeta)$ is asserted for $|\zeta| < 1$. Further, the right-hand side of Eq. (2.8) and hence the right-hand side of Eq. (2.7) is analytic in ζ for $|\zeta| < 1$ and $\mu\in[-1,1]$. Thus $\chi(z,\zeta)$ is analytic in ζ for any $|\zeta| < 1$. We note that

$$|\chi(z,\zeta)| \le \int_{-1}^{+1}\frac{|\psi(z,\mu)|\,d\mu}{|1-2\mu\zeta+\zeta^2|^{\frac{1}{2}}} < \infty, \quad (2.9)$$

since (a) $\int_{-1}^{+1}|\psi(z,\mu)|\,d\mu$ is required to exist if ψ belongs to the solution space and (b) $|1-2\mu\zeta+\zeta^2| \neq 0$ for real $\mu\in[-1,1]$ and $|\zeta| < 1$. Conversely, if

$$\lim_{\zeta\to\zeta_0}\chi(z,\zeta)\to\infty \quad (|\zeta_0| < 1),$$

it follows from (2.9) that condition (a) could not be satisfied. Thus we have:

Theorem 2.1: A necessary condition for $\chi(z,\zeta)$, as defined in (2.6), to be a generating function for the Fourier coefficients $f_n(z)$ of the Legendre expansion of the solutions $\psi(z,\mu)$ of (2.1) is the analyticity of $\chi(z,\zeta)$ in ζ for the entire interior of the unit circle $|\zeta| < 1$.

Applying the definition of the generating function Eq. (2.6) to the recurrence relations of Eq. (2.5), or, equivalently, applying the integral transform (2.7) to (2.1) gives

$$(1+\zeta^2)\frac{\partial^2}{\partial z\partial\zeta}\chi + \zeta\frac{\partial}{\partial z}\chi + 2\zeta\frac{\partial}{\partial\zeta}\chi + \chi(z,\zeta) - c\chi(z,0) = 0. \quad (2.10)$$

Equation (2.10) is of hyperbolic type. By Theorem 2.1, only its solutions which are analytic in ζ for $|\zeta| < 1$ are generating functions for the Fourier coefficients f_n of (2.2). After (2.10) has been solved for χ, the particle density $\rho(z)$ is given by $\chi(z,\zeta)$ at $\zeta = 0$,

$$\chi(z,0) = \int_{-1}^{+1}\psi(z,\mu)\,d\mu = \rho(z), \quad (2.11)$$

as follows immediately from (2.7).

The expansion coefficients $f_n(z)$ of the generating function, if desired, are obtained by evaluating the derivatives at $\zeta = 0$, because

$$f_n(z) = \left(\frac{1}{n!}\frac{\partial^n}{\partial\zeta^n}\chi(z,\zeta)\right)_{\zeta=0}. \quad (2.12)$$

After these remarks we turn to the derivation of the general analytic solution for the generating function $\chi(z, \zeta)$ of (2.10). The method of separation of variables is used. Let an expansion mode of $\chi(z, \zeta)$ be

$$\chi_\nu(z, \zeta) = X_\nu(z)G_\nu(\zeta) \tag{2.13}$$

and denote $G_\nu(0)$ by $G_{0\nu}$.

Then,

$$\chi_\nu(z, 0) = f_{0\nu}(z) = G_{0\nu}X(z).$$

Substituting (2.13) in (2.10) and dividing by $\chi_\nu G_\nu$, we get

$$\frac{1}{X_\nu}\frac{d}{dz}X_\nu\left(\zeta + (1+\zeta^2)\frac{1}{G_\nu}\frac{d}{d\zeta}G_\nu\right) + \frac{2\zeta}{G_\nu}\frac{d}{d\zeta}G_\nu + 1 - \frac{c\,G_{0\nu}}{G_\nu} = 0,$$

which establishes the separation of X_ν and G_ν. Now let

$$\frac{1}{X_\nu(z)}\frac{d}{dz}X_\nu = -\frac{1}{\nu} \quad (\nu \neq 0), \tag{2.14a}$$

where, for the moment, ν is an arbitrary complex constant. Then

$$X_\nu(z) = e^{-z/\nu}. \tag{2.14b}$$

The constant of integration will be absorbed in G_ν, which must satisfy

$$[1 - 2\nu\zeta + \zeta^2]\frac{d}{d\zeta}G_\nu + [\zeta - \nu]G_\nu + \nu c\,G_{0\nu} = 0. \tag{2.15}$$

If $[1 - 2\nu\zeta + \zeta^2] \neq 0$, Eq. (2.15) has the integrating factor $[1 - 2\nu\zeta + \zeta^2]^{-\frac{1}{2}}$ and may be written as

$$\frac{d}{d\zeta}[(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}G_\nu] = -\frac{\nu c\,G_{0\nu}}{(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}. \tag{2.16}$$

For $\nu^2 \neq 1$ (no double root of $1 - 2\nu\zeta + \zeta^2$), the last equation can be integrated directly from zero to ζ to give

$$G_\nu(\zeta) = \frac{G_{0\nu}\left\{1 - c\nu\log\dfrac{\zeta - \nu + [1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}}{1 - \nu}\right\}}{(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}, \tag{2.17a}$$

$$= \frac{G_{0\nu}\left[1 + c\nu\log\dfrac{\nu - \zeta + (1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}{1 + \nu}\right]}{(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}. \tag{21.7b}$$

The case $\nu^2 = 1$ can be treated directly from Eq. (2.16). For $\nu = \pm 1$,

$$\frac{d}{d\zeta}[(1 \mp \zeta)G_{\pm 1}] = \mp\frac{c\,G_{0,\pm 1}}{1 \mp \zeta},$$

and integrating from 0 to ζ yields

$$G_{\pm 1}(\zeta) = \frac{G_{0,\pm 1}}{1 \mp \zeta}[1 + c\log(1 \mp \zeta)], \quad |\zeta| < 1. \tag{2.18}$$

We note that Eq. (2.18) is the limit of Eqs. (2.17a, b) as $\nu \to \pm 1$.

The next step is to examine the values of ν for which the above solutions $G_\nu(\zeta)$ of (2.17) and (2.18) are analytic for $|\zeta| < 1$. The problem is solved if the singularities of $G_\nu(\zeta)$ are known. The numerator of (2.17) is singular if the logarithm is singular. It is easily verified that this can occur only if $\nu^2 = 1$. But, by (2.18), $G_{\pm 1}(\zeta)$ is analytic. Thus we can turn to analyze the zeros of the denominator in (2.17). Instead, we analyze

$$1 - 2\nu\zeta + \zeta^2, \tag{2.19}$$

the zeros of which can be represented as

$$\zeta_\pm = \nu \pm (\nu - 1)\left(\frac{\nu + 1}{\nu - 1}\right)^{\frac{1}{2}} = \left(\left(\frac{\nu + 1}{\nu - 1}\right)^{\frac{1}{2}} \pm 1\right)\Big/\left(\left(\frac{\nu + 1}{\nu - 1}\right)^{\frac{1}{2}} \mp 1\right). \tag{2.20a}$$

Clearly

$$\zeta_+\zeta_- = 1, \quad |\zeta_-| \leq 1 \quad \text{for all } \nu. \tag{2.20b}$$

Equation (2.19) suggests we consider the cases (a) when $\nu \in [-1, 1]$ and (b) when $\nu \notin [-1, 1]$.

(a) For ν real and in the interval $[-1, 1]$, ζ_+ and ζ_- are a pair of conjugate complex numbers. Thus $\zeta_\pm = e^{\pm i\theta}$, where $\cos\theta = \nu$ $(0 \leq \theta \leq \pi)$, and obviously

$$|\zeta_+| = |\zeta_-| = 1. \tag{2.21}$$

Hence, for all $|\zeta| < 1$, (2.19) does not have zeros and the integration leading to the solutions (2.17) and (2.18) is legitimate. The solutions (2.17) and (2.18) are analytic for $|\zeta| < 1$ and hence by Theorem 2.1 are factors of the generating function χ_ν for all $\nu \in [-1, +1]$. For $\nu \to -1$, (2.17a) is used; for $\nu \to +1$, (2.17b) is used to facilitate computation.

(b) For $\nu \notin [-1, 1]$, one of the zeros of (2.19), namely ζ_-, lies inside the unit circle $|\zeta| = 1$. We require $G_\nu(\zeta)$ of (2.17) to be always finite for $|\zeta| < 1$, even when $\zeta = \zeta_-$. Since in this case the denominator of (2.17) vanishes, we require the numerator to vanish also. Thus

$$\lim_{\zeta\to\zeta_-} G_{0\nu}\left[1 - c\nu\log\frac{\zeta - \nu + (1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}{1 - \nu}\right] = 0. \tag{2.22}$$

Since $G_{0\nu} \neq 0$, this equation reduces to

$$1 - c\nu\log\frac{\zeta_- - \nu}{1 - \nu} = 1 - c\nu\log\left(\frac{\nu + 1}{\nu - 1}\right)^{\frac{1}{2}} = 0 \tag{2.22'}$$

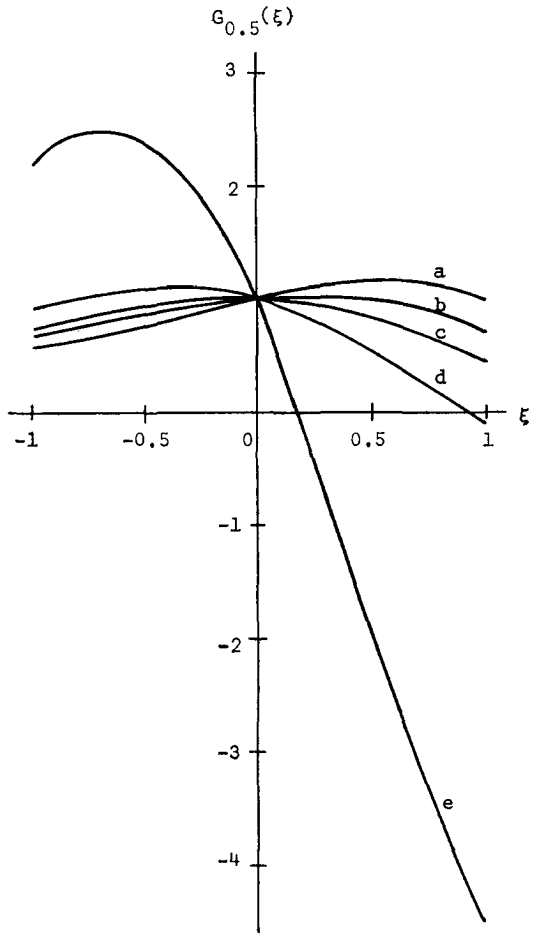


FIG. 1. The function $G_\nu(\xi)$ for $\nu = 0.5 \in [-1, 1]$ and various values of c: (a) $c = 0$, (b) $c = 0.5$, (c) $c = 1$, (d) $c = 2$, (e) $c = 10$.

or

$$\frac{c\nu}{2} \log \frac{\nu + 1}{\nu - 1} = 1. \tag{2.22''}$$

Therefore, in the cut ν plane, only the roots of (2.22'') are admissible. It is known that (2.22'') has exactly two roots, either both real ($c < 1$) or both pure imaginary ($c > 1$) (see Ref. 1). They are denoted by $\pm\nu_0$. That $G_{\pm\nu_0}(\zeta)$ is bounded follows from (2.17a) and (2.22'):

$$\lim_{\zeta\to\zeta_-} G_\nu(\zeta)$$

$$= \frac{G_{0\nu} \lim_{\zeta\to\zeta_-} \left(1 - c\nu \log\left(\frac{\zeta - \nu}{1 - \nu} + \frac{(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}{1 - \nu}\right)\right)}{(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}$$

$$= \frac{G_{0\nu}\left(0 - \lim_{\zeta\to\zeta_-} c\nu \log\left(1 + \frac{[1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}}{\zeta_- - \nu}\right)\right)}{[1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}}$$

$$= G_{0\nu} \frac{c\nu}{\nu - \zeta_-} = c\nu\, G_{0\nu}/(\nu - 1) \left(\frac{\nu + 1}{\nu - 1}\right)^{\frac{1}{2}} < \infty \quad (\nu = \pm\nu_0). \tag{2.23}$$

Furthermore, it can be shown from Eqs. (2.15) and (2.20a) that

$$\lim_{\zeta\to\zeta_-} \frac{d}{d\zeta} G_\nu(\zeta) < \infty, \quad \nu = \pm\nu_0.$$

Thus for $\nu = \pm\nu_0$, $G(\zeta)$ is continuous and has a continuous derivative in ζ for all $|\zeta| < 1$ and hence is analytic in ζ in this region.

Some representative graphs of $G_\nu(\zeta)$ for real argument $\zeta = \xi$ are given in Figs. 1 and 2.

The set of all admissible ν will be called the spectral set S:

$$S = S_P \cup S_c,$$
$$S_P = \{\nu : \nu = \pm\nu_0\}, \tag{2.24}$$
$$S_c = \{\nu : 1 \geq \nu^2 > 0\}.$$

Attention should be given to the fact that $G_\nu(\zeta)$ exists for $\nu \in [-1, 1]$; however, $\nu = 0$ is excluded by (2.14a). Thus, if we restrict $G_\nu(\zeta)$, i.e., the solution (2.17), such that $\nu \in S$, the general solution of the generating function $\chi(z, \zeta)$ of (2.10) is

$$\chi(z, \zeta) = a_+ G_{\nu_0}(\zeta) e^{-z/\nu_0} + a_- G_{-\nu_0}(\zeta) e^{+z/\nu_0} + \int_{(S_c)} G_\nu(\zeta) e^{-z/\nu}\, d\alpha(\nu), \tag{2.25}$$

where χ is analytic in $|\zeta| < 1$ and $|z| < \infty$. If we normalize $G_\nu(0) = G_{0\nu} = 1$, we observe from (2.11)

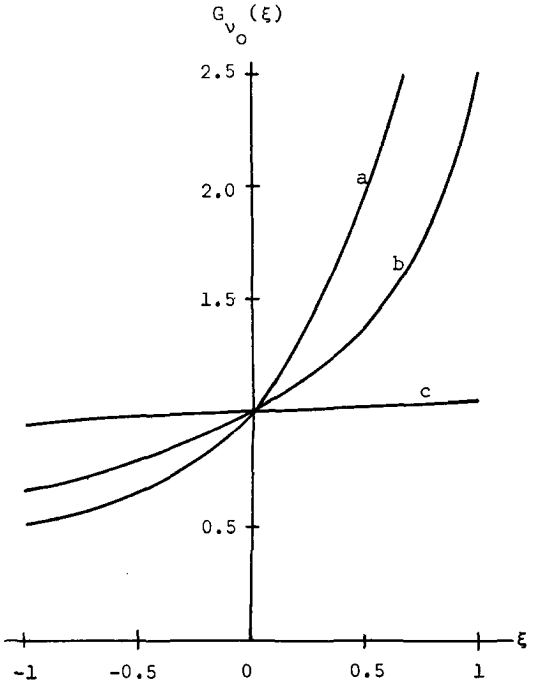


FIG. 2. The function $G_{\nu_0}(\xi)$ for different values of ν_0: (a) $\nu_0 = 1$, $c = 0$, (b) $\nu_0 = 1.0444$, $c = 0.5$, (c) $\nu_0 = 5.8$, $c = 0.99$.

that

$$\rho(z) = \chi(z, 0) = a_+ e^{-z/\nu_0} + a_- e^{z/\nu_0} + \int_{(S_c)} e^{-z/\nu}\, d\alpha(\nu), \tag{2.26}$$

which is a well-known result.[1,3]

3. INVERSION OF THE GENERATING FUNCTION AND THE GENERAL SOLUTION OF (2.1)

The solution of the stationary isotropic transport equation for slab geometry (2.1) is given by (2.2), (2.12), and (2.24). The explicit representation for the solution $\psi(z, \mu)$ in expansion modes will now be developed step by step.

For every admissible ν, the corresponding mode of the generating function

$$\chi_\nu(z, \zeta) = e^{-z/\nu} G_\nu(\zeta) \tag{3.1}$$

is a solution to (2.10). The corresponding expansion mode $\psi_\nu(z, \mu)$ to the solution of the transport equation is then given by

$$\begin{aligned}\psi_\nu(z, \mu) &= \sum_{n=0}^{\infty} \frac{2n+1}{2} \left(\frac{1}{n!} \frac{\partial^n}{\partial \zeta^n} \chi_\nu(z, \zeta)\right)_{\zeta=0} P_n(\mu) \\ &= e^{-z/\nu} \sum_{n=0}^{\infty} \frac{2n+1}{2} \left(\frac{1}{n!} \frac{\partial^n}{\partial \zeta^n} G_\nu(\zeta)\right)_{\zeta=0} P_n(\mu) \\ &\equiv e^{-z/\nu} \Phi_\nu(\mu) \quad (\nu \in S).\end{aligned} \tag{3.2}$$

It follows by inserting (3.2) into (2.1) that $\Phi_\nu(\mu)$ satisfies the equation

$$\left(1 - \frac{\mu}{\nu}\right) \Phi_\nu(\mu) = \frac{c}{2} \int_{-1}^{1} \Phi_\nu(\mu)\, d\mu, \tag{3.3}$$

and, for admissible ν, can be called an eigenfunction of (3.3) corresponding to the eigenvalue ν. Consequently, the general solution of the transport equation for slab geometry (2.1) [corresponding to (2.24) for the generating function $\chi(z, \zeta)$] takes the form

$$\psi(z, \mu) = a_+ \Phi_{\nu_0}(\mu) e^{-z/\nu_0} + a_- e^{z/\nu_0} \Phi_{-\nu_0}(\mu) + \int_{(S_c)} e^{-z/\nu} \Phi_\nu(\mu)\, d\alpha(\mu), \tag{3.4}$$

where the Φ_ν, $\Phi_{\pm\nu_0}$ are defined by (3.2) and satisfy (3.3). As in Refs. 1 and 3, they obey orthogonality relations which can be used to determine $a_\pm$, $A(\nu)$.

Now we relate the eigenfunctions $\Phi_{\nu_0}(\mu)$ to the corresponding eigenfunctions given in Refs. 1 and 3. We assume that $G_\nu(0) = G_{0\nu} = 1$. From (2.17b) and (2.22″),

$$\begin{aligned}G_{\nu_0}(\zeta) &= \left[1 + c\nu_0 \log \frac{\nu_0 - \zeta + (1 + \zeta^2 - 2\nu_0\zeta)^{\frac{1}{2}}}{\nu_0 + 1}\right] \times [1 + \zeta^2 - 2\nu_0\zeta]^{-\frac{1}{2}} \\ &= c\nu_0 \log \frac{\nu_0 - \zeta + (1 + \zeta^2 - 2\nu_0\zeta)^{\frac{1}{2}}}{(\nu_0^2 - 1)^{\frac{1}{2}}} \times (1 + \zeta^2 - 2\nu_0\zeta)^{-\frac{1}{2}}.\end{aligned}$$

Since $\nu_0 \notin (-1, 1)$, it follows from Hobson (Ref. 6, p. 69) that for $|\zeta| < |\zeta_-|$,

$$G_{\nu_0}(\zeta) = c\nu_0 \sum_{k=0}^{\infty} \zeta^k Q_k(\nu_0), \tag{3.5}$$

where $Q_k(\nu_0)$ are Legendre functions of the second kind. Interesting (3.5) into (3.2) yields

$$\Phi_{\nu_0}(\mu) = c\nu_0 \sum_{n=0}^{\infty} \frac{2n+1}{2} Q_n(\nu_0) P_n(\mu). \tag{3.6a}$$

But this is

$$\Phi_{\nu_0}(\mu) = \frac{c\nu_0}{2} \frac{1}{\nu_0 - \mu}, \tag{3.6b}$$

as shown by Hobson (Ref. 6, p. 62). Similarly,

$$\Phi_{-\nu_0}(\mu) = \frac{c\nu_0}{2} \frac{1}{\nu_0 + \mu}. \tag{3.7}$$

Equations (3.6) and (3.7) are identical to those given by Case[3] and Bareiss.[1]

It will now be demonstrated that the eigenfunctions $\Phi_\nu(\mu)$ for $\nu \in [-1, 1]$ defined by (3.2) are equivalent to those given in Refs. 1 and 3. Equation (2.17) can be written in the form

$$G_\nu(\zeta) = \left(c\nu \log \frac{\nu - \zeta + [1 + \zeta^2 - 2\nu\zeta]^{\frac{1}{2}}}{(1 - \nu^2)^{\frac{1}{2}}} + \left\{1 - \frac{c\nu}{2} \log \frac{1+\nu}{1-\nu}\right\}\right) \times [1 + \zeta^2 - 2\nu\zeta]^{-\frac{1}{2}}, \quad \nu \in [S_c].$$

It follows from Hobson (Ref. 6, §37ff.), that this is

$$G_\nu(\zeta) = c\nu \sum_{k=0}^{\infty} \zeta^k Q_k(\nu) + \left\{1 - \frac{c\nu}{2} \log \frac{1+\nu}{1-\nu}\right\} \sum_{k=0}^{\infty} \zeta^k P_k(\nu). \tag{3.8}$$

Hence, the corresponding Φ_ν is given by

$$\Phi_\nu(\mu) = c\nu \sum_{n=0}^{\infty} \frac{2n+1}{2} Q_n(\nu) P_n(\mu) + \left\{1 - \frac{c\nu}{2} \log \frac{1+\nu}{1-\nu}\right\} \sum_{n=0}^{\infty} \frac{2n+1}{2} P_n(\nu) P_n(\mu). \tag{3.9}$$

[6] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, Cambridge, 1931).

But for $\nu \in (-1, 1)$, by definition (Ref. 6, p. 52),

$$Q_n(\nu) = \lim_{\epsilon \to 0} \tfrac{1}{2}\{Q_n(\nu + i\epsilon) + Q_n(\nu - i\epsilon)\}.$$

Consequently, using Heine's expansion (Ref. 6, p. 62) again, the first sum in (3.9) is

$$\sum_{n=0}^{\infty}(2n+1)Q_n(\nu)P_n(\mu) = \lim_{\epsilon \to 0} \frac{1}{2}\left\{\frac{1}{\nu + i\epsilon - \mu} + \frac{1}{\nu - i\epsilon - \mu}\right\} = P\frac{1}{\nu - \mu}, \quad \nu \in [-1, 1], \quad (3.10)$$

where P has been inserted to indicate Cauchy's principal value under integration. The last sum in (3.9) is a δ function, since (see, e.g., Ref. 7, p. 19)

$$\sum_{n=0}^{\infty}\frac{2n+1}{2}P_n(\nu)P_n(\mu) \equiv \delta(\nu - \mu). \quad (3.11)$$

With the definition

$$\lambda(\nu) \equiv 1 - \frac{c\nu}{2}\log\frac{1+\nu}{1-\nu}, \quad \nu \in [-1, 1], \quad (3.12)$$

Eqs. (3.9)–(3.12) lead to the representation

$$\Phi_\nu(\mu) = \frac{c\nu}{2}P\frac{1}{\nu - \mu} + \lambda(\nu)\delta(\nu - \nu), \quad \nu \in [-1, 1]. \quad (3.13)$$

This result (3.13) coincides with the result given by Case.[3] The advantage of this derivation of (3.13) is that principal value and δ function are brought in direct relation to well-established analytical results rather than postulated.

A function space for which the eigenfunctions $\Phi_\nu(\mu)$ form a complete basis for arbitrary functions of μ has been given in Ref. 1. There it is also shown that the orthogonality relation

$$\int_{-1}^{1}\Phi_\nu(\mu)\Phi_{\nu'}(\mu)\mu\,d\mu = \nu M_\nu\delta(\nu - \nu') \quad \left[M_\nu = \lambda^2(\nu) + \left(\frac{\pi c\nu}{2}\right)^2\right] \quad (3.14)$$

taken with proper care, can be used to get the expansion coefficients in (3.4). An alternate way to get the expansion coefficients, presented below, supplies additional properties of the eigenfunctions Φ_ν. From (3.3), (3.14), and the normalization $\int_{-1}^{1}\Phi_\nu(\mu)\,d\mu = 1$, it follows that

$$\int_{-1}^{1}\Phi_\nu(\mu)\Phi_{\nu'}(\mu)\,d\mu = \frac{c}{2} + M_\nu\delta(\nu - \nu'). \quad (3.15a)$$

[7] E. Madelung, *Die Mathematischen Hilfsmittel des Physikers* (Springer-Verlag, Berlin, 1953).

Similarly,

$$\int_{-1}^{+1}\Phi_{\nu'}\Phi_{\pm\nu_0}\,d\mu = \frac{c}{2} \pm M_{\nu_0}\delta_{\nu',\pm\nu_0} \quad (\nu' = \nu, \pm\nu_0), \quad (3.15b)$$

where

$$M_{\nu_0} = \frac{c}{2}[1 + \nu_0^2(c-1)](\nu_0^2 - 1)^{-1}$$

and the normalization

$$\int_{-1}^{+1}\Phi_{\nu_0}\,d\mu = 1$$

is valid. Now, given any function $\psi(\mu)$ of μ in the form

$$\psi(\mu) = a_+\Phi_{\nu_0}(\mu) + a_-\Phi_{-\nu_0}(\mu) + \int_{-1}^{1}A(\nu)\Phi_\nu(\mu)\,d\nu, \quad (3.16)$$

it follows from the normalizations

$$\int_{-1}^{1}\psi(\mu)\,d\mu = a_+ + a_- + \int_{-1}^{1}A(\nu)\,d\nu \quad (3.17)$$

and by (3.15a, b), (3.16), and (3.17) that

$$\int_{-1}^{1}\Phi_{\nu_0}(\mu)\psi(\mu)\,d\mu = \frac{c}{2}\int_{-1}^{1}\psi(\mu)\,d\mu + M_{\nu_0}a_+, \quad (3.18a)$$

$$\int_{-1}^{1}\Phi_{-\nu_0}(\mu)\psi(\mu)\,d\mu = \frac{c}{2}\int_{-1}^{1}\psi(\mu)\,d\mu - M_{\nu_0}a_-, \quad (3.18b)$$

$$\int_{-1}^{1}\Phi_\nu(\mu)\psi(\mu)\,d\mu = \frac{c}{2}\int_{-1}^{1}\psi(\mu)\,d\mu + M_\nu A(\nu). \quad (3.18c)$$

Equations (3.18) can serve to determine the expansion coefficients for a given function $\psi(\mu)$.

As an illustration we evaluate $A(\nu)$ for $\nu = 0$. Let $\psi(\mu)$ be a given flux $\psi(z, \mu)$ at $z = 0$. By (3.12), (3.13), and (3.14) we obtain for $\Phi_0(\mu)$ and M_ν at $\nu = 0$

$$\Phi_0(\mu) = \delta(\mu), \quad M_0 = 1. \quad (3.19)$$

Inserting these values into (3.18c) yields

$$\psi(0) = \frac{c}{2}\int_{-1}^{1}\psi(\mu)\,d\mu + A(0). \quad (3.20)$$

But whenever $\psi(z, \mu)$ has a continuous derivative at $z = 0$ (as is the case when $z = 0$ is in the interior of a reactor cell), it follows from (2.1) that

$$\psi(0) = \frac{c}{2}\int_{-1}^{1}\psi(\mu)\,d\mu. \quad (3.21)$$

The difference between the last two equations gives

$$A(0) = 0. \quad (3.22)$$

Thus, in problems of practical nature, the domain of integration with respect to ν in (3.4) can be extended from S_c to $[-1, 1]$, since $d\alpha(\nu) = A(\nu)\,d\nu = 0$ for

$\nu = 0$. In other words, Riemannian integration, combined with Cauchy's principal value, can be applied instead of Lebesgue integration as was tacitly implied in (3.4).

4. STATIONARY ISOTROPIC TRANSPORT IN GENERAL PLANE GEOMETRY

We relax the requirement of cylindrical symmetry and demonstrate that the generating-function approach leads to a simple representation of the general solution. A proof of completeness is also presented. The stationary isotropic transport equation again has the form

$$\mu \frac{\partial}{\partial z}\psi(z,\mu,\phi) + \psi(z,\mu,\phi) = \frac{c}{4\pi}\int_{-1}^{1} d\mu' \int_{0}^{2\pi} d\phi' \psi(z,\mu',\phi'). \quad (4.1)$$

The angular distribution $\psi(z, \mu, \phi)$ will now be expanded in general spherical harmonics so that

$$\psi(z,\mu,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!} f_{nm}(z) Y_n^m(\mu,\phi), \quad (4.2)$$

where the spherical harmonics[6] are defined as

$$Y_n^m(\mu,\phi) \equiv e^{im\phi} P_n^m(\mu), \quad (4.3)$$

$P_n^m(\mu)$ being the associated Legendre polynomials defined in Ref. 6:

$$P_n^m(\mu) \equiv (-1)^m (1-\mu^2)^{m/2} \frac{d^m}{d\mu^m} P_n(\mu), \quad m \geq 0, \quad (4.4a)$$

$$P_n^{-m}(\mu) \equiv (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(\mu). \quad (4.4b)$$

They satisfy the recurrence relation[6]

$$(2n+1)\mu P_n^m(\mu) = (n-m+1)P_{n+1}^m(\mu) + (n+m)P_{n-1}^m(\mu). \quad (4.5)$$

Applying the orthogonality relation of the spherical harmonics,

$$\int_0^{2\pi} d\phi \int_{-1}^{1} d\mu Y_n^{m*} Y_{n'}^{m'} = \frac{4\pi}{2n+1}\frac{(n+m)!}{(n-m)!}\delta_{nn'}\delta_{mm'}, \quad (4.6)$$

to the expansion (4.2) yields

$$f_{nm}(z) = \int_0^{2\pi} d\phi \int_{-1}^{1} d\mu \psi(z,\mu,\phi) Y_n^{m*}(\mu,\phi), \quad (4.7a)$$

$$f_{nm}(z) \equiv 0 \quad \text{for} \quad n < |m|. \quad (4.7b)$$

If the angular distribution $\psi(z, \mu, \phi)$ is restricted to be real, the expansion (4.2), together with (4.3) and (4.4), implies that

$$f_{n,-m} = (-1)^m \frac{(n-m)!}{(n+m)!} f_{nm}^*. \quad (4.8)$$

Thus it is sufficient to solve for f_{nm} with $m \geq 0$, the remaining expansion coefficients f_{nm} being determined by (4.8).

Substituting the expansion (4.2) into the transport equation (4.1) and applying (4.5) and (4.6) yield the recurrence relation

$$(n+1-m)\frac{d}{dz}f_{n+1,m} + (n+m)\frac{d}{dz}f_{n-1,m} + (2n+1)f_{nm} - cf_{nm}\delta_{m0}\delta_{n0} = 0. \quad (4.9)$$

Since the second subscript m of f_{nm} is the same for all terms in (4.9), this set of equations can be solved separately for each value of m. In particular, for $m = 0$, (4.9) reduces to (2.5) for the axially symmetric component of the angular distribution, and has been solved in the previous sections.

For each m fixed and $m > 0$ we define a generating function

$$\chi_m(z,\zeta) \equiv \sum_{n=m}^{\infty} \zeta^n f_{nm}(z), \quad m > 0. \quad (4.10)$$

Applying (4.7) to (4.10) gives the following transformation of the angular distribution $\psi(z, \mu, \phi)$ to the generating function $\chi_m(z, \zeta)$:

$$\chi_m(z,\zeta) = \int_0^{2\pi} d\phi \int_{-1}^{1} d\mu \psi(z,\mu,\phi) e^{-im\phi} \sum_{n=m}^{\infty} \zeta^n P_n^m(\mu) = \frac{(-1)^m (2m)!\,\zeta^m}{2^m m!}\int_0^{2\pi} d\phi \times \int_{-1}^{1} d\mu \frac{\psi(z,\mu,\phi)e^{-im\phi}(1-\mu^2)^{m/2}}{(1-2\zeta\mu+\zeta^2)^{m+\frac{1}{2}}}. \quad (4.11)$$

The condition $|\zeta| < 1$ is necessary and sufficient for the convergence of the series[6]

$$\sum_{n=m}^{\infty} \zeta^n P_n^m(\mu) = \frac{(-\zeta)^m (2m)!\,(1-\mu^2)^{m/2}}{2^m m!\,(1-2\zeta\mu+\zeta^2)^{m+\frac{1}{2}}}. \quad (4.12)$$

Using the same arguments as in Sec. 2, we conclude that the admissible generating function $\chi_m(z, \zeta)$ must be analytic in ζ for $|\zeta| < 1$.

Multiplying (4.9) by ζ^n, summing over n from m to infinity, and using the definition (4.10) gives

$$(1+\zeta^2)\frac{\partial^2}{\partial z \partial \zeta}\chi_m + \left[(m+1)\zeta - \frac{m}{\zeta}\right]\frac{\partial}{\partial z}\chi_m + 2\zeta\frac{\partial}{\partial \zeta}\chi_m + \chi_m = 0, \quad m > 0. \quad (4.13)$$

The same equation is obtained by applying the integral transform in (4.11) to (4.1). The result (4.13)

is a hyperbolic equation, which can be solved, as was (2.10) of Sec. 2, by the method of separation of variables. Let

$$\chi_{m\nu}(z, \zeta) = X_{m\nu}(z)G_{m\nu}(\zeta). \tag{4.14}$$

Following the steps in Sec. 2, we get

$$X_{m\nu}(z) = e^{-z/\nu} \quad (\nu \neq 0) \tag{4.15}$$

and

$$[1 - 2\nu\zeta + \zeta^2]\frac{d}{d\zeta}G_{m\nu} + \left\{(m+1)\zeta - \frac{m}{\zeta} - \nu\right\}G_{m\nu} = 0. \tag{4.16}$$

Multiplying (4.16) by the integrating factor

$$\zeta^{-m}(1 - 2\nu\zeta + \zeta^2)^{m-\frac{1}{2}}$$

yields

$$\frac{d}{d\zeta}\left(\frac{[1 - 2\nu\zeta + \zeta^2]^{m+\frac{1}{2}}}{\zeta^m}G_{m\nu}\right) = 0. \tag{4.17}$$

Hence,

$$G_{m\nu}(\zeta) = \frac{\text{const} \cdot \zeta^m}{[1 - 2\nu\zeta + \zeta^2]^{m+\frac{1}{2}}}. \tag{4.18}$$

The substitution of (4.18) and (4.15) in (4.14) and expansions in powers of ζ indeed implies $f_{nm} = 0$ for $n < m$ as required by (4.7b). Furthermore, it can be shown that a necessary and sufficient condition for $G_{m\nu}(\zeta)$ of (4.18) to be continuous and analytic inside the unit circle $|\zeta| < 1$ is $\nu \in [-1, 1]$. Hence the $G_{m\nu}$'s are admissible only for $\nu \in [-1, 1]$. The general solution for the generating function $\chi_m(z, \zeta)$ of (4.10) is, in analogy with (2.25),

$$\chi_m(z, \zeta) = \int_{(S_c)} A_m(\nu)G_{m\nu}(\zeta)e^{-z/\nu}\,d\nu, \quad m > 0. \tag{4.19}$$

The solution of the transport equation (4.1) corresponding to (4.19) is obtained from (4.2) and (4.10) by expanding the generating function (4.19). From (4.14), (4.15), (4.18), and Hobson (Ref. 6, p. 105), it follows that

$$\begin{aligned}\chi_{m\nu}(z, \zeta) &= e^{-z/\nu}G_{m\nu}(\zeta) \\ &= \text{const} \cdot \frac{(-2)^m m!}{(2m)!}(1 - \nu^2)^{-m/2} \\ &\quad \times e^{-z/\nu}\sum_{n=m}^{\infty} P_n^m(\nu)\zeta^n.\end{aligned}$$

If we normalize $\chi_{m\nu}$ by letting

$$\text{const} \cdot \frac{(-2)^m m!}{(2m)!}(1 - \nu^2)^{-m/2} = 1,$$

we have simply

$$\chi_m = e^{-z/\nu}\sum_{n=m}^{\infty} P_n^m(\nu)\zeta^n. \tag{4.20}$$

Hence, from (4.2) and (4.10) it follows that

$$\begin{aligned}\psi_{m\nu}(z, \mu) &= e^{-z/\nu}\sum_{l=|m|}^{\infty}\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}P_l^m(\nu)P_l^m(\mu)e^{im\phi} \\ &= e^{-z/\nu}\delta(\nu - \mu)e^{im\phi} \\ &\equiv e^{-z/\nu}\Phi_{m,\nu}(\mu, \phi),\end{aligned} \tag{4.21}$$

where we define

$$\Phi_{m,\nu}(\mu, \phi) = e^{im\phi}\delta(\nu - \mu), \quad m \neq 0. \tag{4.21'}$$

Equation (4.21) was derived for $m > 0$. Using the relation (4.8) between $f_{n,-m}$ and f_{nm}, this result can be shown to hold also for $m < 0$. For $m = 0$ the set of eigenfunctions $\{\Phi_{0,\nu}\}$ is identical to that in Sec. 3 and consists of $\Phi_{\nu_0}(\mu)$, $\Phi_{-\nu_0}(\mu)$, and $\Phi_\nu(\mu)$ as given by (3.6b), (3.7), and (3.13), respectively. The general solution of the transport equation (4.1) is obtained by superposition, similar to (3.4), and is

$$\begin{aligned}\psi(z, \mu, \phi) &= a_+\Phi_{\nu_0}(\mu)e^{-z/\nu_0} + a_-\Phi_{-\nu_0}(\mu)e^{z/\nu_0} \\ &\quad + \int_{-1}^{+1} A(\nu)\Phi_\nu(\mu)e^{-z/\nu}\,d\nu \\ &\quad + \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty}\int_{-1}^{1} A_m(\nu)\Phi_{m,\nu}(\mu, \phi)e^{-z/\nu}\,d\nu.\end{aligned} \tag{4.22a}$$

As we have shown that $A(0) = 0$, we shall show at the end of this section that $A_m(0) = 0$, and therefore we can extend the integration over $-1 \leq \nu \leq 1$. Evaluating the integrals in (4.22a) and using (4.21') and the remark following it, we get the representation

$$\begin{aligned}\psi(z, \mu, \phi) &= a_+\Phi_{\nu_0}(\mu)e^{-z/\nu_0} + a_-\Phi_{-\nu_0}(\mu)e^{z/\nu_0} \\ &\quad + \int_{-1}^{1} A(\nu)\Phi_\nu(\mu)e^{-z/\nu}\,d\nu \\ &\quad + e^{-z/\mu}\sum_{\substack{m=-\infty \\ m\neq 0}}^{+\infty} A_m(\mu)e^{im\phi}.\end{aligned} \tag{4.22b}$$

Completeness of this representation in the extended function space can be established as follows. Any arbitrary function $\psi(\mu, \phi)$ which possesses a Fourier-series expansion in ϕ can be written in the form

$$\psi(\mu, \phi) = \sum_{m=-\infty}^{+\infty} u_m(\mu)e^{im\phi}. \tag{4.23a}$$

But it has already been established[1,3] that arbitrary functions of μ such as $u_0(\mu)$ can be expressed in terms of $\Phi_{\pm\nu_0}$ and Φ_ν; thus

$$u_0(\mu) = a_+\Phi_{\nu_0}(\mu) + a_-\Phi_{-\nu_0}(\mu) + \int_{-1}^{1} A(\nu)\Phi_\nu(\mu)\,d\nu. \tag{4.23b}$$

The representations (4.23) are equivalent to the general expression (4.22b) with $z = 0$. Thus any

function in the space defined in Ref. 1, and with an extended domain to include dependence on ϕ, has a unique and well-defined "initial-value" problem, whenever the function admits an expansion in a Fourier series in ϕ. Thus, the eigenfunctions $\Phi_\nu(\mu)$ as given in Ref. 3 do not represent a complete base for the solution space in the case of homogeneous isotropic slab geometry in the absence of axial symmetry of the solutions.

In addition to the relations (3.14), (3.15a), and (3.15b), the eigenfunctions in (4.22) satisfy the orthogonality relation

$$\int_0^{2\pi}\int_{-1}^{1}\Phi_{m,\nu}(\mu,\phi)\Phi^*_{m',\nu'}(\mu,\phi)\mu\,d\mu\,d\phi = 2\pi\nu M_{m,\nu}\delta(\nu-\nu')\delta_{m,m'} \quad (4.24)$$

for all $\nu, \nu' \in [-1, 1]$ and all integers m, $M_{m,\nu} = 1$ if $m \neq 0$. These orthogonality relations can be employed to determine the expansion coefficients of the representation (4.22) for the solution of the transport equation (4.1) in general slab geometry. In particular, we obtain from (4.22b) and (4.24) for fixed z, say $z = 0$, and given m for $A_m(\mu)$,

$$\int_0^{2\pi}\psi(0,\mu,\phi)e^{-im\phi}\,d\phi = 2\pi A_m(\mu). \quad (4.25)$$

If $\mu = 0$ and $\partial\psi/\partial z$ is continuous (as is the case in the interior of a reactor cell), we obtain from (4.1)

$$0 + \int_0^{2\pi}\psi(0,0,\phi)e^{im\phi}\,d\phi = \frac{c}{4\pi}\int_0^{2\pi}e^{im\phi}\cdot\rho(0)\,d\phi = 0, \quad (4.25')$$

where

$$\rho(z) = \int_{-1}^{+1}d\mu'\int_0^{2\pi}d\phi'\psi(z,\mu',\phi').$$

Use of (4.25′) in (4.25) yields

$$A_m(0) = 0.$$

Thus, the domain of integration in (4.22a) can indeed include $\nu = 0$. Q.E.D.

5. GREEN'S FUNCTION AND PLANE SOURCES

As an application of the expansion for general slab geometry, we derive the Green's function for a plane source and proceed to solve the transport equation for given source distributions. The Green's function ψ_g for a plane source at z_0 satisfies the equation

$$\mu\frac{\partial}{\partial z}\psi_g + \psi_g = \frac{c}{4\pi}\int_0^{2\pi}d\phi\int_{-1}^{1}d\mu\psi_g(z,\mu,\phi;z_0,\mu_0,\phi_0) + \frac{1}{4\pi}\delta(z-z_0)\delta(\mu-\mu_0)\delta(\phi-\phi_0). \quad (5.1)$$

Obviously, the angular distribution is discontinuous at z_0. Integrating Eq. (5.1) from $z_0 - \epsilon$ to $z_0 + \epsilon$ ($\epsilon > 0$) and taking the limit as $\epsilon \to 0$ yields

$$\mu[\psi_g(z_0+,\mu,\phi;z_0,\mu_0,\phi_0) - \psi_g(z_0-,\mu,\phi;z_0,\mu_0,\phi_0)] = \frac{1}{4\pi}\delta(\mu-\mu_0)\delta(\phi-\phi_0). \quad (5.2)$$

If we assume that there are no sources at infinity, the requirement that the solution ψ_g vanishes at $|z| \to \infty$, imposed on the expansion (4.22b), suggests the following representation:

$$\psi_g(z,\mu,\phi;z_0,\mu_0,\phi_0) = a_+\Phi_{\nu_0}(\mu)e^{-(z-z_0)/\nu_0} + \int_0^1 A(\nu)\Phi_\nu(\mu)e^{-(z-z_0)/\nu}\,d\nu + H(\mu)e^{-(z-z_0)/\mu}\sum_{\substack{m=-\infty\\ m\neq 0}}^{+\infty}A_m(\mu)e^{im(\phi-\phi_0)}, \quad z > z_0 \quad (5.3a)$$

$$= -a_-\Phi_{-\nu_0}(\mu)e^{(z-z_0)/\nu_0} - \int_{-1}^0 A(\nu)\Phi_\nu(\mu)e^{-(z-z_0)/\nu}\,d\nu - H(-\mu)e^{-(z-z_0)/\mu}\sum_{\substack{m=-\infty\\ m\neq 0}}^{+\infty}A_m(\mu)e^{im(\phi-\phi_0)}, \quad z < z_0, \quad (5.3b)$$

where H is the Heaviside function

$$H(\mu) = \begin{cases}1, & \text{for } \mu \geq 1,\\ 0, & \text{for } \mu < 0.\end{cases} \quad (5.4)$$

Substituting Eqs. (5.3a) and (5.3b) into Eq. (5.2) and simplifying, we get

$$\frac{1}{4\pi\mu}\delta(\mu-\mu_0)\delta(\phi-\phi_0) = a_+\Phi_{\nu_0}(\mu) + a_-\Phi_{-\nu_0}(\mu) + \int_{-1}^1 A(\nu)\Phi_\nu(\mu)\,d\nu + \sum_{\substack{m=-\infty\\ m\neq 0}}^{+\infty}A_m(\mu)e^{im(\phi-\phi_0)}. \quad (5.5)$$

For $m \neq 0$, multiplying Eq. (5.5) by $(1/2\pi)e^{-im\phi}$ and integrating over ϕ yields

$$\frac{1}{8\pi^2\mu}\delta(\mu-\mu_0) = A_m(\mu). \quad (5.6)$$

Thus, $A_m(\mu)$ is independent of m and ϕ_0. Integrating (5.5) over ϕ and dividing by 2π yields

$$\frac{1}{8\pi^2\mu}\delta(\mu-\mu_0) = a_+\Phi_{\nu_0}(\mu) + a_-\Phi_{-\nu_0}(\mu) + \int_{-1}^1 A(\mu)\Phi_\nu(\mu)\,d\nu. \quad (5.7)$$

The remaining expansion coefficients a_+, a_-, and $A(\nu)$ can be evaluated directly from (5.7) by using the

orthogonality relations (3.14). The result is

$$a_{\pm} = \frac{1}{\pm \nu_0 M_{\nu_0}} \int_{-1}^{1} \frac{\mu \Phi_{\pm\nu_0}(\mu)\delta(\mu - \mu_0)\, d\mu}{8\pi^2\mu} = \frac{1}{8\pi^2} \frac{\Phi_{\pm\nu_0}(\mu_0)}{\pm \nu_0 M_{\nu_0}} \tag{5.8a}$$

and

$$A(\nu) = \frac{1}{8\pi^2} \frac{\Phi_\nu(\mu_0)}{\nu M_\nu}. \tag{5.8b}$$

Since $A_m(\mu)$ of Eq. (5.6) is independent of m, ϕ, and ϕ_0, the summations in (5.3a) and (5.3b) simplify to

$$\frac{\delta(\mu - \mu_0)}{8\pi^2\mu} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} e^{im(\phi-\phi_0)} = \frac{\delta(\mu - \mu_0)}{8\pi^2\mu} [2\pi\delta(\phi - \phi_0) - 1]. \tag{5.9}$$

Hence ψ_g can be expressed as

$$\psi_g(z, \mu, \phi; z_0, \mu_0, \phi_0) = \frac{1}{8\pi^2}\left[\frac{\Phi_{\nu_0}(\mu_0)\Phi_{\nu_0}(\mu)e^{-(z-z_0)/\nu_0}}{\nu_0 M_{\nu_0}} + \int_0^1 \frac{\Phi_\nu(\mu_0)\Phi_\nu(\mu)e^{-(z-z_0)/\nu}}{\nu M_\nu}\, d\nu + e^{-(z-z_0)/\mu} \frac{H(\mu)\delta(\mu - \mu_0)}{\mu} \{2\pi\delta(\phi - \phi_0) - 1\}\right], \quad z > z_0, \tag{5.10a}$$

and

$$\psi_g(z, \mu, \phi; z_0, \mu_0, \phi_0) = -\frac{1}{8\pi^2}\left[\frac{\Phi_{-\nu_0}(\mu_0)\Phi_{-\nu_0}(\mu)e^{(z-z_0)/\nu_0}}{-\nu_0 M_{\nu_0}} + \int_{-1}^0 \frac{\Phi_\nu(\mu_0)\Phi_\nu(\mu)e^{-(z-z_0)/\nu}}{\nu M_\nu}\, d\nu + e^{-(z-z_0)/\mu} \frac{H(-\mu)\delta(\mu - \mu_0)}{\mu} \{2\pi\delta(\phi - \phi_0) - 1\}\right], \quad z < z_0. \tag{5.10b}$$

For a plane-source distribution density $S(z_0, \mu_0, \phi_0)$ the angular distribution takes the form

$$\psi_0(z, \mu, \phi) = \iiint S(z_0, \mu_0, \phi_0)\psi_g(z, \mu, \phi; z_0, \mu_0, \phi_0)\, d\mu_0\, d\phi_0\, dz_0. \tag{5.11}$$

In particular, for an isotropic plane source $S = \delta(z - z_0)$,

$$\psi_{is}(z, \mu, \phi) = \int_0^{2\pi}\int_{-1}^{1} \psi_g(z, \mu, \phi; z_0, \mu_0, \phi_0)\, d\mu_0\, d\phi_0 = \frac{1}{4\pi}\left[\frac{\Phi_{\nu_0}(\mu)e^{-|z-z_0|/\nu_0}}{\nu_0 M_{\nu_0}} + \int_0^1 \frac{\Phi_\nu(\mu)e^{-|z-z_0|/\nu}}{\nu M_\nu}\right] d\nu, \tag{5.12}$$

which is axially symmetric as expected. The corresponding particle density is

$$\rho_{is}(z) = \int_0^{2\pi}\int_{-1}^{1} \psi_{is}(z, \mu, \phi)\, d\mu\, d\phi = \frac{e^{-|z-z_0|/\nu_0}}{2\nu_0 M_{\nu_0}} + \int_0^1 \frac{e^{-|z-z_0|/\nu}}{2\nu M_\nu}\, d\nu. \tag{5.13}$$

These results (5.12) and (5.13) agree with the corresponding angular distribution and particle density given by Case [Ref. 3, Eqs. (53) and (54)].

ACKNOWLEDGMENTS

The authors acknowledge the fine research atmosphere in the Applied Mathematics Division of Argonne National Laboratory under the direction of J. W. Givens, Jr. They express their sincere thanks to Garrett Birkhoff and R. D. Richtmyer for their interest in these investigations. R. F. King carefully edited this report and Mrs. Doris Haight cheerfully transcribed the manuscript.

This paper is based on Argonne National Laboratory, Applied Mathematics Division Tech. Memo 141.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 10 OCTOBER 1968

Algorithm for Reducing Contracted Products of γ Matrices*

JOSEPH KAHANE
Applied Mathematics Department, Brookhaven National Laboratory, Upton, Long Island, New York

(Received 11 December 1967)

A construction algorithm is given which allows any partially contracted product of γ matrices to be reduced in one line to a simple canonical form involving only the uncontracted indices.

I. INTRODUCTION

Because both the facilities and techniques for experimental measurements of electrodynamic effects are improving, an ever increasing demand for doing higher-order electrodynamics calculations is being generated. The procedure of drawing a Feynman diagram, writing down the corresponding matrix element as an integration over the internal momenta, and converting to an integration over the Feynman parameters, increases in complexity as the order of the terms to be calculated increases. The problem of reducing this procedure has been the subject of much investigation and has met with considerable success. Chisholm,[1] Nakanishi,[2] and others[3] have presented rules that allow one to go directly from the Feynman diagram to integration over Feynman parameters. These rules can be understood in terms of a simple and elegant circuit-theory analogy,[4] which allows one immediately to write down the denominator of the integrand in fully reduced form. However, equivalent rules for writing the fully reduced form for the numerator of the integrand do not as yet exist. This latter problem is far more difficult because the complexities of the γ algebra reside in the numerator. The algorithm presented herein is a start toward the solution of this problem.

We shall be concerned with arbitrary products of Dirac matrices γ_μ in which some subsets of the indices are contracted in pairs. Contracted products of this kind arise in nearly all relativistic calculations involving Dirac particles.[5] The usual technique for reducing such expressions[6] is straightforward, namely the successive application of the identities given by Eqs. (A8) and (A9) of the Appendix. Unfortunately, such a technique can be quite cumbersome for higher-order calculations and it produces reduced expressions whose form can, in general, be further simplified. We shall elaborate on this point later.

In this paper we present a simple construction algorithm for reducing such contracted products to a simple canonical form. We define the problem and notation in Sec. II, the algorithm is presented in Sec. III, and Sec. IV is devoted to proving the algorithm.

II. THE PROBLEM

We start by defining the four Dirac γ matrices via their anticommutation relations:

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu}I \quad (\mu, \nu = 1, 2, 3, 4), \tag{2.1}$$

where $\delta_{\mu\nu}$ is the Kronecker delta function. Some of the more relevant results on γ matrices have been collected in an appendix for reference.[7] Equations (A8)–(A14) of this appendix are of particular relevance.

We are interested in reducing expressions of the form[8]

$$M(\sigma; A_0, A_1, A_2, \cdots, A_{2K}) = \sum_{\alpha_1,\alpha_2,\cdots,\alpha_{2K}} A_0\gamma_{\alpha_1}A_1\gamma_{\alpha_2}A_2 \cdots A_{2K-1}\gamma_{\alpha_{2K}}A_{2K} \times \delta_{\alpha_{\sigma_1}\alpha_{\sigma_2}}\delta_{\alpha_{\sigma_3}\alpha_{\sigma_4}} \cdots \delta_{\alpha_{\sigma_{2K-1}}\alpha_{\sigma_{2K}}}. \tag{2.2}$$

Here σ is a partition of $(1, 2, \cdots, 2K)$ into disjoint pairs. That is, $(\sigma_1, \sigma_2, \cdots, \sigma_{2K})$ is a permutation of $(1, 2, \cdots, 2K)$ satisfying

$$\begin{aligned} &\sigma_{2j-1} < \sigma_{2j} \quad (j = 1, 2, \cdots, K), \\ &\sigma_{2j-1} < \sigma_{2j+1} \quad (j = 1, 2, \cdots, K-1). \end{aligned} \tag{2.3}$$

The A_j are products of l_j factors of γ's:

$$A_j = \gamma_{\mu_1^j}\gamma_{\mu_2^j} \cdots \gamma_{\mu_{l_j}^j} \quad (j = 0, 1, 2, \cdots, 2K), \tag{2.4}$$

with free indices μ_i^j. We can, with no loss of generality, assume that

$$l_j \geq 1 \quad (j = 1, 2, \cdots, 2K-1), \tag{2.5}$$

since we can always replace the empty produce, i.e., $A_j = I$, by $A_j = \gamma_1\gamma_1$ (say). Finally, the summation indicated in (2.2) ranges from 1 to 4 for all the α_i.

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] J. S. R. Chisholm, Proc. Cambridge Phil. Soc. **48**, 2, 300 (1952).
[2] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **17**, 401 (1957).
[3] See there view article, N. Nakanishi, Suppl. Progr. Theoret. Phys. (Kyoto) **18**, 1 (1961), see. p. 4 for further references.
[4] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill Book Company, New York, 1965), Chap. 18.4.
[5] R. P. Feynman, Phys. Rev. **76**, 769 (1949).
[6] E. R. Caianiello and S. Fubini, Nuovo Cimento **9**, 1218 (1952).
[7] R. H. Good, Jr., Rev. Mod. Phys. **27**, 187 (1955).
[8] The factors A_0 and A_{2K} are, in fact, superfluous since they never get permuted in the reduction. Their inclusion, however, adds some convenience in proving the algorithm, as will be seen.

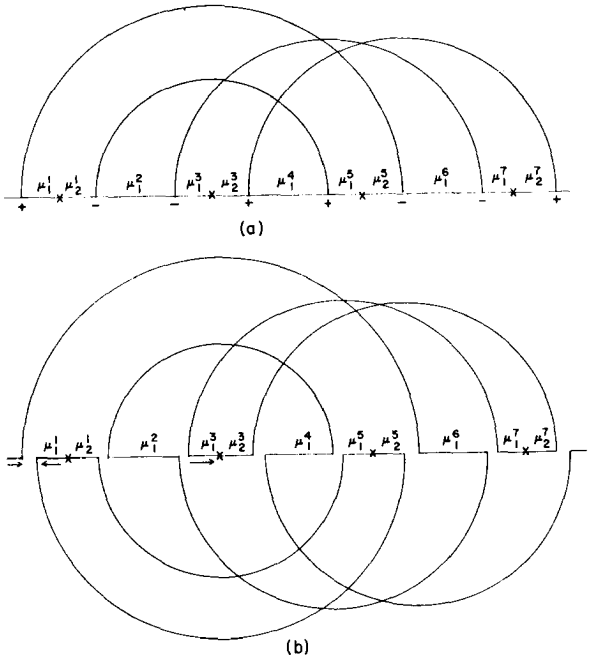


FIG. 1. (a) is the Feynman diagram for the expression

$$M = \sum_{\alpha_1,\alpha_2,\alpha_3,\alpha_4} \gamma_{\alpha_1} A_1 \gamma_{\alpha_2} A_2 \gamma_{\alpha_3} A_3 \gamma_{\alpha_4} A_4 \gamma_{\alpha_2} A_5 \gamma_{\alpha_1} A_6 \gamma_{\alpha_3} A_7 \gamma_{\alpha_4},$$

where $A_1 = \gamma_{\mu_1^1}\gamma_{\mu_2^1}$; $A_2 = \gamma_{\mu_1^2}$; $A_3 = \gamma_{\mu_1^3}\gamma_{\mu_2^3}$; $A_4 = \gamma_{\mu_1^4}$; $A_5 = \gamma_{\mu_1^5}\gamma_{\mu_2^5}$; $A_6 = \gamma_{\mu_1^6}$; $A_7 = \gamma_{\mu_1^7}\gamma_{\mu_2^7}$. The degree-two vertices are designated with an X and we have indicated the parity with a + or − sign below each paired vertex. (b) is the constructed graph g_M obtained from the Feynman diagram in (a). We see that the graph has three disconnected components. By following the rules of the algorithm, we can immediately write down

$$M = +2^4\{\mu_1^1, \mu_2^5, \mu_1^5, \mu_2^1\}\{\mu_1^3, \mu_2^3, \mu_1^7, \mu_1^7\}\gamma_{\mu_1^6}\gamma_{\mu_1^2}\gamma_{\mu_1^4}.$$

To every such expression M, we can uniquely associate a Feynman-type diagram as follows: Place $2K$ vertices on a horizontal base line, leaving the ends free. Connect them in pairs, with arcs drawn above the base line, in accordance with the pairings in σ. The jth vertex represents the factor γ_{α_j}, the jth line segment represents the factor A_j, and the two free ends represent the factors A_0 and A_{2K}, respectively. Next, divide the jth line segment into l_j subsegments by placing $l_j - 1$ division points (degree-two vertices) on it. On the two ends, however, place l_j division points (degree-two vertices), thereby dividing each into l_j subsegments plus a free end. Label these subsegments serially with the symbols $\mu_1^j, \mu_2^j, \cdots, \mu_{l_j}^j$. We illustrate an example of this in Fig. 1(a).

As we already indicated, any M can be reduced by the successive application of either Eq. (A8) or (A9), whichever is appropriate, to each of the K pairings. This technique leads, in general, to a result containing a large number of terms, the form and number of which depends on the order in which the contractions are carried out. One can hope to minimize the number of terms in the result by choosing an order which minimizes the use of (A9), but unfortunately no simple rule is known which gives such an optimal order. The algorithm to be presented here, on the other hand, carries out all the contractions simultaneously and in no way depends on any order. Furthermore, the form of the results are uniform and appear, in general, to have a simpler structure then is achievable by the former technique. We illustrate this by a simple example. Consider

$$M = \sum_{\alpha_1,\alpha_2} \gamma_{\alpha_1}\gamma_{\mu_1}\gamma_{\alpha_2}\gamma_{\mu_2}\gamma_{\mu_3}\gamma_{\alpha_1}\gamma_{\mu_4}\gamma_{\alpha_2}. \tag{2.6}$$

In Fig. 2(a) we give the Feynman diagram for M. Here there are only two pairings and it is immaterial, with regard to choosing an optimal order, which one is contracted first. Doing the summation over α_1 first, we get [by Eq. (A9)]:

$$M = 2\sum_{\alpha_2} (\gamma_{\mu_3}\gamma_{\mu_1}\gamma_{\alpha_2}\gamma_{\mu_2} + \gamma_{\mu_2}\gamma_{\alpha_2}\gamma_{\mu_1}\gamma_{\mu_3})\gamma_{\mu_4}\gamma_{\alpha_2}$$
$$= 4(\gamma_{\mu_3}\gamma_{\mu_1}(\gamma_{\mu_2}\gamma_{\mu_4} + \gamma_{\mu_4}\gamma_{\mu_2}) - \gamma_{\mu_2}\gamma_{\mu_4}\gamma_{\mu_3}\gamma_{\mu_1}). \tag{2.7}$$

Further identities are then required to see that this last expression can be simplified to give

$$M = 4\gamma_{\mu_4}\gamma_{\mu_2}\gamma_{\mu_3}\gamma_{\mu_1}, \tag{2.8}$$

whereas, according to our algorithm the graph of Fig. 2(b) is directly constructable from that of 2(a), and the result (2.8) can then be written down immediately by following the path, from left to right, in Fig. 2(b). This is clarified in the next section.

III. THE ALGORITHM

In this section we present a construction algorithm for the reduction of any M of the form (2.2). We define the notation [cf. Eq. (A10)]:

$$\{\rho_1, \rho_2, \cdots, \rho_{2n}\} = \gamma_{\rho_1}\gamma_{\rho_2}\cdots\gamma_{\rho_{2n}} + \gamma_{\rho_{2n}}\cdots\gamma_{\rho_2}\gamma_{\rho_1}. \tag{3.1}$$

Then the form of the reduced expression will be 2^K

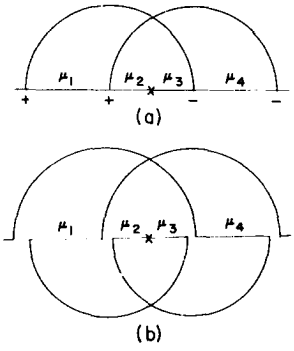


FIG. 2. (a) is the Feynman diagram for the example of Eq. (2.6). (b) is the constructed graph for this expression. The algorithm gives

$$M = +4\gamma_{\mu_4}\gamma_{\mu_2}\gamma_{\mu_3}\gamma_{\mu_1}.$$

multiplying a product of "braces" times a product of γ_μ's, all with an appropriate sign. The indices as they appear in the "braces" and in the γ_μ's are a permutation of all the free indices in M.

Let $M(\sigma; A_0, A_1, \cdots, A_{2K})$ be an expression of the form (2.2) and let $\mathcal{F}_M$ be the Feynman diagram corresponding to M. We construct a new graph $\mathcal{G}_M$ obtained from $\mathcal{F}_M$ as follows:

1. To every paired vertex, $v(i)$ $(i = 1, 2, \cdots, 2K)$, in $\mathcal{F}_M$, assign the parity

$$\epsilon(i) = \begin{cases} +1, & \text{if there are an even number of degree-two vertices preceding } v(i), \\ -1, & \text{if there are an odd number of degree-two vertices preceding } v(i). \end{cases} \tag{3.2}$$

2. For every paired vertex, $v(i)$; $(i = 1, 2, \cdots, 2K)$ in $\mathcal{F}_M$, place a new vertex $w(i)$ on the horizontal line very slightly to the right or left of $v(i)$, accordingly as $\epsilon(i)$ is $+1$ or -1, respectively.

3. Break the connections between $v(i)$ and $w(i)$ and draw K arcs below the horizontal line pairing all the $w(i)$ in accordance with σ.

We now have a graph $\mathcal{G}_M$ with only degree-two vertices and with two free ends. It is well known and obvious that such a graph, if it has k disconnected components, consists of $k - 1$ closed circuits and one chain beginning and ending on the two free ends. We illustrate two examples of this construction in Figs. 1 and 2.

The algorithm for writing down the reduced form of M is:

(a) Starting at the left free end in $\mathcal{G}_M$, we follow the chain and write the factor

$$\gamma_{\nu_1}\gamma_{\nu_2}\cdots\gamma_{\nu_p}\cdots, \tag{3.3}$$

where ν_p is the pth symbol μ_l^j (say) encountered in following the chain.

(b) For each of the $k - 1$ closed circuits, start with the left-most subsegment μ_1^j of that circuit. Traversing the circuit to the right or left along the starting subsegment, according as $\epsilon(j)$ is -1 or $+1$, respectively, where $\epsilon(j)$ is the parity of the vertex immediately to the left of the starting subsegment, we premultiply the previous factor by[9]

$$\{\rho_1\rho_2\cdots\rho_p\cdots\}. \tag{3.4}$$

where ρ_p is the pth symbol $\mu_l^{j'}$ (say) encountered in traversing the circuit and ρ_1 is of course μ_1^j. *Note:* Since, by (A13), any two "braces" commute, the order in which we write them down is immaterial.

[9] That every closed circuit has an even number of γ factors is guaranteed by Corollary 4.1.

(c) Finally, we multiply by

$$(-1)^{K-k+1}2^K. \tag{3.5}$$

Figure 2 is the example [Eq. (2.6)] of the previous section and the result (2.8) is now immediate.

IV. THE PROOF

In this section we present a proof for the construction algorithm. The proof rests on two lemmas.

Lemma 4.1: Let $M(\sigma; A_0, A_1, A_2, \cdots, A_{2K})$ be an expression of the form (2.2) for which

(a) all the pairings pair vertices of opposite parity, and

(b) none of the pairings overlap, i.e., the arcs in the Feynman diagram for M do not intersect.

Then the contruction algorithm, for this M, gives the correct result.

Proof: From hypothesis (b) it is clear that M must be of the form

$$M = A_0 m_1 A_{2r_1} m_2 A_{2r_2}\cdots A_{2r_{t-1}} m_t A_{2K}, \tag{4.1}$$

where $r_1 < r_2 < \cdots < r_t = K$. Here m_i $(i = 1, 2, \cdots, t)$ consists of $\rho_i = r_i - r_{i-1} (r_0 = 0)$ pairings and is of the form

$$m_i = \sum_{\beta_1,\beta_2,\cdots,\beta_{\rho_i}} \gamma_{\beta_1} B_1^i \gamma_{\beta_2} B_2^i \cdots \gamma_{\beta_{\rho_i}} B_{\rho_i}^i \gamma_{\beta_{\rho_i}} \times B_{\rho_{i+1}}^i \cdots \gamma_{\beta_2} B_{2\rho_i-1}^i \gamma_{\beta_1}, \tag{4.2}$$

where $B_k^i = A_{2r_{i-1}+k}$.

In Fig. 3 we draw the Feynman diagram for m_i and M.

Consider the innermost pairing $(i, i+1)$ of m_t, the right-most set of pairings. Since $\epsilon(i) = -\epsilon(i+1)$, there must be an even number of γ terms in A_i. That is,

$$A_i = \gamma_{\mu_1^i}\gamma_{\mu_2^i}\cdots\gamma_{\mu_{l_i}^i} \tag{4.3}$$

with l_i even. Using Eq. (A11), we have

$$\sum_{\alpha_i,\alpha_{i+1}} \gamma_{\alpha_i} A_i \gamma_{\alpha_{i+1}} \delta_{\alpha_i\alpha_{i+1}} = 2\{\mu_{l_i}^i, \mu_1^i, \mu_2^i, \cdots, \mu_{l_i-1}^i\}. \tag{4.4}$$

Suppose $\epsilon(i) = +1(-1)$. Then in $\mathcal{F}_M$ there are an even (odd) number of degree-two vertices to the left of $v(i)$. This implies that there are an even (odd) number of γ terms to the left of the factor (4.4). Hence, by making use of (A12–A14), we get

$$M = b_i M', \tag{4.5}$$

where

$$b_i = \begin{cases} 2\{\mu_1^i, \mu_{l_i}^i, \mu_{l_i-1}^i, \cdots, \mu_2^i\}, & \text{for } \epsilon(i) = +1, \\ 2\{\mu_1^i, \mu_2^i, \cdots, \mu_{l_i}^i\}, & \text{for } \epsilon(i) = -1, \end{cases} \tag{4.6}$$

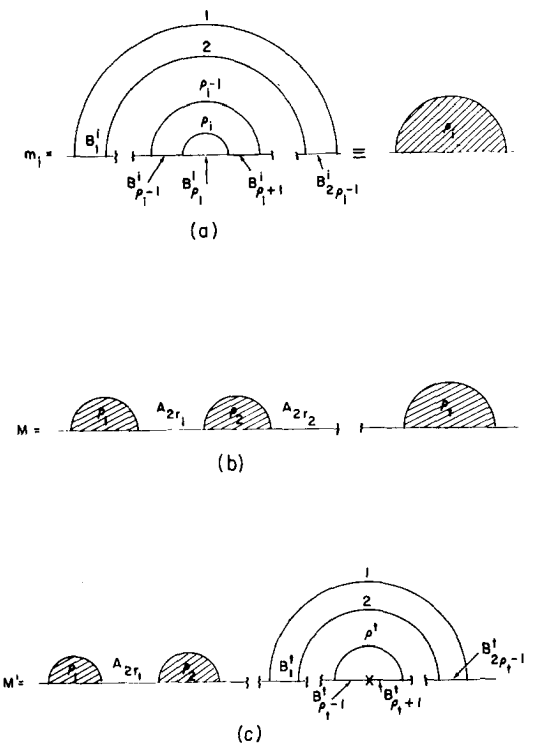


FIG. 3. We exhibit the general structure of m_i as given in (4.2) by drawing its Feynman diagram in (a). We let the shaded semicircle represent the Feynman diagram for m_i. In (b) we display the structure of M of Eq. (4.1) in terms of these shaded semicircles. 3c is the Feynman diagram for M' as defined by (4.5). The X represents the degree-two vertex which replaces the closed circuit of M.

and M' is exactly of the same form as M but with the factor (4.4) deleted.

The Feynman diagram $\mathcal{F}_{M'}$ can be gotten from $\mathcal{F}_M$ by simply shrinking the arc pairing $(v(i), v(i+1))$ and the line representing A_i to a single degree-two vertex. This is illustrated in Fig. 3(c).

Next consider the constructed graphs $\mathcal{G}_M$ and $\mathcal{G}_{M'}$. Since $\epsilon(i) = -\epsilon(i+1)$, it is clear that in $\mathcal{G}_M$ there is a closed loop containing only A_i. Since the vertices of M' have the same parity assignment as they did in M, the constructed graph $\mathcal{G}_{M'}$, is precisely the graph $\mathcal{G}_M$ with this closed loop missing. It then follows that the algorithm for M gives exactly the result (4.5) and (4.6). Hence the lemma will have been proven by induction on K as soon as the case $K = 1$ is taken care of.

For $K = 1$, $M' = A_0 A_2$ and it is clear that the algorithm gives exactly the correct result. Q.E.D.

Let $M(\sigma; A_0, A_1, A_2, \cdots, A_{2K})$ be an expression of the form (2.2) for which at least one of the pairings (i,j), pair two vertices having the same parity. We can apply (A8) to reduce out this pairing and get

$$M = -2M', \tag{4.7}$$

where M' is the "once" reduced expression. M' is obtained from M by reversing the order of all the γ factors between γ_{α_i} and γ_{α_j}, removing the factors γ_{α_i}, γ_{α_j}, and $\delta_{\alpha_i\alpha_j}$, and restricting the summation so as not to include α_i and α_j.

Lemma 4.2: The construction algorithm is valid for M if and only if it is valid for M'.

Proof: Let $\epsilon'(k)$ be the parity in M' of the kth vertex in M and let $v'(k)$ be the vertex in $\mathcal{F}_{M'}$, which corresponds to the kth vertex in $\mathcal{F}_M$. Then

$$\epsilon'(k) = \begin{cases} -\epsilon(k), & \text{for } i < k < j, \\ \epsilon(k), & \text{otherwise.} \end{cases} \tag{4.8}$$

To see this, first suppose that $i < k < j$. If $\epsilon(k) = \epsilon(i)(-\epsilon(i))$, then $\epsilon(k) = \epsilon(j)(-\epsilon(j))$. Hence, in $\mathcal{F}_M$ there are an even (odd) number of degree-two vertices between $v(k)$ and $v(j)$. Therefore, in $\mathcal{F}_{M'}$ there are an even (odd) number of degree-two vertices between the degree-two vertex, which corresponds in M to $v(i)$, and $v'(k)$. Therefore the parity of $v'(k)$ changes by a factor -1 in $\mathcal{F}_{M'}$. On the other hand, if $k > j$, we have inserted two new degree-two vertices to the left on $v'(k)$ in $\mathcal{F}_{M'}$; therefore $\epsilon'(k) = \epsilon(k)$. Finally, for $k < i$ it is obvious that $\epsilon'(k) = \epsilon(k)$.

It is now clear, from (4.8), that the constructed graph $\mathcal{G}_{M'}$ can be obtained from $\mathcal{G}_M$ by the following simple topological transformation: Rotate everything on the base line between the ith and jth vertices 180° about a vertical axis passing through this base line, keeping all connections intact. The two arcs pairing $(v(i), v(j))$ and the displaced vertices $(w(i), w(j))$ are thereby reduced to degree-two vertices.

Hence, the connectivity of $\mathcal{G}_M$ is preserved in $\mathcal{G}_{M'}$ and the lemma is proved as regards (a) and (c) of the algorithm.

It remains only to show that the "braced" factor produced by rule (b) for a particular closed circuit in $\mathcal{G}_M$ is equal to the "braced" factor produced by rule (b) for the corresponding closed circuit in $\mathcal{G}_{M'}$.

A closed circuit can be uniquely characterized by the finite superscripted sequence

$$C = (k_1^{(\beta_1)}, k_2^{(\beta_2)}, \cdots, k_h^{(\beta_h)}), \tag{4.9}$$

where k_r represents the rth A_k encountered in traversing the circuit and the superscript β_r designates the sense in which A_{k_r} is traversed. $\beta_r = +1(-1)$ corresponds to traversing A_{k_r} to the right (left). We always take A_{k_1} to be the left-most A in the circuit and $\beta_1 = +1$.

Let the expression (4.9) represent the circuit C in $\mathcal{G}_M$ under consideration. We can immediately dispose

of the case when $k_1 < i$ or $k_1 \geq j$. Here A_{k_1} remains the left-most A for the corresponding circuit in $\mathfrak{G}_{M'}$ and $\epsilon'(k_1) = \epsilon(k_1)$.

We therefore consider $i \leq k_1 < j$. Define the quantity

$$E_r = \tfrac{1}{2}(\epsilon(k_r) + \epsilon(k_r + 1)). \tag{4.10}$$

Clearly E_r can only take on the values -1, 0, or 1. We next observe two interesting properties of E_r:

(I)

$$E_r = \begin{cases} 0 \Leftrightarrow \text{there are an even number of } \gamma \text{ factors in } A_{k_r}, \\ \pm 1 \Leftrightarrow \text{there are an odd number of } \gamma \text{ factors in } A_{k_r}. \end{cases} \tag{4.11}$$

(II) The circuit or chain containing A_{k_r}, in general follows arcs which can be below or above the base line.

$$E_r = \begin{cases} 0 \Leftrightarrow \text{the path stays on the same side of the base line in traversing } A_{k_r}, \\ \pm 1 \Leftrightarrow \text{the path crosses the base line in traversing } A_{k_r}. \end{cases} \tag{4.12}$$

The latter can be seen by noting that there are four kinds of corners that can be formed by the base line and the pairing arc. They are

N.W. $\lrcorner$, S.E. $\ulcorner$, N.E. $\llcorner$, S.W. $\urcorner$.

The first two always have a positive parity and the last two have a negative parity. A_k is formed by placing an East corner to the left of a West corner. Thus (4.12) follows. [See Corollary 4.1 for an interesting consequence of these observations.]

Let A_{k_r} be the right-most A, to the left of j in C. Let C' be the circuit in $\mathfrak{G}_{M'}$ corresponding to C. We then consider the cases:

Case 1a. $\epsilon(k_1) = +1, \quad \beta_r = +1.$

If $\epsilon(k_r + 1) = +1(-1)$, then the path after traversing A_{k_r} is on the opposite (same) side of the base line as the path before entering A_{k_1}. Hence,

$$E(r) = \sum_{n=1}^{r} E_n \tag{4.13}$$

is odd (even). Therefore the number of γ factors in $A_{k_1} \cdots A_{k_r}$ is odd (even). That is, in the factor for C, which, by using (A12), we can write as

$$b_C = \{\mu_2^{k_1}, \cdots, \mu_{l_{k_1}}^{k_1}, \overbrace{\cdots\cdots\cdots\cdots}^{k_2^{(\beta_2)}, k_3^{(\beta_3)}, \cdots, k_{r-1}^{(\beta_{r-1})}}, \underbrace{\mu_1^{k_r}, \cdots, \mu_{l_{k_r}}^{k_r}}_{k_r^{(+)}}, \overbrace{\cdots\cdots\cdots\cdots}^{k_{r+1}^{(\beta_{r+1})}, \cdots, k_h^{(\beta_h)}}, \mu_1^{k_1}\}, \tag{4.14}$$

there are an even (odd) number of symbols μ in the partial sequence $\mu_2^{k_1} \cdots \mu_{l_{k_r}}^{k_r}$. Therefore, if $\epsilon(k_r + 1) = +1$, then, by (A12) and (A14),

$$b_C = \{\overbrace{\mu_{l_{k_r}}^{k_r}, \cdots, \mu_2^{k_r}, \mu_1^{k_r}}^{k_r^{(-)}}, \overbrace{\cdots\cdots\cdots\cdots}^{k_{r-1}^{(-\beta_{r-1})}, \cdots, k_2^{(-\beta_2)}}, \overbrace{\mu_{l_{k_1}}^{k_1}, \cdots, \mu_2^{k_1}, \mu_1^{k_1}}^{k_1^{(-)}}, \overbrace{\cdots\cdots\cdots\cdots\cdots\cdots}^{k_h^{(-\beta_h)}, k_{h-1}^{(-\beta_{h-1})}, \cdots, k_{r+1}^{(-\beta_{r-1})}}\} \tag{4.15}$$

But this is exactly the factor, according to rule (b), for C'. On the other hand, if $\epsilon(k_r + 1) = -1$, then by (A12) and (A14),

$$b_C = \{\mu_{l_{k_r-1}}^{k}, \cdots, \mu_2^{k_r}, \mu_1^{k_r}, \overbrace{\cdots\cdots\cdots\cdots}^{k_{r-1}^{(-\beta_{r-1})}, \cdots, k_2^{(-\beta_2)}}, \overbrace{\mu_{l_{k_1}}^{k_1}, \cdots, \mu_2^{k_1}, \mu_1^{k_1}}^{k_1^{(-)}}, \overbrace{\cdots\cdots\cdots\cdots\cdots\cdots}^{k_h^{(-\beta_h)}, k_{h-1}^{(-\beta_{h-1})}, \cdots, k_{r+1}^{(-\beta_{r+1})}}, \mu_{l_k}^{k_r}\} \tag{4.16}$$

which is again the correct factor for C'. By a similar analysis we can render the following cases:

Case 1b. $\epsilon(k_1) = -1, \quad \beta_r = +1,$

Case 2a. $\epsilon(k_1) = +1, \quad \beta_r = -1,$

Case 2b. $\epsilon(k_1) = -1, \quad \beta_r = -1,$

The details of which we leave to the reader. Q.E.D.

Corollary 4.1: In the constructed graph $\mathfrak{G}_M$, for any expression M of the form (2.2), every closed circuit contains an even number of γ factors.

Proof: This follows directly from the observations I and II of the proof for Lemma (4.2). We simply note that, in order to close, the path must end on the same side of the base line as where it started. Hence

$$E(l) = \sum_{n=1}^{l} E_n \tag{4.17}$$

must be even. Here l is the total number of A's in the circuit being considered. Q.E.D.

We are now ready to prove the main theorem:

Theorem: Let $M(\sigma; A_0, A_1, A_2, \cdots, A_{2k})$ be an expression of the form (2.2). Then the construction algorithm is valid for M.

Proof: The method of proof is by induction on K. For $K = 1$ the theorem can easily be seen to be true.

For $K > 1$ we distinguish two cases:

Case 1. M has at least one pairing which pairs vertices having the same parity.

Then, by Lemma 4.2, if the construction algorithm is valid for M', the once reduced expression, it must be true for M. The theorem follows.

Case 2. All the K pairings of M pair vertices of opposite parity.

We then have the two subcases:

(a) None of the pairings overlap.

Then the theorem follows directly from Lemma 4.1.

(b) At least one pair of pairings (i_1, j_1) and (i_2, j_2) overlap. That is, $i_1 < i_2 < j_1 < j_2$ (say).

Suppose there are an odd number of degree-two vertices between i_1 and i_2. Then there must be an odd number of degree-two vertices between j_1 and j_2 and an even number of degree-two vertices between i_2 and j_1. Let a be the right-most degree-two vertex between i_1 and i_2, and let b be the left-most degree-two vertex between j_1 and j_2. Consider the expression $\hat{M}$ which is obtained from M by reversing the order of all the γ factors between a and b, inserting the factors γ_a at a, γ_b at b and δ_{ab}, and summing over a and b. Since the number of degree-two vertices between γ_a and γ_b in $\hat{M}$ is even, M is the once-reduced expression from $\hat{M}$. By Lemma 4.2, if the construction algorithm is true for $\hat{M}$, it must be true for M. In $\hat{M}$, however,

$$i_1 < j_1 < i_2 < j_2$$

and $\hat{\epsilon}(i_1) = \hat{\epsilon}(j_1)$; $\hat{\epsilon}(i_2) = \hat{\epsilon}(j_2)$, where $\hat{\epsilon}(i)$ is the parity in $\hat{M}$ of the ith vertex in M. We can now twice reduce $\hat{M}$ to M', where M' has the pairings $(i_1 j_1)$ and $(i_2 j_2)$ reduced. Applying Lemma 4.2 twice, we get that, if the construction algorithm is true for M', it must be true for $\hat{M}$ and therefore for M. The theorem follows from the induction hypothesis by noting that M' has $K - 1$ pairings.

Finally, suppose there are an even number of degree-two vertices between i_1 and i_2. Then there must be an even number of degree-two vertices between j_1 and j_2 and an odd number between i_2 and j_1. Let a be the right-most degree-two vertex between i_2 and j_1, and let b be the left-most degree-two vertex to the right of j_2. Such a vertex can always be assumed to exist since we can always replace A_{2K} by $A_{2K}\gamma_1\gamma_1$. Consider the expression $\hat{M}$ obtained from M as before. Again the number of degree-two vertices between γ_a and γ_b in $\hat{M}$ is even; so if the construction algorithm is valid for $\hat{M}$, it must be valid for M. In $\hat{M}$ we have

$$i_1 < i_2 < j_2 < j_1$$

and

$$\hat{\epsilon}(i_1) = \hat{\epsilon}(j_1); \quad \hat{\epsilon}(i_2) = \hat{\epsilon}(j_2).$$

We can therefore twice reduce $\hat{M}$ to M', where we first contract the inside pairing (i_2, j_2) and then (i_1, j_1), and the theorem follows as before. Q.E.D.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. R. B. Marr for helpful discussions and Dr. Y. Shimamoto and Dr. G. A. Baker, Jr., for their encouragement.

APPENDIX

In this appendix we shall, for the sake of completeness, review some well-known results[2,6] on γ matrices and give some simple extensions needed for the text. We are particularly interested in having Eqs. (A8), (A11)–(A14).

We start again with the defining relations for the γ matrices:

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu}I. \tag{A1}$$

and we form the 16 elements

$$\Gamma_a^{(i)} = \begin{cases} \Gamma_0^{(1)} = I \quad \text{(identity)}, \\ \Gamma_{\mu\nu}^{(2)} = \sigma_{\mu\nu} \equiv \frac{1}{2}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu), \\ \Gamma_\mu^{(3)} = \gamma_\mu, \\ \Gamma_\mu^{(4)} = \gamma_5\gamma_\mu, \\ \Gamma_0^{(5)} = \gamma_5 \equiv \gamma_1\gamma_2\gamma_3\gamma_4. \end{cases} \tag{A2}$$

We state, without proof, six important and well-known theorems about the $\Gamma_a^{(i)}$.

Theorem A1: For $i \neq 1$; $\mathrm{Tr}\ \Gamma_a^{(i)} = 0$, where $\mathrm{Tr}\ (A)$ means the trace of A.

Theorem A2: $(\Gamma_a^{(i)})^2 = (-1)^{i+1}I$.

Theorem A3: The 16 $\Gamma_a^{(i)}$ are linearly independent and complete. They therefore form a basis for the (Clifford) algebra generated by γ_μ, and any element A of the algebra can be expanded as

$$A = \sum_{i,a} A_a^{(i)}\Gamma_a^{(i)},$$

where

$$A_a^{(i)} = \frac{(-1)^{i+1}}{4}\,\mathrm{Tr}\,(\Gamma_a^{(i)}A).$$

Theorem A4: The trace of a product of an odd number of γ_μ's is zero. That is,

$$\mathrm{Tr}\,(\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n+1}}) = 0 \quad (n = 0, 1, \cdots).$$

Theorem A5: In the expansion of

$$\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_n} = \sum A_a^{(i)}\Gamma_a^{(i)},$$

we have

$$A_a^{(1)} = A_a^{(2)} = A_a^{(5)} = 0, \quad \text{for } n \text{ odd},$$

and

$$A_a^{(3)} = A_a^{(4)} = 0, \qquad \text{for } n \text{ even}.$$

Theorem A6: There exists a 4×4 nonsingular matrix B, such that $B\gamma_\mu B^{-1} = \gamma_u^T$, where γ_u^T is the transpose of γ_μ and

$$B\Gamma_a^{(i)}B^{-1} = (-1)^{i+1}(\Gamma_a^{(i)})^T.$$

From Theorem A5 we see directly that we can write

$$\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n+1}} = A^{(3)}\Gamma^{(3)} + A^{(4)}\Gamma^{(4)} \quad \text{(A3)}$$

and

$$\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n}} = A^{(1)}\Gamma^{(1)} + A^{(2)}\Gamma^{(2)} + A^{(5)}\Gamma^{(5)}, \quad \text{(A4)}$$

where

$$A^{(i)}\Gamma^{(i)} = \sum_a A_a^{(i)}\Gamma_a^{(i)}.$$

We then prove the following corollary.

Corollary A1: If even and odd products of γ's are expanded as above, then

$$\gamma_{\alpha_{2n+1}}\gamma_{\alpha_{2n}}\cdots\gamma_{\alpha_2}\gamma_{\alpha_1} = A^{(3)}\Gamma^{(3)} - A^{(4)}\Gamma^{(4)} \quad \text{(A5)}$$

and

$$\gamma_{\alpha_{2n}}\gamma_{\alpha_{2n-1}}\cdots\gamma_{\alpha_2}\gamma_{\alpha_1} = A^{(1)}\Gamma^{(1)} - A^{(2)}\Gamma^{(2)} + A^{(5)}\Gamma^{(5)}. \quad \text{(A6)}$$

Proof: From Theorem A6 we see

$$\begin{aligned}\gamma_{\alpha_n}\gamma_{\alpha_{n-1}}\cdots\gamma_{\alpha_2}\gamma_{\alpha_1} &= B^{-1}\gamma_{\alpha_n}^T\cdots\gamma_{\alpha_2}^T\gamma_{\alpha_1}^T B \\ &= B^{-1}(\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_n})^T B,\end{aligned}$$

and the corollary follows from substituting the appropriate expansions for $\gamma_{\alpha_1}\cdots\gamma_{\alpha_n}$ and applying Theorem A6 again.

Theorem A7:

$$\sum_{\mu=1}^{4}\gamma_\mu\Gamma_a^{(i)}\gamma_\mu = a^{(i)}\Gamma_a^{(i)},$$

where

$$a^{(i)} = (4, 0, -2, 2, -4). \quad \text{(A7)}$$

This can be seen by inspection.

Corollary A2:

$$\sum_{\mu=1}^{4}\gamma_\mu\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n+1}}\gamma_\mu = -2\gamma_{\alpha_{2n+1}}\cdots\gamma_{\alpha_2}\gamma_{\alpha_1}. \quad \text{(A8)}$$

This follows from direct application of Theorem A7 and Corollary A1.

Corollary A3:

$$\sum_{\mu=1}^{4}\gamma_\mu\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n}}\gamma_\mu = 2(\gamma_{\alpha_{2n}}\gamma_{\alpha_1}\cdots\gamma_{\alpha_{2n-1}} + \gamma_{\alpha_{2n-1}}\cdots\gamma_{\alpha_1}\gamma_{\alpha_{2n}}). \quad \text{(A9)}$$

This follows directly from the commutation relations (A1) and the preceding corollary. It can be derived as follows:

$$\begin{aligned}&\sum_{\mu=1}^{4}\gamma_\mu(\gamma_{\alpha_1}\cdots\gamma_{\alpha_{2n}})\gamma_\mu \\ &= \sum_{\mu=1}^{4}\gamma_\mu(A^{(1)}\Gamma^{(1)} + A^{(2)}\Gamma^{(2)} + A^{(5)}\Gamma^{(5)})\gamma_\mu \\ &= 4(A^{(1)}\Gamma^{(1)} - A^{(5)}\Gamma^{(5)}) = 4(A^{(1)}\Gamma^{(1)} - A^{(5)}\Gamma^{(5)})\gamma_{\alpha_{2n}}^2 \\ &= 4\gamma_{\alpha_{2n}}(A^{(1)}\Gamma^{(1)} + A^{(5)}\Gamma^{(5)})\gamma_{\alpha_{2n}} \\ &= 2\gamma_{\alpha_{2n}}(\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n}} + \gamma_{\alpha_{2n}}\cdots\gamma_{\alpha_2}\gamma_{\alpha_1})\gamma_{\alpha_{2n}},\end{aligned}$$

from which (A9) follows. The advantage of deriving it in this way is that it leads immediately to the next corollary. First define the notation

$$\{\alpha_1, \alpha_2, \cdots, \alpha_{2n}\} \equiv \gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n}} + \gamma_{\alpha_{2n}}\cdots\gamma_{\alpha_2}\gamma_{\alpha_1}. \quad \text{(A10)}$$

For $n = 1$, this notation agrees with the standard anticommutator braces. Then we rewrite (A9) as

$$\sum_{\mu=1}^{4}\gamma_\mu\gamma_{\alpha_1}\gamma_{\alpha_2}\cdots\gamma_{\alpha_{2n}}\gamma_\mu = 2\{\alpha_{2n}, \alpha_1, \alpha_2, \cdots, \alpha_{2n-1}\}. \quad \text{(A11)}$$

Also, from the definition (A10) it is obvious that

$$\{\alpha_1, \alpha_2, \cdots, \alpha_{2n}\} = \{\alpha_{2n}, \cdots, \alpha_2, \alpha_1\}. \quad \text{(A12)}$$

Corollary A4:

$$\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_k}\{\alpha_1, \alpha_2, \cdots, \alpha_{2n}\} = \begin{cases}\{\alpha_{2n}, \alpha_1, \alpha_2, \cdots, \alpha_{2n-1}\}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_k} & \text{(for odd } k\text{)}, \\ \{\alpha_1, \alpha_2, \cdots, \alpha_{2n}\}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_k} & \text{(for even } k\text{)}.\end{cases} \quad \text{(A13)}$$

The proof imitates the proof of Corollary A3:

$$\begin{aligned}&\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_k}\{\alpha_1, \alpha_2, \cdots, \alpha_{2n}\} \\ &\quad = 2\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_k}(A^{(1)}\Gamma^{(1)} + A^{(5)}\Gamma^{(5)}) \\ &\quad = 2(A^{(1)}\Gamma^{(1)} + (-1)^k A^{(5)}\Gamma^{(5)})\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_k},\end{aligned}$$

and the corollary follows.

Corollary A5:

$$\{\alpha_1, \alpha_2, \cdots, \alpha_{2n}\} = \{\alpha_3, \cdots, \alpha_{2n}, \alpha_1, \alpha_2\}. \quad \text{(A14)}$$

This follows trivially by applying Corollary A4 twice.