JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 7, NUMBER 11 NOVEMBER 1966

Transformation of Proca-Type System of Partial Differential Equations

C. ABRAHAM
Commonwealth Scientific and Industrial Research Organization, Computing Research Section, Canberra, Australia
(Received 22 March 1966)

The Proca-type system of partial differential equations [$-\text{curl curl } \mathbf{A} + k\mathbf{A} = \tau \text{ grad } \varphi$; $\nabla^2\varphi - K\varphi = \tau \text{ div } \mathbf{A}$, with k, K, and τ constants] is transformed into a new system such that the vector potential $\mathbf{A}$ and the scalar potential φ satisfy separate differential equations. It is also shown that, for the solution of the system, there are only two different modes, $\text{div } \mathbf{A} = 0$ with $\varphi = 0$, and $\text{curl } \mathbf{A} = 0$ with $\varphi \neq 0$, respectively.

INTRODUCTION

A SPECIFIC form of the differential equation to be discussed in this work arises for the vector meson field.[1,2] The general system of equations can be of interest also for other domains of research.

The system to be treated has equations

$$-\nabla \times (\nabla \times \mathbf{A}) + k\mathbf{A} = \tau \text{ grad } \varphi, \tag{1a}$$

$$\nabla^2\varphi - K\varphi = \text{div } \mathbf{A}, \tag{1b}$$

where k, K, τ are constants, and

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}.$$

The aim of this paper is to transform the system of Eqs. (1a, b) into another system for which the function φ and the vector $\mathbf{A}$ will satisfy separate differential equations. In order to get separate equations for different functions, the same technique employed by Abraham[3] in transforming Brown's linear system of partial differential equations[4] for micromagnetics will be used.

DERIVATION OF THE TRANSFORMATION

Taking the divergence of (1a) and using (1b), one obtains

$$k \text{ div } \mathbf{A} = \tau\nabla^2\varphi. \tag{2}$$

Using (1b) and (2),

$$\varphi = \{(k - \tau)/\tau K\} \text{ div } \mathbf{A}. \tag{3}$$

Rewriting Eq. (1a) and using Eq. (3),

$$\nabla^2\mathbf{A} + K^{-1}(\tau - k - K)\nabla(\nabla\cdot\mathbf{A}) + k\mathbf{A} = 0. \tag{4}$$

The divergence of Eq. (4) gives

$$\nabla^2(\nabla\cdot\mathbf{A}) = -kK(\tau - k)^{-1}(\nabla\cdot\mathbf{A}). \tag{5}$$

Acting with the operator ∇^6 on Eq. (4),

$$\nabla^8\mathbf{A} + K^{-1}(\tau - k - K)\nabla^6 \times \nabla(\nabla\cdot\mathbf{A}) + \mathbf{k}\nabla^6\mathbf{A} = 0. \tag{6}$$

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Part I, p. 1891.
[2] G. Wentzel, *Quantum Theory of Fields* (Interscience Publishers, Inc., London, 1949), p. 75.
[3] C. Abraham, Phys. Rev. 140, A144 (1965).
[4] W. F. Brown, Jr., *Micromagnetics* (John Wiley & Sons, Inc., New York, 1963), p. 48.

Using Eq. (4) and acting with the operators ∇ and ∇^2 on (5), one can write Eq. (6) as

$$\nabla^8\mathbf{A} + k\nabla^6\mathbf{A} + \frac{k^3K^3}{(\tau - k)^3}\nabla^2\mathbf{A} + \frac{k^4K^3}{(\tau - k)^3}\mathbf{A} = 0. \quad (7)$$

The solution of Eq. (6) can be represented as

$$\prod_{i=1}^{4}(\nabla^2 - v_i)\mathbf{A} = 0, \quad (8)$$

where the v_i are roots of the fourth-degree polynomial in v,

$$v^4 + kv^3 + k^3K^3(\tau - k)^{-3}v + k^4K^3(\tau - k)^{-3} = 0. \quad (9)$$

It is not difficult to show that these roots are given by

$$v_1 = -k; \quad v_2 = -kK(\tau - k)^{-1};$$
$$v_3 = \tfrac{1}{2}kK(\tau - k)^{-1}(1 + i\sqrt{3});$$
$$v_4 = \tfrac{1}{2}kK(\tau - k)^{-1}(1 - i\sqrt{3}).$$

It follows that, the general solution of Eq. (7) is

$$\mathbf{A} = \sum_{i=1}^{4}\mathbf{A}_i, \quad (10)$$

where $\mathbf{A}_i$ ($i = 1, 2, 3, 4$) are solutions of

$$\nabla^2\mathbf{A}_i = v_i\mathbf{A}_i. \quad (11)$$

The solution of the system of Eq. (1) contains at most eight constants of integration, while the solutions represented by the Eqs. (3), (10), and (12) contain 24 constants of integration. Hence, there must exist some relationship among the components of the vectors $\mathbf{A}_i$. These relations, for each value of i, are obtained by inserting the expressions (3) and (11) into the original system of Eqs. (1).

The relations are

(1) $i = 1$:

$$\varphi_1 = 0, \quad \nabla\cdot\mathbf{A}_i = 0. \quad (12a, b)$$

(2) $i = 2$:

$$\varphi_2 = -(\tau K)^{-1}(\tau - k)\nabla\cdot\mathbf{A}_2, \ \operatorname{curl}\mathbf{A}_2 = 0. \quad (13a, b)$$

(3) $i = 3$:

$$\mathbf{A}_3 = \varphi_3 = 0. \quad (14)$$

(4) $i = 4$:

$$\mathbf{A}_4 = \varphi_4 = 0. \quad (15)$$

However, the solution of the system of Eqs. (1) can be represented by two different modes $\varphi = 0$, with $\operatorname{div}\mathbf{A} = 0$, and $\varphi \neq 0$, with $\operatorname{curl}\mathbf{A} = 0$, as can be seen from Eqs. (12) and Eqs. (13).

The transformed system equivalent to Eqs. (1a, b) is

$$\nabla^2\mathbf{A}_1 = -\mathbf{kA}_1, \quad \varphi_1 = 0, \quad \nabla\cdot\mathbf{A}_1 = 0 \quad (16a, b)$$

and

$$\nabla^2\mathbf{A}_2 = -kK(\tau - k)^{-1}\mathbf{A}_2, \quad (16c)$$

$$\varphi_2 = -(\tau K)^{-1}(\tau - k)\nabla\cdot\mathbf{A}_2; \quad \operatorname{curl}\mathbf{A}_2 = 0. \quad (16d)$$

The advantage Eqs. (16) have over the system of Eqs. (1) is that the former can easily be transformed into an appropriate system of coordinates, according to the given boundary conditions.

ACKNOWLEDGMENT

The author is grateful to Dr. G. N. Lance for valuable suggestions in regard to this manuscript.

Relativistic Effects in Atomic Fine Structure*

LLOYD ARMSTRONG, JR.
Lawrence Radiation Laboratory, University of California, Berkeley, California
(Received 7 April 1966)

Operators are obtained which can be evaluated with respect to nonrelativistic wavefunctions to produce the same result as obtained by evaluating the Breit equation with respect to relativistic wavefunctions. This greatly simplifies calculations involving the Breit equation by allowing the calculations to be made within the more familiar framework of nonrelativistic theory. The operators are classified according to their angular dependence; a comparison with the angular dependence of each fine-structure operator leads to the relativistic equivalents of the fine-structure interactions. The operators are expanded in a power series in $(v/c)^2$, and the lowest nonvanishing terms are shown to be the fine-structure interactions.

I. INTRODUCTION

RECENT advances in computing techniques and machinery have made possible greatly improved Hartree–Fock calculations,[1] with corresponding improvement in calculated fine-structure constants such as a_{SO}, the spin–orbit coupling constant.[2] The recent appearance of relativistic Hartree–Fock calculations[3] raises the possibility of further improvement in calculated fine-structure energies.

Calculation of relativistic fine-structure energies implies, of course, the knowledge of a Hamiltonian describing two-body interactions between relativistic particles. The Breit equation[4] is most often used for this purpose; although only an approximation, it is a good approximation if the Breit operator

$$\left[-\frac{e^2}{2}\frac{\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2}{r_{12}}-\frac{e^2}{2}\frac{(\boldsymbol{\alpha}_1\cdot\mathbf{r}_{12})(\boldsymbol{\alpha}_2\cdot\mathbf{r}_{12})}{r_{12}^3}\right]$$

is treated by use of perturbation theory.[5]

One complication that restricts the usefulness of the Breit equation, however, is the difficulty of extracting from it the explicit form of a particular fine-structure interaction (spin–other–orbit, for example). Sandars and Beck[6] have recently suggested a method of calculating relativistic effects in atoms which can be used to overcome this problem. The method involves obtaining an "equivalent operator" which, when evaluated between nonrelativistic wavefunctions, produces the same results as obtained by evaluating the relativistic operator between relativistic states. This has the great advantage that relativistic effects can be studied within the nonrelativistic scheme—an immense simplification for heavy atoms.

We have obtained equivalent operators for the terms in the Breit equation (Sec. III); these operators are then broken up into groups which correspond to fine-structure interactions (Sec. IV). Finally, these groups are reduced to the nonrelativistic limit in order to obtain the fine-structure interactions. This last step is important because it reveals new operators of the same magnitude as the fine-structure interactions.

II. THE HAMILTONIAN

The analysis is based on the solution by first-order perturbation theory of the Breit equation for two electrons (charge $-e$),[4,5]

$$\mathcal{H}\Psi=\left\{\sum_{i=1,2}\left[\boldsymbol{\alpha}_i\cdot(c\mathbf{p}_i+e\mathbf{A}_i)+\beta_i mc^2-\frac{Ze^2}{r_i}\right]+\frac{e^2}{r_{12}}-\tfrac{1}{2}e^2\frac{\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2}{r_{12}}-\tfrac{1}{2}e^2\frac{(\boldsymbol{\alpha}_1\cdot\mathbf{r}_{12})(\boldsymbol{\alpha}_2\cdot\mathbf{r}_{12})}{r_{12}^3}\right\}\Psi=E\Psi. \quad (1)$$

We assume that the potential terms in Eq. (1) can be approximately replaced by a central field term $\sum_i U(r_i)$. The approximate Hamiltonian is then

$$\mathcal{H}_0=\sum_{i=1,2}[\boldsymbol{\alpha}_i\cdot(c\mathbf{p}_i+e\mathbf{A}_i)+\beta_i mc^2+U(r_i)], \quad (2)$$

and the difference, $\mathcal{H}_1=\mathcal{H}-\mathcal{H}_0$, can be treated

* This work was supported by the United States Atomic Energy Commission.

[1] R. E. Watson and A. J. Freeman, Phys. Rev. **124**, 1117 (1961); **127**, 2058 (1962); F. Herman and S. Skillman, *Atomic Structure Calculations* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963).

[2] M. Blume and R. E. Watson, Proc. Roy. Soc. (London) **A270**, 127 (1962); **A271**, 565 (1963); M. Blume, A. J. Freeman, and R. E. Watson, Phys. Rev. **134**, A320 (1964).

[3] D. Liberman, J. T. Waber, and D. T. Cromer, Phys. Rev. **137**, A27 (1965).

[4] G. Breit, Phys. Rev. **34**, 553 (1929); **36**, 383 (1930); **39**, 616 (1932).

[5] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer-Verlag, Berlin, 1957).

[6] P. G. H. Sandars and J. Beck, Proc. Roy. Soc. (London) **A289**, 97 (1965).

as a perturbation. For the special case in which $\mathbf{A}_i = 0$, the wavefunction satisfying

$$\mathcal{H}_0\Psi_0 = E_0\Psi_0 = (E_0^1 + E_0^2)\Psi_0, \tag{3}$$

where E^i is the energy of electron i, can be written as a product of wavefunctions of the form

$$|ljm) = \begin{pmatrix} F/r\ |ljm\rangle \\ iG/r\ |\bar{l}jm\rangle \end{pmatrix}, \tag{4}$$

where $\bar{l} = l \pm 1$ as $j = l \pm \frac{1}{2}$, and

$$|ljm\rangle = \sum_{m_l m_s} (-)^{l-\frac{1}{2}-m}[j]^{\frac{1}{2}} \times \begin{pmatrix} \frac{1}{2} & l & j \\ m_s & m_l & -m \end{pmatrix} |lm_l\rangle\chi^{\frac{1}{2}}_{m_s}. \tag{5}$$

The term $\chi^{\frac{1}{2}}$ is the usual two-component spinor. Here and in what follows, relativistic wavefunctions are written in the general form $|ljm)$ and nonrelativistic functions as $|ljm\rangle$. Terms written $[a, b, \cdots]$ stand for $(2a + 1)(2b + 1) \cdots$. We restrict our discussion to the configuration l^2.

The radial functions F and G, which can be taken to be real, can be related through Eqs. (2)–(4):

$$\begin{aligned} \left(\frac{d}{dr_i} - \frac{\kappa_i}{r_i}\right)F_i &= \frac{1}{\hbar c}[mc^2 + E_0^i - U(r_i)]G_i, \\ \left(\frac{d}{dr_i} + \frac{\kappa_i}{r_i}\right)G_i &= \frac{1}{\hbar c}[mc^2 - E_0^i + U(r_i)]F_i, \end{aligned} \tag{6}$$

with $\kappa_i = (-)^{j_i+l-\frac{1}{2}}[j_i]$.

The energy, to the first order in the perturbation, is then given by

$$\begin{aligned} &(\Psi_0|\ \mathcal{H}_0 + \mathcal{H}_1\ |\Psi_0) \\ &= (E_0 + E_1)(\Psi_0\ |\ \Psi_0) = E(\Psi_0\ |\ \Psi_0) \\ &= (\Psi_0|\ \boldsymbol{\alpha}\cdot\mathbf{p} + \beta mc^2 + \mathcal{H}_\alpha + \mathcal{H}_\beta + \mathcal{H}_\gamma + \mathcal{H}_\delta\ |\Psi_0), \end{aligned} \tag{7}$$

where

$$\mathcal{H}_\alpha = \sum_i \mathcal{H}_\alpha^i, \qquad \mathcal{H}_\alpha^i = -Ze^2/r_i, \qquad \mathcal{H}_\beta = e^2/r_{12},$$

$$\mathcal{H}_\gamma = -\tfrac{1}{2}e^2(\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2/r_{12}),$$

and

$$\mathcal{H}_\delta = -\tfrac{1}{2}e^2\,\frac{(\boldsymbol{\alpha}_1\cdot\mathbf{r}_{12})(\boldsymbol{\alpha}_2\cdot\mathbf{r}_{12})}{r_{12}^3}.$$

The first two terms on the extreme right-hand side of Eq. (7) are the kinetic energy and mass-effect terms, respectively. In the following sections, we are not directly concerned with these two terms, but rather with the remaining terms in $\mathcal{H}$.

III. EQUIVALENT OPERATORS

We wish to obtain the operator O_e defined by the equation

$$(\Psi_0|\ \mathcal{H}_\alpha + \mathcal{H}_\beta + \mathcal{H}_\gamma + \mathcal{H}_\delta\ |\Psi_0) = \langle\Psi|\ O_e\ |\Psi\rangle, \tag{8}$$

where $|\Psi\rangle$ is the nonrelativistic wavefunction which $|\Psi_0\rangle$ approaches in the nonrelativistic limit. The operator O_e is the "equivalent operator" for the interactions $\mathcal{H}_\alpha$ through $\mathcal{H}_\delta$, and will be obtained below by considering the interactions $\mathcal{H}_\alpha$ through $\mathcal{H}_\delta$ separately.

A. Equivalent Operator for $\mathcal{H}_\alpha$

Evaluation of $\mathcal{H}_\alpha^i$ between relativistic wavefunctions is straightforward, and yields

$$(ljm|\ \mathcal{H}_\alpha^i\ |ljm) = -Ze^2\int \frac{(F_j^2 + G_j^2)_i}{r_i}\,dr_i. \tag{9}$$

The equivalent operator for $\mathcal{H}_\alpha^i$, namely O_α^i, can be written in the general form

$$O_\alpha^i = \sum_{\kappa k K} \alpha^i(\kappa k K)\mathbf{w}_i^{(\kappa k)K}, \tag{10}$$

where the α are constants to be determined, and the $\mathbf{w}^{(\kappa k)K}$ are defined by the relation

$$\mathbf{w}^{(\kappa k)K} = \{\mathbf{t}^\kappa\mathbf{v}^k\}^K, \qquad \langle s||\ \mathbf{t}^\kappa\ ||s\rangle = [\kappa]^{\frac{1}{2}}, \tag{11}$$

and

$$\langle l||\ v^k\ ||l'\rangle = \delta_{ll'}[k]^{\frac{1}{2}}.$$

Because $\mathcal{H}_\alpha$ is a scalar, $K = 0$ in Eq. (10) above, and therefore $\kappa = k$. Taking matrix elements, we obtain

$$\langle ljm|\ O_\alpha^i\ |ljm\rangle = \sum_k \alpha^i(kk)(-)^{k+l+j+\frac{1}{2}}[k]^{\frac{1}{2}}\begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & k \\ l & l & j \end{Bmatrix}. \tag{12}$$

Equating the right-hand sides of Eqs. (9) and (12), and multiplying both sides by

$$\sum_j \begin{Bmatrix} \frac{1}{2} & l & j \\ l & \frac{1}{2} & k \end{Bmatrix}[j](-)^j,$$

we obtain

$$\alpha^i(kk) = [k]^{\frac{1}{2}}(-)^{k+l-\frac{1}{2}}Ze^2 \times \sum_j [j](-)^j\begin{Bmatrix} \frac{1}{2} & l & j \\ l & \frac{1}{2} & k \end{Bmatrix}\int_0^\infty \frac{(F_j^2 + G_j^2)_i}{r_i}\,dr_i. \tag{13}$$

We postpone a discussion of this and subsequent results until Sec. IV.

B. Equivalent Operator for $\mathcal{H}_\beta$

Because $\mathcal{H}_\beta$ is a two-body operator, we must consider matrix elements between relativistic states

composed of two electrons. The final form obtained for O_α does not depend on the type of coupling used for the wavefunction. However, in order to demonstrate more fully the method to be used, we use below wavefunctions of the form $|l^2SLJM)$.

As is apparent from Eq. (4), in relativistic theory, j, and not l, is a good quantum number. The state $|l^2SLJM)$ must then be decomposed into states $|j_1j_2JM)$. which in turn are decomposed in the usual way into a sum of products of $|lj_1m_1)$ and $|lj_2m_2)$. Then

$$(l^2S_1L_1JM|\ \mathcal{H}_\beta\ |l^2S_2L_2JM)$$
$$= \sum_{\substack{j_1j_3\\ j_2j_4}} [S_1, L_1, S_2, L_2, j_1, j_2, j_3, j_4]^{\frac{1}{2}} \times \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_1 \\ l & l & L_1 \\ j_1 & j_2 & J \end{Bmatrix} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_2 \\ l & l & L_2 \\ j_3 & j_4 & J \end{Bmatrix} (j_1j_2JM|\ \mathcal{H}_\beta\ |j_3j_4JM). \tag{14}$$

The term $\mathcal{H}_\beta$ can be expanded as

$$e^2 \sum_K \frac{r_<^K}{r_>^{K+1}} \mathbf{C}_1^K \cdot \mathbf{C}_2^K.$$

The symbol C^K is defined by

$$C_m^K = (4\pi/2K+1)^{\frac{1}{2}} Y_m^K,$$

where Y_m^K is the usual spherical harmonic. In evaluating the matrix element on the right side of Eq. (14), one obtains reduced matrix elements such as

$$(j_1||\ C^K r^K\ ||j_3) = \langle lj_1||\ C^K\ ||lj_3\rangle \int F_{j_1}F_{j_3}r^K\,dr + \langle \bar{l}j_1||\ C^K\ ||\bar{l}j_3\rangle \int G_{j_1}G_{j_3}r^K\,dr. \tag{15}$$

This simplifies to

$$= (-)^{j_3-\frac{1}{2}}[j_1, j_3]^{\frac{1}{2}} \begin{pmatrix} j_1 & K & j_3 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \times \int (F_{j_1}F_{j_3} + G_{j_1}G_{j_3})r^K\,dr$$

for K even, or zero for K odd. We finally obtain, for Eq. (14),

$$(l^2S_1L_1JM|\ \mathcal{H}_\beta\ |l^2S_2L_2JM) = e^2 \sum [S_1, S_2, L_1, L_2]^{\frac{1}{2}}[j_1, j_2, j_3, j_4](-)^{j_1+j_3+J} \times \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_1 \\ l & l & L_1 \\ j_1 & j_2 & J \end{Bmatrix} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_2 \\ l & l & L_2 \\ j_3 & j_4 & J \end{Bmatrix} \begin{Bmatrix} j_3 & j_4 & J \\ j_2 & j_1 & K \end{Bmatrix} \begin{pmatrix} j_1 & K & j_3 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} j_2 & K & j_4 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \times \iint (F_1F_3 + G_1G_3)_1(F_2F_4 + G_2G_4)_2 \frac{r_<^K}{r_>^{K+1}}\,dr_1\,dr_2, \tag{16}$$

where the sum is over j_1, j_2, j_3, j_4, and K, and F_1 has been written for F_{j_1}, etc. Particle assignments are subscripted to the parentheses.

The equivalent operator is written in this case as

$$O_\beta = \sum \beta(k_1K_1k_2K_2k)\mathbf{w}_1^{(k_1K_1)k} \cdot \mathbf{w}_2^{(k_2K_2)k}, \tag{17}$$

where the sum is over k_1, K_1, k_2, K_2, and k. This is the most general form for a scalar two-body interaction. Proceeding as in Sec. IIIA, we evaluate

$$\langle l^2S_1L_1JM|\ O_\beta\ |l^2S_2L_2JM\rangle$$

and equate the results with Eq. (16). The constant β is obtained by utilizing the orthogonality conditions for $6-j$ and $9-j$ symbols. One obtains

$$\beta(k_1K_1k_2K_2k) = e^2 \sum_{\substack{j_1j_3\\ j_2j_4}} (-)^{j_2+j_3+1} \frac{[k_1, K_1, k_2, K_2]^{\frac{1}{2}}}{[k]} [j_1, j_2, j_3, j_4] \times \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & k_1 \\ l & l & K_1 \\ j_1 & j_3 & k \end{Bmatrix} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & k_2 \\ l & l & K_2 \\ j_2 & j_4 & k \end{Bmatrix} \begin{pmatrix} j_1 & k & j_3 \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} j_2 & k & j_4 \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \times \iint (F_1F_3 + G_1G_3)_1(F_2F_4 + G_2G_4)_2 \frac{r_<^k}{r_>^{k+1}}\,dr_1\,dr_2, \tag{18}$$

where k is even. By interchanging j_1 and j_3, j_2 and j_4, we see that β will be zero for either (or both) $k_1 + K_1$ or $k_2 + K_2$ odd.

C. Equivalent Operator for $\mathcal{H}_\gamma$

The derivation of the equivalent operator for $\mathcal{H}_\gamma$ is carried out in essentially the same manner as for the equivalent operator for $\mathcal{H}_\beta$. We first, however, rewrite $\mathcal{H}_\gamma$:

$$\mathcal{H}_\gamma = -\tfrac{1}{2}e^2\frac{\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2}{r_{12}} = -\tfrac{1}{2}e^2\sum_\beta(\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2)(\mathbf{C}_1^\beta\cdot\mathbf{C}_2^\beta)\frac{r_<^\beta}{r_>^{\beta+1}}$$

$$= -\tfrac{1}{2}e^2\sum_{k\beta}(\boldsymbol{\alpha}_1\mathbf{C}_1^\beta)^k\cdot(\boldsymbol{\alpha}_2\mathbf{C}_2^\beta)^k(-)^{1+k+\beta}\frac{r_<^\beta}{r_>^{\beta+1}}. \tag{19}$$

Then

$$(l^2S_1L_1JM|\,\mathcal{H}_\gamma\,|l^2S_2L_2JM) = -\tfrac{1}{2}e^2\sum[j_1,j_2,j_3,j_4,L_1,L_2,S_1,S_2]^{\frac{1}{2}}(-)^{j_3+j_4+J+k}$$
$$\times\begin{Bmatrix}\frac{1}{2}&\frac{1}{2}&S_1\\ l&l&L_1\\ j_1&j_3&J\end{Bmatrix}\begin{Bmatrix}\frac{1}{2}&\frac{1}{2}&S_2\\ l&l&L_2\\ j_2&j_4&J\end{Bmatrix}\begin{Bmatrix}j_3&j_4&J\\ j_2&j_1&k\end{Bmatrix}$$
$$\times\iint(j_1||\,(\boldsymbol{\alpha}\mathbf{C}^\beta)^k\,||j_3)(j_2||\,(\boldsymbol{\alpha}\mathbf{C}^\beta)^k\,||j_4)\frac{r_<^\beta}{r_>^{\beta+1}}\,dr_1\,dr_2. \tag{20}$$

The sum is over j_1, j_2, j_3, j_4, β, and k; the reduced matrix elements are given by

$$(j_1||\,(\boldsymbol{\alpha}\mathbf{C}^\beta)^k\,||j_3) = i[k,j_1,j_3]^{\frac{1}{2}}\Bigg\{\sqrt{2}\,(-)^{l+1}\begin{pmatrix}1&\beta&k\\ -1&0&1\end{pmatrix}\begin{pmatrix}j_1&j_3&k\\ -\frac{1}{2}&-\frac{1}{2}&1\end{pmatrix}(F_1G_3+G_3F_1)$$
$$+(-)^{j_3+\frac{1}{2}}\begin{pmatrix}1&\beta&k\\ 0&0&0\end{pmatrix}\begin{pmatrix}j_1&j_3&k\\ -\frac{1}{2}&\frac{1}{2}&0\end{pmatrix}(F_1G_3-G_1F_3)\Bigg\} \tag{21}$$

for β odd, zero for β even. The equivalent operator is defined as

$$O_\gamma = \sum\gamma(k_1K_1k_2K_2k)\mathbf{w}_1^{(k_1K_1)k}\cdot\mathbf{w}_2^{(k_2K_2)k}, \tag{22}$$

where the sum is over k_1, K_1, k_2, K_2, and k. Solving for γ, we find

$$\gamma(k_1K_1k_2K_2k) = -\tfrac{1}{2}e^2\sum[j_1,j_2,j_3,j_4]^{\frac{1}{2}}\frac{[k_1,K_1,k_2,K_2]^{\frac{1}{2}}}{[k]}(-)^k$$
$$\times\begin{Bmatrix}\frac{1}{2}&\frac{1}{2}&k_1\\ l&l&K_1\\ j_1&j_3&k\end{Bmatrix}\begin{Bmatrix}\frac{1}{2}&\frac{1}{2}&k_2\\ l&l&K_2\\ j_2&j_4&k\end{Bmatrix}\iint(j_1|\,|(\boldsymbol{\alpha}_1\mathbf{C}_1^\beta)^k|\,|j_3)(j_2|\,|(\boldsymbol{\alpha}_2\mathbf{C}_2^\beta)^k|\,|j_4)\frac{r_<^\beta}{r_>^{\beta+1}}\,dr_1\,dr_2. \tag{23}$$

By interchanging j_1 and j_3, j_2 and j_4, we see that γ is zero if either (or both) k_1+K_1 or k_2+K_2 is even.

D. Equivalent Operator for $\mathcal{H}_\delta$

The term $\mathcal{H}_\delta$ can be rewritten in the form

$$\mathcal{H}_\delta = -\tfrac{1}{2}e^2\left[\frac{1}{3}\frac{\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2}{r_{12}}+(5)^{\frac{1}{2}}\frac{((\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2)^2(\mathbf{r}_{12}\mathbf{r}_{12})^2)^0}{r_{12}^3}\right]. \tag{24}$$

The first term on the right above has the same form as $\mathcal{H}_\gamma$; the second term can be evaluated by using the relationship[7]

[7] B. R. Judd, private communication (1965).

$$\frac{(\mathbf{r}_{12}\mathbf{r}_{12})^2}{r_{12}^3}$$
$$=\sum_\beta(-)^\beta\frac{r_<^\beta}{r_>^{\beta+1}}\left\{(\mathbf{C}_1^\beta\mathbf{C}_2^\beta)^2\left[\frac{(8\beta)(\beta+1)(2\beta+1)}{(15)(2\beta-1)(2\beta+3)}\right]^{\frac{1}{2}}\right.$$
$$-(\mathbf{C}_1^{\beta-2}\mathbf{C}_2^\beta)^2\left[\frac{(\beta)(\beta-1)(2\beta-3)(2\beta+1)}{5(2\beta-1)}\right]^{\frac{1}{2}}$$
$$\left.+(\mathbf{C}_1^\beta\mathbf{C}_2^{\beta+2})^2\left[\frac{(\beta+1)(\beta+2)(2\beta+1)(2\beta+5)}{5(2\beta+3)}\right]^{\frac{1}{2}}\right\}. \tag{25}$$

The terms in this expansion can be rewritten

$$F(\beta\gamma)(-)^{\beta}\{(\alpha_1\alpha_2)^2(\mathbf{C}_1^{\beta}\mathbf{C}_2^{\gamma})^2\}^0 = \sum_K (-)^1(5)^{\frac{1}{2}}\begin{Bmatrix}1 & 1 & 2\\ \gamma & \beta & K\end{Bmatrix}((\alpha_1\mathbf{C}_1^{\beta})^K\cdot(\alpha_2\mathbf{C}^{\gamma})^K)F(\beta\gamma), \tag{26}$$

where $\gamma = \beta, \beta \pm 2$, and $F(\beta\gamma)$ is the term multiplying the angular factor $(\mathbf{C}^{\beta}\mathbf{C}^{\gamma})^2$ in Eq. (25). Upon inserting Eq. (26) into Eq. (24), one sees that $\mathcal{H}_\delta$ has the same form as $\mathcal{H}_\gamma$. We write the equivalent operator for $\mathcal{H}_\delta$ as

$$O_\delta = \sum \{\delta^0(k_1K_1k_2K_2k) + \delta^2(k_1K_1k_2K_2k)\} \times (\mathbf{w}_1^{(k_1K_1)k}\cdot\mathbf{w}_2^{(k_2K_2)k}), \tag{27}$$

where the sum is over k_1, K_1, k_2, K_2, and k. The expression δ^0 corresponds to the first term on the right of Eq. (24), δ^2 to the second. These two expressions are easily evaluated by comparison with Eqs. (19) and (23). One obtains

$$\delta^0(k_1K_1k_2K_2k) = \tfrac{1}{3}\gamma(k_1K_1k_2K_2k) \tag{28}$$

and

$$\delta^2(k_1K_1k_2K_2k) = -\tfrac{1}{2}[2]e^2\sum\frac{[j_1, j_2, j_3, j_4, k_1, K_1, k_2, K_2]^{\frac{1}{2}}}{[k]}(-)^{\beta} \times \begin{Bmatrix}\frac{1}{2} & \frac{1}{2} & k_1\\ l & l & K_1\\ j_1 & j_3 & k\end{Bmatrix}\begin{Bmatrix}\frac{1}{2} & \frac{1}{2} & k_2\\ l & l & K_2\\ j_2 & j_4 & k\end{Bmatrix}\begin{Bmatrix}1 & 1 & 2\\ \gamma & \beta & k\end{Bmatrix}F(\gamma\beta) \times \iint (j_1|\,|(\alpha_1C_1^{\beta})^k|\,|j_3)(j_2|\,|(\alpha_2C_2^{\gamma})^k|\,|j_4)\frac{r_<^{\beta}}{r_>^{\beta+1}}\,dr_1\,dr_2. \tag{29}$$

The sum is over j_1, j_2, j_3, j_4, β, and γ. Both δ^0 and δ^2 are zero if β is even, and if either (or both) $k_1 + K_1$ or $k_2 + K_2$ is even.

Further simplification can be obtained for particular cases: let $\delta^2 = \delta^{21} + \delta^{22} + \delta^{23}$, where δ^{21} stands for the case in which $\gamma = \beta$, δ^{22} for $\gamma = \beta + 2$, and δ^{23} for $\gamma = \beta - 2$. For k odd, $\delta^{21} = \frac{2}{3}\gamma$, δ^{22} and δ^{23} are zero. In this case $\delta^0 + \delta^2 = \gamma$. For k even and $k = \beta + 1$,

$$\delta^0 + \delta^{21} + \gamma = [2(k+1)/2k+1]\gamma;$$

for k even and $k = \beta - 1$,

$$\delta^0 + \delta^{21} + \gamma = [2k/(2k+1)]\gamma.$$

No analogous simplifications are possible for δ^{22} or δ^{23}.

IV. INTERPRETATION OF THE OPERATORS

The terms in O_δ having the same angular dependence as the fine-structure interactions can be identified as relativistic fine-structure interactions. These relativistic interactions can be expanded in a power series in orders of $(v/c)^2$; the lowest nonvanishing terms will, in most instances, be just the usual fine-structure interactions. We consider now the terms according to their angular dependence.

A. Terms with No Angular Dependence

The only term of interest here is $\alpha(00)$; $\beta(00000)$, the only other nonzero term having no angular dependence, will be seen to be the first term in the expansion of the operator e^2/r_{12}:

$$\alpha^i(00)W^{(00)0} = -\frac{Ze^2}{2[l]}\left([l+\tfrac{1}{2}]\int\frac{(F_+^2+G_+^2)_i}{r_i}\,dr_i + [l-\tfrac{1}{2}]\int\frac{(F_-^2+G_-^2)_i}{r_i}\right)dr_i, \tag{30}$$

where $F_\pm$ stands for $F_{j=l\pm\frac{1}{2}}$, etc.

The expansion of Eq. (30) in orders of $(v/c)^2$ is based on Eq. (6). We define $E_0^i = W^i + mc^2$, and write Eq. (6) as

$$G_i = \frac{\hbar}{2mc}\left\{1 + \frac{W^i - U(r_i)}{2mc^2}\right\}^{-1}\left(\frac{d}{dr_i} - \frac{\kappa}{r_i}\right)F_i. \tag{31}$$

The expansion of the expression in braces in powers of $(W - U)/2mc^2$ is roughly equivalent to an expansion in orders $(v/c)^2$. We need to consider only the first term in the expansion

$$G_i = \frac{\mu_0}{e}\left(\frac{d}{dr_i} - \frac{\kappa}{r_i}\right)F_i, \tag{32}$$

where $\mu_0 = e\hbar/2mc$. To this order, F satisfies the equation

$$\left[-\frac{\hbar^2}{2m}\left(\frac{d^2}{dr^2} - \frac{l(l+1)}{r^2}\right) + U(r)\right]F_i = W^iF_i \tag{33}$$

for both $j = l + \frac{1}{2}$ and $j = l - \frac{1}{2}$ states; Eq. (33) is just the radial Schrödinger wave equation for a particle in a central field. The normalization used in this limit is $\int F^2\,dr = 1$.

In this order of approximation, the term containing F^2 in Eq. (30) becomes

$$-Ze^2\int\frac{F_i^2}{r_i^2}\,dr_i. \tag{34}$$

The term in G^2 can be obtained by use of a general relationship obtained from Eq. (32),

$$\int GVG\,dr=\frac{\mu_0^2}{e^2}\int F\Big\{-\frac{dV}{dr}\frac{dF}{dr}+\Big[\frac{\kappa}{r}\frac{dV}{dr}-V\Big(\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2}\Big)\Big]F\Big\}\,dr, \tag{35}$$

where V is any function of r. The term containing G^2 then becomes

$$\frac{\mu_0^2}{e^2}\int F\Big[\tfrac{1}{2}\nabla^2\Big(-\frac{Ze^2}{r_i}\Big)+\frac{Ze^2}{r_i}\Big(\frac{d^2}{dr_i^2}-\frac{l(l+1)}{r_i^2}\Big)\Big]F_i\,dr_i. \tag{36}$$

This term is discussed further in the next section.

B. Coulomb-Repulsion Terms

The Coulomb-repulsion Hamiltonian, e^2/r_{12}, can be written as

$$2e^2\sum_K\begin{pmatrix}l & K & l\\ 0 & 0 & 0\end{pmatrix}^2\frac{[l]^2}{[K]}\frac{r_<^K}{r_>^{K+1}}(\mathbf{w}_1^{(0K)K}\cdot\mathbf{w}_2^{(0K)K}). \tag{37}$$

Only O_β has terms with this angular dependence; the equivalent operator for this interaction, $O_{\text{C.r.}}$, can therefore be written

$$O_{\text{C.r.}}=\sum_K\beta(0K0KK)(\mathbf{w}_1^{(0K)K}\cdot\mathbf{w}_2^{(0K)K}). \tag{38}$$

The first nonvanishing term in the expansion of $O_{\text{C.r.}}$ is exactly Eq. (37). The second nonvanishing term is

$$\frac{\mu_0^2}{e^2}\iint\Big\{F_1^2F_2^2\tfrac{1}{2}\nabla^2U'-U'\times\sum_{i\neq j=1,2}F_i^2F_j\Big(\frac{d^2}{dr_j^2}-\frac{l(l+1)}{r_j^2}\Big)F_j\Big\}\,dr_1\,dr_2, \tag{39}$$

where $U'=e^2/r_{12}$ and $\nabla^2=\nabla_1^2+\nabla_2^2$.

When evaluated in this limit, the matrix element of the term $\sum_i(E^i-\beta_imc^2)$ contains, in addition to the nonrelativistic energy, a component of the order μ_0^2/e^2. This component is given by

$$\sum_i(\mu_0^2/e^2)(W^i+E_1^i)\Big(\frac{d^2}{dr_i^2}-\frac{l(l+1)}{r_i^2}\Big). \tag{40}$$

Combining this expression with Eqs. (36) (summed over i) and (39), one obtains

$$\iint F_1^2F_2^2\Big(\tfrac{1}{2}\nabla^2V-\tfrac{1}{8}p^4\frac{1}{m^3c^2}\Big)\,dr_1\,dr_2, \tag{41}$$

where

$$p^4=(p_1^4+p_2^4)\quad\text{and}\quad V=-\frac{Ze^2}{r_1}-\frac{Ze^2}{r_2}+\frac{e^2}{r_{12}}.$$

To obtain Eq. (41), we have made the approximation that

$$W^i+E^i+\frac{Ze^2}{r_i}-\frac{e^2}{r_{ij}}=\frac{P_i^2}{2m}.$$

The first term in Eq. (41) is the Darwin term[8] for two electrons; the second, the mass correction term.

C. Spin–Orbit Terms

The spin–orbit Hamiltonian can be written as

$$\mathcal{H}_{\text{s.o.}}=-a_{\text{s.o.}}[\tfrac{1}{2}l(l+1)(2l+1)]^{\frac{1}{2}}\mathbf{w}^{(11)0}, \tag{42}$$

where

$$a_{\text{s.o.}}=\frac{\hbar^2}{2m^2c^2}\frac{1}{r}\frac{dU(r)}{dr}.$$

Because

$$(\mathbf{w}_1^{(11)0}\cdot\mathbf{w}_2^{(00)0})=(2[l])^{-\frac{1}{2}}\mathbf{w}_1^{(11)0},$$

both O_α and O_β contain terms having the angular dependence $\mathbf{w}^{(11)0}$. The relativistic spin–orbit constant is then given by

$$\begin{aligned}a_{\text{s.o.}}^{\text{rel}}(i)&=-\Big[\frac{2}{l(l+1)(2l+1)}\Big]^{\frac{1}{2}}\\&\quad\times[\alpha^i(110)+(2[l])^{-\frac{1}{2}}\beta(11000)]\\&=\frac{2}{[l]}\Big[\int F_+V_{\text{rel}}F_++G_+V_{\text{rel}}G_+)_i\,dr_i\\&\quad-\int(F_-V_{\text{rel}}F_-+G_-V_{\text{rel}}G_-)_i\,dr_i\Big].\end{aligned} \tag{43}$$

V_{rel} is a "relativistic potential energy" given by

$$V_{\text{rel}}(r_1)=-\frac{Ze^2}{r_1}+\frac{e^2}{2[l]}\int_0^\infty[(2l+2)(F_+^2+G_+^2)_2+2l(F_-^2+G_-^2)_2]\frac{1}{r_>}\,dr_2, \tag{44}$$

where $r_>$ is the larger of r_1, r_2. In the limit discussed above, the second term on the right of (44) becomes the integral over r_2 of the potential energy of a charge at r_1 due to a spherically averaged charged shell at r_2. The relativistic spin–orbit term reduces to $a_{\text{s.o.}}$ in the nonrelativistic limit.

[8] A. Messiah, *Quantum Mechanics*, translated by J. Potter (North-Holland Publishing Company, Amsterdam, 1963), Vol. II.

D. Orbit–Orbit Terms

The orbit–orbit interaction can be written as[9]

$$\mathcal{H}_{o.o.} = -16\mu_0^2 \sum_K \frac{(2K+1)}{(K+2)} \langle l||\, C^K \,||l\rangle^2 \times (l)(l+1)(2l+1)\begin{Bmatrix} K & K+1 & 1 \\ l & l & l \end{Bmatrix}^2 \times \int_0^\infty \int_{r_2}^\infty R_1^2 R_2^2 \frac{r_2^K}{r_1^{K+3}}\, dr_1\, dr_2\, (\mathbf{w}_1^{(0K+1)K+1} \cdot \mathbf{w}_2^{(0K+1)K+1}). \tag{45}$$

The equivalent operator for this interaction, $O_{o.o.}$, is given by the terms in O_γ and O_δ with the same angular dependence as $\mathcal{H}_{o.o.}$:

$$O_{o.o.} = \sum_K \{\gamma(0\ K{+}1\ 0\ K{+}1\ K{+}1) + \delta(0\ K{+}1\ 0\ K{+}1\ K{+}1)\} \times (\mathbf{w}_1^{(0K+1)K+1} \cdot \mathbf{w}_2^{(0K+1)K+1}). \tag{46}$$

Only the terms in this sum with K even will be nonzero. In expanding $O_{o.o.}$, one finds that the first nonvanishing term is just $\mathcal{H}_o$.

E. Spin–Other–Orbit Terms

The spin–other–orbit interaction can be written[10]

$$\mathcal{H}_{s.o.o.} = 2 \sum_K [(K+1)(2l+K+2)(2l-K)]^{\frac{1}{2}} \times (-)^{K+1}[K+1]^{-\frac{1}{2}} (\mathbf{w}^{(0K+1)K+1} \cdot \mathbf{w}^{(1K)K+1}) \times \{M^{K-1}\langle l||\, C^{K+1} \,||l\rangle^2 + 2M^K \langle l||\, C^K \,||l\rangle^2\} + (-)^K [K]^{-\frac{1}{2}} (\mathbf{w}^{(0K)K} \cdot \mathbf{w}^{(1K+1)K}) \times \{M^K \langle l||\, C^K \,||l\rangle^2 + 2M^{K-1}\langle l||\, C^{K+1} \,||l\rangle^2\}], \tag{47}$$

where the M^K are the angular integrals of Marvin.[11] The sum over K falls into two parts, the sum over K even and the sum over K odd. For K even, terms in the equivalent operator, $O_{s.o.o.}$, with the angular dependence $(\mathbf{w}^{(0K+1)K+1} \cdot \mathbf{w}^{(1K)K+1}$, will arise from O_γ and O_δ; with $(\mathbf{w}^{(0K)K} \cdot \mathbf{w}^{(1K+1)K})$, from O_β. For K odd the situation is reversed. The equivalent operator is given by

$$O_{s.o.o.} = \sum_K \{[\beta(0\ K\ 1\ K{+}1\ K) + \gamma(0\ K\ 1\ K{+}1\ K) + \delta(0\ K\ 1\ K{+}1\ K)](\mathbf{w}_1^{(0K)K} \cdot \mathbf{w}_2^{(1K+1)K}) + [\beta(0\ K{+}1\ 1\ K\ K{+}1) + \gamma(0\ K{+}1\ 1\ K\ K{+}1) + \delta(0\ K{+}1\ 1\ K\ K{+}1)](\mathbf{w}_1^{(0K+1)K+1} \cdot \mathbf{w}_2^{(1K)K+1})\}. \tag{48}$$

The first nonvanishing term in the expansion of Eq. (48) is $\mathcal{H}_{s.o.o.}$.

F. Spin–Spin Terms

The spin-spin Hamiltonian is given by[10]

$$\mathcal{H}_{s.s.} = 2(5)^{\frac{1}{2}}\mu_0^2 \sum_K [(2K+4)(2K+3)(2K+2)]^{\frac{1}{2}} \begin{Bmatrix} 1 & 1 & 2 \\ K+2 & K & K+1 \end{Bmatrix} \times \langle l||\, C^K \,||l\rangle \langle l||\, C^{K+2} \,||l\rangle \int_0^\infty \int_0^{r_2} R_1^2 R_2^2 \frac{r_1^K}{r_2^{K+3}}\, dr_1\, dr_2\, (\mathbf{w}_1^{(1K+2)} \cdot \mathbf{w}_2^{(1K)K+1}). \tag{49}$$

The equivalent operator for this Hamiltonian, $O_{s.s.}$, comes from O_γ and O_δ, and is given by

$$O_{s.s.} = \sum_K \{\gamma(1\ K{+}2\ 1\ K\ K{+}1) + \delta(1\ K{+}2\ 1\ K\ K{+}1)\}(\mathbf{w}_1^{(1K+2)K+1} \cdot \mathbf{w}_2^{(1K)K+1}). \tag{50}$$

The only nonzero terms in this sum will occur for K even.

Upon expanding the expression for $O_{s.s.}$, we find that the first nonvanishing term is given by Eq. (49) plus the additional term

$$4\mu_0^2 \frac{[(K+1)(K+2)]^{\frac{1}{2}}}{(2K+3)} \langle l||\, C^K \,||l\rangle \langle l||\, C^{K+2} \,||l\rangle \int_0^\infty \frac{F_1^4}{r_1^2}\, dr_1\, (\mathbf{w}_1^{(1K+2)K+1} \cdot \mathbf{w}_2^{(1K)K+1}). \tag{51}$$

The radial part of this additional expression is of the form of a delta-function between $\mathbf{r}_1$ and $\mathbf{r}_2$; this term is discussed further in the next section.

---

[9] C. W. Ufford and H. B. Callen, Phys. Rev. **110**, 1352 (1958).
[10] B. R. Judd, Physica (to be published).
[11] H. H. Marvin, Phys. Rev. **71**, 102 (1947).

G. Spin–Spin Contact Terms

The spin–spin contact Hamiltonian[12] is given by

$$\mathcal{H}_{s.s.c.} = -\tfrac{1}{3}32\pi\mu_0^2(\mathbf{s}_1\cdot\mathbf{s}_2)\,\delta(\mathbf{r}_1-\mathbf{r}_2)$$
$$= \frac{4\mu_0^2}{3r^2}\,\delta(r_1-r_2)\sum_{K\beta}(-)^{K+\beta}\langle l||\,C^K\,||l\rangle^2$$
$$\times(\mathbf{w}^{(1K)\beta}\cdot\mathbf{w}^{(1K)\beta}), \tag{52}$$

where we have used[13]

$$\delta(\mathbf{r}_1-\mathbf{r}_2) = \delta(r_1-r_2)\frac{1}{4\pi r^2}\sum_K[K](C_1^K\cdot C_2^K).$$

Again, the equivalent operator for this interaction $O_{s\,s.c.}$ comes from O_γ and O_δ,

$$O_{s.s.c.} = \sum_{K\beta}\{\gamma(1\,K\,1\,K\,\beta) + \delta(1\,K\,1\,K\,\beta)\}(\mathbf{w}^{(1K)\beta}\cdot\mathbf{w}^{(1K)\beta}).$$

The only nonzero terms in this expansion occur for K even.

Upon expanding $O_{s.s.c.}$, we find that the first nonvanishing term is given by $\mathcal{H}_{s.s.s.}$ plus some additional terms whose values depend on β. The additional terms are for $\beta = K+1$,

$$\frac{2K\mu_0^2}{3(2K+3)}\langle l||\,C^K\,||l\rangle^2 \times\int\frac{F_1^4}{r_1^2}\,dr_1\,(\mathbf{w}_1^{(1K)K+1}\cdot\mathbf{w}_2^{(1K)K+1}); \tag{53a}$$

for $\beta = K-1$,

$$\frac{2(K+1)\mu_0^2}{3(2K-1)}\langle l||\,C^K\,||l\rangle^2 \times\int\frac{F_1^4}{r_1^2}\,dr_1\,(\mathbf{w}_1^{(1K)K-1}\cdot\mathbf{w}_2^{(1K)K-1}); \tag{53b}$$

and for $\beta = K$,

$$\tfrac{2}{3}\mu_0^2\langle l||\,C^K\,||l\rangle^2\int\frac{F_1^4}{r_1^2}\,dr_1\,(\mathbf{w}_1^{(1K)K}\cdot\mathbf{w}_2^{(1K)K}). \tag{53c}$$

The additional contributions to the spin–spin Hamiltonian found by expanding the equivalent operators in powers of $(v/c)^2$ [Eqs. (51) and (53)] can be included in the Hamiltonian by adding the term

$$\mathcal{H}'_{s.s.c.} = -\frac{16\pi}{3}\mu_0^2\,\delta(\mathbf{r}_1-\mathbf{r}_2) \times\left[\mathbf{s}_1\cdot\mathbf{s}_2-\frac{3(\mathbf{s}_1\cdot\mathbf{r})(\mathbf{s}_2\cdot\mathbf{r})}{r^2}\right]. \tag{54}$$

This operator has not been obtained in previous treatments[5,14] of the spin–spin interaction because earlier results have depended on the assumed shape of the infinitesimal region in which the electrons overlap. The situation is highly analogous to that which exists with respect to the Fermi contact term[15] in hyperfine structure. Judd[7] has found that $\mathcal{H}'_{s.s.c.}$ can be obtained by use of classical electromagnetic theory if the electron spin moments are replaced by currents, as suggested by Casimir.[16] If one uses this method, the result does not depend on the shape of the infinitesimal volume surrounding one of the electrons. Judd[7] has also obtained $\mathcal{H}'_{s.s.c.}$ by the method of Bethe and Salpeter,[5] assuming that electron 1 is excluded from, and electron 2 confined between, two concentric spheres which collapse, in the limit, to a common radius.

Unfortunately, $\mathcal{H}'_{s.s.c.}$, which can be written as

$$\mathcal{H}'_{s.s.c.} = \frac{4(5)^{\frac{1}{2}}\mu_0^2}{r^2}\sum_{K\delta}(-)^{K+\delta}[K,\,\delta] \times\begin{pmatrix}1&1&2\\0&0&0\end{pmatrix}\begin{pmatrix}K&\delta&2\\0&0&0\end{pmatrix}((\mathbf{s}_1\mathbf{s}_2)^2(\mathbf{C}_1^K\mathbf{C}_2^\delta)^2)^0, \tag{55}$$

can be shown to always give zero total contribution to the energy. That is, when the matrix element of $\mathcal{H}'_{s.s.c.}$ is taken between the states $|SL\rangle$ and $|S'L'\rangle$, the sum over K and δ can be performed, producing a result which depends on the product

$$\begin{Bmatrix}\frac{1}{2}&\frac{1}{2}&1\\ \frac{1}{2}&\frac{1}{2}&1\\ S&S'&2\end{Bmatrix}\begin{pmatrix}L'&l&l\\0&0&0\end{pmatrix}\begin{pmatrix}L&l&l\\0&0&0\end{pmatrix}.$$

For this product not to be trivially zero, $S = S' = 1$, and L, L' must be even; such a state, however, would violate the Pauli principle. It can also be shown that $\mathcal{H}'_{s.s.c.}$ makes zero contribution when evaluated between wavefunctions arising from mixed configurations.[7]

H. Other Terms

There are three more distinct operators in O_s which have not been discussed. These are

$$O_1 = \sum_K\beta(1\,K{+}1\,1\,K{+}1\,K) \times(\mathbf{w}_1^{(1K+1)K}\cdot\mathbf{w}_2^{(1K+1)K}),$$

[12] J. C. Slater, *Quantum Theory of Atomic Structure* (McGraw-Hill Book Company, Inc., New York, 1960), Vol. II.
[13] D. M. Brink and G. R. Satchler, *Angular Momentum* (Clarendon Press of Oxford University Press, Inc., New York, 1962).
[14] A. M. Sessler and H. M. Foley, Phys. Rev. **92**, 1321 (1953).
[15] E. Fermi, Z. Physik **60**, 370 (1930).
[16] H. B. G. Casimir, *On the Interaction Between Atomic Nuclei and Electrons* (W. H. Freeman and Company, San Francisco, 1963).

TABLE I. Terms in O_e classified according to corresponding fine-structure interaction. Numbers in first column are KK as defined in Sec. III A. Numbers in second and third columns are k_1, K_1, k_2, K_2, k as defined in Secs. III B–D.

O_α	O_β	O_γ and O_δ	Interaction
0 0			$-Ze^2/r$
1 1	1 1 0 0 0		spin–orbit
	0 K 0 K 0 (K even)		e^2/r_{12}
		1 K 1 K+2 K+1 (K even)	spin–spin
		0 K+1 0 K+1 K+1 (K even)	orbit–orbit
		0 K+1 1 K K+1 (K even)	spin–other–orbit
		0 K 1 K+1 K (K odd)	spin–other–orbit
	0 K 1 K+1 K (K even)		spin–other–orbit
	0 K+1 1 K K+1 (K odd)		spin–other–orbit
		1 K 1 K K+1 (K even)	spin–spin contact
		1 K 1 K K (K even)	spin–spin contact
		1 K 1 K K−1 (K even)	spin–spin contact
	1 K+1 1 K+1 K (K even)		
	1 K+1 1 K−1 K (K even)		
	1 K−1 1 K−1 K (K even)		

$$O_2 = \sum_K \beta(1\ K{+}1\ 1\ K{-}1\ K) \times (\mathbf{w}_1^{(1K+1)K} \cdot \mathbf{w}_2^{(1K-1)K}),$$

and

$$O_3 = \sum_K \beta(1\ K{-}1\ 1\ K{-}1\ K) \times (\mathbf{w}_1^{(1K-1)K} \cdot \mathbf{w}_2^{(1K-1)K}).$$

Upon expanding these expression, we find that none has any nonvanishing terms to order μ_0^4/e^4.

V. DISCUSSION

Table I reviews some of the results of the preceding section. In it, the terms in O_e are classified according to the type of fine-structure interaction produced. In the parts of the spin–spin, spin–other–orbit, and orbit–orbit interactions arising from O_γ and O_δ, the angular dependence of each electron is given by $W^{(\alpha\beta)K}$, where K is odd. As was shown in Secs. IIIC and D, in this case $O_\gamma = O_\delta$. In the nonrelativistic limit, the contributions from O_γ and O_δ to the spin–spin contact terms are also equal; this is not the case in the relativistic limit, however.

As mentioned in Sec. IIIC, the values of O_e do not depend on the particular type of coupling assumed; this implies that the equations for O_e are valid for any two electrons in a configuration l^n. This in turn implies that the equivalent operator for the configuration l^n can be obtained by replacing the indices 1, 2 in O_e by i, j and performing the sums $\sum_{i=1}^{n} O_\alpha^i$ and $\sum_{i>j} (O_\beta + O_\gamma + O_\delta)$.

Using the operators obtained above and the relativistic Hartree–Fock wavefunctions, then, one can calculate in a straightforward manner the value of a particular fine-structure interaction in the configuration l^n. The evaluation of the angular terms is carried out in the nonrelativistic scheme, where the powerful tensor techniques of Racah[17] can be easily utilized. The methods used to obtain these operators can also be used to obtain operators valid for application to mixed configurations.

ACKNOWLEDGMENT

I wish to thank Dr. Brian Judd for stimulating discussions and for his critical reading of the manuscript for this paper.

[17] G. Racah, Phys. Rev. **76**, 1352 (1949); B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, Inc., New York, 1963).

The Validity of Perturbation Series with Zero Radius of Convergence*

GEORGE A. BAKER, JR.
Applied Mathematics Department, Brookhaven National Laboratory, Upton, Long Island, New York
AND
ROY CHISHOLM
University of Kent, Canterbury, England, and Texas A. & M. University, College Station, Texas
(Received 22 March 1966)

Although the perturbation expansion for the S-matrix of the Peres-model field theory has zero radius of convergence, it uniquely defines the S-matrix and is easily summable by the method of Padé approximants.

1. INTRODUCTION

RECENTLY, the use of Padé approximants for analytically continuing power series has been applied extensively and has become better understood theoretically.[1,2] In particular, the technique has been shown to be applicable in potential theory for certain classes of potentials.[3] It would be of immense value if the Padé method could be applied in strong-coupling field theory. The technique seems particularly well-suited to the calculation of resonances and perhaps form factors; one cannot hope to calculate scattering-matrix elements directly from the perturbation series, because these need to be evaluated just where Padé approximants will simulate branch cuts, but it might be possible to calculate the discontinuities across these cuts.

The likelihood of the nonconvergence of field-theory perturbation series was established by Hurst.[4] Dyson[5] pointed out that a zero radius of convergence of the electrodynamic perturbation series was consistent with physical instability obtained by allowing the electric charge to be analytically continued to a purely imaginary value. The same phenomenon was noticed by one of us for the many-fermion problem, where it corresponds physically to collapse.[6] Peres[7] has invented a simple soluble "model field theory" which resembles realistic field theories both in the form of interaction and in allowing the analog of pair creation; these characteristics are not present in potential theory, so that the model exhibits some properties of realistic field theories which have not yet been treated by the Padé method.

Peres has shown that one matrix element in the model theory is a Bessel function whose expansion in powers of the coupling constant g has zero radius of convergence. His deduction from this fact, that the perturbation series is useless, is however wrong; we show that this series is in fact a series of Stieltjes and is therefore summable by the Padé method; the Padé approximants have poles in the g-plane only along the imaginary axis, where the Bessel function is singular. We show further that every matrix element in the model theory is the sum of terms whose perturbation expansions have the same properties. Thus, any matrix element can be calculated by the Padé method, and it is established that the only singularities in the g-plane lie on the imaginary axis, and correspond physically to the "Dyson disaster".

2. MATRIX ELEMENTS FOR PERES' MODEL

The two "fields" in the model are harmonic oscillators of the same frequency, described by operators x and y. The interaction is taken as

$$g\delta(t)x^2y, \tag{2.1}$$

analogous to an instantaneous interaction of form $\bar{\psi}\psi\varphi$ between a fermion field ψ and a boson field φ. The operators x and y can cause single quantum jumps (creations or annihilations) between the oscillator states, which are of the form

$$\Psi_m(x) = C_m e^{-\frac{1}{2}x^2} H_m(x), \tag{2.2}$$

where H_m is a Hermite polynomial and C_m are constants. The general S-matrix element is given by Eq. (6) of Ref. 7 and is a real multiple of

$$I = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, H_r(x)H_s(y)H_m(x)H_n(y) \times \exp(-x^2 - y^2 - igx^2y). \tag{2.3}$$

* This research was sponsored in part by the European Office of Aerospace Research, United States Air Force, Grant No. AF EOAR 64-61, and in part by the U. S. Atomic Energy Commission.

[1] G. A. Baker, Jr., *Progress in Theoretical Physics* (Academic Press Inc., New York, 1965), Vol. I, p. 1.
[2] J. C. R. Chisholm, J. Math. Phys. 7, 39 (1966).
[3] J. S. R. Chisholm, J. Math. Phys. 4, 1506 (1963), Ref.1, Sec. IV B.
[4] C. A. Hurst, Phys. Rev. 85, 920L (1952).
[5] F. J. Dyson, Phys. Rev. 85, 631 (1952).
[6] G. A. Baker, Jr., Phys. Rev. 131, 1869 (1963).
[7] A. Peres, J. Math. Phys. 4, 332 (1963).

This integral is zero unless r and m are either both even or both odd; in this case

$$I = \sum \alpha_{pq} I_{pq'}, \tag{2.4}$$

where α_{pq} are real numbers, and

$$I_{pq} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, x^{2p} y^q \times \exp(-x^2 - y^2 - igx^2 y), \tag{2.5}$$

with p and q nonnegative integers. A particular integral I depends on integrals I_{pq} which all have q even or all have q odd. Performing the integration over x in (2.5), we have

$$I_{pq} = \pi^{\frac{1}{2}} 2^{-p} (2p-1)!! \times \int_{-\infty}^{\infty} dy \exp(-y^2) y^q (1+igy)^{-p-\frac{1}{2}}.$$

Integrating by parts p times,

$$I_{pq} = \frac{\pi^{\frac{1}{2}}}{(ig)^p} \times \int_{-\infty}^{\infty} dy \frac{d^p}{dy^p} (\exp(-y^2) y^q)(1+igy)^{-\frac{1}{2}}. \tag{2.6}$$

The integral in (2.6) is a sum of terms, with real coefficients, of the form

$$J_{l\lambda} \equiv \int_{-\infty}^{\infty} dy \exp(-y^2) y^l (1+igy)^{-\lambda}$$

$$= \int_0^{\infty} dy \exp(-y^2) y^l [(1+igy)^{-\lambda} + (-1)^l (1-igy)^{-\lambda}]. \tag{2.7}$$

The values of l arising in a particular integral (2.6) are either all even or all odd, depending on the parity of q.

The integrand in (2.7) can be expanded by the binomial theorem and integrated term by term. The result is always a series with zero radius of convergence; for example, the ground-state-to-ground-state matrix element, discussed by Peres, has $l = 0$ in (2.7). The series expansion is then

$$J_0 = \pi^{\frac{1}{2}} \sum_{r=0}^{\infty} (-g^2)^r (4r!/2^{6r} r! 2r!). \tag{2.8}$$

With $l = 1$, (2.7) gives

$$J_1 = -ig\pi^{\frac{1}{2}} \sum_{r=0}^{\infty} (-g^2)^r \frac{(4r+2)!}{2^{6r} r!(2r+1)!}. \tag{2.9}$$

When l is odd, J_l always contains an odd imaginary factor ig.

We note that the series in (2.8) is very similar to the series derived from the Euler function:

$$\int_0^{\infty} \frac{e^{-t}\, dt}{1+g^2 t} = \sum_{r=0}^{\infty} (-g^2)^r r!. \tag{2.10}$$

The latter series is a series of Stieltjes, so that the diagonal Padé approximants converge monotonically from above to the Euler function on the positive real axis and converge for all complex values of g^2 except on the line $-\infty < g^2 \leq 0$, where the function is singular. Monotonically increasing lower bounds on the positive real axis are given by the $[N, N-1]$ approximants. We now show that these properties are shared by the series expansions of the functions J_l. The rate of divergence is only $n!$, and so the function defined is unique (Theorem 7, Ref. 1).

3. IDENTIFICATION AS STIELTJES SERIES

Following Wall,[8] one can show by using Cauchy's theorem that sufficient conditions for a series expansion of a function $f(g^2)$ in powers of g^2 to be a series of Stieltjes are:

(i) $f(g^2)$ is regular except for the real range of values $-\infty < g^2 \leq 0$, on which line Im $f(g^2)$ may be discontinuous, but finite,

(ii) Im $f(g^2) \lesseqgtr 0$ if Im $g^2 \gtreqless 0$,

(iii) $|f(g^2)| \to 0$ uniformly as $g^2 \to \infty$ in the cut plane.

The functions (2.7) are clearly regular unless g is purely imaginary, where a rotation of the integration contour sufficies to prove finiteness, and so satisfy (i). It is not difficult to check that condition (iii) is also satisfied.

Remembering that J_l contains an unpaired factor ig when l is odd, as in (2.9), condition (ii) must be checked for

$$f(g^2) = J_l = \int_0^{\infty} dy \exp(-y^2) y^l [(1+igy)^{-\frac{1}{2}} + (1-igy)^{-\frac{1}{2}}] \quad (l \text{ even}), \tag{3.1}$$

$$f(g^2) = iJ_l/g = \int_0^{\infty} dy \exp(-y^2) y^l i[(1+igy)^{-\frac{1}{2}} - (1-igy)^{-\frac{1}{2}}]/g \quad (l \text{ odd}). \tag{3.2}$$

Taking $0 < \arg g < \frac{1}{2}\pi$, so that $\frac{1}{2}\pi < \arg(igy) < \pi$, we have

$$|(1+igy)^{-\frac{1}{2}}| > |(1-igy)^{-\frac{1}{2}}|$$

[8] H. S. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1948), Sec. 66.

and

$$0 < \arg (1 - igy)^{-\frac{1}{2}} < -\arg (1 + igy)^{-\frac{1}{2}} < \tfrac{1}{2}\pi.$$

Thus the imaginary parts of the integrands of (3.1) and (3.2) are each negative, and condition (ii) is satisfied.

For (3.2) we may simply verify this condition for the boundary of the cut plane, and note that because of (i) and (iii) Im(f) reaches its maximum (and minimum) values on the boundary and so is of definite sign throughout each half-plane (in g^2). Thus the series in g^2 derived by expanding (2.7) are series of Stieltjes, for all integers $l \geq 0$. It is to be noted that the sum of two series of Stieltjes is again a series of Stieltjes.

In fact, one need only consider the integral (2.7) for $l = 0$ and $l = 1$, since J_l for $l \geq 2$ can be expressed in terms of J_0 and J_1 through the recurrence relation

$$I_{l+2} = ig^{-1}[I_{l+1} - \tfrac{1}{2}lI_{l-1}] + \tfrac{1}{4}(2l + 1)I_l.$$

4. CONCLUSIONS

For Peres model field theory, we have shown that:

(a) each matrix element is a linear combination of integrals of the form (2.7), with l a nonnegative integer;

(b) each matrix element is singular for pure imaginary values of the coupling constant g, and for no other values;

(c) the expansion of every integral (2.7) in powers of g has zero radius of convergence, but is a multiple of a series of Stieltjes in g^2; and thus

(d) the diagonal Padé approximants formed from the series expansion of (2.7) in powers of g^2 will converge to the function (2.7), and will have singularities only for g^2 real and negative.

The singularity structure of a realistic field theory is vastly more complex than that of this model. Nevertheless, this simple example shows that use of a perturbation series can be reconciled with the occurrence of the "Dyson disaster" and the resultant zero radius of convergence of the series.

APPENDIX

In order to illustrate the smooth and relatively rapid convergence of the Padé approximants, we give in Table I the first few which can be formed from the coefficients through g^{12} for $S_{0,0}$. For comparison we have included the partial sums for $g^2 = 1$. The $[N, N]$ Padé approximants from monotonically-decreasing upper bounds and the $[N, N - 1]$ form monotonically-increasing lower bounds over the range $0 \leq g^2 < +\infty$. Both bounds converge to the limiting function in the cut $(-\infty, 0)$ complex plane. It will be observed that even with only 6 terms we can obtain an accuracy at $g^2 = 1$ of better than one percent, even though the Taylor series is diverging drastically.

TABLE I. Padé analysis of $S_{0,0}$.

	$g^2 = 1$	3	10	Partial sum ($g^2 = 1$)
[0, 0]	1.0	1.0	1.0	1.0
[1, 0]	0.842105	0.64	0.347826	0.8125
[1, 1]	0.910448	0.868613	0.842932	1.01758
[2, 1]	0.883198	0.740263	0.477577	0.594604
[2, 2]	0.897712	0.831876	0.781703	1.88335
[3, 2]	0.890418	0.772340	0.543369	−3.31997
[3, 3]	0.894720	0.817125	0.748844	22.85923

Unitary and Antiunitary Ray Representations of the Product of n Commuting Parity Operators*

A. O. BARUT AND S. KOMY†
Department of Physics, University of Colorado, Boulder, Colorado
(Received 1 March 1966)

The unitary and antiunitary ray representations of the group of n commuting operators Γ_i, $i = 1, 2, \cdots, n$, satisfying $\Gamma_i^2 = 1$, $\Gamma_i\Gamma_j = \Gamma_j\Gamma_i$, have been determined explicitly. The ray representations are shown to be homomorphic to the representations of a class of generalized Clifford algebras defined in the text. The enumeration of different algebras gives a complete classification of inequivalent ray representations. Whereas the irreducible vector representations of the group are all one-dimensional, the irreducible ray representations lead to various multiplet structure of states ranging in dimensions from 1 to $2^{\frac{1}{2}n}$ (n even), or $2^{\frac{1}{2}(n-1)}$ (n odd), in powers of two. The arbitrary antiunitary case can be reduced to the case with only one antiunitary Γ.

I. INTRODUCTION

BY a general parity operator we mean an operator whose square is the identity operator. The full inhomogeneous Lorentz group has two abstractly defined parity operators: a unitary parity operator corresponding to space reflections and an antiunitary parity operator corresponding to time reflections. The physical interpretation of these operators may depend on the physical situation. For example, the space reflection operator may be interpreted as P or as PC.[1] In addition, we have other parities, like charge conjugation C, G parity (isospin parity). The P, C, and T operators may even be different in strong, weak, and electromagnetic interactions.[2] Thus, we have the problem of a general quantum-mechanical treatment of the representations of a number of parity operators. The product of two commuting unitary parities have been considered in the literature.[3,4] We discuss in this paper the ray representations, both unitary and antiunitary, of n parity operators.

It is remarkable that the ray representations of this simple finite group lead to such a richness of different classes and each class to different *types* of representations. It is also remarkable that one gets higher-dimensional irreducible representations, that is, multiplet structure of states, much the same way that one is used to obtain from the representations of compact symmetry groups. The present paper deals with the mathematical theory of the subject; the physical applications will be treated separately.[5]

II. UNITARY RAY REPRESENTATIONS

Let the group G be generated by n elements $\Gamma_1, \Gamma_2, \cdots, \Gamma_n$ such that

$$\Gamma_i^2 = 1, \tag{1}$$

$$\Gamma_i\Gamma_j = \Gamma_j\Gamma_i. \tag{2}$$

The group consists of I, Γ_i, and all the possible distinct products of Γ_i's, and is Abelian. Its elements are denoted collectively by g_i. We have

$$g_i^2 = 1, \tag{1'}$$

$$g_ig_j = g_jg_i. \tag{2'}$$

Let $U(g_i)$ denote the unitary ray representations. They satisfy

$$U(g_i)U(g_j) = \omega(i, j)U(g_ig_j),$$
$$U(1) = I. \tag{3}$$

Two representations $U(g)$ and $\alpha U(g)$ lead to equivalent factor systems. Hence, because of (1') the diagonal phases $\omega(i, i)$ can be chosen to be the identity so that

$$[U(g_i)]^2 = I. \tag{4}$$

In order to determine the remaining phases, we pass from the group law (3) to the commutation relations

$$U(g_i)U(g_j) = \omega(i, j)U(g_ig_j) = \omega(i, j)U(g_jg_i)$$
$$= \omega(i, j)\omega(j, i)^{-1}U(g_j)U(g_i) \equiv C_{ij}U(g_j)U(g_i),$$
$$C_{ij} = \omega(i, j)/\omega(j, i). \tag{5}$$

* Supported in part by the United States Air Force Office of Aerospace Research under AFOSR Grant AF-AFOSR-30-65.

† On leave from Edinburgh University, Edinburgh, Scotland.

[1] E. P. Wigner, *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gürsey, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1964).

[2] T. D. Lee, *Proceedings of the Oxford International Conference on Elementary Particles* (Rutherford High-Energy Laboratory, Cambridge, England, 1966).

[3] V. I. Ogievetskii and C. Kuang-Chao, Zh. Eksperim. i Teor. Fiz. **36**, 264 (1959) [English transl.: Soviet Phys.—JETP **9**, 179 (1959)]. N. Tarimer, Phys. Rev. **140**, B977 (1965).

[4] The general theory of projective representations of finite groups goes back to I. Schur, J. Reine u. Angew. Math. **132**, 85 (1907); **139**, 155 (1911).

[5] For an application to leptons see A. O. Barut, Phys. Rev. **147**, 978 (1966).

We refer to the phase factors C_{ij} as the "commutation factors." These factors C_{ij} cannot be absorbed into the U's because the multiplication of U by any factor α occurs on both sides of Eq. (5) and hence the factor α cancels. But the commutation factors C_{ij} must be restricted to ± 1. This follows from Eq. (5), for, if we multiply Eq. (5) from the left by $U(g_i)U(g_i)$, we obtain

$$\begin{aligned} I &= C_{ij}U(g_i)U(g_i)U(g_j)U(g_i) \\ &= C_{ij}C_{ij}[U(g_i)]^2[U(g_i)]^2 = C_{ij}^2, \end{aligned}$$

or,

$$C_{ij} = \pm 1. \tag{6}$$

How many independent commutation factors C_{ij} are there? Because the group is generated by the n elements $\Gamma_1, \cdots, \Gamma_n$, the commutation factors of these generating elements, i.e.,

$$\begin{aligned} U(\Gamma_i)U(\Gamma_j) &= \mathring{C}_{ij}U(\Gamma_j)U(\Gamma_i), \\ [U(\Gamma_i)]^2 &= I, \end{aligned} \tag{7}$$

determine all other factors C_{ij}. In fact, Eqs. (7) generate a *generalized Clifford algebra* with the $2n$ basis elements:

$$I;\ U(\Gamma_i)U(\Gamma_j);\ \cdots;\ U(\Gamma_1)U(\Gamma_2)\cdots U(\Gamma_n), \qquad i<j, \text{ etc.}, \tag{8}$$

which is just the basis of the representation of the group (Frobenius) algebra of the group G, Eq. (1). The ordinary Clifford algebra corresponds to all $\mathring{C}_{ij} = -1$. Indeed, for any two group elements g_i and g_j, we have, if we express g_i and g_j in terms of the products of Γ_i,

$$\begin{aligned} &U(g_i)U(g_j) \\ &= U(\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_N})U(\Gamma_{j_1}\Gamma_{j_2}\cdots\Gamma_{j_M}) \\ &= \Big(\prod_{l=1}^{M}\prod_{k=1}^{N}\mathring{C}_{i_k j_l}\Big)U(g_j)U(g_i) = C_{ij}U(g_j)U(g_i), \end{aligned} \tag{9}$$

which is exactly the same as the commutation relations of the two elements $U(\Gamma_{i_1})\cdots U(\Gamma_{i_N})$ and $U(\Gamma_{j_1})\cdots U(\Gamma_{j_M})$ of the algebra (8). We have thus proved the following theorem:

Theorem: The unitary ray representations of the group (1′) and (2′) satisfy the generalized Clifford algebra generated by (7) with the basis (8).

We can therefore determine first the representations of the algebra and then pass to the representations of the group. The representations of the commutation relations (7) differ from the ray representations of the group as follows: To a given representation of the Clifford algebra there corresponds as many ray representations of the group as there are phase factors ω_{ij} satisfying

$$\omega_{ij}/\omega_{ji} = C_{ij}.$$

This freedom in the choice of the ω_{ij} is, however, drastically restricted by the associative law of the group which we have not yet used. Indeed, if one multiplies Eq. (3) from the left by $U(g_ig_j)^{-1}$ and uses Eq. (4) and Eq. (2′), one gets

$$\begin{aligned} U(g_i)U(g_j)U(g_ig_j) &= \omega_{ij}, \\ U(g_i)U(g_j)U(g_jg_i) &= \omega_{ij}, \end{aligned}$$

or

$$\omega_{ij}^{-1}U(g_i)U(g_j)U(g_j)U(g_i) = \omega_{ij},$$

or

$$\omega_{ij}^{-1} = \omega_{ji}; \qquad C_{ij} = C_{ji}^{-1}. \tag{10}$$

Consequently,

$$C_{ij} = \omega_{ij}^2 = \pm 1, \qquad \omega_{ij} = \pm 1, \pm i. \tag{11}$$

It is now a relatively simple matter to pass from the representations of the commutation relations (i.e., Clifford algebra) to the ray representations of the group: Take a representation of the algebra with $\mathring{C}_{ij}$ fixed. From (11) the corresponding ω_{ij} take only two possible values $\pm(\mathring{C}_{ij})^{\frac{1}{2}}$. Now, quite generally, if $U(g)$ is a representation of the commutation relations, so is $\{\alpha_iU(g_i)\}$; and if $\{U(g_i)\}$ corresponds to the phase system ω_{ij} of the ray representations of the group, $\{\alpha_iU(g_i)\}$ corresponds to the equivalent phase system

$$\omega_{ij}(\alpha_i\alpha_i/\alpha_{ij}).$$

In our case, because of (4), we have $\alpha_i = \pm 1$. Note the difference between the *equivalence of phase systems* and the *equivalence of representations*. Two representations U and αU belong to equivalent phase systems, but they are in general not equivalent representations, that is, there exists no matrix S such that

$$SU(g)S^{-1} = \alpha U(g), \quad \text{for all } g. \tag{12}$$

The concept of equivalent phase systems tells us simply that if we have found one representation the other is trivially obtained by multiplication with a phase factor, like $U(\Gamma)$ and $-U(\Gamma)$. But the eigenvalues of the operator Γ in the two representations are of course opposite of each other.

We now discuss all the representations of the commutation relations (7), or the representations of the algebra with the basis (8). Some special cases

of this algebra are well known: If all $\mathring{C}_{ij} = +1$ we have the trivial case of an Albelian group and all irreducible representations are one-dimensional. If all $\mathring{C}_{ij} = -1$, the representation $U(g)$ forms a *bona fide* Clifford algebra whose representations are known. All other mixed cases take an intermediary position between these two extreme cases.

Because of Eq. (10), there are

$$\tfrac{1}{2}n(n-1) = K$$

distinct factors $\mathring{C}_{ij}$, hence *a priori* there are 2^K different types of ray representations, depending on which of the $\mathring{C}_{ij}$ are equal to $+1$ and which are equal to -1. If k of these factors are -1, they can be distributed in $\binom{K}{k}$ different ways among the K factors, and

$$\binom{K}{1} + \binom{K}{2} + \cdots + \binom{K}{k} + \cdots + \binom{K}{K} = 2^K.$$

Consider now the class with k of the $\mathring{C}_{ij}$ being equal to -1. We distinguish two cases: (i) The k minus signs are so distributed that we have a sub-Clifford algebra with r generating elements, $k = \frac{1}{2}r(r-1)$; (ii) there is no subalgebra which is a *bona fide* Clifford algebra. For example, if $k = 3$, the case (i) corresponds to $\mathring{C}_{12} = \mathring{C}_{13} = \mathring{C}_{23} = -1$, then Γ_1, Γ_2, Γ_3 form a Clifford algebra; case (ii) corresponds, say, to $\mathring{C}_{12} = \mathring{C}_{13} = \mathring{C}_{14} = -1$, that is, Γ_1 and Γ_2 anticommuting, but Γ_2 and Γ_3 commuting. It is clear that we have the case (i) only if k is of the form $k = \frac{1}{2}r(r-1)$, $r = 1, 2, 3, 4, \cdots n$.

Case (i)

The representations are determined by those of the Clifford subalgebra with r generating elements. The remaining elements of the algebra commute with each other and commute with this subalgebra and therefore do not increase the dimensionality of the representation.

The representations of the Clifford algebra with r generating elements ($r = 1, 2, \cdots, n$) have been first determined by Jordan and Wigner[6] for even r. For even r there is a single irreducible representation (up to a unitary equivalence) of the Clifford subalgebra of degree $2^{\frac{1}{2}r}$. Every other representation is completely reducible, faithful and of degree which is a multiple of $2^{\frac{1}{2}r}$. For odd r a similar analysis exists.[7] In this case there are two inequivalent, but nonfaithful, representations of degree $2^{\frac{1}{2}(r-1)}$; the faithful representation is the direct sum of these two and, therefore, of degree $2^{\frac{1}{2}(r+1)}$. Again every other representation is completely reducible and its degree is a multiple of $2^{\frac{1}{2}(r-1)}$.

Case (ii)

Except the cases where, by relabeling the elements, we may obtain a Clifford subalgebra as in case (i), the generalized algebra with mixed signs $\mathring{C}_{ij} = \pm 1$ leads, in general, to new types of representations different from those of Clifford algebra. As a matter of fact, all of these algebras (including the Clifford algebra) are special instances of a much larger algebra with arbitrary $\mathring{C}_{ij}$ (forming a group) whose irreducible representations have been determined in the succeeding paper.[8] In our case, the prescription of determining the irreducible representations is as follows: Consider the group $\mathcal{E}$ consisting of the 2^n elements given in Eq. (8) and their negatives. Thus the order of $\mathcal{E}$ is 2^{n+1}. The representations of the commutation relations (7) are also the representations of $\mathcal{E}$. The factor group $\mathcal{E}/C_2$, where C_2 is the group of two elements $(+1, -1)$, is Abelian and has 2^n one-dimensional representations. If the K *additional* representations are of dimensions $l_1, l_2, \cdots, l_K$, we have

$$l_1^2 + l_2^2 + \cdots + l_K^2 = 2^{n+1} - 2^n = 2^n. \qquad (13)$$

The number K is equal simply to the number of basis elements in the set (8) *which commute with all the 2^n elements.* Thus, if all $\mathring{C}_{ij} = +1$, $K = 2^n$, $l_1 = \cdots = l_K = 1$; and if all $\mathring{C}_{ij} = -1$, $K = 1$ or 2 depending on whether $\tau =$ even or odd, and $l = 2^{\frac{1}{2}n}$ (n even) or $l_1 = l_2 = 2^{\frac{1}{2}(n-1)}$ (n odd). Finally, to find the number of commuting elements in the set (8) we look at the table of $\mathring{C}_{ij}$ and determine how many $\mathring{C}_{ij}$, $\mathring{C}_{ij}\mathring{C}_{ik}$, $\mathring{C}_{ij}\mathring{C}_{ik}\mathring{C}_{il}, \cdots$, are $+1$, for fixed i, all j, all (j, k), $\cdots$. It follows from this that the dimensions of irreducible representations are determined by those of the subalgebra containing $-$ signs for $\mathring{C}_{ij}$; the additional $\mathring{C}_{ij} = +1$ terms do not increase the dimensionality as in case (i). Thus, all irreducible representations of dimensions 1, 2, 4, 8, $\cdots$, $2^{\frac{1}{2}n}$ occur. This completes the enumeration of all irreducible representations. Explicit forms of the matrices are given in the Appendix.

III. ANTIUNITARY AND MIXED UNITARY-ANTIUNITARY RAY REPRESENTATIONS

We wish now to represent n_1 generating elements by unitary operators $U(\Gamma_i)$, $i = 1, 2, \cdots n_1$, and

[6] P. Jordon and E. P. Wigner, Z. Physik **47**, 631 (1928).
[7] H. Boerner, *Representations of Groups* (North-Holland Publishing Company, Amsterdam, 1963), p. 269.
[8] See A. O. Barut, J. Math. Phys. **7**, 1908 (1966).

n_2 generating elements by antiunitary operators $A(\Gamma_j)$, $j = 1, 2, \cdots n_2$. $n = n_1 + n_2$. Because the product of two unitary operators is unitary and the product of two antinunitary operators is again unitary, the whole algebra with 2^n basis elements splits into, say, m antiunitary and $2^n - m$ unitary operators. In particular, the case where all generating elements are represented by antiunitary operators corresponds to $n_1 = 0$, in which case, because

$$m = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots = 2^{n-1},$$

the number of unitary and antiunitary operators is the same as those of the unitary representations of $(n - 1)$ Γ's and the antiunitary representation of *one* Γ. Thus we have the following theorem.

Theorem: The ray representations of G generated by n elements Γ_i in which any number of the Γ's are represented by antiunitary operators can, by relabeling the elements and the proper adjustment of phases, be made to coincide with the ray representations in which *only one* Γ is represented by an antiunitary operator. The representation algebra is thus always generated by a direct product of the form

$$[I, U(\Gamma_1), \cdots, U(\Gamma_{n-1})] \times [I, A(\Gamma_n)].$$

Among all the discrete commuting symmetry transformations there can only be *one* antiunitary *generating* operator (the others are the products).

We now discuss, in general, the phases between antiunitary and unitary operators.

As in Sec. II we first normalize the unitary operators in such a way that the product of two unitary operators satisfy

$$[U(g_i)]^2 = I, \qquad \omega_{ii} = 1.$$

$$U(g_i)U(g_j) = C_{ij}U(g_j)U(g_i); \qquad C_{ij} = \pm 1. \tag{14}$$

Next we look at the products of two antiunitary operators. The determination of the phase factors here is slightly more complicated. For in the relation[9]

$$A(g_i)A^*(g_i) = \varepsilon_{ii}U(g_i^2) = \varepsilon_{ii}, \tag{15}$$

the diagonal phases ε_{ii} now cannot be normalized to unity by multiplying A with a phase factor. From the associativity law, however, it follows that these phases are equal to ± 1, for the multiplication of the above equation by $A(g_i)$ from the left gives

$$A(g_i)A(g_i)^*A(g_i) = \varepsilon_{ii}^*A(g_i), \tag{16}$$

hence

$$\varepsilon_{ii} = \varepsilon_{ii}^*. \tag{17}$$

To determine the off-diagonal phase factors in

$$A(g_i)A^*(g_j) = \varepsilon_{ij}U(g_ig_j) \tag{18}$$

and

$$A(g_j)A^*(g_i) = \varepsilon_{ji}U(g_jg_i) = \varepsilon_{ji}U(g_ig_j), \tag{19}$$

we multiply these two equations to obtain

$$A(g_j)A^*(g_i)A(g_i)A^*(g_j) = \varepsilon_{ij}\varepsilon_{ji}$$

or

$$\varepsilon_{ij}\varepsilon_{ii} = \varepsilon_{ij}\varepsilon_{jj}. \tag{20}$$

Because ε_{kk} are real and equal to ± 1, the *product* $\varepsilon_{ij}\varepsilon_{ji}$ is also equal to ± 1. Let us pass again to the commutation relations. From (18) and (19), we obtain

$$A(g_i)A^*(g_j) = d_{ij}A(g_j)A^*(g_i), \tag{21}$$

where

$$d_{ij} = \varepsilon_{ij}/\varepsilon_{ji} = d_{ji}^{-1}. \tag{22}$$

In contrast to the unitary case, the associativity law does not allow us to determine the ε_{ij}. To see this, let us multiply Eq. (18) by $A(g_i)$ from the left,

$$A(g_i)A(g_i)^*A(g_j) = \varepsilon_{ij}^*A(g_i)U^*(g_ig_j),$$

and replace the last factor $U(g_ig_j)^*$ from Eq. (19) and use (21), (16), and (17):

$$\begin{aligned}\varepsilon_{ij}A(g_j) &= \varepsilon_{ij}^*A(g_i)\varepsilon_{ji}^{*-1}A(g_j)^*A(g_i)\\ &= \varepsilon_{ij}^*\varepsilon_{ji}^{*-1}\,d_{ij}\,\varepsilon_{ii}A(g_j)\end{aligned}$$

or

$$d_{ij} = \varepsilon_{ij}^*/\varepsilon_{ji}^* = 1/d_{ji}^*, \tag{22'}$$

i.e., the associativity law gives nothing new.

We can, however, pass to equivalent phase systems by multiplying in (18) $A(g_i)$ and $A(g_j)$ by phase factors so that $\varepsilon_{ij} = +1$. But then ε_{ji} cannot always be equal to $+1$, but from (20) it is equal to ± 1. Thus, in an equivalent phase system,

$$\varepsilon_{ij} = +1, \qquad \varepsilon_{ji} = \pm 1. \tag{23}$$

It follows from (22) and (23) that the commutation factors d_{ij} are real and equal to ± 1. Consequently, the representations are determined up to these arbitrary phase factors in the antiunitary operators.

Finally, we consider the product of one unitary and one antiunitary operator. Here we have only the off-diagonal phase factors

[9] (*) means complex conjugation. The rules of right and left multiplications with antiunitary operators follow from the fact that every antiunitary operator is a product of a unitary operator times the operation of complex conjugation.

$$U(g_i)A(g_j) = \rho_{ij}A(g_ig_j),$$
$$A(g_j)U^*(g_i) = \rho_{ji}A(g_jg_i) = \rho_{ji}A(g_ig_j). \quad (24)$$

The commutation relations are given by

$$U(g_i)A^*(g_j) = f_{ij}A(g_j)U^*(g_i) \quad (25)$$

with

$$f_{ij} = \rho_{ij}/\rho_{ji} = f_{ji}^{-1}. \quad (26)$$

Again, we make use of the associativity law as before and obtain

$$U(g_i)^2A(g_j) = \rho_{ij}U(g_i)A(g_ig_j),$$
$$A(g_j) = \rho_{ij}U(g_i)\rho_{ji}^{-1}A(g_j)U^*(g_i)$$
$$= \rho_{ij}\rho_{ji}^{-1}f_{ij}A(g_j)U^*(g_i)^2,$$

or

$$f_{ij}^2 = 1, \qquad f_{ij} = \pm 1. \quad (27)$$

To determine ρ_{ij} we again pass, as in the previous case, to an equivalent phase system by multiplying the first equation of (24) by phase factors such that

$$\rho_{ij} = +1, \quad (23')$$

but then

$$\rho_{ji} = \pm 1$$

gives us two distinct types, as in Eq. (23). Both in (23) and (23') we can choose $\varepsilon_{ij} = +1$ (or $\rho_{ij} = +1$) for $i < j$, then ε_{ij} (or ρ_{ij}) $= \pm 1$ holds for $i > j$.

Now we discuss the explicit representations of the group. We bring first G, by relabeling the group elements, to the form given in the theorem at the beginning of this section, i.e., $(n-1)$ unitary generators $U(\Gamma_i)$, $i = 1, \cdots, n-1$ and *one* antiunitary generator $A(\Gamma_n)$. Because every antiunitary operator is of the form $A(\Gamma_n) = U(\Gamma_n)K$, where K is the complex conjugation, we have from (15)

$$U(\Gamma_n)U(\Gamma_n)^* = \varepsilon_{nn} = \pm 1, \quad (28)$$

in addition to the unitarity condition

$$U(\Gamma_n)U(\Gamma_n)^\dagger = U(\Gamma_n)^\dagger U(\Gamma_n) = I. \quad (29)$$

Thus the unitary matrix is symmetric or antisymmetric depending on whether $\varepsilon_{nn} = +1$ or -1, respectively. We have then from (25) and (27)

$$U(\Gamma_i)U(\Gamma_n)^* = f_{in}U(\Gamma_n)U(\Gamma_i). \quad (30)$$

Suppose we have a representation of $G(n-1)$, $U(\Gamma_i)$, as discussed in Sec. II, then Eqs. (28) and (30) can be satisfied by real or pure imaginary matrices (for example, $\sigma_1\sigma_2^* = +\sigma_2\sigma_1$ but $\sigma_1\sigma_3^* = -\sigma_3\sigma_1$). We may then take for $\varepsilon_{nn} = \mp 1$, $U(\Gamma_n)^* = \mp U(\Gamma_n)$, $\mp f_{in} = C_{in}$, then the representation of (30) is reduced again to the cases of Sec. I, but with the additional restriction $U(\Gamma_n)^* = \mp U(\Gamma_n)$.

APPENDIX: EXPLICIT FORM OF THE REPRESENTATION MATRICES

In all cases the representation matrices can be written as a direct product of Pauli matrices and the 2-by-2 identity matrix I. Because the dimensions of the irreducible representations are determined by the noncommutative subalgebra, consider first a subalgebra with r generating elements. The representation with the *highest* dimension for all $\mathring{C}_{ij} = -1$ can be chosen in some standard form, for example,

(a) r = even:

$$U(\Gamma_1) = \sigma_1 \times I \times I \cdots \times I \ (\tfrac{1}{2}r \text{ factors})$$
$$U(\Gamma_2) = \sigma_2 \times I \times I \cdots \times I$$
$$U(\Gamma_3) = \sigma_3 \times \sigma_1 \times I \cdots \times I$$
$$U(\Gamma_4) = \sigma_3 \times \sigma_2 \times I \cdots \times I$$
$$U(\Gamma_5) = \sigma_3 \times \sigma_3 \times \sigma_1 \cdots \times I$$
$$U(\Gamma_6) = \sigma_3 \times \sigma_3 \times \sigma_2 \cdots \times I$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$U(\Gamma_{r-1}) = \sigma_3 \times \sigma_3 \times \cdots \times \sigma_3 \times \sigma_1$$
$$U(\Gamma_r) = \sigma_3 \times \sigma_3 \times \cdots \times \sigma_3 \times \sigma_2. \quad (A1)$$

(b) r = odd: There are two inequivalent representations:

(i) $U(\Gamma_{2i-1})$ $= U(\Gamma_{2i-1})$ of case (a), $U(\Gamma_{2i})$ $= U(\Gamma_{2i})$ of case (a), $\Big\}\ j = 1 \cdots \tfrac{1}{2}(r-1)$

and (A2)

$$U(\Gamma_n) = \sigma_3 \times \sigma_3 \times \cdots \times \sigma_3 \times [\tfrac{1}{2}(r-1) \text{ factors}].$$

(ii) The negative of the matrices in (i).

It is sufficient to consider then all other representations with the same dimension of $2^{\frac{1}{2}r}$ (case a) or $2^{\frac{1}{2}(r-1)}$ (case b)—then of course we vary r. These are obtained simply by replacing one Pauli matrice in (A1) by another as the case may be. For example, if

$$C_{ij} = -1, \quad j = 2, 3, \cdots r; \quad C_{kj} = +1, \quad k \neq 1,$$

—a case which also gives an irreducible representation of the same dimension as above—we replace in (A1) all σ_3's by σ_2 and all σ_2's by I.

On Irreducible Representation of a Class of Algebras and Related Projective Representations*

A. O. BARUT
Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado
(Received 1 March 1966)

The number and the dimensions of irreducible representations of a general class of algebras occurring in the projective representations of finite groups have been determined. The Lie subalgebras have been found, and the isomorphism between the quantum-mechanical ray representations of finite groups and the fundamental representations of Lie algebras is shown.

I. THE ALGEBRA

We consider an algebra $\mathfrak{A}$ generated by n elements α_i, $i = 1, 2 \cdots n$, satisfying the condition

$$\alpha_i\alpha_j = C_{ij}\alpha_j\alpha_i. \tag{1}$$

Clearly, $C_{ii} = 1$, $C_{ji} = C_{ij}^{-1}$. In general, no other conditions on α_i are required. We refer in particular to a special case as *the finite case*, where we suppose that α_i, in addition, satisfy a number of polynomial restrictions of the form

$$P_\nu(\alpha_i) = a_\nu, \qquad \nu = 1 \cdots N, \tag{2}$$

where a_ν are arbitrary numbers and P_ν are polynomials in α_i of order n_ν. We may assume for simplicity that there are no other numerical coefficients in P_ν except the a_ν.

The problem of quantum-mechanical ray representations of some finite groups lead to very special cases of this algebra,[1] in particular, the Clifford algebra corresponds to all $C_{ij} = -1$ and Eq. (2) is replaced by $\alpha_i^2 = 1$, $i = 1, 2 \cdots n$. Furthermore, the subalgebras of this algebra are isomorphic to Lie algebras (see Sec. IV) so that the problem is also of interest in the study of enveloping algebras of Lie algebras.

Because of Eq. (1) the algebra $\mathfrak{A}$ has the basis elements consisting of the ordered products

$$1;\ \alpha_i;\ \alpha_i\alpha_j(i \le j);\ \alpha_i\alpha_j\alpha_k(i \le j \le k);\ \cdots;$$

$$\alpha_i\alpha_j \cdots \alpha_l\alpha_m \cdots$$

$$(i \le j \le \cdots \le l \le m \cdots);\ \cdots. \tag{3}$$

Note that the equality signs in $i \le j$, etc., give us also all the terms of the form $\alpha_1^{n_1}\alpha_2^{n_2} \cdots$.

In the case of conditions (2) the series (3) breaks after some power, otherwise it is infinite. In the former case the commutation factors C_{ij} in (1) cannot be completely arbitrary; they satisfy polynomial restrictions similar to (2). For example, if $\alpha_i^{n_i} = a_i$, we find, multiplying (1) from left by $\alpha_i^{n_i-1}$, that $(C_{kj})^{n_i} = 1$, all j, and $(C_{ij})^{n_i} = 1$, all i. In other words, we assume that C_{ij} in (1) and any possible restrictions (2) are *consistently* chosen.

The representations of (1) are determined up to numerical multiples of α_j, for, together with α_j, the multiples $\lambda_j\alpha_j$ also satisfy Eq. (1). The complex numbers λ_j are, in the finite case, restricted by the conditions in (2), otherwise not. We may adjust them so that the coefficients a_ν in (2) are equal to unity.

II. IRREDUCIBLE REPRESENTATIONS

To find the irreducible representations, we interpret Eq. (1) as the group composition law of a finite or denumerable infinite group G consisting of the direct product of the set (3), the basis elements of $\mathfrak{A}$, and the Abelian group $\mathfrak{B}$ generated by the coefficients C_{ij} [and possible by the numerical coefficients in Eq. (2)]. In the finite case, both G and $\mathfrak{B}$ are finite groups, and the order of G, h, is equal to ab, where a is the dimension of the algebra and b the order of the Abelian group $\mathfrak{B}$. Example: For the Clifford algebra generated by n elements $a = 2^n$, $b = 2$, $h = 2^{n+1}$.

The factor group $\mathfrak{F} = G/\mathfrak{B}$ is the group generated by n *commuting* α_i's and consists of the set (3) which form a group if α_i's commute, otherwise not. In the finite case the order of $\mathfrak{F}$ is also a. For example, if $\alpha_i^{n_i} = 1$, $\mathfrak{F}$ is the direct product of n cyclic groups.

The representations of $\mathfrak{F}$ are also representations of G, although they do not satisfy the required relations (1). Our problem then is to find the additional representations of G which do satisfy (1).

Let K be the number of the desired additional irreducible inequivalent representations of G, in the finite case, and $l_1, l_2, \cdots, l_K$ their dimensions. Be-

* Supported in part by the U. S. Air Force Office of Scientific Research.

[1] See A. O. Barut and S. Komy, J. Math. Phys. 7, 1903 (1966).

cause the order of the group is equal to the sum of squares of the dimensions of inequivalent irreducible representations, we have

$$h = a + l_1^2 + l_2^2 + \cdots + l_K^2 = ab$$

or

$$l_1^2 + l_2^2 + \cdots + l_K^2 = (b-1)a. \tag{4}$$

It remains to determine the integer K; then the solution of (4) is unique. The total number of irreducible representations of G is equal to the number f of conjugate classes, $f = a + K$. Let us look at two limiting cases: (i) If all $C_{ij} = +1$, then each element forms a separate class, and

$$f_{\max} = ab, \qquad K = (b-1)a, \tag{5}$$

hence, only one-dimensional representations occur. (ii) The minimum number of classes is obtained if every element x_0 of the set (3) is in the same class with $\mathfrak{B}x_0$. That is, if we form xx_0x^{-1} and using (1), we obtain

$$xx_0x^{-1} = C(x, x_0)x_0, \tag{6}$$

where $C(x_1x_0) \in \mathfrak{B}$. Thus case (ii) corresponds to the situation where Eq. (6) generates all elements C of $\mathfrak{B}$ when x varies over the set (3). The elements of $\mathfrak{B}$ always form separate classes because they commute with all other elements of G (numbers), hence

$$f_{\min} = a + (b-1), \qquad K_{\min} = b - 1. \tag{7}$$

The dimensions of the irreducible representations satisfy

$$l_1^2 + l_2^2 + \cdots + l_{b-1}^2 = a(b-1). \tag{8}$$

This case gives the smallest number of irreducible representations of highest dimensions. In the general case, the number of irreducible representations, K, lies between these two limits

$$(b-1) \leq K \leq a(b-1). \tag{9}$$

In fact, K is determined by the number of commuting elements of the set (3). Thus, for a given C_{ij}, we first determine the number of elements among the basis set (3) of the algebra which commute with all other basis elements. Let this number be k ($k \geq 1$, because the identity commutes with all other elements), then

$$K = k(b-1). \tag{10}$$

In the infinite case, Eq. (1) without the restrictions in Eq. (2), the number of irreducible representations is in general infinite. But if the given C_{ij} are such that $\mathfrak{B}$ is a finite group, it follows from (8) that there may be cases where we have a finite number irreducible representations, some of them infinite dimensional.

III. PROJECTIVE REPRESENTATIONS

The discussion given in the previous section allows one to determine in a simple way all the *additional* ray representations of some finite groups. Any finite group can be generated by a (small) number of generating elements α_i, i.e., from all possible products of α_i with polynomial conditions of the form of Eq. (2). The symmetric group S_n is generated by *two* elements $\alpha_1 = (12)$ and $\alpha_2 = (12 \cdots n)$. However, these do not commute. S_n is also generated by $(n-1)$ partly commuting elements, namely, the $(n-1)$ transpositions $\alpha_1 = (12)$, $\alpha_2 = (23)$, $\cdots$, $\alpha_{n-1} = (n-1, n)$. If the generating elements commute, then the ray representations defined by

$$D(\alpha_i)\, D(\alpha_j) = \omega_{ij}\, D(\alpha_i\alpha_j) \tag{11}$$

satisfy

$$D(\alpha_i)\, D(\alpha_j) = C_{ij}\, D(\alpha_j)\, D(\alpha_i); \qquad C_{ij} = \omega_{ij}/\omega_{ji}. \tag{12}$$

This is our Eq. (1). Thus the problem of determining ray representations is reduced to the one treated in the previous section. If the generating elements of the finite group do not commute, then the method of Sec. II can, in principle, still be generalized, but the counting of conjugate classes is more complicated.

IV. LIE SUBALGEBRAS

We determine the Lie algebras generated by the subalgebras of $\mathfrak{A}$ with the basis (3). It is well known that[2] in the case of Clifford algebra with n generating elements (2^n basis elements), a subalgebra of dimension $n(n+1)/2$ satisfy, in addition to the multiplication law of the algebra, commutation relations of a Lie algebra which is isomorphic to the Lie algebra of SO_{n+1}. The irreducible representations of the Clifford algebra are the irreducible fundamental representations of the Lie algebra of SO_{n+1}. But, of course, the Lie algebra has infinitely many higher dimensional irreducible representations. We now generalize these results. Let us take the special case that the restrictions in Eq. (2) are of the form

$$\alpha_i^{n_i} = 1, \qquad i = 1, 2 \cdots n, \tag{13}$$

[2] H. Boerner, *Representations of Groups* (North-Holland Publishing Company, Amsterdam, 1963), p. 269.

then

$$(C_{ij})^{n_i} = (C_{ij})^{n_j} = 1 \tag{14}$$

and forms

$$[\alpha_i^{\nu_i}, \alpha_j^{\nu_j}] = (C_{ij}^{\nu_i\nu_j} - 1)\alpha_j^{\nu_j}\alpha_i^{\nu_i} \equiv \alpha_{ij}^{\nu_i\nu_j} \tag{15}$$

and

$$[\alpha_{ij}^{\nu_i\nu_j}, \alpha_{kl}^{\nu_k\nu_l}] = (C_{il}^{\nu_i\nu_l}C_{ik}^{\nu_i\nu_k}C_{jl}^{\nu_j\nu_l}C_{jk}^{\nu_j\nu_k} - 1)\alpha_{ij}^{\nu_i\nu_j}\alpha^{\nu_k\nu_l},$$
$$[\alpha_{ij}^{\nu_i\nu_j}, \alpha_k^{\nu_k}] = (C_{ij}^{\nu_i\nu_j} - 1)(C_{ik}^{\nu_i\nu_j}C_{jk}^{\nu_j\nu_k} - 1)\alpha_k^{\nu_k}\alpha_i^{\nu_i}\alpha_j^{\nu_j}. \tag{16}$$

Let us first choose $\nu_i = \frac{1}{2}n_i$. It follows then from equations (13), (15), and (16) that the $\frac{1}{2}n(n+1)$ quantities $\alpha_i^{n_i}$ and $\alpha_{ij}^{n_in_j}$ form a Lie algebra. In fact, this is essentially the Lie algebra of SO_{n+1}. If we choose the integers ν_i less than $\frac{1}{2}n_i$, the commutation relations (16) are not closed. We have to continue the process of taking the commutation relations p steps such that $p\nu_i = n_i$, or $\nu_i = n_i/p$, whenever admissible. Thus, we obtain a set of Lie subalgebras whose fundamental irreducible representations of higher dimensions are also the irreducible representations of the algebras of Sec. I. This is because the generating set α_i also belongs to the Lie algebra.

We thus have the general, and also physically interesting, conclusion that the irreducible ray representations of finite groups are isomorphic to the fundamental irreducible representations of Lie algebras.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

Invariance and Conservation Laws in Classical Mechanics. II

HARRY H. DENMAN
Department of Physics, Wayne State University, Detroit, Michigan
(Received 14 April 1966; revised manuscript received 4 May 1966)

In this paper, the invariances of the equation of motion of a classical particle, to coordinate and time translations, to scale transformations and inversions, and to Galilean transformations, are considered individually. Resultant conditions on the equation of motion are given, and, for invariance to the one-parameter continuous transformations, it is shown that the equation of motion can be reduced from second to first order. Associated with each such reduction is a conservation law. The implications of the invariance of the system Lagrangian to these transformations are indicated, and the conservation laws, if any, associated with them. Some requirements on the Lagrangian for invariant equations of motion are also presented, and it is shown that the invariance of an equation of motion derived from a Lagrangian does not imply the invariance of that Lagrangian to the transformation. It is also shown that time-translation invariance of the equation of motion does *not* always require conservation of the Hamiltonian.

1. INTRODUCTION

IN a previous paper,[1] the following question was considered. If the invariance of a physical system (taken in paper I to be a classical particle moving in one dimension) to a transformation implies the invariance of an observable (assumed in I to be the equation of motion[2]), what restrictions does the invariance place on the equation of motion and on the Lagrangian describing the system, and what, if any, conservation law is generated by this invariance?

The equation of motion (EM) for the particle is taken as the general second-order ordinary differential equation

$$\ddot{q} = f(\dot{q}, q, t). \tag{1.1}$$

In I, the linear coordinate transformations

$$q' = \alpha q + a, \qquad \alpha \neq 0, \tag{1.2}$$

were considered. If $\alpha = 1$, one has the coordinate translations; if $a = 0$, $\alpha > 0$, the coordinate-scale transformations (or dilations); if $\alpha = -1$, $a = 0$, coordinate inversion. Using techniques essentially equivalent to certain applications of one-parameter

[1] H. H. Denman, J. Math. Phys. **6**, 1611 (1965), hereafter referred to as I. A misprint occurred in Ref. 10 of this paper: for $r = (\partial G/\partial t)/(\partial G/\partial v)$, read $r = -(\partial G/\partial t)/(\partial G/\partial v)$.

[2] The importance of the equation of motion as a classical observable, and the fact that the invariance of the equation of motion does not necessarily have the same implications as the invariance of the Lagrangian, has been stressed by E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965); Progr. Theoret. Phys. (Kyoto) **11**, 437 (1954); and private communication. Also, see L. J. Tassie and H. A. Buchdahl, Australian J. Phys. **17**, 431 (1964).

then

$$(C_{ij})^{n_i} = (C_{ij})^{n_j} = 1 \tag{14}$$

and forms

$$[\alpha_i^{\nu_i}, \alpha_j^{\nu_j}] = (C_{ij}^{\nu_i\nu_j} - 1)\alpha_j^{\nu_j}\alpha_i^{\nu_i} \equiv \alpha_{ij}^{\nu_i\nu_j} \tag{15}$$

and

$$[\alpha_{ij}^{\nu_i\nu_j}, \alpha_{kl}^{\nu_k\nu_l}] = (C_{il}^{\nu_i\nu_l}C_{ik}^{\nu_i\nu_k}C_{jl}^{\nu_j\nu_l}C_{jk}^{\nu_j\nu_k} - 1)\alpha_{ij}^{\nu_i\nu_j}\alpha^{\nu_k\nu_l},$$
$$[\alpha_{ij}^{\nu_i\nu_j}, \alpha_k^{\nu_k}] = (C_{ij}^{\nu_i\nu_j} - 1)(C_{ik}^{\nu_i\nu_j}C_{jk}^{\nu_j\nu_k} - 1)\alpha_k^{\nu_k}\alpha_i^{\nu_i}\alpha_j^{\nu_j}. \tag{16}$$

Let us first choose $\nu_i = \frac{1}{2}n_i$. It follows then from equations (13), (15), and (16) that the $\frac{1}{2}n(n+1)$ quantities $\alpha_i^{n_i}$ and $\alpha_{ij}^{n_in_j}$ form a Lie algebra. In fact, this is essentially the Lie algebra of SO_{n+1}. If we choose the integers ν_i less than $\frac{1}{2}n_i$, the commutation relations (16) are not closed. We have to continue the process of taking the commutation relations p steps such that $p\nu_i = n_i$, or $\nu_i = n_i/p$, whenever admissible. Thus, we obtain a set of Lie subalgebras whose fundamental irreducible representations of higher dimensions are also the irreducible representations of the algebras of Sec. I. This is because the generating set α_i also belongs to the Lie algebra.

We thus have the general, and also physically interesting, conclusion that the irreducible ray representations of finite groups are isomorphic to the fundamental irreducible representations of Lie algebras.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

Invariance and Conservation Laws in Classical Mechanics. II

HARRY H. DENMAN
Department of Physics, Wayne State University, Detroit, Michigan
(Received 14 April 1966; revised manuscript received 4 May 1966)

In this paper, the invariances of the equation of motion of a classical particle, to coordinate and time translations, to scale transformations and inversions, and to Galilean transformations, are considered individually. Resultant conditions on the equation of motion are given, and, for invariance to the one-parameter continuous transformations, it is shown that the equation of motion can be reduced from second to first order. Associated with each such reduction is a conservation law. The implications of the invariance of the system Lagrangian to these transformations are indicated, and the conservation laws, if any, associated with them. Some requirements on the Lagrangian for invariant equations of motion are also presented, and it is shown that the invariance of an equation of motion derived from a Lagrangian does not imply the invariance of that Lagrangian to the transformation. It is also shown that time-translation invariance of the equation of motion does *not* always require conservation of the Hamiltonian.

1. INTRODUCTION

IN a previous paper,[1] the following question was considered. If the invariance of a physical system (taken in paper I to be a classical particle moving in one dimension) to a transformation implies the invariance of an observable (assumed in I to be the equation of motion[2]), what restrictions does the invariance place on the equation of motion and on the Lagrangian describing the system, and what, if any, conservation law is generated by this invariance?

The equation of motion (EM) for the particle is taken as the general second-order ordinary differential equation

$$\ddot{q} = f(\dot{q}, q, t). \tag{1.1}$$

In I, the linear coordinate transformations

$$q' = \alpha q + a, \qquad \alpha \neq 0, \tag{1.2}$$

were considered. If $\alpha = 1$, one has the coordinate translations; if $a = 0$, $\alpha > 0$, the coordinate-scale transformations (or dilations); if $\alpha = -1$, $a = 0$, coordinate inversion. Using techniques essentially equivalent to certain applications of one-parameter

[1] H. H. Denman, J. Math. Phys. **6**, 1611 (1965), hereafter referred to as I. A misprint occurred in Ref. 10 of this paper: for $r = (\partial G/\partial t)/(\partial G/\partial v)$, read $r = -(\partial G/\partial t)/(\partial G/\partial v)$.

[2] The importance of the equation of motion as a classical observable, and the fact that the invariance of the equation of motion does not necessarily have the same implications as the invariance of the Lagrangian, has been stressed by E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965); Progr. Theoret. Phys. (Kyoto) **11**, 437 (1954); and private communication. Also, see L. J. Tassie and H. A. Buchdahl, Australian J. Phys. **17**, 431 (1964).

Lie groups to ordinary differential equations,[3] it was shown that:

A. The EM (1.1) is coordinate-translation invariant[4] if and only if

$$\partial f/\partial q \equiv 0 \tag{1.3a}$$

or

$$f = f(\dot{q}, t). \tag{1.3b}$$

B. (1.1) is coordinate-scale invariant if and only if

$$\dot{q}(\partial f/\partial \dot{q}) + q(\partial f/\partial q) \equiv f, \tag{1.4a}$$

i.e., f is homogeneous of degree 1 in $\dot{q}$ and q. Alternately,

$$f = qg(\dot{q}/q, t). \tag{1.4b}$$

C. (1.1) is coordinate-inversion invariant if and only if

$$f(-\dot{q}, -q, t) \equiv -f(\dot{q}, q, t). \tag{1.5}$$

If (1.1) is coordinate-translation invariant, it can be written, letting $v = \dot{q}$,

$$\dot{v} = f(v, t), \tag{1.6}$$

which is first order. The solution of (1.6) is

$$v = v(t, v_0, t_0), \tag{1.7}$$

where v_0 is the velocity when $t = t_0$. Writing (1.6) in the form

$$dv - f(v, t)\, dt = 0, \tag{1.8}$$

an integrating factor $\mu(v, t)$ can normally be found[5] such that

$$\mu\, dv - \mu f\, dt \tag{1.9}$$

is an exact differential. Thus there exists a conserved quantity $F(v, t)$, where

$$F(v, t) = F(v_0, t_0), \tag{1.10}$$

and (1.10) may be regarded as a conservation law generated solely by the coordinate-translation invariance of the EM. From (1.7),

$$dq = v(t, v_0, t_0)\, dt,$$

and the motion $q(t)$ can be found by quadrature.

If (1.1) is coordinate-scale invariant, it becomes, using (1.4b),

$$\ddot{q} = qg(\dot{q}/q, t). \tag{1.11}$$

If $z = \dot{q}/q$, (1.11) can be written in the first-order form

$$\dot{z} + z^2 = g(z, t) \tag{1.12}$$

or

$$dz + [z^2 - g(z, t)]\, dt = 0. \tag{1.13}$$

There exists an integrating factor which makes the left side of (1.13) an exact differential. Thus coordinate-scale invariance of the EM generates the conservation law

$$G(z, t) = G(z_0, t_0), \tag{1.14}$$

where $z_0 = v_0/q_0$ and q_0 is the coordinate when $t = t_0$. If (1.13) or (1.14) is solved for $z = z(t, z_0, t_0)$, then $q(t)$ can be found by quadrature, i.e.,

$$\ln q = \int z(t, z_0, t_0)\, dt + C. \tag{1.15}$$

Coordinate-inversion invariance of the EM does not appear to generate any conservation law.

In the following sections, the effects both on the EM and on the system Lagrangian of time translations, time-scale transformations, and time inversion, as well as Galilean transformations, are discussed.

2. TIME TRANSFORMATIONS

Consider the linear time transformations

$$t' = \beta t + b, \qquad \beta \neq 0. \tag{2.1}$$

For $(\beta, b) = (1, b)$, $(\beta, 0)$, $(-1, 0)$, one has, respectively, the time translations, time-scale transformations, and time inversion. Using the techniques of I, or those of group theory,[3] one finds:

D. The EM (1.1) is time-translation invariant[4] if and only if

$$\partial f/\partial t \equiv 0 \tag{2.2a}$$

or

$$f = f(\dot{q}, q). \tag{2.2b}$$

E. (1.1) is time-scale invariant if and only if

$$\dot{q}(\partial f/\partial \dot{q}) - t(\partial f/\partial t) \equiv 2f \tag{2.3a}$$

or

$$f(\dot{q}, q, t) = h(t\dot{q}, q)/t^2. \tag{2.3b}$$

---

[3] A. Cohen, *An Introduction to the Lie Theory of One-Parameter Groups* (G. Stechert and Co., New York, 1931). The author is indebted to Professor E. S. Northam for this reference. Also, E. L. Ince, *Ordinary Differential Equations* (Dover Publications, Inc., New York, 1953), Chap. IV.

[4] By invariant, as in I, it is understood that the equation of motion is to be of the same *form* in both coordinate systems, i.e., if (1.1) holds in the (q, t) space–time, then $\ddot{q}' = f(\dot{q}', q', t')$ in the (q', t') space–time.

[5] E. L. Ince, Ref. 3, p. 27.

F. (1.1) is time-inversion invariant if and only if

$$f(-\dot{q}, q, -t) \equiv f(\dot{q}, q, t). \tag{2.4}$$

Conserved quantities generated by the first two of these time invariances can be developed in the following ways. For time-translation invariance, (1.1) becomes, using (2.2b),

$$\ddot{q} = f(\dot{q}, q). \tag{2.5}$$

Letting $\dot{q} = v$, (2.5) can be written

$$v\, dv/dq = f(v, q), \tag{2.6}$$

which is first order. Integration of (2.6) gives the conservation law

$$H(v, q) = H(v_0, q_0). \tag{2.7}$$

Solution of (2.6) or (2.7) yields

$$dq/dt = v(q, v_0, q_0), \tag{2.8}$$

which gives $q(t)$ by the quadrature

$$\int dq/v(q, v_0, q_0) = t + C. \tag{2.9}$$

For time-scale invariance, (1.1) must have the form

$$t^2\ddot{q} = f(t\dot{q}, q). \tag{2.10}$$

Letting $u = t\dot{q}$, (2.10) can be written

$$du/dq = 1 + u^{-1}f(u, q), \tag{2.11}$$

which is first order. Further, (2.11) implies the existence of a conserved quantity

$$J(t\dot{q}, q) = J(t_0\dot{q}_0, q_0). \tag{2.12}$$

From (2.11) or (2.12)

$$t\, dq/dt = f(q, t_0\dot{q}_0, q_0), \tag{2.13}$$

from which the motion may be determined by the quadrature

$$\int dq/f = \ln t + C. \tag{2.14}$$

Time-inversion invariance (like coordinate-inversion invariance) does not seem to generate any conservation law.

3. GALILEAN TRANSFORMATION

Consider the Galilean transformation

$$q' = q - Vt, \qquad t' = t. \tag{3.1}$$

If the EM (1.1) is to be invariant with respect to this transformation for all V, then

$$f(\dot{q} - V, q - Vt, t) = f(\dot{q}, q, t)$$

for all f such that (1.1) is satisfied. This condition is satisfied if and only if

$$\partial f/\partial\dot{q} + t(\partial f/\partial q) \equiv 0 \tag{3.2a}$$

or (see Appendix A)

$$f(\dot{q}, q, t) = f(t\dot{q} - q, t). \tag{3.2b}$$

Letting $w = t\dot{q} - q$, (1. 1) and (3.2b) yield

$$\dot{w} = f(w, t), \tag{3.3}$$

which yields a conservation law

$$K(w, t) = K(w_0, t_0). \tag{3.4}$$

Since $t\dot{q} - t = t^2 d(q/t)/dt$, $q(t)$ may be found by quadrature.

4. LAGRANGIAN OF SYSTEM

Some implications of these coordinate-transformation invariances of the EM for a Lagrangian $L(\dot{q}, q, t)$ describing the particle were discussed in I. While the nature of L for each invariance of the EM could not be specified, it was demonstrated that invariance of the EM to coordinate translations does not require the same invariance of L (while the inverse implication is valid).

If L is to be coordinate-scale *invariant*,[6] it must be of the form (see Appendix A)

$$L(\dot{q}/q, t) = L(z, t). \tag{4.1}$$

The EM derived from this Lagrangian is

$$\frac{\partial z}{\partial\dot{q}}\frac{d}{dt}\left(\frac{\partial L}{\partial z}\right) + \frac{\partial L}{\partial z}\left[\frac{d}{dt}\left(\frac{\partial z}{\partial\dot{q}}\right) - \frac{\partial z}{\partial q}\right] = 0. \tag{4.2}$$

But z satisfies Lagrange's equation identically, since it is an exact derivative; thus

$$d(\partial L/\partial z)/dt = 0, \tag{4.3}$$

since $\partial z/\partial\dot{q} = 1/q \neq 0$. Therefore coordinate-scale invariance of L implies the conservation law

$$\partial L/\partial z = r(z, t) = \text{constant}, \tag{4.4}$$

which in turn implies, algebraically, $z = z(t)$. This procedure is to be contrasted with that for the EM, where the general form of the coordinate-scale in-

[6] Invariance of L means that it has the same *values* as a function of $\dot{q}'$, q', t' as L has as the *same* function of $\dot{q}$, q, t. However, as is well known, if $L(\dot{q}', q', t') = aL(\dot{q}, q, t) + dR(q, t)/dt$, where a is a constant, the equations of motion are unchanged. This is a weaker but perhaps more reasonable condition on L, which will be called the gauge-invariance of L. Tassie and Buchdahl have discussed a form of this weaker invariance requirement on L, with $a = 1$ (see Ref. 2). That such gauge transformations on L may be more significant than simple invariance has also been suggested by Professor A. Katz (private communication). If L is gauge invariant to a transformation, the EM is invariant to that transformation (except for a multiplicative constant).

variant EM (1.11) must be transformed and solved as a first-order differential equation to obtain the conservation law (1.14), while the integration of (4.3) immediately yields the conservation law (4.4). But there may exist Lagrangians which are not coordinate-scale invariant, while the EM derived from them are. An example is given in Ref. 9 of I, where $L = \dot{q} \ln \dot{q} + \gamma q$. While L is not invariant to coordinate-scale transformations, it is gauge-invariant, and therefore the EM ($\ddot{q} = \gamma\dot{q}$) is unchanged.[7] Letting $\dot{q} = v$, $\partial L/\partial v = L_{,}$, $\partial^2 L/\partial q \partial v = L_{,,q}$, etc., a necessary and sufficient condition that $L(v, q, t)$ produce the coordinate scale-invariant EM (1.11) is

$$L_q - vL_{,,q} - L_{,,t} = qf(v/q, t)L_{,,,}. \tag{4.5}$$

Lastly, coordinate-inversion invariance of L implies

$$L(-\dot{q}, -q, t) = L(\dot{q}, q, t).$$

Considering the time transformations, the invariance of L to time translations implies that L is not a function of t, and generates the conservation of the Hamiltonian

$$H = -L + \dot{q}(\partial L/\partial\dot{q}), \tag{4.6}$$

if Lagrange's EM is satisfied.

However, the EM may be time-translation invariant without the corresponding invariance of L. Consider, for example,

$$L = Q(\dot{q}, q)T(t). \tag{4.7}$$

The EM can then be written

$$\frac{1}{\partial Q/\partial\dot{q}}\left[\frac{d}{dt}\left(\frac{\partial Q}{\partial\dot{q}}\right) - \frac{\partial Q}{\partial q}\right] + \frac{1}{T}\frac{dT}{dt} = 0. \tag{4.8}$$

If (4.8) is to be time-translation invariant, from (2.2a),

$$T^{-1}(dT/dt) = \gamma,$$

where γ is a constant; then T is proportional to $e^{\gamma t}$. Thus if L is separable in the form (4.7), it must have the form

$$L = e^{\gamma t}Q(\dot{q}, q), \tag{4.9}$$

if the EM is to be time-translation invariant; (4.9) is also sufficient for such invariance. [The Lagrangian (4.9) is gauge-invariant to time translation.] In this case,

$$H = e^{\gamma t}[\dot{q}(\partial Q/\partial\dot{q}) - Q]. \tag{4.10}$$

[7] Another example obtains if L has the form $g(q)\, f(z, t)$. Then it can be shown that, for coordinate-scale invariance of the EM, $g = q^\gamma$, or $L = q^\gamma f(z, t)$, where γ is a constant. This L is not coordinate-scale invariant, but it is gauge invariant.

This H is *not* time-translation invariant, although the Lagrange EM is.[8]

As shown in Appendix A, L is time-scale invariant if and only if $L = L(u, q)$, where $u = t\dot{q}$. The resulting EM is

$$d[t(\partial L/\partial u)]/dt = \partial L/\partial q, \tag{4.11}$$

which is also time-scale invariant. Since (4.11) can also be written as

$$u\frac{d}{dq}\left(\frac{\partial L}{\partial u}\right) + \frac{\partial L}{\partial u} = \frac{\partial L}{\partial q}, \tag{4.12}$$

which is first order, it implies the existence of a conserved quantity

$$M(u, q) = M(u_0, q_0), \tag{4.13}$$

i.e., a conservation law.

If the EM is to be time-scale invariant, it must be of the form (2.10). Expressing the general Lagrangian for the system as $L(v, q, t)$, a necessary and sufficient condition on L for time-scale invariance of its EM is therefore

$$L_q - vL_{,,q} - L_{,,t} = t^{-2}L_{,,}f(tv, q). \tag{4.14}$$

While the general solution to this partial differential equation is not known, one can exhibit a Lagrangian which is not time-scale invariant, while its EM is. Let $L = T(t)U(u, q)$. The associated EM is

$$T't\frac{\partial U}{\partial u} + T\frac{d}{dt}\left(t\frac{\partial U}{\partial u}\right) = T\frac{\partial U}{\partial q}, \tag{4.15}$$

which, if it is to be time-scale invariant, requires

$$T = t^\gamma. \tag{4.16}$$

Thus, if $L = T(t)U(t\dot{q}, q)$, a necessary and sufficient condition that the EM be time-scale invariant is that L have the gauge-invariant form

$$L = t^\gamma U(t\dot{q}, q). \tag{4.17}$$

Time-inversion invariance of L requires

$$L(-\dot{q}, q, -t) = L(\dot{q}, q, t).$$

If L is to be invariant to Galilean transformations, it must have the form (see Appendix A) $L = L(w, t)$,

[8] For multi-degree of freedom systems with generalized coordinates q_i, if

$$L = e^{\alpha t}Q(\dot{q}_i, q_i), \quad \text{then} \quad H = e^{\alpha t}[\textstyle\sum \dot{q}_i(\partial Q/\partial\dot{q}_i) - Q],$$

which is not time-translation invariant, while the Lagrange equations of motion are. However, Hamilton's equations are not in general invariant to this transformation. (One must, of course, first express H in terms of p_i and coordinates q_i.) This example is in contradiction to a statement by A. Messiah, *Quantum Mechanics*, G. M. Termmer., Tr. (North-Holland Publishing Company, Amsterdam, 1962), Vol. 2, p. 664.

where $w = t\dot{q} - q$. Then the EM is

$$d[t(\partial L/\partial w)]/dt = -\partial L/\partial w, \tag{4.18}$$

which is Galilean invariant. Rearranging (4.18), it can be put in the form

$$d[t^2(\partial L/\partial w)]/dt = 0, \qquad t \neq 0, \tag{4.19}$$

which implies the conservation law

$$t^2(\partial L/\partial w) = \text{constant}. \tag{4.20}$$

A necessary and sufficient condition on L that its EM be Galilean invariant is that

$$L_q - vL_{v,q} - L_{v,t} = L_{v,v}f(tv - q, t). \tag{4.21}$$

While the general solution to (4.21) is unknown, one can exhibit a Lagrangian which is not Galilean invariant, while its EM is. Consider

$$L = W(w, t) + g(q, t). \tag{4.22}$$

The associated EM is

$$\frac{d}{dt}\left(t\frac{\partial W}{\partial w}\right) + \frac{\partial W}{\partial w} = \frac{\partial g}{\partial q}. \tag{4.23}$$

In order for (4.23) to be Galilean invariant, $\partial g/\partial q$ must be a function of t only. Thus, if L has the structure (4.22), a necessary and sufficient condition that its EM be Galilean invariant is that

$$L = W(w, t) + f(t)q; \tag{4.24}$$

such L's are not invariant, but are gauge-invariant, to the Galilean transformation.

5. DISCUSSION

Some implications of the invariance of the EM of a classical one-dimensional system to coordinate and time translations, scale transformations and inversions, and Galilean transformations, have been investigated in this paper and in I. Except for the space–time inversions, each of these transformations forms a one-parameter, sectionally continuous group. In such cases,[9] it is possible to reduce the EM from second to first order by a change of variables. Since integrating factors normally exist for first-order ordinary differential equations, invariance of the EM to each of the continuous transformations above implies a conservation law involving only the invariants of that transformation. Further, each of these conservation laws yields the motion of the system $q = q(t, v_0, q_0, t_0)$ by quadrature.

[9] That the invariance of a second-order ordinary differential equation to a one-parameter continuous group generally implies that the equation may be reduced to first order is shown by A. Cohen, Ref. 3, pp. 88–89.

Since the general solution of the EM involves only two arbitrary parameters, the conserved quantities will be independent if the number of invariant transformations are two or less, and will be dependent if the number is greater than two.

Considering the Lagrangian of the system, the invariance of L to a continuous transformation is a stringent requirement which leads immediately to a simple conservation law in the cases of space and time translations, coordinate scale and Galilean transformations, and to a first-order ordinary differential equation for time-scale transformations (which in turn implies a conservation law). However, in each of these cases, one can exhibit a Lagrangian which is *not* invariant to the transformation, while the associated equation of motion is. (In each of these examples, the Lagrangian is gauge invariant.) Also, time-translation invariance of the EM does not invariably imply conservation of the Hamiltonian. However, it is gratifying that each conservation law, whether obtained from the invariance of the EM or the Lagrangian, involves the same invariant combinations of the variables of the motion.

Thus, invariances to the continuous space–time transformations and the Galilean transformation have strong effects on the equation of motion, and on the Lagrangian. Both approaches give conservation laws involving the same variables, which considerably simplify finding the motion. While these effects on the Lagrangian are usually both simple and immediate, the approach to invariance requirements via the equation of motion is appealing, as this equation can be regarded as a direct manifestation of our attempt to describe the system. The two approaches may give quite different results.

ACKNOWLEDGMENTS

The author would like to thank Professor David Fradkin for some very helpful suggestions; Professor Boris Podolsky and Professor Eugene P. Wigner for their continued encouragement during this work.

APPENDIX

The classical transformations considered here (except for coordinate and time inversions) are one-parameter, sectionally continuous transformations of the general form

$$q' = q + \epsilon u(q, t), \tag{A 1}$$

$$t' = t + \delta v(t), \tag{A 2}$$

where u and v are at first arbitrary functions, and ϵ and δ may be regarded as generators of the groups

of transformations discussed. Since, physically, these transformations may be made independently, the above transformations are not carried out simultaneously, and may be separated into two cases.

Case I:

$$q' = q + \epsilon u(q, t), \tag{A 3}$$

$$t' = t. \tag{A 4}$$

This case includes the coordinate translations and scale transformations, and the Galilean transformation.

The velocity transformation is then ($\dot{q}' \equiv dq'/dt'$)

$$\dot{q}' = dq'/dt = \dot{q} + \epsilon \dot{u}. \tag{A 5}$$

Solving (A 3) and (A 5) simultaneously to eliminate ϵ,

$$u\dot{q}' - \dot{u}q' = u\dot{q} - \dot{u}q. \tag{A 6}$$

Regarding u and $\dot{u}$ as functions of t' on the left side of (A 6), the quantity $u\dot{q} - \dot{u}q$ is an invariant of the transformations (A 3) and (A 4).

For coordinate translations, $u = 1$ may be chosen (then ϵ becomes the generator of the translations). Then $\dot{q}$ is an invariant quantity, and if (and only if) the Lagrangian $L = L(\dot{q}, t)$, it is invariant with respect to this transformation (as discussed in detail in I). Considering the EM, $\ddot{q}$ is also invariant, and thus f in (1.1) must be a function only of $\dot{q}$ and t for invariance of the EM to coordinate translations.

For coordinate-scale transformations, $u = q$ may be chosen ($\epsilon > -1$ then generates the scale transformations). The invariant quantity $u\dot{q} - \dot{u}q$ is identically 0. However, (A 3) and (A 5) are, in this case,

$$q' = (1 + \epsilon)q, \qquad \dot{q}' = (1 + \epsilon)\dot{q},$$

which yield the invariant $\dot{q}/q$. Thus if and only if L can be expressed as $L(\dot{q}/q, t)$, it will be invariant with respect to the coordinate-scale transformations. Then $\ddot{q}/q$ is also invariant, and the EM must have the form (1.4b) to be coordinate-scale invariant.

For the Galilean transformations, $u = -Vt/\epsilon$, and the invariant is proportional to $t\dot{q} - q$. Thus, if and only if $L = L(t\dot{q} - q, t)$, it is invariant with respect to Galilean transformations. Since $\ddot{q}$ is also invariant to this transformation, it is necessary and sufficient that the EM be of the form

$$\ddot{q} = f(t\dot{q} - q, t) \tag{A 7}$$

for Galilean invariance.

Case II:

$$q' = q, \tag{A 8}$$

$$t' = t + \delta v(t). \tag{A 9}$$

These transformations include the time transformations discussed herein. The resultant velocity transformation is

$$\dot{q}' = (1 + \delta\dot{v})^{-1}\dot{q}. \tag{A 10}$$

Combining (A 9) and (A 10) to eliminate δ yields

$$\dot{q}' + t'(\dot{v}/v)\dot{q}' = \dot{q} + t(\dot{v}/v)\dot{q}', \tag{A 11}$$

which is not identical in the primed and original variables. However, for the time translations, $\dot{v} = 0$, and in this case $\dot{q}$ and q are invariants. Thus, if and only if $L = L(\dot{q}, q)$, it will be invariant with respect to time translation. Also, $\ddot{q}$ is invariant in this case, and therefore it is necessary and sufficient that the EM be of the form

$$\ddot{q} = f(\dot{q}, q) \tag{A 12}$$

for time-translation invariance.

For time-scale transformations, $v = t$, and (A 11) reduces to

$$\dot{q}'t' = \dot{q}t, \tag{A 13}$$

so that $\dot{q}t$ is an invariant for this transformation. Thus, L will be invariant to the time-scale transformations if and only if

$$L = L(\dot{q}t, q). \tag{A 14}$$

From (A 13) and (A 9), it also follows, in this case, that

$$\ddot{q}'t'^2 = \ddot{q}t^2. \tag{A 15}$$

Thus the EM must be of the form (2.10) for time-scale invariance. Alternately, from (A 10),

$$\ddot{q}'/(\dot{q}')^2 = \ddot{q}/\dot{q}^2, \tag{A 16}$$

so that $\ddot{q}/\dot{q}^2$ is invariant, and an EM of the form

$$\ddot{q} = \dot{q}^2 g(\dot{q}t, q) \tag{A 17}$$

will also be invariant, but multiplication by t^2 converts it to the form (2.10).

Relativistic Spinor Formulation of Stokes Parameters with Application to the Inverse Compton Effect*

DAVID S. DE YOUNG
Department of Physics
and
Center for Radiophysics and Space Research,
Cornell University, Ithaca, New York
(Received 9 February 1966; revised manuscript received 29 April 1966)

A formulation of the Stokes Parameters for light is found in terms of second-rank antisymmetric spinors, and arbitrary Lorentz transformations are made by using the unimodular representation of the Lorentz group. A three-component spinor form is established along with its accompanying representation of the Lorentz group, and the connection with the photon spin is explicitly shown. Differential Compton cross sections for any Lorentz frame are calculated. These are seen to be invariant only for particular polarization reference directions.

I. INTRODUCTION

THE use of four parameters to completely characterize the state of polarization of a beam of light was first introduced by Stokes[1] in 1852. More recent applications have been made by Fano,[2] Tolhoek,[3] Lipps and Tolhoek,[4] and a comprehensive summary has been given by McMaster.[5] The use of such parameters for determining polarization effects in scattering processes has been considered, particularly with regard to Compton scattering.[2,5] Recently interest has been directed toward the production of high-energy photons by Compton scattering of light from energetic electrons, several authors[6-9] having considered this process with a laser as a photon source. This effect has also entered into considerations of the energy loss experienced by interstellar electrons.[10] Milburn[6] has obtained differential cross sections in the extreme relativistic limit for a specific initial state of purely plane-polarized light. Arutyunyan[9] *et. al.* use the Stokes parameters to obtain the degree of polarization of Compton scattered light in the extreme relativistic limit, although no cross sections are given.

In many astrophysical applications, the incident light is partially polarized or unpolarized, and the electron energies cover a wide range. Hence a general formalism which provides a simple method of calculating polarization effects for all energies and geometries is desirable. Such a formalism must of course include a method of calculating the transformation of polarization effects from one arbitrary Lorentz frame to another. The spinor formulation of the Stokes parameters presented here provides a completely general system for the calculation of cross sections and polarization effects for all energies and polarization conditions, along with the transformation of all quantities from one Lorentz frame to another. In addition to providing a simple and straightforward method of calculation, such a formalism provides insight into the relation among the relevant spinors, the spin of the photon, and the polarization.

The four Stokes parameters commonly used are P_0, the intensity of the beam, P_1, the degree of plane polarization with respect to two arbitrarily oriented orthogonal axes which are perpendicular to the beam, P_2, the degree of plane polarization with respect to two orthogonal axes rotated 45° to the right of those used for P_1, and P_3, the degree of circular polarization. The degree of polarization is given by

$$P = [P_1^2 + P_2^2 + P_3^2]^{\frac{1}{2}}/P_0. \tag{1.1}$$

More detailed construction of the P_i is given in Refs. 1–5. As a pure polarization state may be written in terms of two orthogonal "basis" states, $\phi = c_1\phi_1 + c_2\phi_2$, it is possible to describe a partially polarized beam as a super-position of such states and to form a density matrix

* This work was supported in part by the United States Air Force Grant AFOSR 321-63.
[1] G. G. Stokes, Trans. Cambridge Phil. Soc. 9, 399 (1852).
[2] U. Fano, J. Opt. Soc. Am. 39, 859 (1949).
[3] H. A. Tolhoek, Rev. Mod. Phys. 28, 277 (1956).
[4] F. Lipps and H. A. Tolhoek, Physica 20, 85 (1954).
[5] W. H. McMaster, Am. J. Phys. 22, 351 (1954).
[6] R. H. Milburn, Phys. Rev. Letters 10, 75 (1963).
[7] C. Bemporad, Cambridge Electron Accelerator Report CEAL-1014.
[8] F. R. Arutyunyan and V. A. Tumanyan, Zh. Eksperim. i Teor. Fiz. 44, 2100 (1963) [English transl.: Soviet Phys.—JETP 17, 1412 (1963)].
[9] F. R. Arutyunyan, I. I. Gol'dman, and V. A. Tumanyan, Zh. Eksperim. i Teor. Fiz. 45, 312 (1963) [English transl.: Soviet Phys.—JETP 18, 218 (1964)].
[10] E. Feenberg and H. Primakoff, Phys. Rev. 73, 449 (1948).

$$\rho = \begin{pmatrix} c_1c_1^* & c_1c_2^* \\ c_2c_1^* & c_2c_2^* \end{pmatrix}. \tag{1.2}$$

The Stokes parameters are given by

$$\begin{aligned} P_0 &= |c_1|^2 + |c_2|^2, \quad P_1 = |c_1|^2 - |c_2|^2, \\ P_2 &= c_1c_2^* + c_2c_1^*, \quad P_3 = i(c_2c_1^* - c_1c_2^*). \end{aligned} \tag{1.3}$$

Some authors choose to regard the ϕ_1 and ϕ_2 as creating a two-component first-rank spinor and thus utilize Pauli-type spin matrices. It should be mentioned that, for photons, this is not strictly true, since a Lorentz transformation of first-rank spinors does not lead to correct results. Although there are only two independent polarization states for a plane wave, it will be seen that the correct form to use is that of a second-rank spinor.

In Secs. II and III, a formulation of the Stokes parameters in the correct second-rank spinor form is given, and the transformation properties of such quantities via the unimodular representation of the Lorentz group is noted. The spinor form is also used to generate a 3×3 representation of the Lorentz group which bears a close relation to the photon spin, and the photon polarization is characterized in this form. The quantum mechanical form of the Stokes parameters using the photon annihilation and creation operators is discussed briefly. Section IV considers the application to Compton scattering, the differential cross sections being obtained for arbitrary polarization and electron energies, together with the degree of polarization of the scattered photons.

II. SPINOR FORMALISM

In the following, we denote the first-rank two-component spinors which are elements of a complex two-dimensional space by $\phi = (\phi_1, \phi_2)$. The set of unimodular linear transformations on this space form the two-dimensional group C_2,

$$\phi_1' = a\phi_1 + b\phi_2, \ \phi_2' = c\phi_1 + d\phi_2, \ ad - bc = 1. \tag{2.1}$$

Spinors which transform according to the complex conjugate of Eq. (2.1) form an associated space which is distinguishable from the space spanned by $\phi = (\phi_1, \phi_2)$ due to the difference in the complex conjugate of the transformation matrix. (Such a distinction could not be made if unitarity were required in addition to unimodularity.) We denote spinors that transform according to the complex conjugate of the transformation in Eq. (2.1) by a dot over the index;

$$\begin{aligned} \phi_{\dot\alpha}' &= \bar a\phi_{\dot\alpha} + \bar b\phi_{\dot\beta}, \quad \phi_{\dot\beta}' = \bar c\phi_{\dot\alpha} + \bar d\phi_{\dot\beta}, \\ \bar a\bar d - \bar b\bar c &= 1; \quad \alpha = 1, \ \beta = 2. \end{aligned} \tag{2.2}$$

Spinors of higher rank are of course generated from the direct product and may or may not have mixed dotted and undotted indices, depending upon the spaces involved in the product, e.g., $\phi_{\alpha\dot\beta} = \phi_\alpha \times \phi_{\dot\beta} = (\phi_{\dot\alpha\alpha}, \phi_{\dot\alpha\beta}, \phi_{\dot\beta\alpha}, \phi_{\dot\beta\beta})$; $\alpha = 1, \beta = 2$.

As a classical plane electromagnetic wave may be characterized either by the vector potential or by the fields $\mathbf{E}$ and $\mathbf{H}$, a spinor representation of such quantities should serve as a starting point for the formation of the Stokes parameters. It can easily be shown[11,12] that a correspondence exists between any four-vector such as A_μ and a second-rank spinor of the form $\phi_{\dot\alpha\beta}$. However, as a more intuitive appreciation of the polarization of light is obtained from the $\mathbf{E}$ and $\mathbf{H}$ fields, we proceed from the field tensor $F^{\mu\nu}$. One can show[12] that to any second-rank tensor $T^{\mu\nu}$ there corresponds a fourth-rank spinor $\phi_{\dot\alpha\beta\dot\gamma\delta}$, but due to the antisymmetry of the field tensor $F^{\mu\nu}$ it is possible also to describe the fields in terms of second-rank spinors by first forming the antisymmetric self-dual tensor $G^{\mu\nu} = F^{\mu\nu} + F^{\mu\nu *}$. This procedure was first introduced by LaPorte and Uhlenbeck,[11] who obtain the following second-rank spinor components:

$$\begin{aligned} \phi_{\dot\alpha\alpha} &= 2(k_2 + ik_1), \quad \phi_{\dot\beta\beta} = 2(k_2 - ik_1), \\ \phi_{\dot\alpha\beta} &= \phi_{\dot\beta\alpha} = -2ik_3, \quad \alpha = 1, \ \beta = 2, \end{aligned} \tag{2.3}$$

where $\mathbf{k} = \mathbf{H} - i\mathbf{E}$. Spinor components with undotted indices are obtained by taking the complex conjugate of Eq. (2.3).

Hence Eq. (2.3) and its complex conjugate give the fields in terms of the second-rank spinor components. For a plane polarized wave traveling along the z axis, application of $\mathbf{E}\cdot\mathbf{H} = 0$ and $E^2 - H^2 = 0$ gives

$$\phi_{\dot\alpha\beta} = \phi_{\dot\beta\alpha} = \phi_{\dot\beta\beta} = 0$$

and

$$\begin{aligned} E_x &= \tfrac{1}{8}(\phi_{\dot\alpha\alpha} + \phi_{\alpha\alpha}), \\ E_y &= \tfrac{1}{8}(\phi_{\dot\alpha\alpha} - \phi_{\alpha\alpha}), \end{aligned} \tag{2.4}$$

with $\alpha = 1$, $\beta = 2$. The relative phases of E_x and E_y are not included in Eq. (2.4). For any other set of axes the fields will not assume such a simple form, but in any case the spinor components $\phi_{\alpha\beta}$ and their complex conjugates serve to completely specify all components of $\mathbf{E}$ and $\mathbf{H}$.

[11] O. LaPorte and G. Uhlenbeck, Phys. Rev. **37**, 1380 (1931).

[12] W. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953).

One may now regard the polarization properties of a light beam as being completely characterized by a second-rank spinor, and all relevant operations may be carried out in the actual spinor space. Such a point of view will be seen to have certain advantages in that the spinor form more directly reflects the helicity states of the photon, 3-space rotations and Lorentz transformations may be carried out by simple and (of more importance) irreducible representations of the Lorentz group, and such representations may be used to generate the transformations of the photon annihilation and creation operators.

One may form a 4×4 density matrix in the spinor space from the product $\phi_{\alpha\beta}\phi_{\dot\alpha\dot\beta}$ or a more conventional 2×2 matrix using any two mutually orthogonal field components and Eq. (2.3). This will then correspond to the conventional form of Eq. (1.2). For the special choice of axes used in obtaining Eq. (2.4), applying Eqs. (1.2) and (1.3) gives the Stokes parameters in this case as

$$\begin{aligned} P_0 &= \tfrac{1}{16}\,|\phi_{\alpha\alpha}|^2, \\ P_1 &= \tfrac{1}{32}(\phi^2_{\dot\alpha\dot\alpha} + \phi^2_{\alpha\alpha}), \\ P_2 &= \tfrac{1}{32}(\phi^2_{\dot\alpha\dot\alpha} - \phi^2_{\alpha\alpha})\cos\delta, \\ P_3 &= \tfrac{1}{32}(\phi^2_{\dot\alpha\dot\alpha} - \phi^2_{\alpha\alpha})\sin\delta, \end{aligned} \tag{2.5}$$

where the relative phase δ between the two orthogonal components of the field has been explicitly shown. For any other choice of reference axes the forms in (2.5) will be more complex, involving terms in $\phi_{\dot\alpha\dot\beta}$, $\phi_{\dot\beta\dot\alpha}$, etc.

Arbitrary Lorentz transformations of the fields and polarizations are readily obtained by applying the spinor representation of the Lorentz group. Such a representation is directly obtained in terms of second-rank spinors from the fact that the four-dimensional group of proper rotations is homomorphic to the direct product of the group C_2 with itself, i.e., $C_2' \times C_2$ is a universal covering group of O_4. The prime on the first factor indicates that in general the parameters of the two groups C_2 are independent. The proper homogeneous Lorentz transformations, with the time coordinate written as $ict = x_4$, form a subgroup of O_4 since the form $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is left invariant, and moreover it can be shown[12,13] that the parameters of C_2' are completely determined by $C_2(C_2' = C_2^*)$ so that in fact this restricted Lorentz group is homomorphic to C_2. The spinor representation must have elements of determinant 1 (unimodularity) and must leave the form $\phi_{\alpha\alpha}\phi_{\beta\beta} - \phi_{\alpha\beta}\phi_{\beta\alpha}(\alpha = 1, \beta = 2)$ invariant.[13] This then allows the transformation of any type of polarization, for once the transformation of the density matrix for a pure state is obtained the same transformation will apply to any arbitrary superposition of states. Consider, for example, two observers S and S' such that the light travels along the z axis in S [i.e., Eq. (2.5) applies]. Let S' be cutting across the beam at an arbitrary angle θ with respect to z and with arbitrary speed v relative to S. Further let the motion of S' be contained in the z–y plane of S. For spinor frames where a 4-vector x_μ is written as the spinor $\phi_{\dot\alpha\beta} = (z + ct, x + iy, x - iy, -z + ct)$, a unimodular Lorentz transformation for a first-rank 2-component spinor is given, in the case of z axis velocity, by[14]

$$L = \begin{pmatrix} A & 0 \\ 0 & A^{-1} \end{pmatrix} \tag{2.6}$$

and a 4-component $\phi_{\dot\alpha\beta}$ transforms according to

$$L^* \times L = \begin{pmatrix} A^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & A^{-2} \end{pmatrix}, \tag{2.7}$$

where

$$A^2 = [(1 - \beta)/(1 + \beta)]^{\frac{1}{2}} \quad \text{and} \quad \beta = v/c.$$

Rotations are performed via the usual 2×2 half Euler angle matrices U, the 4×4 representation being obtained via the outer product. A short calculation shows that $L^* \times L^*$ and $L \times L$ along with the corresponding rotation matrices also satisfy unimodularity and leave the form $\phi_{\alpha\alpha}\phi_{\beta\beta} - \phi_{\alpha\beta}\phi_{\beta\alpha}$; $\alpha = 1, \beta = 2$ (or $\phi_{\dot\alpha\dot\alpha}\phi_{\dot\beta\dot\beta} - \phi_{\dot\alpha\dot\beta}\phi_{\dot\beta\dot\alpha}$) invariant and hence form a representation of the Lorentz group for $\phi_{\alpha\beta}$ and $\phi_{\dot\alpha\dot\beta}$. Fields in S' are then obtained by the usual sequence of rotation to bring z and z' parallel, application of Eq. (2.6), and inverse rotation.

If one requires that the polarization of a beam is always measured with the apparatus oriented exactly along the beam (which appears to be the only meaningful measurement to make), then the inverse rotation is not through $-\theta$ but through the angle given by

$$\cos\theta' = [\cos\theta - (v/c)]/[1 - (v/c)\cos\theta],$$

i.e., abberation in S' must be considered. Applying the above-mentioned sequence of three operations to the column spinor with elements $(\phi_{\dot\alpha\dot\alpha}, \phi_{\dot\alpha\dot\beta}, \phi_{\dot\beta\dot\alpha}, \phi_{\dot\beta\dot\beta})$

[13] J. Aharoni, *The Special Theory of Relativity* (Oxford University Press, London, 1959).

[14] Reference 13, pp. 195–201. The sign of $\beta = v/c$ in this book is the negative of that conventionally used. We use the conventional sign here.

(where only $\phi_{\dot{a}\dot{a}} \neq 0$ in S) and its complex conjugate, gives the spinors as seen in S'. Insertion into Eq. (2.5) and (1.1) shows that for any velocity and any angle θ, S, and S' observe exactly the same type and degree of polarization, although the light is of course Doppler-shifted. The real value of the above result occurs in cases where no inverse rotation is desired. Such conditions frequently occur in scattering experiments and in astrophysical cases, for example, in the measurement of amplitudes polarized perpendicular to and contained in some arbitrary plane. Here the formalism allows an easy calculation of the polarization amplitudes.

III. PHOTON SPIN

The relation between the photon spin and its associated spinors is more clearly seen when transformations are made according to a 3×3 representation of the Lorentz group generated from the spinors ϕ_α and $\phi_{\dot{a}}$.[15] This can be done through the standard method of forming monomials of the spinors as basis vectors in the representation space. However, as opposed to the method for 3-space rotations using the unitary group U_2, the spaces spanned here by ϕ_α and $\phi_{\dot{a}}$ are distinguishable, and thus the monomials must include powers of ϕ_α and $\phi_{\dot{a}}$.

Let u and v be two independent integers. Then form the monomials

$$f_{s,t} = K(\phi_\alpha)^{u-s}(\phi_{\dot{a}})^{v-t}(\phi_\beta)^s(\phi_{\dot{\beta}})^t, \tag{3.1}$$

where K is a normalization constant, s and t are integers and

$$0 \leq s \leq u, \qquad 0 \leq t \leq v. \tag{3.2}$$

Clearly there exist $[(u+1)(v+1)]$ monomials generated by the independent variation of s and t; hence the forms in Eq. (3.1) span a $[(u+1)(v+1)]$-dimensional space. As the transformations of Eqs. (2.1) and (2.2) are linear, the transformation of $f_{s,t}$ by application of Eqs. (2.1) and (2.2) will result in a polynomial contained in the space spanned by the original $f_{s,t}$. Calculation shows that the product of two transformations equals the successive application of them, and that the unit transformation [$a = d = 1$ and $b = c = 0$ in Eqs. (2.1) and (2.2)] yields the original $f_{s,t}$. Hence Eq. (3.1) does provide a representation space, and the matrices of the representation are found by applying Eqs. (2.1) and (2.2) to the $f_{s,t}$ and regrouping in terms of the original monomials:

$$\begin{aligned} f'_{s,t} &= \sum_{i=0}^{u} \sum_{j=0}^{v} D_{s,i} D_{t,j} f_{ij} \\ &= \sum_{i} \sum_{j} D_{s,i,t,j} f_{ij}. \end{aligned} \tag{3.3}$$

The matrix elements D are formed from the direct product of irreducible representations of C_2^* and C_2; hence the above representation is irreducible. If unitarity as well as unimodularity could be imposed,[16] the elements D would be derived from U_2^* and U_2. As U_2 is a universal covering group for O_3, the three-dimensional rotation group, the form of D would be $D^{J^*} \times D^j$, where the D^J are the $(2J+1)$-dimensional matrices of the irreducible representations of O_3. With this similarity in mind, let us set

$$u = 2J, \quad v = 2j, \quad s = J - M, \quad t = j - m. \tag{3.4}$$

The normalization constant K in Eq. (3.1) may be evaluated from the condition of unimodularity or may be seen directly from the fact that U_2 is a subgroup of C_2, and thus the same (real) normalization constant is used in representations of both groups. Insertion of Eq. (3.4) into Eq. (3.1) yields

$$\begin{aligned} f^{Jj}_{Mm} = {} & [(J+M)!\,(J-M)!\,(j+m)!\,(j-m)!]^{-\frac{1}{2}} \\ & \times (\phi_\alpha)^{J+M}(\phi_{\dot{a}})^{j+m}(\phi_\beta)^{J-M}(\phi_{\dot{\beta}})^{j-m}; \end{aligned} \tag{3.5}$$

$$\alpha = 1, \qquad \beta = 2.$$

Equation (3.2) becomes

$$-J \leq M \leq J, \qquad -j \leq m \leq j$$

and the forms of Eq. (3.5) span a $[(2J+1)(2j+1)]$-dimensional space. Hence the matrices D will be $[(2J+1)(2j+1)]$-dimensional, and we now find their explicit form for a general Lorentz transformation.

We consider a general transformation composed of a rotation through the two Euler angles θ and ω and a velocity transformation along the new z axis. Uninteresting rotations about the new z axis are not considered. In terms of the two-dimensional spinor transformations of Eqs. (2.1) and (2.2), the velocity transformation is given by Eq. (2.6) and

[15] For a slightly different approach to three-component spinors and a complete discussion of their properties, see A. Peres, J. Math. Mech. 11, 61 (1962).

[16] This cannot be done because the noncompactness of the Lorentz group prohibits the existence of finite-dimensional matrices which would transform a given finite matrix into a unitary matrix.

the rotation by the usual 2×2 half Euler angle matrices of the U_2 representation:

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} e^{i\frac{1}{2}\omega} \cos \frac{1}{2}\theta & e^{-i\frac{1}{2}\omega} \sin \frac{1}{2}\theta \\ -e^{i\frac{1}{2}\omega} \sin \frac{1}{2}\theta & e^{-i\frac{1}{2}\omega} \cos \frac{1}{2}\theta \end{pmatrix}. \quad (3.6)$$

Equation (3.6) and Eq. (2.6) are applied to the spinors ϕ_α, and the result is substituted into Eq. (3.5). In a manner entirely analogous to that used for three-dimensional rotations,[17] the terms involving ϕ_α and $\phi_{\dot{\alpha}}$ are expanded via the binomial theorem and regrouped in terms of the original f^{Ji}_{Mm}. When this is done one finds

$$f'^{Ji}_{Mm} = \sum_{M'} \sum_{m'} D^{Ji}_{M'Mm'm} f_{M'm'}, \quad (3.7)$$

where

$$D^{Ji}_{M'Mm'm} = D^{J}_{M'M}(\omega, \theta, 0) \times D^{i*}_{m'm}(\omega, \theta, 0) A^{2(M+m)}. \quad (3.8)$$

Here $D^J_{M'M}$ and $D^i_{m'm}$ are the standard $(2J+1) \times (2J+1)$ and $(2j+1) \times (2j+1)$ matrices representing the three-dimensional rotation group:

$$D^J_{M'M}(\omega, \theta, \lambda) = \sum_k \frac{[(J+M)!\,(J-M)!\,(J+M')!\,(J-M')!]^{\frac{1}{2}}}{(J+M-k)!\,k!\,(J-M'-k)!\,(M'-M+k)!} \times (-1)^{M'-M-k} e^{iM'\omega} [\cos \tfrac{1}{2}\theta]^{2J+M-M'-2k} \times [\sin \tfrac{1}{2}\theta]^{2k+M'-M} e^{iM\lambda}. \quad (3.9)$$

This representation is unimodular but not unitary. Explicit matrices may be calculated from Eq. (3.8) by using the expression

$$D^{J*}_{M'M} = (-1)^{M-M'} D^J_{-M'-M}. \quad (3.10)$$

For the photon case, the above representation connects the photon spin and a classical circularly polarized wave via the spinor formalism. A three-dimensional representation will be obtained from either $D^{(10)}$ or $D^{(01)}$. That these are in fact the correct values for J and j has been shown by Weinberg,[18] who has demonstrated that, for massless particles of helicity λ, the condition $j - J = \lambda$ must be met. Thus for photons of helicity ± 1, the relevant forms are $D^{(10)}$ and $D^{(01)}$. Equation (3.5) becomes, with $J = 1$, $j = 0$, $M = +1$, say,

$$f^{10}_{10} = 1/\sqrt{2}\,(\phi_\alpha)^2 = 1/\sqrt{2}\,(\phi_{\alpha\alpha}). \quad (3.11)$$

The last step in Eq. (3.11) is allowed because any second-rank symmetric tensor may be written as $\phi_{\alpha\beta} = \frac{1}{2}(\chi_\alpha \psi_\beta + \chi_\beta \psi_\alpha)$,[19] and in this case our $\phi_{\alpha\beta}$ is generated from the outer product of a two-component spinor with itself. Equation (3.8) therefore shows $D^{(10)}$ gives the transformation of $\phi_{\alpha\beta}$ under a rotation plus z axis Lorentz transformation, while $D^{(01)}$ transforms $\phi_{\dot{\alpha}\dot{\beta}}$. The case $M = 0$ or $m = 0$ gives the transformation of $\frac{1}{2}(\phi_{12} + \phi_{21})$, which in our case is equal to ϕ_{12} as we always have $\phi_{12} = \phi_{21}$ [Eq. (2.3)].

Classically, a left circularly or elliptically polarized beam is characterized, in the case of z-axis propagation, by $E_x + iE_y$. Applying Eq. (2.4) gives $E_x + iE_y = \frac{1}{4}\phi_{11}$, and therefore $D^{(10)}$ transforms a left circularly or elliptically polarized beam. In this case, the other two components ϕ_{12} and ϕ_{22} of the column 3-vector to which $D^{(10)}$ is applied are both equal to zero. ϕ_{12} will always be zero unless the z axis is rotated, and $\phi_{11} = 0$, $\phi_{22} \neq 0$ corresponds to the case of a change in sign of the z axis. The relation to photon spin is evident, with $D^{(10)}$ giving the transformation of a photon with spin parallel to $\mathbf{k}$, the direction of the beam. $\phi_{11} \neq 0$ corresponds to the z axis being oriented parallel to $\mathbf{k}$, and $\phi_{22} \neq 0$ corresponds to the z axis antiparallel to $\mathbf{k}$; i.e., the two possible "m" components of the photon spin. Similarly, a left circularly polarized wave is given by $E_x - iE_y = \frac{1}{4}\phi_{\dot{\alpha}\dot{\alpha}}(\alpha = 1)$ and transforms according to $D^{(01)}$, corresponding to a negative helicity photon with spin antiparallel to $\mathbf{k}$. Stokes parameters using the two circular polarization states as a basis are easily formed by a similarity transformation and are given as

$$P_0 = |e_+|^2 + |e_-|^2, \quad P_1 = (e_- e_+^* + e_+ e_-^*), \quad P_2 = i(e_- e_+^* - e_+ e_-^*), \quad P_3 = |e_+|^2 - |e_-|^2, \quad (3.12)$$

where e_+ and e_- are the positive and negative helicity amplitudes. The relative phase between e_+ and e_- is suppressed here.

For example, a transformation between the observers S and S' mentioned earlier involves a rotation through the Euler angle θ and gives, via Eq. (3.8),

[17] See, for example, E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959), pp. 163–167; or M. Hammermesh, *Group Theory and its Application to Physical Problems* (Addison–Wesley Publishing Company, Inc., Reading, Mass., 1962), pp. 350–356. The orientation of the 3-space rotation used here is in the same sense as that of Hammermesh.

[18] S. Weinberg, Phys. Rev. **134**, B882 (1964).

[19] E. M. Corson, *Introduction to Tensors, Spinors, and Relativistic Wave Equations* (Blackie & Son Ltd., London, 1953).

$$D^{(10)}_{M'Mm'm} = \begin{bmatrix} (1/2)A^2(1+\cos\theta) & (-1/\sqrt{2})\sin\theta & (1/2)A^{-2}(1-\cos\theta) \\ (1/\sqrt{2})A^2\sin\theta & \cos\theta & (-1/\sqrt{2})A^{-2}\sin\theta \\ (1/2)A^2(1-\cos\theta) & (1/\sqrt{2})\sin\theta & (1/2)A^{-2}(1+\cos\theta) \end{bmatrix}. \tag{3.13}$$

This allows one to compute the polarization amplitudes parallel and perpendicular to some arbitrary plane. Inverse rotation through the angle $\theta' = \cos^{-1}[(\cos\theta - \beta)/(1 - \beta\cos\theta)]$ and the application of a column vector $(\phi_{11}, 0, 0)$ gives

$$\begin{bmatrix} \gamma(1+\beta\cos\theta)\phi_{11} \\ 0 \\ 0 \end{bmatrix}. \tag{3.14}$$

Application of Eq. (3.12) gives the same polarization in S and in S'. Eq. (3.14) shows directly that the helicity is unchanged, corresponding to the fact that the photon spin is parallel or antiparallel to its direction of motion for all Lorentz observers.

In the case of the quantized photon field, Stokes parameters may be formed quadratic forms of the annihilation and creation operators a and a^+. One suggested[20] form is

$$P_0 = a_1^+a_1 + a_2^+a_2, \qquad P_1 = a_1^+a_1 - a_2^+a_2,$$
$$P_2 = a_1^+a_2 + a_2^+a_1, \qquad P_3 = i(a_2^+a_1 - a_1^+a_2),$$

where a_1 and a_2 correspond to the transverse components of the field. Transformations of a and a^+ have been shown by Weinberg[18] to again be generated by the representations $D^{(10)}$ and $D^{(01)}$ but not by the 4-vector representation $D^{(\frac{1}{2}\frac{1}{2})}$. Hence the present formalism may also be applied in the quantized case.

IV. APPLICATIONS–COMPTON EFFECT

The formalism developed in the preceding sections may be applied to all processes involving polarization effects and especially to scattering phenomena. Here the above results provide a convenient method of determining the polarization dependence of the cross sections, of transforming the cross sections between any two Lorentz frames, and of finding the state of polarization of the scattered beam all in one operation. The particular example chosen will be that of the inverse Compton effect. The method applies for any conditions on the incident and scattered light and on the scattering electrons. A useful means of calculating the polarization dependence of Compton cross sections has been given by Fano[2] wherein a matrix describing the scattering process is applied to a four-component column "vector" composed of the four Stokes parameters. Cross sections are then obtained by applying a similar four-component row "vector" which describes the type of polarization to which the analyzer is sensitive. Fano's matrix is set up for incident polarization basis vectors lying in and perpendicular to the plane of scattering. Hence, for incident plane polarized light polarized parallel to the scattering plane, the incident normalized column vector with elements (P_0, P_1, P_2, P_3) becomes $(1, 1, 0, 0)$. After averaging over electron spins, Fano's matrix becomes

$$T_{ik} = \tfrac{1}{2}r_0^2(k/k_0)^2 \times \begin{bmatrix} 1+\cos^2\theta+(k_0-k)(1-\cos\theta) & -\sin^2\theta & 0 & 0 \\ -\sin^2\theta & 1+\cos^2\theta & 0 & 0 \\ 0 & 0 & 2\cos\theta & 0 \\ 0 & 0 & 0 & 2\cos\theta+(k_0-k)(1-\cos\theta)\cos\theta \end{bmatrix}, \tag{4.1}$$

where k_0 is the incident photon momentum, k the scattered photon momentum, θ the angle of scattering, and r_0 the classical electron radius. Equation (4.1) is in units of $m_ec^2 = 1$, which will be retained throughout. The differential cross section is then given by

$$\frac{d\sigma}{d\Omega} = \frac{1}{2}\sum_{i,k=0}^{3} T_{ik}w_iv_k, \tag{4.2}$$

where v_k are the components of the incident Stokes "vector" and w_i are those of the analyzer. Equation (4.1) is derived for the process seen in the rest frame of the electron, a frame which is not always that of the observer, particularly in the astrophysical case. The formalism developed in Secs. II

[20] J. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1955), p. 45.

and III now allows us to readily calculate the cross sections in any Lorentz frame moving in any direction with respect to the electron rest frame. Let us assume head-on collision between the electron and photon, i.e., the electron is moving into the beam of light, and that the photon is scattered at an angle θ as measured in the electron rest frame. Transformation of cross sections to the laboratory frame of the scattered light is seen thus to be exactly that of the example treated in Sec. II, where θ is now the scattering angle. Hence the Stokes parameters as we have written them are unchanged in going from S to S' (except for an over-all normalization factor), and therefore Eq. (4.2) applies in the laboratory frame. Hence the differential cross sections for scattered photons polarized parallel and perpendicular to the scattering plane are known immediately to be the same in all frames. Such cross sections are of course dependent on the electron energy due to the Doppler shifting of the incident light.

In many cases it is experimentally more convenient, and perhaps necessary, to measure the cross sections for light scattered parallel and perpendicular to the plane of polarization of an incident beam. Here the differential cross sections are not invariant but differ in S and S', and it is in cases such as this that the present formalism provides a much simpler method of calculation. The extreme relativistic limit of this case has been treated by Milburn,[6] but, if the electron energy is not always high, such an approximation may break down and a more general result is desirable. The exact form of the transformed cross sections in this case is obtained by applying the results of Secs. II and III in the following manner. We consider first the process as seen in the electron rest frame, using polarization basis vectors appropriate to Eq. (4.1). For a linearly polarized plane wave polarized at an angle ϕ with respect to the scattering plane, the incident Stokes "vector" becomes

$$\begin{pmatrix} 1 \\ \cos 2\phi \\ -\sin 2\phi \\ 0 \end{pmatrix}. \tag{4.3}$$

Equation (4.1) is applied to this. The resulting column vector must now be transformed to a system 2 of whose axes lie along the direction of incident polarization and along the incident beam. Following Milburn,[21] we take the z axis parallel to the beam and the x axis parallel to the direction of incident polarization. Application to this new Stokes vector of an analyzer (1, 1, 0, 0) will yield the differential cross section for scattered photons parallel to the plane of polarization of the incident photon, while application of (1, −1, 0, 0) will give the differential cross section for outgoing photons polarized perpendicular to the incident reference plane. Such a transformation gives $P_0' = P_0$, as expected, and

$$P_1' = -P_0 \sin^2\theta \cos^2\phi - P_1 \sin^2\theta \sin^2\phi + 2P_2 \cos\theta \cos\phi \sin\phi, \tag{4.4}$$

where θ is the scattering angle. P_2' is not needed in the cross section calculation and P_3 is always zero in this case. So far everything has been performed in the electron rest frame.

In the laboratory frame, Eq. (4.3) is unchanged as the photon beam and the electron are moving antiparallel to each other. We are measuring the polarization relative to our originally chosen axes at this point, hence we can apply Eq. (4.1) to Eq. (4.3). This gives the scattered Stokes vector for polarizations perpendicular and parallel to the scattering plane. This has been seen to be invariant, and thus we have the same form in S'. A change of reference axes may now be performed in S', using the transformed equation (4.4). This is the only noninvariant form, as θ must be replaced by $\theta' = \cos^{-1}[(\cos\theta - \beta)/(1 - \beta\cos\theta)]$, the laboratory scattering angle. Inserting this change and applying (1, 1, 0, 0) to the transformed Stokes vector gives the differential cross section in the laboratory frame for scattered photons polarized parallel to the plane of incident polarization. One obtains, for arbitrary velocity and with $x = \cos\theta$,

$$\begin{aligned}\left(\frac{d\sigma}{d\Omega}\right)_{\text{lab},\,\parallel} &= \tfrac{1}{4}r_0^2\left(\frac{k}{k_0}\right)^2\left(\frac{k_0^2(1-x)^2}{1+k_0(1-x)}\right.\\
&\times\left[1-\frac{1}{\gamma^2}\frac{1-x^2}{(1+\beta x)^2}\cos^2\phi\right]+(1+x^2)\\
&\times\left\{1+\tfrac{1}{2}\cos^2 2\phi\left[1+\frac{(x-\beta)^2}{(1-\beta x)^2}\right]\right.\\
&\left.-\frac{1}{2\gamma^2}\left[\frac{1-x^2}{(1-\beta x)^2}\right](2\cos 2\phi+1)\right\}+(x^2-1)\\
&\times\left\{\cos 2\phi\left[1-\cos^2\phi\,\frac{1}{\gamma^2}\frac{1-x^2}{(1-\beta x)^2}+\frac{(x-\beta)^2}{2(1-\beta x)^2}\right]\right.\\
&\left.\left.-\frac{1}{2\gamma^2}\left[\frac{1-x^2}{(1-\beta x)^2}\right]\right\}-2x\,\frac{x-\beta}{1-\beta x}\sin^2 2\phi\right).\end{aligned} \tag{4.5}$$

Similarly, application of (1, −1, 0, 0) gives the laboratory cross section for photons perpendicular to the plane of incident polarization:

[21] R. H. Milburn, private communication.

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{lab},\perp} = \tfrac{1}{4}r_0^2\left(\frac{k}{k_0}\right)^2\left(\frac{k_0^2(1-x)^2}{1+k_0(1-x)}\right.$$
$$\times\left[1+\frac{1}{\gamma^2}\frac{1-x^2}{(1-\beta x)^2}\cos^2 2\phi\right]+(1+x^2)$$
$$\times\left\{1-\tfrac{1}{2}\cos^2 2\phi\left[1+\left(\frac{x-\beta}{1-\beta x}\right)^2\right.\right.$$
$$\left.\left.+\frac{1}{2\gamma^2}\frac{1-x^2}{(1-\beta x)^2}(2\cos 2\phi+1)\right]\right\}+(x^2-1)$$
$$\times\left\{\cos 2\phi\left[\cos^2\phi\,\frac{1}{\gamma^2}\frac{1-x^2}{(1-\beta x)^2}-\frac{1}{2}\left(\frac{x-\beta}{1-\beta x}\right)^2\right]\right.$$
$$\left.\left.+\frac{1}{2\gamma^2}\frac{1-x^2}{(1-\beta x)^2}\right\}+2x\,\frac{x-\beta}{1-\beta x}\sin^2 2\phi\right). \quad (4.6)$$

Taking the extreme relativistic limit $\beta \to 1$, $\gamma \to \infty$ and averaging over the angle ϕ, one obtains exactly the result given by Milburn in Ref. (6).

In astrophysical applications, the incident light is usually unpolarized,[10,22] e.g., starlight. For cross sections measuring the components parallel and normal to the plane of scattering and for incident unpolarized light ($P_0 = 1$, $P_2 = P_3 = P_1 = 0$), a direct application of Eq. (4.1) gives

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{lab},\parallel} = \tfrac{1}{4}r_0^2\left(\frac{k}{k_0}\right)^2\left[\frac{k_0^2(1-x)^2}{1+k_0(1-x)}+2x^2\right] \quad (4.7)$$

and

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{lab},\perp} = \tfrac{1}{4}r_0^2\left(\frac{k}{k_0}\right)^2\left[\frac{k_0^2(1-x)^2}{1+k_0(1-x)}+2\right]. \quad (4.8)$$

Application of Eq. (1.1) gives the degree of polarization of the scattered light in both cases as

$$P = (1-x^2)\{k_0^2(1-x)^2$$
$$\times[1+k_0(1-x)]^{-1}+x^2+1\}^{-1}. \quad (4.9)$$

The dependence of these quantities upon the relative velocities of the two frames is obtained from the value of k_0 in the electron rest frame, given in value k_i in the observer's frame. The scattering angle in the laboratory frame is transformed to $x = \cos\theta$, the electron rest-frame cosine, via the standard formula. The electron energy is $E = \gamma mc^2 = \gamma$ in our units, and $k_0 = \gamma(1+\beta)k_i$ for our geometry.

ACKNOWLEDGMENTS

The author would like to thank Dr. P. Morrison for his suggestion of the problem that stimulated this work and also Dr. K. Barnes and Dr. D. Yennie for helpful discussions.

[22] F. Hoyle, Phys. Rev. Letters 15, 131 (1965).

The Solution of a Stark-Effect Model as a Dynamical Group Calculation

G. FLAMAND
Laboratoire de Physique Théorique et Hautes Energies, Orsay, France
(Received 22 March 1966)

A model for the Stark mixing of orbital states in the annihilation of particles at rest is solved by using the dynamical group techniques. The Schrödinger equation for the problem is given meaning in a representation of the three-dimensional rotation group and reduced to a two-dimensional linear homogeneous differential equation.

I. INTRODUCTION

IN attempting to give firmer theoretical grounds to the Stark mixing of orbital states in the annihilation of mesons and antibaryons *at rest* (the so called Day–Snow–Sucher effect), Leon and Bethe considered a semi-classical model.[1] It is the purpose of this paper to give a solution to this model by using dynamical group techniques.

Let us briefly describe the model. The basic physical process involved in this Stark mixing is in the close encounter of the mesic atom (or antibaryon proton atom) with the protons of the hydrogen molecules in the bubble chamber. Leon and Bethe idealized this process as the collision of a hydrogenlike mesic atom moving (on a straight line with constant velocity) through the field Φ of a fixed proton screened by an electron. The internal states of the mesic atom are described quantum mechanically by the standard Schrödinger equation. The field Φ is supposed to induce *dipolar* transitions among the internal states *only* at a given energy $-\frac{1}{2}n^{-2}$.

Let the internal wavefunction of the mesic atom at time t be

$$\psi(t) = \sum_{\alpha=1}^{n^2} a_\alpha(t)\ |n\alpha\rangle.$$

Here α is some set of internal quantum numbers. The physically relevant set we eventually use consists of the orbital numbers l and m. $|n\alpha\rangle$ is the bound-state wavefunction of energy $E_n = -\frac{1}{2}n^{-2}$ and internal quantum number α. In appropriate rotating axes, $\psi(t)$ obeys the equation (which is the mathematical description of the model)

$$\dot{a}_\alpha(t) = \sum_{\beta=1}^{n^2} \{\Phi(t)\langle n\alpha|\ R_z\ |n\beta\rangle + G(t)\langle n\alpha|\ L_y\ |n\beta\rangle\} a_\beta(t). \qquad \text{(A)}$$

For a more complete description of the physical problem, and for the precise forms of $\Phi(t)$ (which represents the action of the field Φ) and of $G(t)$ (which is linked to the rotation of the axes), we refer the reader to Ref. 1. However, as far as this paper is concerned, they are of little significance. The solution of the system (A), which we present in Sec. IV, is a useful computational tool. Indeed, antibaryon annihilation requires the consideration of values of n up to 20–30, which makes (A) a system of 400 to 900 coupled linear differential equations, while we need only solve a simple system of two or three such equations.

However, what interests us most here is the group-theoretic aspect of the model. It provides us with an example of a *dynamical* problem, the solution of which can be obtained by pure powerful group-theoretic techniques. It is of interest to recapitulate the successive steps of this approach. We first identify the states of the physical system as an irreducible representation R of some group G. Secondly, we give a meaning to the various terms in the Schrödinger equation relative to this group G (here we identify the terms with representatives of the Lie algebra $\mathfrak{G}$ of G within R). Then the Schrödinger equation can be given a purely abstract meaning, and can be solved by group techniques. It is noteworthy that the second step involves the identification of $\mathbf{R}$ (the usual position operator) in terms of the elements of the Lie algebra $\mathfrak{G}$. Thus, more generally, we should expect that, in more difficult calculations of the same type, not only the expressions of the elements of G or $\mathfrak{G}$ in terms of the observables $\mathbf{R}$, $\mathbf{P}$ (and functions thereof) are required, but also the knowledge of the observables in terms of elements of G or $\mathfrak{G}$.

This program is carried out in the next sections. In Sec. II, we collect a few hydrogen-atom properties, well known since Pauli's early work.[2] In particular, the needed dynamical group G here is the group O_4 (rotations in a four-dimensional Euclidean

[1] M. Leon and H. A. Bethe, Phys. Rev. 127, 636 (1962).

[2] W. Pauli, Z. Physik 36, 336 (1926).

space). At the end of Sec. II and in Sec. III, we identify the matrices $\langle n\alpha\ |\mathbf{R}|\ n\beta\rangle$ and $\langle n\alpha\ |\mathbf{L}|\ n\beta\rangle$ as the representatives of the Lie algebra of O_4 in the n^2-dimensional irreducible representation of O_4. We refer the reader to the Appendix for the proof of similar properties on the infinite multiplet of the scattering states at a given energy of the hydrogen atom. In Sec. IV, we present the announced group-theoretic solution. To carry it through, we had to develop a new mathematical technique for handling a certain kind of linear homogeneous first-order differential systems having a Lie algebra structure.

II. THE HYDROGEN ATOM AND THE GROUP O_4

In this section, we recall some well-known facts and introduce notations. $\mathbf{R}$ and $\mathbf{P}$ are the usual position and momentum operators satisfying the canonical rule $[R_k, P_l] = i\delta_{kl}$. Choosing $\hbar = 1$, then the Hamiltonian of the hydrogen atom is

$$H = p^2/2m - e^2/R,$$

$\mathbf{L}$ is the angular momentum operator

$$L_i = \varepsilon_{ikl}R_kP_l = (\mathbf{R} \times \mathbf{P})_i,$$

and

$\mathbf{A}$ is the Lenz vector

$$\mathbf{A} = (2me^2)^{-1}(\mathbf{L} \times \mathbf{P} - \mathbf{P} \times \mathbf{L}) + \hat{R} \qquad \text{with} \quad \hat{R} = \mathbf{R}/R.$$

The following commutation rules hold:

$$[\mathbf{L}, H] = 0 = [\mathbf{A}, H],$$

$$[L_i, L_k] = i\varepsilon_{ikl}L_l,\ [L_i, A_k] = i\varepsilon_{ikl}A_l$$

since $\mathbf{A}$ is a vector with respect to $\mathbf{L}$; and finally

$$[A_k, A_l] = \frac{-2i}{me^4}\varepsilon_{ijk}HL_k.$$

We now consider these commutation rules on the multiplet of bound states of energy $H = -|E_n|$, and we introduce $\mathbf{K} = \mathbf{A}(me^4/2\ |E_n|)^{\frac{1}{2}}$. On this multiplet, the above commutation rules become those of the Lie algebra of the group O_4.

$$[L_i, L_k] = i\varepsilon_{ikl}L_l,$$

$$[K_i, K_k] = i\varepsilon_{ikl}L_l,\ [L_i, K_k] = i\varepsilon_{ikl}K_l.$$

It is well known that this Lie algebra is isomorphic to $O_3 \times O_3$. This is easily recognized through the substitution

$$\mathbf{F} = \tfrac{1}{2}(\mathbf{L} + \mathbf{K}), \qquad \mathbf{G} = \tfrac{1}{2}(\mathbf{L} - \mathbf{K}),$$

so that

$$[F_i, G_j] = 0 \quad \text{for all} \quad i, j = 1, 2, 3,$$

$$[F_i, F_k] = i\varepsilon_{ikl}F_l, \qquad [G_i, G_k] = i\varepsilon_{ikl}G_l.$$

The irreducible representations of this Lie algebra are labeled in the standard way by the values of F^2 and G^2,

$$F^2 = \lambda(\lambda + 1),$$

$$G^2 = \mu(\mu + 1), \quad \lambda \text{ and } \mu \text{ are half integers } \geq 0.$$

To distinguish between O_4 and $O_3 \times O_3$, a further algebraic relation is needed. It is provided by the property $\mathbf{L}\cdot\mathbf{A} = 0$ so that $\mu = \lambda$ and $F^2 = G^2 = \lambda(\lambda + 1)$ on the multiplet of bound states $H = -|E_n|$. Then we want to relate λ to a physical quantity. This identification is carried through using the algebraic relation

$$1 - A^2 = (-2/me^4)H(L^2 + 1),$$

whence we compute

$$|E_n| = \tfrac{1}{2}(me^4/n^2) \quad \text{with} \quad n = 2\lambda + 1.$$

Therefore, every bound-state multiplet of the hydrogen atom with energy $-|E_n|$ is a realization of an irreducible unitary representation of the group O_4 labeled by the index λ, $(2\lambda + 1 = n)$, and there exists a one-to-one correspondence between the multiplets of bound states and the irreducible representation of O_4. The restriction of $\mathbf{L}$ to a multiplet is a set of three matrices which represent three of the infinitesimal generators of O_4.

III. IDENTIFICATION OF R

To identify $\mathbf{R}$ within a multiplet, we calculate the commutation relations of $\mathbf{R}$ with $\mathbf{L}$ and $\mathbf{A}$. Since $\mathbf{R}$ is a vector with respect to $\mathbf{L}$, we have

$$[L_i, R_k] = i\varepsilon_{ikl}R_l.$$

Then we compute the commutators of $\mathbf{A}$ and $\mathbf{R}$ in such a way as to separate out the nonzero part on a multiplet of bound states:

$$[A_i, R_j] = \frac{3}{2}\frac{i}{me^2}\varepsilon_{ijk}L_k + \frac{1}{2e^2}[H, \delta_{ij}R^2 - R_iR_j].$$

Hence we have the following lemma.

Lemma 1:

$$[A_i, R_j] = \frac{3}{2}\frac{i}{me^2}\varepsilon_{ijk}L_k$$

on a multiplet of states of negative energy $E_n = -me^4/2n^2$.

We now introduce the vector $\mathbf{\Omega}$,

$$\mathbf{\Omega} = \mathbf{R} - \tfrac{3}{2}(e^2/2\,|E_n|)\mathbf{A},$$

which has the properties,

$$[L_i,\ \Omega_k] = i\varepsilon_{ikl}\Omega_l, \quad [K_i,\ \Omega_j] = 0, \quad i, j = 1, 2, 3.$$

We are thus led to Lemma 2.

Lemma 2: Three $n^2 \times n^2$ matrices V_i, commuting or not commuting among themselves, but satisfying $[L_i, V_k] = i\varepsilon_{ijk}V_k$, $[K_i, V_j] = 0$ on the n^2-dimensional unitary representation space of O_4, are three null matrices, $V_i = 0$.

Proof: Immediately $[F_i, V_k] = \frac{1}{2}i\varepsilon_{ikl}V_l$, $[G_i, V_k] = \frac{1}{2}i\varepsilon_{ikl}V_l$, where $\mathbf{F}$ and $\mathbf{G}$ were introduced in Sec. II.

Thus $\mathbf{V}$ is the *same* vector with respect to two commuting rotation groups, and as such must be zero. An explicit proof is trivial, e.g., using Jacobi identity

$$[G_2, [F_1, V_2]] = [F_1, [G_2, V_2]] = 0.$$

On the other hand,

$$4[G_2, [F_1, V_2]] = -V_1,$$

thus showing that $\mathbf{V}$ is zero. Note that this proof is valid for any representation of $SO_3 \times SO_3$.

We can now state the theorem.

Theorem 1: On the multiplet of the n^2 states of negative energy $E_n = (-me^4/2n^2)$ of the hydrogen atom, the position operator $\mathbf{R}$ and the Lenz vector $\mathbf{A}$ are to be identified with each other as follows:

$$\mathbf{R} = \frac{3}{2}\frac{e^2}{2\,|E_n|}\mathbf{A} = \tfrac{3}{2}na_0\mathbf{K},$$ where a_0 is the Bohr radius.

Remark: Between two different multiplets, $\mathbf{R}$ has much more complicated properties and has to be identified with other operators no longer in $SO(4)$ but in $SO(4, 1)$, which is a dynamical group for the whole set of H-atom bound states.

Let us state some of the consequences of this property.

We are going to use the preceding theorem to give a group interpretation of some properties of the H atom. We remark that, with respect to O_4, it is natural to consider two bases in the n^2-dimensional representation space.

Basis I: This is the basis $|fg\rangle$. F_z, and G_z are diagonal,

$$F_z\,|fg\rangle = f\,|fg\rangle,$$
$$G_z\,|fg\rangle = g\,|fg\rangle.$$

Basis II: L^2 and L_z diagonal ($\mathbf{L} = \mathbf{F} + \mathbf{G}$). The states will be labeled $|lm\rangle$, they are essentially the usual angular momentum eigenstates.

To go from one basis to the other is a standard problem in the recoupling of angular momenta. We adopt the following phase normalization[3]:

$$|lm\rangle = \sum_{f,g} (2l+1)^{\frac{1}{2}}(-1)^m \begin{pmatrix} \lambda & \lambda & l \\ f & g & -m \end{pmatrix} |fg\rangle,$$

$$|fg\rangle = \sum_{lm} (2l+1)^{\frac{1}{2}}(-1)^m \begin{pmatrix} \lambda & \lambda & l \\ f & g & -m \end{pmatrix} |lm\rangle.$$

We already can understand Basis I: That F_z and G_z are diagonal means K_z and L_z, i.e., R_z and L_z are diagonal. Thus, Basis I essentially corresponds to the so called parabolic coordinates (or "Stark" coordinates) up to phase normalizations. It is also clear that the coefficients connecting angular momentum and Stark eigenstates (Appendix of Ref. 1) are essentially Clebsch–Gordan coefficients. We do not enter any more detailed analysis of these facts.

Now, we verify that Basis II exactly corresponds to the usual angular momentum eigenstates $\psi_{nlm}(\mathbf{r}) = Y_l^m(\hat{r})R_{nl}(r)$, and that the unitary scalar product is the usual $\int d_r^3\,\varphi^*_{(r)}\psi_{(r)}$. Because of the properties of the spherical harmonics $Y_l^m(\hat{r})$, we already know this to be true with respect to ordinary $\mathbf{L}$ rotations. We just have to check that the action of $\mathbf{K}$ on $|lm\rangle$ is identical, up to the overall constant factor $\frac{3}{2}na_0$, with the action of $\mathbf{R}$ on $\psi_{nlm}(\hat{r})$, which we can compute directly. The following formula tells us how K_α acts on $|lm\rangle$.

$$\begin{aligned} K_\alpha\,|lm\rangle = {} & 2(-1)^{2\lambda-m}\begin{pmatrix} l+1 & 1 & l \\ -m' & \alpha & m \end{pmatrix}[(2l+1)(2l+3)]^{\frac{1}{2}} \\ & \times [\lambda(\lambda+1)(2\lambda+1)]^{\frac{1}{2}}\begin{Bmatrix} l & 1 & l+1 \\ \lambda & \lambda & \lambda \end{Bmatrix}|l+1\ m'\rangle \\ & + 2(-1)^{2\lambda-m}\begin{pmatrix} l-1 & 1 & l \\ -m' & \alpha & m \end{pmatrix}[(2l+1)(2l-1)]^{\frac{1}{2}} \\ & \times [\lambda(\lambda+1)(2\lambda+1)]^{\frac{1}{2}}\begin{Bmatrix} l & 1 & l-1 \\ \lambda & \lambda & \lambda \end{Bmatrix}|l-1\ m'\rangle. \end{aligned}$$

We compute the Racah coefficient by using Racah's formula[4]

$$\begin{Bmatrix} l & 1 & l-1 \\ \lambda & \lambda & \lambda \end{Bmatrix} = (-1)^{l+2\lambda}\left\{\frac{l[(2\lambda+1)^2 - l^2]}{4\lambda[2\lambda+1][\lambda+1][4l^2-1]}\right\}^{\frac{1}{2}}.$$

[3] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957), Chap. 3.

[4] Reference 3, p. 99.

We then derive

$$K_0 |lm\rangle = \left[\frac{(l+1)^2 - m^2}{4(l+1)^2 - 1}\right]^{\frac{1}{2}} \times [(2\lambda+1)^2 - (l+1)^2]^{\frac{1}{2}} |l+1, m\rangle + \left[\frac{l^2 - m^2}{4l^2 - 1}\right]^{\frac{1}{2}} [(2\lambda+1)^2 - l^2]^{\frac{1}{2}} |l-1, m\rangle.$$

Now we compute $z\psi_{lnm}$,[5]

$$z\psi_{lnm} = \tfrac{3}{2}na_0\left\{\left[\frac{(l+1)^2 - m^2}{4(l+1)^2 - m^2}\right]^{\frac{1}{2}} \times [n^2 - (l+1)^2]^{\frac{1}{2}}\psi_{n,l+1,m} + \left[\frac{l^2 - m^2}{4l^2 - 1}\right]^{\frac{1}{2}} [n^2 - l^2]^{\frac{1}{2}}\psi_{n,l-1,m}\right\}.$$

Comparing these last two formulas gives us the announced result.

IV. THE GROUP-THEORETIC SOLUTION OF THE STARK MODEL

Theorem 1 enables us to rewrite the system (A) mentioned in Sec. I as

$$\dot{a}_\alpha(t) = \sum_{\beta=1}^{n^2} \{F(t)\langle n\alpha| K_z |n\beta\rangle + G(t)\langle n\alpha| L_y |n\beta\rangle\} a_\beta(t), \qquad (A')$$

where

$$F(t) = \tfrac{3}{2}n\Phi(t).$$

We now write

$$a_\alpha(t) = \sum_{\beta=1}^{n^2} U_{\alpha\beta}(t)a_\beta(0).$$

Therefore, the matrix $U(t)$ is the solution of the equation

$$\dot{U}_{\alpha\gamma}(t) = \sum_{\beta=1}^{n^2} \{F(t)\langle n\alpha| K_z |n\beta\rangle + G(t)\langle n\alpha| L_y |n\beta\rangle\} U_{\beta\gamma}(t), \qquad (A'')$$

and $U(0) = 1$.

Now, we notice that together K_z, L_y, K_z form a Lie subalgebra of O_4 which is isomorphic to the standard angular momentum algebra,

$$[K_z, L_y] = iK_z, \quad [L_y, K_z] = iK_x, \quad [K_z, K_x] = iL_y.$$

The set of states $|n\alpha\rangle$ also forms a reducible representation space for this Lie subalgebra. Thus, we can give a more general formulation of the problem. We consider a linear vector space B which is a representation space for the angular momentum Lie algebra. We assume that all the topological operations we perform are valid, which is certainly the case for finite-dimensional spaces. In particular, we assume that the exponentials of the operators J_i, $i = 1, 2, 3$, which represent the Lie algebra, can be defined; and so can the products of such an exponential with the others or with the J's. Besides, differentiating an operator with respect to t is allowed. On B, the system (A'') generalizes to the equation (B) below.

With these provisos, we purport to reduce the study of Eq. (B) to the solution of a much simpler problem. We now write the system (B):

$$\dot{U}(t) = \left\{\sum_{i=1}^{3} \alpha_i(t)J_i\right\} U(t), \qquad U(0) = 1. \qquad (B)$$

$U(t)$ is an operator acting on the vector space B; $\alpha_i(t)$ are three arbitrary numerical functions with "good" properties, and the precise form of the boundary condition is not important.

Following Ref. 6, we parametrize the solution $U(t)$ with three arbitrary functions

$$U(t) = \exp[f(t)J_+] \exp[g(t)J_-] \exp[h(t)J_3].$$

It is a simpler problem to find the three functions $f(t)$, $g(t)$, and $h(t)$.

We use the following definition for $J_\pm$:

$$J_\pm = (2)^{-\frac{1}{2}}(J_1 \pm iJ_2),$$

thereby redefining the commutation rules

$$[J_3, J_3] = -\varepsilon J_\varepsilon, \qquad [J_\varepsilon, J_{-\varepsilon}] = \varepsilon J_3 \quad (\varepsilon = \pm 1).$$

Now we compute $\dot{U}U^{-1}$,

$$\dot{U}U^{-1} = \dot{f}J_+ + \dot{g} \exp[fJ_+]J_- \exp[-fJ_+] + \dot{h} \exp[fJ_+] \exp[gJ_-]J_3 \exp[-gJ_-] \exp[-fJ_+].$$

Assuming that the Baker–Hausdorff formula

$$\exp[\lambda A]B \exp[-\lambda A] = B + \lambda[A, B] + \frac{\lambda^2}{2!}[A, [A, B]] + \cdots$$

is valid, enables us to calculate $\dot{U}U^{-1}$ as an element of the Lie algebra spanned by J_i, $i = 1, 2, 3$. Moreover, since the J's are linearly independent (provided $J_i \neq 0$), we can write the following system of equations for f, g, and h.

$$\dot{f} - \dot{g}\tfrac{1}{2}(f^2) - \dot{h}[f + \tfrac{1}{2}(f^2 g)] = \alpha_+,$$
$$\dot{g} + \dot{h}g = \alpha_-,$$
$$\dot{g}f + \dot{h}(1 + gf) = \alpha_3,$$

where $\alpha_\varepsilon = (2^{-\frac{1}{2}})(\alpha_1 - i\varepsilon\alpha_2)$. (1)

[5] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press Inc., New York, 1957), p. 253, Eqs. (60)–(67); p. 263, Eqs. (63)–(65).

[6] J. Wei and E. Norman, J. Math. Phys. 4, 575 (1963).

Then the derivatives of the unknowns are

$$\begin{aligned} \dot{h} &= \alpha_3 - \alpha_- f, \\ \dot{g} &= -\alpha_3 g + \alpha_-(1 + fg), \\ \dot{f} &= \alpha_3 f - \alpha_-(\tfrac{1}{2}f^2) + \alpha_+, \end{aligned} \tag{2}$$

where we decided to compute h as an integral over $\alpha_3 - \alpha_- f$. Next, let $A = g$, $B = 1 + fg$, and $C = f + \frac{1}{2}f^2 g$. Clearly A, B, and C are not independent; indeed, they verify the relation $B^2 - 2AC = 1$. However, the great virtue of this change of unknowns is that (2) can be linearized in a most interesting way. Thus we get

$$\begin{aligned} \dot{A} &= -\alpha_3 A + \alpha_- B, \\ \dot{B} &= \alpha_- C + \alpha_+ A, \\ \dot{C} &= \alpha_3 C + \alpha_+ B. \end{aligned} \tag{3}$$

We can rewrite (3) in matrix form.

$$\begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \end{pmatrix} = [\alpha_3 M_3 + \alpha_- M_- + \alpha_+ M_+] \begin{pmatrix} A \\ B \\ C \end{pmatrix},$$

where

$$M_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_- = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

and

$$M_+ = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

In the M matrices, we recognize the standard three-dimensional irreducible representation of the angular momentum Lie algebra. We thus have the result that, in order to solve Eq. (B) in the space B, it is enough to solve the same system on the *three-dimensional* space of the standard three-dimensional irreducible representation.

Going one step further, we easily check that $w(A, B, C) \equiv B^2 - 2AC$ is constant in time ($\dot{w} = 0$), thereby confirming the fact that ABC are *not* independent. This leads us to the question of whether there exists a change of functions $u(A, B, C)$, $v(A, B, C)$, [and $w(A, B, C)$] such that u and v satisfy Eq. (B) on the two-dimensional spin representation of the angular momentum algebra. We prove that the answer is *no* because the constant $w(ABC)$ has the value 1. u and v would verify the conditions (up to an irrelevant constant equivalence)

$$\begin{aligned} \dot{u} &= \alpha_3(\tfrac{1}{2}u) + \alpha_+(2)^{-\frac{1}{2}}v, \\ \dot{v} &= -\alpha_3\tfrac{1}{2}v + \alpha_-(2)^{-\frac{1}{2}}u. \end{aligned} \tag{4}$$

Now, the derivative with respect to t of any function $\varphi(A, B, C)$ can be written as

$$\dot{\varphi} = [\alpha_3\, \partial_3 + \alpha_-\, \partial_- + \alpha_+\, \partial_+]\varphi,$$

where

$$\partial_3 = -A\, \partial_A + C\, \partial_c, \qquad \partial_- = B\, \partial_A + C\, \partial_B,$$
$$\partial_+ = A\, \partial_B + B\, \partial_c.$$

(These three operators verify the angular momentum rules up to obvious factors -1.)

Since we want system (4) to be satisfied independently of the precise forms of the functions α, we must have

$$\begin{aligned} &\partial_3 u = \tfrac{1}{2}u, \quad \partial_- u = 0, \quad \partial_+ u = (2^{-\frac{1}{2}})v, \\ &\partial_3 v = -\tfrac{1}{2}v, \quad \partial_- v = (2^{-\frac{1}{2}})u, \quad \partial_+ v = 0. \end{aligned}$$

From these equations, one easily derives the *algebraic* relation

$$uvw = 0. \tag{5}$$

Now, since system (3) is linearly homogeneous, it has two types of boundary conditions,

$$w \equiv B^2 - 2AC = 1$$

and

$$B^2 - 2AC = 0.$$

In the first type, we deduce $uv = 0$, which indicates that there does not exist any appropriate change of functions; this, unfortunately, is our case. In the second type, condition (5) is automatically satisfied and the proper changes of unknowns are readily found to be

$$A = \tfrac{1}{2}u^2, \qquad B = (2)^{-\frac{1}{2}}uv, \qquad C = \tfrac{1}{2}v^2.$$

We are going to present *another way* of solving the same problem. Here, we use the parametrization and the notations used in Ref. 6.

$$U(t) = \exp[h(t)H] \exp[g(t)E] \exp[f(t)F],$$

where

$$E = J_y - iJ_x, \qquad F = -J_y - iJ_x,$$
$$H = -2J_z.$$

The only significant difference between our parametrization and that of Ref. 6 is the fact that, in Ref. 6, $\exp[hH]$ appears on the left instead of on the right of the other two factors. We carry

out a similar calculation, and reduce the system (B) eventually to the following nonlinear system:

$$\begin{aligned} \dot{f} &= ce^{-2h}, \\ \dot{g} &= be^{2h} - ce^{-2h}g^2, \\ \dot{h} &= a - cge^{-2h}, \end{aligned} \tag{6}$$

where $a=-\frac{1}{2}\alpha_z$, $b=\frac{1}{2}(i\alpha_x+\alpha_y)$, and $c=\frac{1}{2}(i\alpha_x-\alpha_y)$. Now, the authors of Ref. 6 reduce system (6) to a Ricatti equation. Here we depart from their approach. We introduce the following successive changes of functions: First, let

$$\begin{aligned} h = \tfrac{1}{2}\,\mathrm{Log}\, y \quad \text{to get} \quad y\dot{f} &= c, \\ y\dot{g} &= by^2 - cg^2, \\ \dot{y} &= 2ay - 2cg. \end{aligned} \tag{7}$$

We compute f in terms of the other functions, $f = \int (c/y)\, dt$, and concentrate on the last two equations of (7). Setting

$$g = uw, \qquad y = v^2,$$

then (7) has the form

$$\dot{u} = -au + bv, \qquad \dot{v} = av - cu.$$

Finally, we set $X = u + v$, $Y = u - v$, and $Z = \binom{X}{Y}$, a two-dimensional column vector. We then have the formula

$$\dot{Z} = [\alpha_x(t)\tfrac{1}{2}\sigma_x + \alpha_y(t)\tfrac{1}{2}\sigma_y + \alpha_z(t)\tfrac{1}{2}\sigma_z]Z,$$

where the σ_i's are the standard 2×2 Pauli matrices. Hence, solving the system (B) on the space B is reduced to solving a homomorphic system on the *two-dimensional* representation space of the standard representations of the angular momentum Lie algebra. The price we pay here for reducing the number of dimensions to that of the lowest faithful representation is in the logarithm needed to compute $h = \frac{1}{2}\,\mathrm{Log}\, y = \mathrm{Log}\, v$.

From a purely mathematical standpoint, one can guess that interesting problems are lurking behind these simple calculations. These questions will be discussed in a separate paper to be published later.

Before closing, let us comment briefly on the application of the solution obtained. The physical problem is to investigate the effect of the Stark mixing on the relative rates of annihilation in the orbital S and P states of the initial system (proton–antiproton, etc.). Thus, we have to work in the angular momentum basis $|nlm\rangle$, and we face a recoupling problem in an irreducible representation of O_4. This is a matter of standard Clebsch–Gordan and Racah coefficients.

In view of the nature of this paper, we defer numerical calculations to another less abstracted paper, where we also study the effect of absorption, and try to assess the value and the limitations of the model.

ACKNOWLEDGMENTS

We thank Professor Ph. Meyer and Professor C. Bouchiat for reading the manuscript.

APPENDIX. IDENTIFICATION OF R WITHIN A MULTIPLET OF SCATTERING STATES OF THE HYDROGEN ATOM

We present here another instance of the identification of an observable in terms of the elements of the underlying dynamical group. We identify $\mathbf{R}$ with $-\frac{2}{3}(2mE)^{\frac{1}{2}}\mathbf{K}$ (note the minus sign) on the multiplet of *standard regular* scattering functions of the hydrogen atom at energy E (we call these states the multiplet E). Thus, this Appendix is an extension of Sec. III.

On the multiplet E, the commutation rules of the components of the Lenz vector become

$$[A_i, A_j] = -\frac{2i}{me^4}EL_k.$$

We are thus led to set

$$\mathbf{A} = \left(\frac{2E}{me^4}\right)^{\frac{1}{2}}\mathbf{K}$$

so that

$$[K_i, K_j] = -iL_k.$$

Thus $\mathbf{L}$ and $\mathbf{K}$ obey the commutation rules of the Lie algebra of the proper homogeneous Lorentz group $\mathfrak{L}_+^\uparrow$.

A. Identification of the Multiplet E with an Irreducible Unitary Representation of $\mathfrak{L}_+^\uparrow$

This is a very straightforward task, since a detailed account of the structure of the Lorentz group is presented in Naimark's book.[7] We use his notations:

$$\begin{vmatrix} H_\pm = L_1 \pm iL_2 \\ H_3 = L_3 \end{vmatrix} \qquad \begin{vmatrix} F_\pm = K_1 \pm iK_2 \\ F_3 = K_3. \end{vmatrix}$$

An irreducible unitary representation R of $\mathfrak{L}_+^\uparrow$ is infinite-dimensional. It is labeled by two numbers [c (pure imaginary or real $0 \le c \le 1$) and k_0 (positive integer or half integer)]. k_0 is the lowest weight of the representation of the three-dimensional group

[7] M. A. Naimark, *Les représentations linéaires du groupe de Lorentz* (Dunod Cie., Paris, 1962).

contained in R. Since S-wave Coulomb scattering exists, we already know that k_0 must be zero. We now use the Casimir operators to identify c. Consider first

$$\Delta' \equiv H_+F_- + H_-F_+ + F_+H_- + F_-H_+ + 4H_3F_3 = 4\mathbf{K}\mathbf{L}$$

on R;

$$\Delta' = -4ik_0c$$

(Ref. 7, p. 138). We know that $\mathbf{K}\cdot\mathbf{L} = 0$, therefore

$$k_0c = 0.$$

Consider next

$$\Delta \equiv F_+F_- + F_-F_+ + 2F_3^2 - (H_+H_- + H_-H_+ + 2H_3^2) = 2(K^2 - L^2),$$

on R,

$$\Delta = -2(k_0^2 + c^2 - 1).$$

Since

$$1-(2E/me^4)K^2 = (-2E/me^4)(L^2 + 1),$$

we conclude that

$$k_0^2 + c^2 = -me^4/2E.$$

Thus, we consider the multiplet E as an irreducible unitary representation of $\mathcal{L}_+^1$ belonging to the *principal series* and is *labeled* by

$$k_0 = 0, \qquad c^2 = -me^4/2E.$$

Now we finish the proof of the above statement while analyzing this representation R in more concrete term. We show that the standard *regular* wavefunctions given on p. 22 of Ref. 5 can be identified with a standard basis f_m^l of the representation space of R (in the sense of Theorem 2 on p. 98 of Ref. 7), up to normalization factors we compute. The mass is put equal to one for convenience. Let us call

$$\psi_m^l(E) \equiv Y_m^l(\hat{r})(-i\rho)^l \frac{a_l}{(2l+1)!} e^{-\frac{1}{2}\rho} \times F(l+1-n, 2l+2, \rho) \equiv Y_m^l(\hat{r})f_l(\rho), \qquad \text{(A1)}$$

where $Y_m^l(\hat{r})$ are the standard spherical harmonics, $\rho = 2i(2E)^{\frac{1}{2}}r$, $n = -ie^2/(2E)^{\frac{1}{2}}$, and

$$F(\alpha, \beta, z) = 1 + \frac{\alpha}{\beta}z + \frac{\alpha(\alpha+1)}{\beta(\beta+1)}\frac{z^2}{2!} + \cdots.$$

Since $\mathbf{K}$ is a vector with respect to $\mathbf{L}$, it is enough to compare $K_3\psi_m^l(E)$ with formula (54) on p. 98 of Ref. 7, which, in our notations, has the form

$$K_3\psi_m^l = i[(l+m)(l-m)]^{\frac{1}{2}}\left[\frac{l^2-n^2}{4l^2-1}\right]^{\frac{1}{2}}\psi_m^{l-1} - i[(l+m+1)(l-m+1)]^{\frac{1}{2}}\left[\frac{(l+1)^2-n^2}{4(l+1)^2-1}\right]^{\frac{1}{2}}\psi_m^{l+1}.$$

Now, we know K_3 is an explicit differential operator

$$K_3 = in\left[\frac{1}{2e^2}\left(L^2\frac{\partial}{\partial z} - \frac{\partial}{\partial z}L^2\right) + \cos\theta\right].$$

We now gather a few formulas from Appendices A-37 and A-22 of Ref. 5:

$$\frac{\partial}{\partial z}Y_m^l(\hat{r})f_l(r) = \left[\frac{(l+m+1)(l-m+1)}{4(l+1)^2-1}\right]^{\frac{1}{2}}Y_m^{l+1}\left(\frac{d}{dr} - \frac{l}{r}\right)f_l(r) + \left[\frac{(l+m)(l-m)}{4l^2-1}\right]^{\frac{1}{2}}Y_m^{l-1}\left[\frac{d}{dr} + \tfrac{1}{2}(l+1)\right]f_l(r),$$

$$\cos\theta Y_m^l(\hat{r})f_l(r) = \left[\frac{(l+m+1)(l-m+1)}{4(l+1)^2-1}\right]^{\frac{1}{2}}Y_m^{l+1}f_l(r) + \left[\frac{(l+m)(l-m)}{4l^2-1}\right]^{\frac{1}{2}}Y_m^{l-1}f_l(r).$$

We can now write

$$K_3\psi_m^l(E) = i\left[\frac{(l+m)(l-m)}{4l^2-1}\right]^{\frac{1}{2}}Y_m^{l-1}(\hat{r}) \times \left[-2l\left(\frac{d}{d\rho} + \frac{l+1}{\rho}\right) + n\right]f_l(\rho) + i\left[\frac{(l+m+1)(l-m+1)}{4(l+1)^2-1}\right]^{\frac{1}{2}}Y_m^{l+1}(\hat{r}) \times \left[2(l+1)\left(\frac{d}{d\rho} - \frac{l}{\rho}\right) + n\right]f_l(\rho).$$

The problem is thus reduced to finding the normalization a_l in (A1) in such a way that the set of $f(\rho)$'s fulfill the equations

$$(l^2-n^2)^{\frac{1}{2}}f_{l-1}(\rho) = \left\{-2l\left[\frac{d}{d\rho} + \frac{l+1}{\rho}\right] + n\right\}f_l(\rho)$$

and

$$[(l+1)^2-n^2]^{\frac{1}{2}}f_{l+1}(\rho) = \left\{-2(l+1)\left[\frac{d}{d\rho} - \frac{l}{\rho}\right] - n\right\}f_l(\rho).$$

We spare the reader the term-by-term comparison of the series expansions, just stating the result that

$$a_l = \xi(-i)^l\{(-n^2)(1-n^2)\cdots(l^2-n^2)\}^{\frac{1}{2}},$$

where ξ is some arbitrary number. The standard basis in Naimark's sense can be written as

$$\psi_m^l(E) = Y_m^l(\hat{r})(-\rho)^l \frac{[-n^2(1-n^2)\cdots(l^2-n^2)]^{\frac{1}{2}}}{(2l+1)!} \times e^{-\frac{1}{2}\rho}F(l+1-n, 2l+2, \rho).$$

B. Identification of R on the Multiplet E

Lemma 1 in Sec. II can be used once more to yield, between states of energy E,

$$\left[K_i, \left(\frac{2E}{me^4}\right)^{\frac{1}{2}}\tfrac{2}{3}me^2R_i\right] = i\varepsilon_{ijk}L_k.$$

Since

$$[K_i, K_j] = -i\varepsilon_{ijk}L_k,$$

it is clear that the vector $\mathbf{\Omega} = \mathbf{K} + \frac{2}{3}(2mE)^{\frac{1}{2}}\mathbf{R}$ satisfies the properties that (1) it is a vector with respect to $\mathbf{L}$ and (2) it commutes with $\mathbf{K}$,

$$[L_i, \Omega_j] = i\varepsilon_{ijk}\Omega_k,$$

$$[K_i, \Omega_j] = 0, \quad i, j = 1, 2, 3.$$

Theorem 1A: Such a vector operator is represented by the null operator on an irreducible representation of $\mathfrak{L}_+^\dagger$ such that $k_0 = 0$.

Proof: An irreducible representation R of $\mathfrak{L}_+^\dagger$ is, of course, a reducible representation for its subgroup $SO(3)$ with infinitesimal generators $\mathbf{L}$. As such, R is built up as a direct sum of irreducible representation spaces $\mathfrak{M}_k$ for $SO(3)$, $(2k+1)$-dimensional. If $k_0 = 0$ (i.e., if $\mathbf{L}\cdot\mathbf{K} = 0$), then R contains all $\mathfrak{M}_k$'s for $k = 0, 1 \cdots \kappa$ (each $\mathfrak{M}_k$ once); and if R is unitary, then $\kappa = \infty$.

We call f_ν^k, $\nu = -k, -k+1, \cdots, +k$ a standard basis for $\mathfrak{M}_k$. Then Naimark proves that[8] (our $\mathbf{K}$ coincides with his $\mathbf{F}$)

$$K_3 f_\nu^k = (k^2-\nu^2)^{\frac{1}{2}}C_k f_\nu^{k-1} - [(k+1)^2-\nu^2]^{\frac{1}{2}}C_{k+1}f_\nu^{k+1}$$

with

$$C_k = i\left(\frac{k^2-c^2}{4k^2-1}\right)^{\frac{1}{2}}$$

($c^2 \leq 0$, or $-1 \leq c \leq +1$ for R unitary).

Morever, a careful scrutiny of the derivation of the above formula reveals that any vector operator $\mathbf{V}$ (with respect to $\mathbf{L}$), *independently* of its commutation relations with $\mathbf{K}$ or with itself, verifies the formula[9]

$$V_3 f_\nu^k = (k^2-\nu^2)^{\frac{1}{2}}C_k(V)f_\nu^{k-1} - \nu A_k(V)f_\nu^k - [(k+1)^2-\nu^2]^{\frac{1}{2}}D_{k+1}(V)f_\nu^{k+1}$$

and analogous formulas for V_+, V_-.

Now we want to examine the consequences of $[V_i, K_j] = 0$ for all $i, j = 1, 2, 3$. Since $\mathbf{V}$ and $\mathbf{K}$ are both vectors, it is enough to consider $[V_3, K_3] = 0$ and $[V_3, K_+] = 0$.

Let us apply $[V_3, K_3] = 0$ on f_ν^k. A straightforward calculation shows that

$$C_k(V)C_{k-1} = C_kC_{k-1}(V), \qquad A_k(V)C_k = C_kA_{k-1}(V),$$

$$D_{k+1}(V)C_{k+2} = C_{k+1}D_{k+2}(V),$$

$$A_k(V)C_{k+1} = C_{k+1}A_{k+1}(V),$$

$$(k^2-\nu^2)C_k[C_k(V) - D_k(V)] = [(k+1)^2-\nu^2][C_{k+1}(V) - D_{k+1}(V)].$$

From the above relations, one easily derives

$$A_k(V) = a, \qquad C_k(V) = bC_k = D_k(V).$$

Thus, we now write

$$V_3 f_\nu^k = -a\nu f_\nu^k + b\{(k^2-\nu^2)^{\frac{1}{2}}C_k f_\nu^{k-1} - [(k+1)^2-\nu^2]^{\frac{1}{2}}C_{k+1}f_\nu^{k+1}\}.$$

Then we recognize $V_3 = -aL_3 + bK_3$.

It is now clear that

$$[V_3, K_+] = 0$$

implies $a = 0 = b$, i.e., $\mathbf{V} = 0$, as stated.

Remark: It should be obvious that the proof we have just given, also goes through for an irreducible representation of $SO(4)$. Indeed, the only thing to be changed is the value of C_k which becomes

$$C_k = (\pm)[(n^2-k^2)/(4k^2-1)]^{\frac{1}{2}}.$$

This does not affect the proof.

We can also remark the interesting fact that one goes from an irreducible representation of $SO(4)$ to one of $\mathfrak{L}_+^\dagger$ by a proper analytic continuation of a Racah coefficient [$n^2 \equiv (2\lambda+1)^2$ becomes c^2].

Theorem 2A: On the multiplet E, $E > 0$, $\mathbf{R} = -3/(8mE)^{\frac{1}{2}}\mathbf{K}$.

[8] Reference 7, p. 98, Eq. (55), and p. 100, Theorem 3.

[9] Reference 7, p. 94, Eq. (36).

On the Connection between External and Internal Symmetries of Strongly Interacting Particles

MOSHÉ FLATO AND DANIEL STERNHEIMER
Institut Henri Poincaré, Paris
(Received 11 October 1965)

The relation between the mass and spin-parity J^P, and internal quantum numbers of elementary particles, hints at a nontrivial connection between the external symmetry, namely the Poincaré group or its Lie algebra, and a so-called internal symmetry. Studying this connection mathematically, we find that any extension of (resp. by) the Poincaré Lie algebra $\mathfrak{P}$ by (resp. of) a semisimple Lie algebra $\mathfrak{X}$ is equivalent to the trivial one, $\mathfrak{P} \oplus \mathfrak{X}$. Moreover, if we are looking for a Lie algebra containing $\mathfrak{P}$ and $\mathfrak{X}$ in an economical and nontrivial way, namely what we call a (nontrivial) unification of $\mathfrak{P}$ and $\mathfrak{X}$, we find restrictions on the possibilities of choice of $\mathfrak{X}$, which exclude compact internal symmetries. An explicit treatment of $SL(3, \mathbb{C})$ as internal symmetry, related to the external symmetry by the unification process of Lie algebras, gives a mass formula which is in very good accordance with the experimental data, and can be theoretically interpreted by means of a so-called "classification principle".

1. INTRODUCTION

VARIOUS authors have recently investigated the possible connections between external and internal symmetries, or the possibility of interpreting the isotopic-spin space as reflecting "real internal motions" in Minkowski space. From external formalism treatments, one knows that external motions of elementary particles are invariant under the universal covering $\boldsymbol{P}^{\#}$ of the Poincaré (inhomogeneous Lorentz) group $\boldsymbol{P}$ (because of half-integer spin values, we consider $\boldsymbol{P}^{\#}$). On the other hand, internal quantum numbers (such as I, I_3, B, S) have been phenomenologically introduced, which are conserved by strong interactions, and we know from Noether's theorem that (nondiscrete) conservation laws are related to (continuous) symmetry groups. Naturally, the question arises: what is the possible connection between isospace, or, more generally, "internal space" and Minkowski space? Are $\boldsymbol{P}^{\#}$ and the so-called internal symmetry group independant, or does there exist a certain connection between the two? Physically speaking, the question is that of finding relations between external spin J (or possibly spin-parity J^P) and the squared physical mass of a particle $-g^{\mu\nu}p_\mu p_\nu$ on one hand, and internal quantum numbers on the other hand. Mathematically, the problem can be formulated as follows: given the Poincaré group $\boldsymbol{P}$ (or given $\boldsymbol{P}^{\#}$) and an internal symmetry group $\boldsymbol{X}$, which, in general, is taken (for simplicity) as semisimple, can we find a group which contains both and connects them in a nontrivial way (in a sense to be explained later on), and which is physically interesting with respect to the above problem?

One can try and treat this problem by using, as Michel[1] did, the theory of group extensions. Michel studied the central extensions of $\boldsymbol{P}$ and $\boldsymbol{P}^{\#}$ by an Abelian gauge, or by an internal symmetry group, from the algebraic point of view. The only nonnegative results he obtains are phase relations of the kind $(-1)^{B+L} = (-1)^{2J}$ (obtained by studying the central extensions of $\boldsymbol{P}$ by an Abelian gauge).

As we are looking for a connection of nondiscrete type between internal and external formalisms (in view, for instance, of a mass formula), and as $-g^{\mu\nu}p_\mu p_\nu$ belongs to the (center of the) enveloping algebra of the Lie algebra $\mathfrak{P}$ of the Poincaré group $\boldsymbol{P}$, and also for other reasons, the details of which we do not discuss now, it will be worthwhile working, at least as a first step, with Lie algebras instead of Lie groups—but of course with Lie algebras on the real field.

Assuming that the internal quantum numbers are related to a semisimple internal Lie algebra $\mathfrak{X}$, several authors[2] obtained, under somewhat restrictive assumptions, negative results concerning the connection between $\mathfrak{P}$ and $\mathfrak{X}$. We also see in Sec. 2 that any extension of $\mathfrak{P}$ by (semisimple) $\mathfrak{X}$, or of $\mathfrak{X}$ by $\mathfrak{P}$, gives the direct sum $\mathfrak{P} \oplus \mathfrak{X}$. According to these results, it would seem that "internal" and external motions of elementary particles are independent. But what we see later in Sec. 3 (concerning the unification of $\mathfrak{P}$ and an internal Lie algebra $\mathfrak{X}$) and the following remarks hint very strongly at the nonindependance of these motions.

[1] F. Lurcat and L. Michel, Nuovo Cimento 21, 574 (1961). L. Michel, *Extensions centrales du groupe de Poincaré*, preprint, École Polytechnique (1964); Nucl. Phys. 57, 356 (1964).
[2] W. D. McGlinn, Phys. Rev. Letters 12, 467 (1964); see also E. C. G. Sudarshan, J. Math. Phys. 6, 1329 (1965), and references therein.

If we consider elementary particles in the $SU(3)$ classification (for instance), we see that to every supermultiplet (set of particles classified in the same representation) a certain spin-parity J^P is assigned. Physically, this may be justified by the fact that it is in accordance with experience (or by semi-empirical treatments). But there exists no mathematical reason of this fact, since, as we see later, each unification of $\mathcal{P}$ and of the Lie algebra $\mathfrak{su}(3)$ is trivial, $\mathcal{P} \oplus \mathfrak{su}(3)$, that is to say that any spin can be attributed to any classification-representation, and that for instance there is no group-theoretical reason why there should not be an octet of particles of spin-parity $\frac{3}{2}^+$(!).

But, from the phenomenological point of view, it seems that (for baryons) the sequence (J^P, I, I_3, Y) of quantum numbers defines at most one baryon. This shows that there must be a close link between internal and external statistics, and that, because of this, there must also exist a link between "internal" and external motions: the baryon, a fermion with respect to external statistics, is probably also a "fermion" with respect to the internal formalism. Therefore there must also be, in internal statitsics, a kind of *exclusion principle for baryons.*

The Okubo–Gell-Mann mass formula, or all similar formulas obtained in compact groups, is in fact purely phenomenological: this formalism does not explain the connection between $-g^{\mu\nu}p_\mu p_\nu$ (the squared relativistic physical mass of a particle) and the strong mass splitting. Indeed, if the link between $\mathcal{P}$ and the internal algebra $\mathfrak{X}$ is the direct sum $\mathcal{P} \oplus \mathfrak{X}$ (and it is so if the internal group X is compact, or even under somewhat less restrictive assumptions, as we see in Sec. 3), there is incompatibility between the notion of mass in external formalism and Gell–Mann-Okubo-like formulas giving what is supposed to be a strong mass splitting; for in that case, $g^{\mu\nu}p_\mu p_\nu$ commutes with every element of $\mathfrak{X}$, and therefore all the particles classified in the same representation of $\mathfrak{X}$ should have the same mass: supposedly those representations that give a "mass spectrum" give a *degenerate* spectrum. And on the other hand, the method used contains a mysterious process of "symmetry breaking", the significance of which is that it works. Actually, all that is done in the works of this type consists in considering the mass (or any function of the mass, if it seems more convenient) as a tensorial operator, and then in applying the Wigner–Eckart theorem. But if, for us, the use of this theorem is clear and well justified in problems concerning angular and magnetic momenta of particles, the consideration of an operator like the mass operator as a Wigner–Eckart operator, and moreover the precise specification of its tensorial type, is an *ad hoc* hypothesis without any theoretical justification.

We consider the (squared) mass operator as $-g^{\mu\nu}p_\mu p_\nu$. Our problem is therefore first to find a Lie algebra $\mathfrak{U}$ containing both $\mathcal{P}$ and $\mathfrak{X}$ in an economical and nontrivial way (in a sense to be explained), so that $g^{\mu\nu}p_\mu p_\nu$ will not be an invariant of $\mathfrak{U}$—and this excludes the trivial connection, $\mathcal{P} \oplus \mathfrak{X}$. In order to get an invariant, we have to add to $-g^{\mu\nu}p_\mu p_\nu$ a "correction", which, in some "good" cases, can be expressed as a nice function of internal quantum numbers (such as I, I_3, Y). This invariant is constant in every (irreducible) representation of $\mathfrak{U}$, and the constant is one of those that characterizes the representations of $\mathfrak{U}$, and therefore also related to representations of $\mathfrak{X}$ (e.g., by induced representation technique). The correction (in nice cases) gives the strong mass splitting in the supermultiplet [and the constant may then be considered as the squared mass of a (sometimes hypothetic) singlet, related to the supermultiplet]. We can also look directly for the spectrum of the operator representing $-g^{\mu\nu}p_\mu p_\nu$ in irreducible representations of $\mathfrak{U}$ (but then, the expression of the mass-splitting in terms of internal quantum numbers is less clear). Within this frame, we have therefore no mystery of broken symmetry: the concept of mass is a geometrical one, given by the connection between external and internal formalisms.

However, this interpretation of mass raises some difficulties—but mass must be somehow interpreted in some external formalism, and the most natural interpretation is the usual relativistic one. In the finite-dimensional representations of $\mathcal{P}$ (and also of the "unification" $\mathfrak{U}$), the p_μ's are nilpotent (they are even such that $p_\mu^2 = 0$ in 4-dimensional representations of $\mathcal{P}$); therefore we cannot hope to get a real mass spectrum (by taking the eigenvalues) except in some infinite-dimensional representations (skew-Hermitian Lie algebra representations on Hilbert space, for instance) of some unifications $\mathfrak{U}$ (of $\mathcal{P}$ and of some well-chosen internal Lie algebras $\mathfrak{X}$), because we need (real) eigenvalues not only for the invariant, but also for the "correction term". Nevertheless, the mass formula, which gives the strong and electromagnetic mass splitting obtained in the case of $SL(3, \mathbf{C})$, is in very good accordance with the experimental data if we give it a formal meaning without considering the justification of the passage to the eigenvalues. And in any case, if we do not classify strong particles in suitably chosen infinite-dimensional representations,

the problem arises of connecting the two facts that (1) we classify particles according to some finite-dimensional representations of the internal group, and that (2) we can hope to give an exact meaning to the mass formula only in some infinite-dimensional representations. We therefore propose the introduction of a *classification principle*: we *classify* particles by means of *finite-dimensional representations* of the strong group (considering a vector space of fields on which the representation transformations operate), and for each one we give a *formal sense* to the *mass formula* by *relations between the masses of the particles classified* in that representation. This might be related with some mathematical connection between finite-dimensional representations of the internal symmetry and infinite-dimensional representations of the unification (acting, e.g., on states).

It is then possible that the partial success (from the phenomenological point of view) of unitary symmetries is due to the fact that the results we deduce from them are a good approximation of what would be obtained from infinite-dimensional unitary representations of the "good" group (noncompact), and that they reflect a symmetry of states rather than of a Hamiltonian. This is, in a way, analogous to nonrelativistic atomic spectroscopy, where the state of orbital momentum L is invariant under $SU(2(2L+1))$, whereas the Hamiltonian is only $SO(3)$-invariant [$SO(3)$ has only a geometrical meaning]. In "elementary-particles-spectroscopy", in which states of higher and higher energy are measured, one can fit experimental results by using (phenomenologically) unitary symmetries of higher and higher dimension—whereas, according to the results to be developed in Sec. 3 concerning the unification process of Lie algebras, and to geometrical considerations developed elsewhere [cf. Refs. 3 and 4; the "internal P, C, T operations" can be defined in internal symmetries containing the Lorentz group in a way similar to that developed in Ref. 3], it is most probable that the "Hamiltonian" symmetry of strong interactions is "space–time-like".

We conclude this introduction by two remarks. The first one is to insist on the fact that the internal quantum numbers, with which we want to get a mass-splitting expression, have been phenomenologically introduced, by generalizing what was known. For instance, the isospin I has been introduced in a way analogous to ordinary spin, on the basis of Heisenberg's idea that the proton p and the neutron n are two states of the same particle (the nucleon N), their masses being very close; the hypercharge Y has been in fact defined by the Gell-Mann–Nishijima formula, which generalizes the charge formula for nucleons. And although it has been claimed that the only (nondiscrete) exactly conserved internal quantities are those already introduced (I and I_3, B, S), this does not necessarily imply that the mass is a function of those quantities alone. It may well be that, for instance, other quantities, not conserved in the nowadays-observed strong interactions (i.e., they are not "good quantum numbers"), occur in a mass formula. Maybe there exist other quantum numbers that we have not noticed yet, because they are not conserved in strong interactions (and until now we studied the quantum numbers mainly from the point of view of conservation laws), and with which the classification or the mass formula, for instance, can be written in a very simple way. It might also turn out that the mass should be taken as an independant feature characterizing the particles. (As to the mass, we take these possibilities into account implicitly when we look directly for the spectrum of the operator representing $-g^{\mu\nu}p_\mu p_\nu$).

The second remark is not less basic. Until now, we implicitly assumed that our internal Lie algebra was a finite-dimensional one—and, as a matter of fact, very little is known on infinite-dimensional Lie algebras. But let us look more closely at Yang and Mills's argument (developed by Sakurai[5]) as compared to the usual theory. In the latter, we know that, e.g., the conservation of baryonic number is equivalent to the invariance under a *phase* transformation of the kind

$$\psi_\alpha \rightarrow \exp(ib\beta)\psi_\alpha$$

of the complex field ψ_α (β is a real parameter; b the baryonic charge); the same applies to isospin and hypercharge, e.g.; from that, one gets the conservation laws by direct use of Noether's theorem, and those phase transformations can be mathematically expressed by the fact that the Lie algebra of the corresponding Lie group is finite-dimensional. On the other hand, if we examine Yang and Mills's

[3] M. Flato, G. Rideau, and J. P. Vigier, Nucl. Phys. **61**, 250 (1965); see also M. Flato, *Symétries de type Lorentzien et interactions fortes* (Gauthier-Villars, Paris, 1966), Chaps. I and III, Sec. 3.

[4] D. Bohm, M. Flato, F. Halbwachs, P. Hillion, and J. P. Vigier, Nuovo Cimento **36**, 672 (1965); M. Flato and D. Sternheimer, Compt. Rend. **260**, 3532 (1965).

[5] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954); J. J. Sakurai, Ann. Phys. (N. Y.) **11**, 1 (1960).

argument closely, the local nature of field theory implies here that we have a *gauge* transformation

$$\psi_\alpha \to \exp(ib\beta(x))\psi_\alpha$$

of the first kind, where $\beta(x)$ is a certain real function of the x_μ's (i.e., on Minkowski space). We therefore get not only the existence of a vectorial field undergoing a gauge transformation of the second kind, but also that the Lie algebra corresponding to the internal group may be infinite-dimensional. Although we agree with a great part of Sakurai's philosophy (especially on his criticism of broken symmetries), we use other mathematical tools and introduce some new physical principles (connection between internal and external symmetries, etc. . .), for his study can no longer account for the many experimental data already known. But, in view of Yang and Mills's argument, the internal finite-dimensional Lie algebra we consider have to be in a certain sense an "approximation" (more for what concerns its structure than its dimension) of the infinite-dimensional Lie algebra implied by their argument. Our work deals mainly with the choice and the study of a suitable internal Lie algebra (of finite dimension).

2. EXTENSION THEORY AND THE POINCARÉ LIE ALGEBRA

We choose our internal Lie algebra according to several criteria; we already studied some geometrical ones,[3,4] and our object in Secs. 2 and 3 is to develop "algebraic" criterions, namely the limitations on the possibilities of choice of an internal Lie algebra $\mathfrak{X}$ imposed by the necessary existence of a suitable connection between $\mathcal{P}$ and $\mathfrak{X}$. The first idea in this way is to study the extensions of $\mathcal{P}$ by $\mathfrak{X}$, or of $\mathfrak{X}$ by $\mathcal{P}$. We recall the following:

Definition[6]: Let $\mathfrak{a}$ and $\mathfrak{b}$ be two Lie algebras (L.a.) on the same commutative field K. An *extension* of $\mathfrak{b}$ by $\mathfrak{a}$ is an exact sequence:

$$0 \to \mathfrak{a} \xrightarrow{\lambda} \mathfrak{g} \xrightarrow{\mu} \mathfrak{b} \to 0, \tag{2.1}$$

where $\mathfrak{g}$ is a L.a. on K, μ a homomorphism of $\mathfrak{g}$ on $\mathfrak{b}$, λ an isomorphism of $\mathfrak{a}$ onto the kernel of μ ($\ker \mu = \operatorname{Im} \lambda \approx \mathfrak{a}$). Two extensions $\mathfrak{a} \xrightarrow{\lambda} \mathfrak{g} \xrightarrow{\mu} \mathfrak{b}$ and $\mathfrak{a} \xrightarrow{\lambda'} \mathfrak{g}' \xrightarrow{\mu'} \mathfrak{b}$ are said to be *equivalent* if there exists a (necessarily bijective) homomorphism f of $\mathfrak{g}$ into $\mathfrak{g}'$ such that $f \circ \lambda = \lambda'$ and $\mu' \circ f = \mu$. An extension is said to be *inessential* or a *semidirect product*, and we write $\mathfrak{g} = \mathfrak{b} \cdot \mathfrak{a}$ if there exists in $\mathfrak{g}$ a supplementary subspace of $\ker \mu$ which is a subalgebra; if there is such a subspace which is an ideal, the extension is called *trivial*; it is called *central* if $\ker \mu$ is contained in the center of $\mathfrak{g}$.

A semidirect product can also be defined[6] by the existence of an isomorphism ν of $\mathfrak{b}$ into $\mathfrak{g}$ such that $\mu \circ \nu$ is the identity on $\mathfrak{b}$; it is equivalent to a semidirect product constructed in a canonical way with an homomorphism φ of $\mathfrak{b}$ into der ($\mathfrak{a}$), the derivation algebra of $\mathfrak{a}$ (the L.a. of the K-linear operations D of $\mathfrak{a}$ such that

$$D[a_1, a_2] = [Da_1, a_2] + [a_1, Da_2] \tag{2.2}$$

$\forall a_1, a_2 \in \mathfrak{a}$), by defining the commutators on the product vector space $\mathfrak{a} \times \mathfrak{b}$ as

$$[(a, b), (a', b')] = ([a, a'] + \varphi(b)\cdot a' - \varphi(b')\cdot a, [b, b']). \tag{2.3}$$

Here, $\varphi(b)$ is the derivation of $\mathfrak{a}$, image of b by φ. Let us recall also the following results.

Proposition 1: Any extension by a semisimple Lie algebra $\mathfrak{a}$ is equivalent to the trivial one (and then $\mathfrak{g} \approx \mathfrak{a} \oplus \mathfrak{b}$, the direct sum). Any extension of a semisimple L.a. $\mathfrak{b}$ is inessential.

The first part is a consequence of a result of Bourbaki[6] Corollary 1 to Proposition 1, Chap. 6, No. 1, and the second is Corollary 3 to Theorem 5 (Chap. 6, No. 6), which is the *Levi–Malcev theorem*, (see Ref. 6, especially work by Chevalley), according to which any L.a. is semidirect product of a Levi (maximal semisimple) subalgebra by the radical (maximal solvable ideal). Moreover, it is easily seen that any extension by a complete Lie algebra (i.e., where all derivations are inner, and with {0} center) is also trivial.

Therefore any extension of $\mathcal{P}$ by $\mathfrak{X}$ semisimple (or complete) is equivalent to the trivial one ($\mathcal{P} \oplus \mathfrak{X}$). In order to study the inessential extensions of $\mathfrak{X}$ by $\mathcal{P}$ (the only one for any $\mathfrak{X}$, as can be seen), we must first study the structure of der ($\mathcal{P}$).

Structure of der ($\mathcal{P}$)

One knows that $\mathcal{P}$ is a semidirect product of the Lorentz L.a. $\mathfrak{L}$ by a four-dimensional commutative ideal $\mathcal{P}_0$ (corresponding to the translations), the defining homomorphism $\varphi : \mathfrak{L} \to \text{der } (\mathcal{P}_0) \approx \mathfrak{gl}(4, \mathbf{R})$ being the "Naïmark representation" D_N of $\mathfrak{L}$, corresponding to the realization of $\mathfrak{L}$ by the basic representation of $\mathfrak{so}(3, 1)$ in 4 dimensions (cf. Appendix). We denote by small German letters the Lie algebras, except for those most used in this paper; see Ref. 7

[6] N. Bourbaki, *Algèbres de Lie* (Hermann & Cie, Paris, 1960); C. Chevalley, *Théorie des groupes de Lie* (Hermann & Cie, Paris, 1951, 1955), Vols. II and III.

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

for the notations $\mathfrak{so}(p, q)$, $\mathfrak{su}(p, q)$, etc. Let $p \in \mathcal{P}$: it can be written $p = l + p_0$, with $l \in \mathfrak{L}$ and $p_0 \in \mathcal{P}_0$. Let $p_1 = l_1 + p_{01}$, $p_2 = l_2 + p_{02}$ be two elements of $\mathcal{P}$. Then we have

$$[p_1, p_2] = [l_1, l_2] + ([l_1, p_{02}] + [p_{01}, l_2]). \tag{2.4}$$

If $p = l + p_0$, we can write $l = L(p)$ and $p_0 = T(p)$ (L and T are linear operations). Let $D \in \text{der}\,(\mathcal{P})$, $p \in \mathcal{P}$, and write

$$Dp = dp + Ap + ap + \Delta p, \tag{2.5}$$

where $dp = L(Dl)$, $\Delta p = L(Dp_0)$, $ap = T(Dl)$, and $Ap = T(Dp_0)$ if $p = l + p_0$. As one knows (Bourbaki, Ref. 6, Chap. 5, No. 5, Proposition 6) that any derivation of a Lie algebra sends its radical into itself (more precisely, into its greatest nilpotent ideal—but, as far as $\mathcal{P}$ is concerned, these two coincide), $\Delta = 0$; one can also derive this last result directly by an elementary computation. If $p, p' \in \mathcal{P}$, one sees immediately that

$$\begin{aligned} d[p, p'] &= L(D[l, l']) = [l, dp'] + [dp, l'], \\ a[p, p'] &= T(D[l, l']) \\ &= [ap, l'] + [l, ap'] = [ap, p'] + [p, ap'], \\ A[p, p'] &= T(D[l, p_0'] + D[p_0, l']) \\ &= [l, Ap'] + [Ap, l'] + [dp, p_0'] + [p_0, dp']. \end{aligned} \tag{2.6}$$

Therefore $a \in \text{der}\,(\mathcal{P})$, and consequently $d + A \in \text{der}\,(\mathcal{P})$. The restriction to $\mathfrak{L}$ of the derivation $d + A$ of $\mathcal{P}$ is also a derivation of $\mathfrak{L}$, and it is exactly d. $\mathfrak{L}$ being semisimple, any derivation is inner (Bourbaki, Ref. 6, Chap. 6, No. 2, Corollary 3 of Proposition 1), i.e., there exists an element $\theta \in \mathfrak{L}$ such that $d = \text{ad}_{\mathfrak{L}}\,\theta$ ($\text{ad}_{\mathfrak{L}}$ standing for the adjoint mapping of $\mathfrak{L}$). But $\text{ad}\,\theta = \text{ad}_{\mathcal{P}}\,\theta$ is an (inner) derivation of $\mathcal{P}$, and it coincides with d on $\mathfrak{L}$. Therefore we can put

$$A' = d + A - \text{ad}\,\theta, \tag{2.7}$$

and $A' \in \text{der}\,(\mathcal{P})$, $A' : \mathfrak{L} \to 0$, $\mathcal{P}_0 \to \mathcal{P}_0$ (exactly as A did).

Thus we have decomposed D in a sum of 3 derivations [with $\theta(D) \in \mathfrak{L}$]:

$$D = a(D) + A'(D) + \text{ad}\,\theta(D), \tag{2.8}$$

where we wrote explicitly the dependence of each term on D. Let us now take a basis $(a_k; b_k)$ of $\mathfrak{L}$ and (p_k, p_4) of $\mathcal{P}_0 (k = 1, 2, 3)$, the commutators being the usual one (cf. Appendix). We have $[a_k, p_4] = 0$ and therefore $[a_k, A'p_4] = 0$, $\forall k$, whence $A'p_4 = \alpha p_4$ with $\alpha = \alpha(D) \in \mathbf{R}$. But we have also $[b_k, p_k] = p_4$, and $A'\mathfrak{L} = 0$; therefore $A'p_k = \alpha p_k$ with the same α. Now let π denote the derivation

$$p = l + p_0 \to p_0 \tag{2.9}$$

(the projection on the radical—always uniquely determined—$\mathcal{P}_0$); one checks easily that this is a derivation of $\mathcal{P}$. We have then $A'(D) = \alpha(D)\pi$, for some $\alpha(D) \in \mathbf{R}$.

Let us now write

$$a \cdot a_k = \Sigma_\mu \alpha_k^\mu p_\mu, \qquad a \cdot b_k = \Sigma_\mu \beta_k^\mu p_\mu \tag{2.10}$$

($\mu = 1, 2, 3, 4$; $k = 1, 2, 3$) and use the commutation relations of $\mathcal{P}$

$$[a_k, b_k] = 0, \quad [a_i, b_i] = b_k, \quad [a_i, a_i] = a_k \tag{2.11}$$

(ijk cyclic); we get straightforward $a(D) = \text{ad}\,p_0(D)$, where

$$p_0(D) = \alpha_2^3 p_1 + \alpha_3^1 p_2 + \alpha_1^2 p_3 - \beta_i^1 p_4,$$

[We could have given a more abstract proof: $f = d + a$ is a $\mathbf{R}$-linear mapping of $\mathfrak{L}$ in the vector space $\mathcal{P}$ of the representation $\sigma : l \to \text{ad}_{\mathcal{P}}\, l$ of $\mathfrak{L}$; therefore, in view of a lemma due to Whitehead (cf. Bourbaki, Ref. 6, Chap. 6, No. 2, Remark 2), and because

$$f([l, l']) = D([l, l']) = [l, f(l')] - [l', f(l)],$$

there exists $-p(D) = \theta + p_0$ in $\mathcal{P}$ such that

$$Dl = -\sigma(l)p = [p, l];$$

thus $D = \text{ad}\,p + A'$, where $A' = A - (\text{ad}\,p) \mid \mathcal{P}_0$ (the last symbol means ad p restricted to $\mathcal{P}_0$) is also a derivation of $\mathcal{P}$, which is zero on $\mathfrak{L}$ [cf. also G. P. Hochschild, Am. J. Math. **64**, 677 (1942)]; but $\sigma \mid \mathcal{P}_0$ is the (complex) irreducible representation D_N of $\mathfrak{L}$, and

$$[A', \text{ad}\,l] = \text{ad}\,(A'l) = 0;$$

therefore $A' = \alpha\pi$ (with $\alpha \in \mathbf{R}$ since A' is a $\mathbf{R}$-linear mapping of $\mathcal{P}$).]

We thus obtained

$$D = \text{ad}\,(\theta(D) + p_0(D)) + \alpha(D)\pi,$$

that is,

$$Dp = [\theta(D) + p_0(D), p] + \alpha(D)p_0 \qquad (\text{if } p = p_0 + l). \tag{2.12}$$

Moreover, it is clear that, if $D_1, D_2 \in \text{der}\,(\mathcal{P})$, then $\alpha([D_1, D_2]) = 0$, and one can check easily that:

$$\theta([D_1, D_2]) = [\theta(D_1), \theta(D_2)],$$

and because $[D, \text{ad}\,p] = \text{ad}\,(Dp) : p_0([D_1, D_2]) = D_1 p_0(D_2) - D_2 p_0(D_1)$, which is equivalent to the

fact that the mapping $D \to \text{ad } p_0(D)$ is a derivation of der $(\mathcal{P})$ [i.e., $\in$ der (der $(\mathcal{P})$) = der $(\mathcal{P})$]. It is not difficult now to see the structure of the L.a. der $(\mathcal{P})$. Let us denote by $\bar{\mathcal{P}}_0$ the (5-dimensional) solvable L.a. generated by $\mathcal{P}_0$ and a monodimensional L.a. u(1), a generator of which we denote by π', with the communtation law $[\pi', p_0] = p_0$, $p_0 \in \mathcal{P}_0$. $\bar{\mathcal{P}}_0$ is a semidirect product $(\mathrm{u}(1) \cdot \mathcal{P}_0)$, a defining homomorphism $\mathrm{u}(1) \to \text{der } (\mathcal{P}_0) = \mathfrak{gl}(4, \mathbf{R})$ of which is given by

$$\pi' \to I_4 = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{pmatrix}.$$

One checks easily that der $(\bar{\mathcal{P}}_0)$ is composed of 5×5 matrices $(h_{\alpha\beta})$ with $h_{5\beta} = 0$, the other being any real number. If $\tilde{D}_N$ denotes the representation $D_N \oplus O$ (O is the trivial one-dimensional representation) of $\mathfrak{L}$, we have

$$\text{der } (\mathcal{P}) \approx \mathfrak{L} \cdot \bar{\mathcal{P}}_0, \tag{2.13}$$

a defining homomorphism of this semidirect product being $\tilde{D}_N$. Also,

$$\text{der } (\mathcal{P}) \approx \mathrm{u}(1) \cdot \mathcal{P}, \tag{2.14}$$

a defining homomorphism here being $\mathrm{u}(1) \to \mathbf{R}\pi \subset$ der $(\mathcal{P})$, and

$$\text{der } (\mathcal{P}) \approx (\mathfrak{L} \oplus \mathrm{u}(1)) \cdot \mathcal{P}_0, \tag{2.15}$$

a defining homomorphism being $\mathfrak{L} \to D_N(\mathfrak{L})$, $\pi' \to I_4$ [π' is the generator of the u(1)]. Moreover, one gets a *five-dimensional representation of* $\mathcal{P}$ [and of der $(\mathcal{P})$] with

$$\mathfrak{L} \to (D_N \oplus O)(\mathfrak{L}), p_\mu \to (\text{ad } p_\mu) \mid \bar{\mathcal{P}}_0 = E_{\mu 5}$$

(in the canonical basis of the $\mathfrak{gl}$ algebras), and for der $(\mathcal{P})$, $\pi \to I_4 \oplus O = (\text{ad } \pi) \mid \bar{\mathcal{P}}_0$ (for instance).

Let us now pass to the following.

Study of Inessential Extensions by $\mathcal{P}$

$$0 \to \mathcal{P} \to \mathfrak{R} \to \mathfrak{X} \to 0.$$

They are the only extensions if $\mathfrak{X}$ is semisimple—and also the only one, since $\mathcal{P}$ has $\{0\}$ center and der $(\mathcal{P})$ = u(1). $\mathcal{P}$, in any case. We must therefore seek all possible homomorphisms $\varphi : \mathfrak{X} \to$ der $(\mathcal{P})$, and see what semidirect products they define. We have, $\forall x \in \mathfrak{X}$,

$$\varphi(x) = \text{ad } p(\varphi(x)) + \alpha(\varphi(x))\pi, \tag{2.16}$$

from the structure of der $(\mathcal{P})$. If $[\mathfrak{X}, \mathfrak{X}] = \mathfrak{X}$ ($\mathfrak{X}$ is equal to its derived algebra), any element of $\mathfrak{X}$ is a linear combination of commutators of elements of $\mathfrak{X}$. But φ is a homomorphism, and the α's of commutators of derivations are 0. Therefore $\varphi(\mathfrak{X}) \subset$ ad $\mathcal{P}$; as the center of $\mathcal{P}$ is $\{0\}$, $\psi = \text{ad}^{-1} \circ \varphi$ is a homomorphism $\mathfrak{X} \to \mathcal{P}$, and therefore $\varphi = 0$ (i.e., the extension is necessarily trivial) if $\mathfrak{X}$, supposed semisimple, has no direct factor isomorphic to $\mathfrak{L}$, or to $\mathfrak{so}(3)$ or $\mathfrak{sl}(2, \mathbf{R})$. But the triviality of the extensions holds in a more general context, as a consequence of the following result.

Proposition 2. Any inessential extension (2.1) defined by a homomorphism $\varphi = \text{ad} \circ \psi$, where ψ is a homomorphism $\mathfrak{b} \to \mathfrak{a}$, is equivalent to the trivial extension $\mathfrak{a} \oplus \mathfrak{b}$.

From the definition of commutators on the product-vector-space $\mathfrak{a} \times \mathfrak{b}$, we have

$$[(a, b), (a', b')] = ([a + \psi(b), a' + \psi(b')] - \psi([b, b']), [b, b']), \tag{2.17}$$

for ψ is a homomorphism. Now it is clear that the mapping $f : \mathfrak{g} \to \mathfrak{a} \oplus \mathfrak{b}$ defined by

$$\mathfrak{g} \ni (a, b) \xrightarrow{f} (a + \psi(b), b) \in \mathfrak{a} \oplus \mathfrak{b} \tag{2.18}$$

is a homomorphism (the commutators of $\mathfrak{a} \oplus \mathfrak{b}$ being the usual one of a direct sum). Further, the diagram

$$0 \to \mathfrak{a} \xrightarrow{\lambda} \mathfrak{g} \xrightarrow{\mu} \mathfrak{b} \to 0, \qquad \mathfrak{a} \xrightarrow{\lambda} \mathfrak{a} \oplus \mathfrak{b} \xrightarrow{\mu} \mathfrak{b}, \quad \mathfrak{g} \xrightarrow{f} \mathfrak{a} \oplus \mathfrak{b} \tag{2.19}$$

where $\lambda : a \to (a, 0)$ and $\mu : (a, b) \to b$, is commutative, i.e., $f \circ \lambda = \lambda$, $\mu \circ f = \mu$. Therefore[6] the two extensions are equivalent. Q.E.D.

Any element $(a, b) \in \mathfrak{g}$ can be written

$$(a, b) = (a + \psi(b), 0) + (-\psi(b), b),$$

and it is easy to check that this expresses the decomposition of $\mathfrak{g}$ into the direct sum of two orthogonal (i.e., with zero commutator) ideals $\lambda(\mathfrak{a})$ and $\nu'(\mathfrak{b}) = (-\psi(\mathfrak{b}), \mathfrak{b})$; the set $\nu(\mathfrak{b}) = \{(0, b) \mid b \in \mathfrak{b}\}$ was only a supplementary subalgebra of $\lambda(\mathfrak{a})$ in $\mathfrak{g}$; thus, with nonzero ψ (if there exists such a ψ), we get what can be called an "oblique direct sum", because then $\lambda(\mathfrak{a})$ is orthogonal to $\nu'(\mathfrak{b})$ but not to $\nu(\mathfrak{b})$, and this fact may hide the triviality of the extension.

Corollary. If $\mathfrak{a}$ is a L.a. with $\{0\}$ center, any inessential extension of a L.a. $\mathfrak{b}$ by $\mathfrak{a}$, corresponding to a homomorphism $\varphi : \mathfrak{b} \to$ der $(\mathfrak{a})$, the image $\bar{\varphi}$ of which in der $(\mathfrak{a})$/ad $\mathfrak{a}$ (ad $\mathfrak{a}$ is the L.a. of inner derivations of $\mathfrak{a}$) is $\{0\}$, is equivalent to the trivial extension.

As a conclusion of this study (for which a direct approach is just as short as the transcription of

Hochschild's general cohomological theory of extensions), we may write down the following:

Theorem. *Any extension of $\mathcal{P}$ by $\mathfrak{X}$ semisimple or complete, or of $\mathfrak{X}$ semisimple by $\mathcal{P}$, is equivalent to the trivial one. Any (inessential) extension of a L.a. $\mathfrak{X}$ such that $[\mathfrak{X}, \mathfrak{X}] = \mathfrak{X}$ by $\mathcal{P}$ is equivalent to the trivial extension.*

For instance, any (inessential) extension of $\mathcal{P}' \approx \mathcal{P}$ by $\mathcal{P}$ is equivalent to the trivial one. The result obtained is somewhat comforting, in a sense, because of the *fundamental dissymmetry* of the notion of (nontrivial) extension; nontrivial extensions of $\mathfrak{b}$ by $\mathfrak{a}$, or of $\mathfrak{a}$ by $\mathfrak{b}$, are very different. And if we could get nontrivial results in extending $\mathcal{P}$ by an internal L.a. $\mathfrak{X}$, or $\mathfrak{X}$ by $\mathcal{P}$, there would remain to understand why Nature prefers to extend $\mathcal{P}$ by $\mathfrak{X}$, rather than $\mathfrak{X}$ by $\mathcal{P}$, or the reverse.

We can now get some information on group extensions. More precisely:

Proposition 3. Any topological extension of the Poincaré group $\boldsymbol{P}$, or of its universal covering $\boldsymbol{P}^{\#}$, by an analytic group $\boldsymbol{X}$ whose Lie algebra $\mathfrak{X}$ has but inner derivations, is a central extension (and therefore a trivial one, in the case of $\boldsymbol{P}^{\#}$).

From the condition on $\mathfrak{X}$, $\mathrm{Aut}(\mathfrak{X})/\mathrm{Int}(\mathfrak{X})$, the factor group of all automorphisms of $\mathfrak{X}$ by the adjoint group $\mathrm{Int}(\mathfrak{X})$, which is isomorphic to $\mathrm{Aut}(\boldsymbol{X}^{\#})/\mathrm{Int}(\boldsymbol{X})$ (where $\boldsymbol{X}^{\#}$ is the universal covering of $\boldsymbol{X}$, and $\mathrm{Aut}(\boldsymbol{X}^{\#})$ the group of its analytic automorphisms) is totally disconnected, and therefore cannot contain a non-trivial homomorphic image of $\boldsymbol{P}^{\#}$ or $\boldsymbol{P}$; hence any (topological) extension is central. The last part follows from the work of Michel.[1]

It seems that a similar result is true in the case of extensions of $\boldsymbol{X}$ by $\boldsymbol{P}$ (with $[\mathfrak{X}, \mathfrak{X}] = \mathfrak{X}$ in order to get the triviality of the extensions); it depends mainly on the topological nature of the adjoint group $\mathrm{Int}(\mathcal{P}) = \mathrm{Ad}\ \boldsymbol{P}$—whether it is a closed subgroup of $GL(10, \mathsf{R})$ or not.

Remarks on Field Extension

Let $\mathfrak{a}$ be a L.a. on a (commutative) field K. The following commutation law, on the product vector space $\mathfrak{a} \times \mathfrak{a}$, defines a Lie algebra $\mathfrak{a}'$:

$$[(x_1, y_1), (x_2, y_2)]$$
$$= ([x_1, y_1] + [x_2, y_2], [x_1, y_2] + [x_2, y_1]); \quad (2.20)$$

$\mathfrak{a}'$ can be obtained from $\mathfrak{a}$ by extension to $\mathsf{K}[k]$ of the scalars' ring, k being an algebraic element of order 2 on K (i.e., $k \notin \mathsf{K}$, $k^2 = 1$), and then restriction to K. It is a direct sum of two orthogonal ideals $(\frac{1}{2}(1 \pm k)\mathfrak{a})$, both isomorphic to $\mathfrak{a}$. If now $\mathsf{K} = \mathsf{R}$, and if we denote by $\mathfrak{b}$

$$\mathfrak{b} = \mathfrak{a}(\mathsf{C}) \mid_{\mathsf{R}} = \mathfrak{a} \otimes \mathsf{C}, \quad (2.21)$$

the "twin-form" of $\mathfrak{a}$ (obtained by extension to C, the complex field, of the scalars, and then restriction to R, the real field), we have $\mathfrak{b}(\mathsf{C}) \approx \mathfrak{a}'(\mathsf{C})$ (cf. Bourbaki, Ref. 6, Chap. 1, Exercise 4).

Therefore, as $\mathfrak{L} \approx \mathfrak{a}(\mathsf{C}) \mid_{\mathsf{R}}$, where $\mathfrak{a}$ is any simple L.a. of order 3 [i.e., $\mathfrak{sl}(2, \mathsf{R})$, or $\mathfrak{so}(3) = \mathfrak{su}(2)$], $\mathfrak{L}(\mathsf{C}) \approx \mathfrak{a}(\mathsf{C}) \oplus \mathfrak{a}(\mathsf{C})$ and thus

$$\mathfrak{L}(\mathsf{C}) \mid_{\mathsf{R}} = \mathfrak{L} \otimes \mathsf{C} \approx \mathfrak{L} \oplus \mathfrak{L}.$$

The same thing is not true for $\mathcal{P}$, which, on the contrary of $\mathfrak{L}$, does not have a complex structure (e.g., see Ref. 7). In $\mathcal{P} \otimes \mathsf{C}$, $g^{\mu\nu}p_\mu p_\nu$ is no more an invariant; we must indeed consider $g^{\mu\nu}(p_\mu + ip_\mu)(p_\nu + ip_\nu)$, where $ip_\mu = q_\mu$ stands for the element $p_\mu \otimes i$ of $\mathcal{P} \otimes \mathsf{C}$, and we get the invariants $g^{\mu\nu}p_\mu q_\nu$ and $g^{\mu\nu}(p_\mu p_\nu - q_\mu q_\nu)$. We have therefore here a "correction" to $g^{\mu\nu}p_\mu p_\nu$, and a relation of the kind $m^2 = C + m'^2$, where C is an invariant, $m^2 = -g^{\mu\nu}p_\mu p_\nu$ and $m'^2 = -g^{\mu\nu}q_\mu q_\nu$. But in such an hypothesis, it is not quite clear which is the internal Lie algebra, and for this reason also one does not see how to express the "correction" in functions of internal quantum numbers.

We have thus seen that the extension theory of Lie algebras failed to give us the desired connection between the external algebra $\mathcal{P}$ and an internal semisimple algebra $\mathfrak{X}$. For what concerns group extensions, their possible nontriviality (in cases when the Lie algebras extensions are trivial) is due to the relative positions of a discrete subgroup, as it is in the case of $U(2) = U(1) \cdot SU(2)$ [semidirect product, isomorphic to $(U(1) \times SU(2))/\mathsf{Z}_2$], from where one can deduce the phase relation $(-1)^{2I} = (-1)^Y$; thus, in such a case, the consideration of group extensions gives us only phase relations—similar to those obtained by Lurçat and Michel[1]—but one cannot in such a way get "nondiscrete" relations, such as a mass formula. The theory of field extension also does not seem to give any interesting result. We are then naturally led to the considerations of Sec. 3.

3. UNIFICATION OF LIE ALGEBRAS

The most natural solution—and which, contrary to the extension theory, is symmetric—is to find a L.a. containing both $\mathcal{P}$ and an internal L.a. $\mathfrak{X}$ as subalgebras. Formulated in that way, this problem, studied by various authors (cf., for instance, Gar-

diner,[8]) is highly indeterminate, and cannot give any criterion for $\mathfrak{X}$, as it is always possible to imbed $\mathcal{P}$ and $\mathfrak{X}$ in another algebra of sufficiently high order [and even to do it in such a way that $\mathcal{P}$ and $\mathfrak{X}$ do not commute]. In order to get criterions for the choice of $\mathfrak{X}$, minimality conditions have to be imposed on the Lie algebra containing $\mathcal{P}$ and $\mathfrak{X}$. And besides, if the algebra containing $\mathcal{P}$ and $\mathfrak{X}$ is too large, the physical meaning of the superfluous generators (those of a supplementary subspace of the subspaces $\mathcal{P}$ and $\mathfrak{X}$) will not be clear.

Though one of us (M.F.) announced as early as May 1964 (in seminars, in Marseilles and Torino) the main features of the results of this part, the first results published in this direction were negative.[2] In such attempts, it was supposed that we are given a Lie algebra, the underlying vector space of which is the direct sum of those of $\mathcal{P}$ (with generators $M_{\mu\nu}$ and p_μ) and of a semisimple L.a. $\mathfrak{X}$ (with generators X_α), and such that $[\mathfrak{L}, \mathfrak{X}] = 0$, i.e., $[M_{\mu\nu}, X_\alpha] = 0 \ \forall \ \mu, \nu, \alpha$, or with apparently weaker (but in fact equivalent) hypotheses (such as $[M_{\mu\nu}, X_\alpha] = 0, \ \forall\alpha$, for some $M_{\mu\nu}$). A calculation on structure constants, eventually simplified by use of a Weyl basis (use which is somewhat delicate when dealing with real L.a., with which we must deal if we want, e.g., to look at topological considerations) gives $[\mathcal{P}, \mathfrak{X}] = 0$. We give here a simple proof of this result, in a somewhat more general context; this lemma will be useful in what follows.

Lemma 1. Let $\mathfrak{X}$ be a semisimple Lie algebra (on $\mathbf{R}$, e.g., or on any commutative field with characteristic zero) of finite dimension, and suppose there exist isomorphisms φ and ψ of (respectively) $\mathcal{P}$ and $\mathfrak{X}$ into a Lie algebra $\mathfrak{R}$ such that $\mathfrak{R} = \varphi(\mathcal{P}) + \psi(\mathfrak{X})$, (not necessarily direct) sum of vector spaces, and that we have $[\varphi(\mathfrak{L}), \psi(\mathfrak{X})] = 0$. Then $\varphi(\mathcal{P}) \cap \psi(\mathfrak{X}) = \{0\}$ and $\mathfrak{R} \approx \mathcal{P} \oplus \mathfrak{X}$.

Here, $\mathfrak{L}$ stands for a Levi subalgebra of $\mathcal{P}$ (isomorphic to the Lorentz L.a.), and it is always possible to select a basis $(M_{\mu\nu}, p_\mu)$ of $\mathcal{P}$ such that $M_{\mu\nu}$ is a basis of $\mathfrak{L}$, and the commutators are the usual ones. Any element of $\varphi(\mathcal{P}) \cap \psi(\mathfrak{X})$ commutes with the whole of $\varphi(\mathfrak{L})$ by hypothesis; but, in $\mathcal{P}$, only $\{0\}$ commutes with the whole of $\mathfrak{L}$; therefore $\varphi(\mathcal{P}) \cap \psi(\mathfrak{X}) = \{0\}$ and the sum of vectors spaces is direct. Now, given any $p_0 \in \mathcal{P}_0$, there exist (not uniquely determined, but it does not matter) $l \in \mathfrak{L}$ and $p_0' \in \mathcal{P}_0$ such that $p_0 = [l, p_0']$ (indeed, we have $p_\mu = [M_{\mu\nu}, -g_{\mu\rho}p_\rho]$); whence

$$[\psi(x), \varphi(p_0)] = [\psi(x), \varphi([l, p_0'])]$$
$$= [\varphi(l), [\psi(x), \varphi(p_0')]] \qquad (3.1)$$

because φ is a homomorphism, applying the Jacobi identity and the hypothesis. From (3.1) we get, again by the hypothesis, $[\psi(x), \varphi(p_0)] \in \varphi(\mathcal{P})$, for all $x \in \mathfrak{X}$, $p_0 \in \mathcal{P}_0$, and thus $[\psi(\mathfrak{X}), \varphi(\mathcal{P})] \subset \varphi(\mathcal{P})$: $\varphi(\mathcal{P})$ is therefore an ideal of $\mathfrak{R}$, which is then an extension of $\mathfrak{X}$ by $\mathcal{P}$, and, by the results of Sec. 2, isomorphic to the direct sum $\mathcal{P} \oplus \mathfrak{X}$. This last result can also be proved directly, using the Levi–Malcev theorem, which shows that $\mathfrak{R}$ has as Levi subalgebra $\varphi(\mathfrak{L}) \oplus \psi(\mathfrak{X})$, and the fact that $\mathfrak{gl}(4, \mathbf{R}) = \mathfrak{sl}(4, \mathbf{R}) \oplus \mathfrak{u}(1)$ contains no subalgebra isomorphic to $\mathfrak{L} \oplus \mathfrak{X}_1$, $\forall \mathfrak{X}_1$ simple L.a., which shows that every homomorphism from $\varphi(\mathfrak{L}) \oplus \psi(\mathfrak{X})$ to der $(\mathcal{P}_0)$, coinciding with the representation D_N on $\varphi(\mathfrak{L})$ [in order to get $\varphi(\mathcal{P})$ by the semidirect product], is trivial on $\psi(\mathfrak{X})$. We have, moreover, that $\mathfrak{R} = \varphi(\mathcal{P}) \oplus \psi(\mathfrak{X})$.

Remark. Let us now suppose that we have $\mathfrak{R} = \varphi(\mathcal{P}) + \psi(\mathfrak{X})$, direct sum of vector spaces, and that $\psi(\mathfrak{X})$ commutes with one regular element of $\varphi(\mathfrak{L})$, say $\varphi(l_1)$. Then, by a similar argument, $\mathfrak{R} \approx \mathcal{P} \oplus \mathfrak{X}$. Indeed, there exist elements l_2, l_3, l_1', l_2', l_3' that, together with l_1, make up a basis of $\mathfrak{L}$, such that

$$[l_i, l_j] = \pm l_k, \qquad [l_i, l_i'] = \pm l_{k'},$$
$$[l_i', l_j'] = \pm l_k, \qquad [l_i, l_i'] = 0 \quad (ijk \text{ cyclic}).$$

As $[\psi(x), \varphi(l_1)] = 0 \ \forall \ x \in \mathfrak{X}$, we get, by the Jacobi identity, $[\psi(x), \varphi(l_k)] \in \varphi(\mathcal{P})$ $(k = 2, 3)$ and $[\psi(x), \varphi(l_k')] \in \varphi(\mathcal{P})$ $(k = 2, 3, 1)$, and therefore

$$[\psi(x), \varphi(p_0)] \in \varphi(\mathcal{P}) \ \forall p_0 \in \mathcal{P}_0 \ (\text{as } p_0 = [l, p_0']);$$

hence $\varphi(\mathcal{P})$ is an ideal in $\mathfrak{R}$, which is therefore an extension of $\mathfrak{X}$ by $\mathcal{P}$, and thus isomorphic to $\mathcal{P} \oplus \mathfrak{X}$ ($\mathfrak{X}$ being semisimple). In a similar way, if $\mathfrak{R} = \varphi(\mathcal{P}) + \psi(\mathfrak{X})$ without intersection, and if a regular (or a semisimple) element Q of $\psi(\mathfrak{X})$ commutes with all $\varphi(\mathcal{P})$, we find, by use of a Weyl basis, that $\psi(\mathfrak{X})(\mathbf{C})$ is (See Ref. 2) an ideal in $\mathfrak{R}(\mathbf{C})$, and thus $\mathfrak{R}(\mathbf{C})$ is isomorphic to $\mathcal{P}(\mathbf{C}) \oplus \mathfrak{X}(\mathbf{C})$; therefrom, passing to real forms (and because $\mathfrak{L}$ is already a twin form), $\mathfrak{R} \approx \mathcal{P} \oplus \mathfrak{X}$. However, it must be emphasized that the assumption Q semisimple *does* restrict generality (as long as $\mathfrak{X}$ is noncompact)—and this, regardless of the interpretation of the commutation or noncommutation of Q with $\varphi(\mathcal{P})$. The negative results of Ref. 2 are thus seen to be special cases of our study in Sec. 2. We see later that positive results can be obtained; but first let us set our problem more precisely, and, for this purpose, introduce the following definition.

[8] C. W. Gardiner, Phys. Rev. Letters 11, 3 (1964).

Definition 1. Let $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ be n finite-dimensional Lie algebras, over the same commutative field $\mathbf{K}$. A Lie algebra $\mathfrak{u}$ over $\mathbf{K}$ is called a *unifying algebra* of $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ if there exist isomorphisms λ_k of $\mathfrak{a}_k$ into $\mathfrak{u} (k = 1, \cdots, n)$ such that

$$\mathfrak{u} = \lambda_1(\mathfrak{a}_1) + \cdots + \lambda_n(\mathfrak{a}_n), \tag{3.2}$$

not necessarily the direct sum of vector spaces. We write $\mathfrak{u} = U(\mathfrak{a}_1, \cdots, \mathfrak{a}_n)$.

This definition calls for a few remarks. First, we do not suppose that all possible pairs of algebras $\lambda_k(\mathfrak{a}_k)$ have null-intersection. We thus say that we have a unification with (specifying if necessary which algebras have such intersection) or without intersection. In the case of two algebras with $\{0\}$ intersection, we get as a special case the notion of inessential extension (in that case, one of the algebras is an ideal in the unification). In any case, we have always a (trivial) unification with $\{0\}$ intersection, namely the direct sum $\mathfrak{a}_1 \oplus \cdots \oplus \mathfrak{a}_n$. It may happen that this is the only unification (with or without intersection). For instance, it may be seen, by trying to define the missing structure constants, that there is no $U(\mathfrak{so}(3), \mathfrak{so}(3))$ with one-dimensional intersection, and from this (using the Levi–Malcev theorem, and showing that the Lie algebra of the inhomogeneous rotation group in 3 dimensions is not a unification) that the only $U(\mathfrak{so}(3), \mathfrak{so}(3))$, apart of $\mathfrak{so}(3)$ itself, is the trivial one, $\mathfrak{so}(3) \oplus \mathfrak{so}(3)$. Moreover, if we try to define, *a priori*, the structure constants of a unification, we see that the Jacobi identity implies (in general, quadratic) conditions, the number of which is much greater (in general) than the number of the missing constants. Therefore the existence of a nontrivial unification is a somewhat rare, and therefore an interesting, phenomenon.

We see also that the notion of "minimality" is here realized: we have $\dim \mathfrak{u} \leq \dim \mathfrak{a}_1 + \cdots + \dim \mathfrak{a}_n$ (being equal only in case of $\{0\}$ intersection). If two algebras (for instance) are imbedded in a higher-dimensional L.a., the latter being not a unification of the two former L.a., it can always be considered as a unification of the two former and of a suitable number of one-dimensional L.a. $\mathfrak{u}(1)$ (this notation being considered as having, *a priori*, no topological implications on the compactness or noncompactness of the corresponding one-parameter subgroup of a Lie group corresponding to the unification). If we limit that number *a priori*, we can make more precise the wanted degree of minimality.

It is quite banal to state that any L.a. can be considered as a unification of itself and of any subalgebra. Moreover, if two L.a. have a common direct factor, their direct sum, divided by this direct factor, is a (quite trivial) unification, with intersection on this common direct factor (that is, so to speak, "put in common"). We are therefore led to the following:

Definition 2. (i) A *unification* $\mathfrak{u} = U(\mathfrak{a}, \mathfrak{b})$ is said to be *trivial* if, while $\mathfrak{a} \approx \mathfrak{a}_1 \oplus \mathfrak{a}'$, $\mathfrak{b} \approx \mathfrak{a}_1 \oplus \mathfrak{b}'$, one has $\mathfrak{u} = \mathfrak{a}_1 \oplus \mathfrak{a}' \oplus \mathfrak{b}'$. In the same way, a unification of n L.a. is said to be trivial if it is isomorphic to the direct sum, possibly up to a common (to two or more L.a.) direct factor.

(ii) A unification $\mathfrak{u} = U(\mathfrak{a}, \mathfrak{b})$ is said to be *banal* if $\mathfrak{b} \subset \mathfrak{a}$ and $\mathfrak{u} \approx \mathfrak{a}$ (or if $\mathfrak{a} \subset \mathfrak{b}$ and $\mathfrak{u} \approx \mathfrak{b}$). Accordingly, a unification of n L.a. is said to be banal if one L.a. contains all the others (possibly up to a direct factor, or to a trivial unification).

We introduced the notion of banal unification and distinguished it from that of a trivial one for the following reason: In a trivial unification, any invariant of one L.a. is an invariant of the unification, while that is not (in general) the case for a banal unification. We see for instance that $\mathcal{P} \subset \mathfrak{su}(2, 2)$, but $g^{\mu\nu}p_\mu p_\nu$ is not an invariant of $\mathfrak{su}(2, 2)$ (banal unification of $\mathcal{P}$ and of itself).

Before we pass to the general study, we first treat two particular cases (interesting in themselves), one with positive and one with negative results.

Example 1. A nontrivial and nonbanal unification $U(\mathcal{P}, \mathfrak{L}, \mathfrak{u}(1) \oplus \mathfrak{u}(1))$ with a 2-dimensional intersection.

[We construct later a unification $U(\mathcal{P}, \mathfrak{L})$ with such an intersection]. Let us consider the semidirect product $\mathfrak{R} = \mathfrak{L} \cdot \mathfrak{R}_0$, where $\mathfrak{R}_0$ is a 10-dimensional commutative ideal, and a defining homomorphism $\varphi : \mathfrak{L} \to \mathfrak{gl}(10, \mathbf{R}) = \operatorname{der}(\mathfrak{R}_0)$ $\varphi = D_N \oplus \operatorname{ad}$ (ad standing for the adjoint 6-dimensional representation of $\mathfrak{L}$). Denote by (q_k, q'_k) $(k = 1, 2, 3)$ a basis of the space of the representation ad; (p_k, p_4) being a basis of $\mathcal{P}_0 \subset \mathfrak{R}_0$, the space of D_N, (p_k, p_4, q_k, q'_k) is a basis of $\mathfrak{R}_0$, and $\mathfrak{R} \supset (\mathfrak{L} \cdot \mathcal{P}_0) = \mathcal{P}$. In this basis, the commutation relations of $\mathfrak{R}$ are those of $\mathcal{P}$, in addition, with $\mathfrak{R}_0$ commutative and

$$[a_i, q_j] = \epsilon_{ijk} q_k, \qquad [a_i, q'_j] = \epsilon_{ijk} q'_k,$$
$$[b_i, q_j] = \epsilon_{ijk} q'_k, \qquad [b_i, q'_j] = -\epsilon_{ijk} q_k. \tag{3.3}$$

$\mathfrak{L}$ is a Levi subalgebra of $\mathfrak{R}$, any other being of the form $(I + \operatorname{ad}(p_0 + q_0))\mathfrak{L}$, where I is the identity,

$p_0 \in \mathcal{P}_0$, and q_0 is a linear combination of the q_k's and the q_k''s. We are looking for a subalgebra $\mathcal{L}'$ of $\mathcal{R}$ isomorphic to $\mathcal{L}$, with $\mathcal{L}' \cap \mathcal{P}$ minimal. For this problem, it is enough to look for a $\mathcal{L}'$ of the form $(I + \text{ad } q_0)\mathcal{L}$. If $q_0 = \sum_k (\alpha_k q_k + \alpha_k' q_k')$, we get $\mathcal{L}'$ with a (canonical) basis (a_k', b_k') where

$$a_i' = a_i + \alpha_i q_k - \alpha_k q_i + \alpha_i' q_k' - \alpha_k' q_i'$$

and

$$b_i' = b_i - \alpha_i' q_k + \alpha_k' q_i + \alpha_i q_k' - \alpha_k q_i'$$

(i, j, k cyclic here).

Now, in general, we have $\mathcal{L} \cap \mathcal{L}' = \{\mathfrak{u}(1) \oplus \mathfrak{u}(1)\}$, as the general element $\sum_k (\lambda_k a_k + \mu_k b_k) \in \mathcal{L}$ belongs also to $\mathcal{L}'$ if and only if 6 linear homogeneous equations (the expression of which is evident from the a_i''s and the b_i''s) in the λ_k''s and μ_k''s are verified, a two-parameter solution of which is readily seen; moreover, we see that if (λ_k, μ_k) is a solution, $(\mu_k, -\lambda_k)$ is also a solution. Therefore the intersection is a (maximal) commutative 2-dimensional subalgebra of $\mathcal{L}$ (or the whole of $\mathcal{L}$ if $q_0 = 0$), for

$$[\sum_k (\lambda_k a_k + \mu_k b_k), \sum_k (\mu_k a_k - \lambda_k b_k)] = 0.$$

In the nondegenerate case, we have

$$\mathcal{R} = \mathcal{P} + \mathcal{L}' + (\mathfrak{u}(1) \oplus \mathfrak{u}(1)),$$

where $\mathcal{P} = \mathcal{L} \cdot \mathcal{P}_0$, non-direct sum of vector spaces, and so $\mathcal{R} = U(\mathcal{P}, \mathcal{L}, (\mathfrak{u}(1) \oplus \mathfrak{u}(1)))$ with a 2-dimensional intersection.

Remark. Here, $g^{\mu\nu} p_\mu p_\nu$ commutes with $\mathcal{L}$, but also with $\mathcal{R}_0$ (commutative), and is therefore an invariant of $\mathcal{R}$. We see that this is a general phenomenon in all unifications of $\mathcal{P}$ and of a Lorentz L.a. $\mathcal{L}$. In order to avoid this, we have to consider, as an internal algebra, a L.a. containing strictly $\mathcal{L}$ (and not as a direct factor).

Proposition 1. Every unifying algebra $U(\mathcal{P}, \mathfrak{su}(3))$ is isomorphic to $\mathcal{P} \oplus \mathfrak{su}(3)$. Every unifying algebra $U(\mathcal{P}, \mathfrak{su}(3), \mathfrak{u}(1))$ is isomorphic either to $\mathcal{P} \oplus \mathfrak{su}(3)$ [1-dimensional intersection, between $\mathfrak{u}(1)$ and $\mathcal{P}$, or $\mathfrak{su}(3)$], to $\mathfrak{us}(3) \oplus \mathcal{P} \oplus \mathfrak{u}(1)$, or else to the pseudo-trivial unification $\mathfrak{us}(3) \oplus \text{der}(\mathcal{P})$.

We are looking for a unification $\mathcal{R}$—a L.a. such that there exist injective homomorphisms φ, ψ (and possibly ψ'):

$$0 \to \mathcal{P} \xrightarrow{\varphi} \mathcal{R}, \quad 0 \to \mathfrak{su}(3) \xrightarrow{\psi} \mathcal{R}$$

$$[\text{and possibly } \psi' : \mathfrak{u}(1) \xrightarrow{\psi'} \mathcal{R}], \qquad (3.4)$$

such that

$$\mathcal{R} = \varphi(\mathcal{P}) + \psi(\mathfrak{su}(3)) \text{ [and possibly } \psi'(\mathfrak{u}(1))].$$

Let us denote by $\mathcal{B}$ a Levi subalgebra of $\mathcal{R}$: $\mathcal{R} = \mathcal{B} \cdot \mathcal{R}_0$ ($\mathcal{R}_0$ is the radical); $\mathcal{B}$ must contain $\approx \mathcal{L}$ (i.e., a subalgebra isomorphic to $\mathcal{L}$) and $\approx \mathfrak{su}(3)$. Moreover, $\dim \mathcal{B} \leq \dim \mathcal{R} \leq 18$ [or possibly 19, if we add a $\mathfrak{u}(1)$], and, if $\dim \mathcal{B} > \dim \mathcal{L} + \dim \mathfrak{su}(3) = 14$ (or possibly 15), we must have $\mathcal{B} \supset \approx \mathcal{P}$, because with every $p_0 \in \mathcal{P}_0$, $[\mathcal{L}, p_0]$ gives $\mathcal{P}_0$. $\mathcal{B}$ cannot be a real form of types $\mathfrak{a}_2$ or $\mathfrak{g}_2$ (as none contains $\mathcal{L}$), nor of type $\mathfrak{b}_2$ [none contains $\mathfrak{su}(3)$]; $\mathcal{B}$ cannot be $\mathfrak{sl}(3, \mathbf{C})$ (type $\mathfrak{a}_2 \otimes \mathbf{C}$), the latter being of dimension 16 and not containing $\mathcal{P}$. The only real form of type $\mathfrak{a}_3$ containing $\mathfrak{su}(3)$ and $\mathcal{L}$ is $\mathfrak{su}(3, 1)$, which does not contain $\mathcal{P}$ [and has no real 4-dimensional representation, a fact that eliminates the possibility of constructing $\mathcal{P}$ by a semidirect product of $\mathfrak{su}(3, 1)$ and a 4-dimensional—at most—commutative ideal, because we are dealing with L.a. over $\mathbf{R}$]. Therefore $\mathcal{B}$ cannot be a simple L.a. If now $\mathcal{B} = \mathcal{B}_1 \oplus \cdots \oplus \mathcal{B}_k$, with $\mathcal{B}_j$ ($j = 1, \cdots, k$) simple, by the same argument, a single $\mathcal{B}_j$ cannot contain both $\approx \mathcal{L}$ and $\approx \mathfrak{su}(3)$; therefore $\mathcal{L} \oplus \mathfrak{su}(3)$ is a semisimple subalgebra of $\mathcal{B}$ (and of $\mathcal{R}$) [and, by Lemma 1, $U(\mathcal{P}, \mathfrak{su}(3))$ is isomorphic to $\mathcal{P} \oplus \mathfrak{su}(3)$]. This is then a Levi subalgebra, because it is contained by no other semisimple L.a. of dimension ≤ 19, containing $\mathcal{P}$ if its dimension is ≥ 16. As $\mathfrak{su}(3)$ has no real 4- or 5-dimensional representation, the homomorphism $\mathcal{L} \oplus \mathfrak{su}(3) \to \text{der}(\mathcal{R}_0)$, defining the semi-direct product $\mathcal{B} \cdot \mathcal{R}_0$ (with $\mathcal{R}_0$ of dimension 4 or possibly 5), and coinciding with D_N on $\mathcal{L}$ ($\mathcal{R}_0$ must here contain a 4-dimensional commutative subalgebra, $\mathcal{P}_0$) is 0 on $\mathfrak{su}(3)$. Therefore

$$U(\mathcal{P}, \mathfrak{su}(3)) \approx \mathcal{P} \oplus \mathfrak{su}(3) \qquad (3.5)$$

and also

$$U(\mathcal{P}, \mathfrak{su}(3), \mathfrak{u}(1)) \approx \mathfrak{su}(3) \oplus U(\mathcal{P}, \mathfrak{u}(1)). \qquad (3.6)$$

The proposition follows then from the lemma:

Lemma 2. $U(\mathcal{P}, \mathfrak{u}(1)) \approx \mathcal{P}$ (banal unification), or $\mathcal{P} \oplus \mathfrak{u}(1)$ (trivial), or der $(\mathcal{P})$.

Indeed, $\mathcal{L}$ is a Levi subalgebra of such a unification $\mathcal{R}$, which is therefore $\mathcal{R} = \mathcal{L} \cdot \mathcal{R}_0$, $\mathcal{R}_0$ being a solvable 5-dimensional ideal (if it is not the banal unification); but $\mathcal{L}$ has no 5-dimensional real representation containing D_N, except $D_N \oplus O$ (O is the trivial 1-dimensional representation). Therefore, $\mathcal{R} \approx (\mathcal{L} \oplus \mathfrak{u}(1)) \cdot \mathcal{P}_0$ (since $\mathcal{R}_0$ must contain $\mathcal{P}_0$, 4-dimensional commutative L.a.). D_N being irreducible, the only matrices of order 4 commuting with $D_N(\mathcal{L})$ are scalar matrices, and therefore the image of $\mathfrak{u}(1)$ by a defining homomorphism is either $\{0\}$ or all the scalar matrices, whence the result.

Before we go to the general study, let us recall two theorems, that we need later on (see Ref. 7 for the first one, and Ref. 9 for the second).

Theorem A (Iwasawa). Let $\mathfrak{g}_0$ be a real semisimple L.a. There exists a decomposition (*the Iwasawa decomposition*)

$$\mathfrak{g}_0 = \mathfrak{k}_0 + \mathfrak{a}_0 + \mathfrak{n}_0 \tag{3.7}$$

(direct sum of vector spaces), where $\mathfrak{k}_0$ is a maximal compactly imbedded subalgebra, $\mathfrak{a}_0$ a maximal Abelian subspace (with semisimple elements—cf. Ref. 6 for this notion) of the subspace $\mathfrak{p}_0$ corresponding to the Cartan decomposition $\mathfrak{g}_0 = \mathfrak{k}_0 + \mathfrak{p}_0$ ($\mathfrak{k}_0 + i\mathfrak{p}_0$ is then a compact real form of the complex L.a. $\mathfrak{g}$, of which $\mathfrak{g}_0$ is a real form), and $\mathfrak{n}_0$ a nilpotent subalgebra. There exists moreover a basis of $\mathfrak{g} = \mathfrak{g}_0(\mathbf{C})$ such that the matrices of ad $(\mathfrak{k}_0 + i\mathfrak{p}_0)$ are skew-Hermitian, those of ad $\mathfrak{n}_0$ lower triangular, and those of ad $\mathfrak{a}_0$ real diagonal.

Theorem B (Malcev). In a complex simple L.a., the maximal dimension of commutative subalgebras (with nilpotent elements then) is $[\frac{1}{4}(n+1)^2]$ for $\mathfrak{a}_n$, $\frac{1}{2}n(n-1)+1$ for $\mathfrak{b}_n$ $(n \geq 4)$, $\frac{1}{2}n(n+1)$ for $\mathfrak{c}_n$, $\frac{1}{2}n(n-1)$ for $\mathfrak{d}_n$, and 3, 5, 9, 16, 27, 36 for $\mathfrak{g}_2$, $\mathfrak{b}_3$, $\mathfrak{f}_4$, $\mathfrak{e}_6$, $\mathfrak{e}_7$, and $\mathfrak{e}_8$, respectively, and they are all conjugate by automorphisms except for $\mathfrak{b}_4$, $\mathfrak{d}_4$ (2 classes), $\mathfrak{g}_2$ (3 classes) (and for $\mathfrak{a}_1$, $\mathfrak{a}_2$).

For real forms of those complex simple L.a., the maximal dimension of commutative subalgebras with nilpotent elements is at most that of the corresponding complex L.a.; even for noncompact forms, it may be smaller. For instance, one checks that for $\mathfrak{so}(p, 1)$ this dimension is $p-1$, which is definitely smaller than the above number for $p \geq 9$.

We begin the **general study** by the case of the simple L.a. that are contained in $\mathcal{P}$ (and therefore in $\mathfrak{L}$), i.e., $\mathfrak{so}(3)$ and $\mathfrak{sl}(2, \mathbf{R})$ (type $\mathfrak{a}_1$). We denote by D_4 a 4-dimensional real representation of $\mathfrak{L}$, nonequivalent to D_N, and nonreducible by real transformations [one gets it by considering subalgebras of $\mathfrak{sl}(4, \mathbf{R})$ isomorphic to $\mathfrak{L}$ in a basic representation of $\mathfrak{sl}(4, \mathbf{R})$, or by considering the basic representation of $\mathfrak{so}(3, 2)$ as $\mathfrak{sp}(2, \mathbf{R})$, in 4 dimensions; cf. Appendix]. We prove now the following lemma, on $U(\mathcal{P}, \mathfrak{X})$ and $U(\mathcal{P}, \mathfrak{X}, \mathrm{u}(1))$ for $\mathfrak{X}$ of type $\mathfrak{a}_1$—here, as in the whole study, we study the influence of the addition of a one-dimensional algebra on the triviality of the unifications, because of the possible physical interpretation of this addition. [Moreover, the results are very similar for $U(\mathcal{P}, \mathfrak{X}, \mathrm{u}(1) \oplus \mathrm{u}(1))$.]

Lemma 3. Every $U(\mathcal{P}, \mathfrak{so}(3))$ is either trivial $[\mathcal{P} \oplus \mathfrak{so}(3)]$ or banal $(\mathcal{P})$. Every $U(\mathcal{P}, \mathfrak{so}(3), \mathrm{u}(1))$ is either trivial or banal $[\mathcal{P}, \mathcal{P} \oplus \mathfrak{so}(3), \mathcal{P} \oplus \mathfrak{so}(3) \oplus \mathrm{u}(1)$, or $\mathcal{P} \oplus \mathrm{u}(1)]$, or pseudo-trivial $[\mathfrak{so}(3) \oplus \mathrm{der}\,(\mathcal{P})]$, or else a semidirect product $(\mathfrak{L} \cdot \mathcal{R}_0)$ defined by the representation $D_N \oplus D_4$ of $\mathfrak{L}$, $\mathcal{R}_0$ being an 8-dimensional commutative ideal. The same holds for $\mathfrak{sl}(2, \mathbf{R})$—with the obvious transposition.

Indeed, the unification $\mathcal{R}$ cannot be a simple L.a. because it must be of dimension ≤ 13 [or possibly 14, if adding a u(1)] and contain $\mathcal{P}$. Now, $\mathfrak{so}(4, 1) = U(\mathfrak{L}, \mathfrak{so}(3), \mathrm{u}(1))$, but we cannot form a semidirect product of that algebra with $\mathcal{P}_0$ so as to get $\mathcal{P}$ (it has no 4-dimensional real representation). The same holds for $\mathfrak{sl}(2, \mathbf{R})$ with $\mathfrak{so}(3, 2)$ (the 4-dimensional representation of that algebra giving D_4 on $\mathfrak{L}$, and not D_N). Therefore any Levi subalgebra $\mathcal{B}$ of $\mathcal{R}$ is isomorphic either to $\mathfrak{X} \oplus \mathfrak{L}$ or to $\mathfrak{L}$ ($\mathfrak{X}$ being here of type $\mathfrak{a}_1$). If it is $\mathfrak{X} \oplus \mathfrak{L}$, then $U(\mathcal{P}, \mathfrak{X}) \approx \mathcal{P} \oplus \mathfrak{X}$ from our lemma, and $U(\mathcal{P}, \mathfrak{X}, \mathrm{u}(1)) \approx (\mathfrak{L} \oplus \mathfrak{X}) \cdot \mathcal{R}_0$, with $\mathcal{R}_0$ solvable and 4- or 5-dimensional, containing $\mathcal{P}_0$; the defining homomorphism must then be null on $\mathfrak{X}$ [for $\mathfrak{a}_1 \oplus \mathfrak{a}_1 \oplus \mathfrak{a}_1 \not\subset \mathfrak{a}_4$, from Ref. 10] and $\mathcal{R} \approx \mathfrak{X} \oplus U(\mathcal{P}, \mathrm{u}(1))$. If it is $\mathfrak{L}$, $\mathcal{R} = (\mathfrak{L} \cdot \mathcal{R}_0)$ with solvable $\mathcal{R}_0$ of dimension ≤ 8, and we must be able to "translate" a subalgebra of $\mathfrak{L}$ isomorphic to $\mathfrak{X}$ so that to obtain $\mathcal{R} = \mathfrak{L} + \mathfrak{X} +$ (possibly) u(1). Further, the defining homomorphism must contain D_N as a "direct factor" (so as to obtain $\mathcal{P} = \mathfrak{L} \cdot \mathcal{P}_0$); therefore, $\mathcal{R}_0$ must be 8-dimensional-commutative, and the defining homomorphism $D_N \oplus D_4$ (or a similar expression, with D_4 replaced by a representation equivalent to D_4—possibly by complex transformations), because one checks that $D_N \oplus D_N$ gives a (too large) intersection between $\mathfrak{L}$ and $\mathfrak{X}$. It is now easy to check that this is a unification, as described here above [for instance, if we take a basis p_μ, q_μ of $\mathcal{R}_0$ and $\mathfrak{X} \approx \mathfrak{X}' \subset \mathfrak{L}$, then $(I + 2 \operatorname{ad} \alpha q_4)\mathfrak{X}' = \mathfrak{X}$ will do, the added u(1) being, e.g., generated by q_4].

Proposition 2. Let $\mathfrak{X}$ be a simple compact L.a., distinct from $\mathfrak{so}(3)$. Then every $U(\mathcal{P}, \mathfrak{X})$ is trivial $(\mathcal{P} \oplus \mathfrak{X})$, and every $U(\mathcal{P}, \mathfrak{X}, \mathrm{u}(1))$ banal or trivial $[\mathcal{P} \oplus \mathfrak{X}$ or $\mathcal{P} \oplus \mathfrak{X} \oplus \mathrm{u}(1)]$ or pseudo-trivial $[\mathfrak{X} \oplus \mathrm{der}\,(\mathcal{P})]$.

[9] A. I. Malcev, Transl. Am. Math. Soc., Ser. 1, 9, 214 (1962) (original in Russian, 1945).

[10] E. B. Dynkin, Transl. Am. Math. Soc., Ser. 2, 6, 111, 245 (1957) (original in Russian, 1952).

It follows from such a result that compact simple L.a. are excluded by our criterion on possible internal symmetries (we see later that all compact symmetries are excluded).

Let $\mathfrak{R}$ be the unification, $\mathfrak{G}$ a Levi subalgebra of $\mathfrak{R}$, containing the image $\psi(\mathfrak{X})$ of $\mathfrak{X}$ in $\mathfrak{R}$; $\mathfrak{G} \supset \psi(\mathfrak{X})$ strictly, for $\mathfrak{G} \supset \approx\mathfrak{L}$ also. Then either $\mathfrak{G}$ is simple, or $\mathfrak{G}$ is semisimple–nonsimple, in which case $\mathfrak{G} \approx \mathfrak{X} \oplus \mathfrak{L}$ because of dimension considerations. In that case, by Lemma 1, and by the same considerations than in the case of $\mathfrak{su}(3)$, $U(\mathfrak{P}, \mathfrak{X}) \approx \mathfrak{P} \oplus \mathfrak{X}$ and $U(\mathfrak{P}, \mathfrak{X}, \mathfrak{u}(1)) \approx \mathfrak{X} \oplus U(\mathfrak{P}, \mathfrak{u}(1))$. Let us therefore suppose $\mathfrak{G}$ simple; we have $\dim \mathfrak{G} \leq \dim \mathfrak{R} \leq \dim \mathfrak{X} + 10$ (or possibly 11). But we learn from Dynkin[10] that maximal semisimple subalgebras of greatest dimension are in $\mathfrak{a}_n$, $\mathfrak{a}_{n-1}$; in $\mathfrak{b}_n$, $\mathfrak{b}_n$; in $\mathfrak{c}_n$, $\mathfrak{c}_{n-1} \oplus \mathfrak{a}_1$; in $\mathfrak{d}_n$, $\mathfrak{b}_{n-1}$; in $\mathfrak{g}_2$, $\mathfrak{a}_2$; in $\mathfrak{f}_4$, $\mathfrak{e}_6$, $\mathfrak{e}_7$, $\mathfrak{e}_8$, subalgebras, the dimension of which is smaller by more than 12 than the dimension of the algebra. Therefore $\mathfrak{G} \approx \mathfrak{L} \oplus \mathfrak{X}$, except perhaps in case $\mathfrak{X}$ is of types $\mathfrak{g}_2$, $\mathfrak{su}(n)$ ($n = 2, 3, 4$, and possibly 5), $\mathfrak{so}(n)$ ($n = 5, 7, 8, 9, 10$, and possibly 11), because of the difference of dimension between the above-mentioned algebras, and the fact that the noncompact real form of $\mathfrak{g}_2$ (that, following Freudenthal[11] we denote by $\mathfrak{g}_{2,2}$) does not contain $\mathfrak{L}$ (we denote the compact form by $\mathfrak{g}_{2,0}$).

As no real form of $\mathfrak{b}_3$ contains both $\mathfrak{L}$ and $\mathfrak{g}_{2,0}$ (compact), the hypothesis $\mathfrak{X} = \mathfrak{g}_{2,0}$ fails. We already treated the cases of $\mathfrak{su}(2) = \mathfrak{so}(3)$, and $\mathfrak{su}(3)$ (in that case, we had the announced result). For $\mathfrak{X} = \mathfrak{su}(4) = \mathfrak{so}(6)$, $\mathfrak{G}$ simple can be only either $\mathfrak{so}(6, 1)$ or $\mathfrak{su}(4, 1)$; but then $\mathfrak{G}$ must be the whole unification $\mathfrak{R}$, which it is not, since those algebras do not contain $\approx \mathfrak{P}$. [The "noncompact part" of $\mathfrak{so}(6, 1)$ is of dimension 6, and that of $\mathfrak{P}$ is of dimension 7; for $\mathfrak{su}(4, 1)$, we check it in a basic representation. Besides, if it were a unification, the intersection between $\varphi(\mathfrak{P})$ and $\psi(\mathfrak{X})$ would be at least 3-dimensional—a compact subalgebra of $\mathfrak{P}$—and not, as imposed by dimension considerations, 1- or possibly 2-dimensional.]

We have the same result for $\mathfrak{X} = \mathfrak{so}(5) = \mathfrak{sp}(2)$, because $\mathfrak{so}(5, 1) = \mathfrak{su}^*(4) \supset \approx\mathfrak{P}$, and has no real 4 (or 5)-dimensional representation. For $\mathfrak{X} = \mathfrak{so}(7)$, it is clear that $\mathfrak{g}_2 \otimes \mathbf{C}$ does not fit; but $\mathfrak{so}(7, 1)$ also does not fit, since it does not contain $\mathfrak{P}$: the nilpotent subalgebra of $\mathfrak{so}(7, 1)$ in the Iwasawa decomposition is 6-dimensional but commutative, while $\mathfrak{P}$ contains a 6-dimensional noncommutative nilpotent subalgebra. Besides, one sees from a basic representation of $\mathfrak{so}(p, 1)$ that these algebras (for $p \geq 4$) contain the "inhomogeneous $\mathfrak{so}(4)$" L.a., but not the "inhomogeneous $\mathfrak{so}(3, 1)$" L.a., for which we need, among the algebras $\mathfrak{so}(p, q)$, $p \geq 4$ and $q \geq 2$. Thus $\mathfrak{so}(7)$ fails.

In the case of $\mathfrak{X} = \mathfrak{so}(8)$, the (only) possibility $\mathfrak{G} = \mathfrak{R} = \mathfrak{so}(8, 1)$ fails for the same reason (and, besides, the nilpotent subalgebra is 7-dimensional, but Abelian). The cases of $\mathfrak{so}(9)$, $\mathfrak{so}(10)$, $\mathfrak{so}(11)$ are similar (and, besides, the intersection of a supposed unification $\mathfrak{R} = \mathfrak{G}$ must be at least 3-dimensional, which is impossible). Hence we have $\mathfrak{G} \approx \mathfrak{L} \oplus \mathfrak{X}$. Q.E.D.

Proposition 3. If $\mathfrak{X}$ is semisimple and no simple ideal of $\mathfrak{X}$ has a real representation of dimension at most 10 [resp. 11], then either $\mathfrak{X} \supset \approx \mathfrak{P}$, in which case $U(\mathfrak{P}, \mathfrak{X}) \approx \mathfrak{P} \oplus \mathfrak{X}$ [resp. $U(\mathfrak{P}, \mathfrak{X}, \mathfrak{u}(1)) \approx \mathfrak{X} \oplus U(\mathfrak{P}, \mathfrak{u}(1))$], or $\mathfrak{X} \supset \approx\mathfrak{P}$, in which case there is also the banal unification $\mathfrak{X}$ [resp. $\mathfrak{X}$, or $\mathfrak{X} \oplus \mathfrak{u}(1)$].

It is sufficient to prove the proposition for simple $\mathfrak{X}$. From the Levi–Malcev theorem, the unification $\mathfrak{R} = \mathfrak{G} \cdot \mathfrak{R}_0$ ($\mathfrak{G}$ is a Levi subalgebra); $\mathfrak{G} \supset \approx\mathfrak{X}$. If $\mathfrak{G}$ is not simple, $\mathfrak{G} \approx \mathfrak{L} \oplus \mathfrak{X}$, and the result follows. If $\mathfrak{G}$ is simple, it has, *a fortiori*, no real representation of dimension at most 10 [resp. 11]; so $\mathfrak{R} = \mathfrak{G} \oplus \mathfrak{R}_0$; but $\mathfrak{R} \supset \approx\mathfrak{P}$ and $\mathfrak{G} \supset \approx\mathfrak{L}$, so that $\mathfrak{G} \supset \approx\mathfrak{P}$ from the structure of $\mathfrak{P}$ (which is not a direct sum of two subalgebras), and therefore $\mathfrak{R}_0 = \{0\}$ [resp. $\{0\}$ or $\mathfrak{u}(1)$]. Q.E.D.

Among the remaining simple algebras, all contain $\mathfrak{L}$ except for the following three (noncompact): $\mathfrak{sl}(3, \mathbf{R})$, $\mathfrak{su}(2, 1)$, and $\mathfrak{g}_{2,2}$.

Lemma 4. Every $U(\mathfrak{P}, \mathfrak{sl}(3, \mathbf{R}))$ is $\mathfrak{P} \oplus \mathfrak{sl}(3, \mathbf{R})$ and every $U(\mathfrak{P}, \mathfrak{sl}(3, \mathbf{R}), \mathfrak{u}(1))$ is $U(\mathfrak{P}, \mathfrak{u}(1)) \oplus \mathfrak{sl}(3, \mathbf{R})$, up to isomorphism.

Let $\mathfrak{R} = \mathfrak{G} \cdot \mathfrak{R}_0$ be a unification, $\mathfrak{G}$ being a Levi subalgebra. $\mathfrak{G} \supset \approx\mathfrak{L}$ and $\approx \mathfrak{sl}(3, \mathbf{R})$, and therefore cannot be any real form of $\mathfrak{b}_2$ or $\mathfrak{g}_2$. But $\dim \mathfrak{G} \leq 18$ (or possibly 19) and $\mathfrak{sl}(3, \mathbf{C}) \supset \approx\mathfrak{P}$; the only remaining possibility is thus $\mathfrak{sl}(4, \mathbf{R})$. However, the inhomogeneous $\mathfrak{sl}(4, \mathbf{R})$ L.a. is not a $U(\mathfrak{P}, \mathfrak{sl}(3, \mathbf{R}), \mathfrak{u}(1))$, as we can verify easily by considering its structure (which we read immediately from Table I.a. in the Appendix and from the basic representation of that algebra, written in the Appendix). Q.E.D.

Lemma 5. We have $\mathfrak{su}(2, 2) = U(\mathfrak{P}, \mathfrak{su}(2, 1))$ and, in a similar way, $\mathfrak{so}(4, 3) = U(\mathfrak{P}, \mathfrak{g}_{2,2})$, the intersection being the (only) nilpotent noncommutative algebra of order 3 (denoted by $\mathfrak{g}_3$ in Bourbaki, Ref. 6, Chap. 4, Exercise 9).

[11] H. Freudenthal, Math. Ann. **156**, 263 (1964).

Let us take as generators of $\mathfrak{su}(2, 2)$ the $(a_k, a_k', c_2; b_k, b_k', c_1, c_3)$ with which Table II is written (cf. Appendix). $\mathfrak{su}(2, 1)$ can be represented in $\mathfrak{su}(2, 2)$ by

$$(\tfrac{1}{2}(a_k + a_k'), \tfrac{1}{2}(a_1' - a_1 + 2c_2); \tfrac{1}{2}(b_1' + c_1), \tfrac{1}{2}(b_2' + b_3), \tfrac{1}{2}(b_2 - b_3'), \tfrac{1}{2}(b_1 + c_3)), \quad (3.8)$$

as is easily seen in a suitable 4-dimensional representation of $\mathfrak{su}(2, 2)$, in which $\mathfrak{su}(2, 1)$ is represented by 3rd-order matrices. We can represent $\mathcal{P}$ in $\mathfrak{su}(2, 2)$ by

$$(a_k; b_k, p_k = (a_k' + b_k'), p_4 = (c_1 + c_2)). \quad (3.9)$$

We have also other ways (equivalent for our purpose) to represent these subalgebras in $\mathfrak{su}(2, 2)$. With $\mathcal{P}$ and $\mathfrak{su}(2, 1)$ thus represented in $\mathfrak{su}(2, 2)$, and with the aid of the above-mentioned 4-representation, one checks easily that their intersection is the only nilpotent noncommutative L.a. (isomorphic to that of lower triangular matrices, with zero diagonal, of order 3) generated by

$$\begin{aligned} \tfrac{1}{2}x_3 &= (a_1' + c_2) + (b_1' + c_1) = p_1 + p_4, \\ (a_2 + a_2') + (b_2' + b_3) &= p_2 + (a_2 + b_3) = x_1, \\ (a_3 + a_3') - (b_2 - b_3') &= p_3 + (a_3 - b_2) = x_2, \end{aligned} \quad (3.10)$$

with commutators

$$[x_1, x_2] = x_3, \qquad [x_2, x_3] = [x_1, x_3] = 0. \quad (3.11)$$

[We notice that this is the Lie algebra of the canonical commutation relations.] Since

$$(\mathfrak{L} \oplus \mathfrak{u}(1)) \cdot \mathcal{P}_0 \approx \mathrm{der}\,(\mathcal{P})$$

is a (nonsemisimple) maximal subalgebra of $\mathfrak{su}(2, 2)$ (generated by a_k, b_k, p_k, p_4, c_3), and $\mathfrak{su}(2, 1)$ is a (simple) maximal subalgebra, the described unification (up to conjugacy) is the only one; this can also be seen from the Iwasawa decompositions of $\mathfrak{su}(2, 2)$, $\mathfrak{su}(2, 1)$, and a similar decomposition of $\mathcal{P}$, decompositions in which we specify the commutative part of the nilpotent subalgebra, and the classification of all 3-dimensional (order-3) L.a. (cf. Bourbaki, Ref. 6, Chap. 6, Exercise 23), by a close study of the nature of the intersection needed to obtain the wanted unification.

The case of $\mathfrak{g}_{2,2}$ and $\mathfrak{so}(4, 3)$ is very similar; indeed, we have the decomposition:

$\mathfrak{g}_{2,2}$ = ($\mathfrak{su}(2) + \mathfrak{su}(2)$) + [(Abelian subalgebra with semisimple elements of order 2) + (nilpotent subalgebra of order 6, and with order-3 commutative part)] = $\mathfrak{su}(2, 1)$ + {subspace generated by two "compact generators", one noncompact semisimple and three nilpotent generators}.

$\mathfrak{g}_{2,2}$ is irreducible in $\mathfrak{so}(4, 3)$ (with respect, e.g., to the basic 7-dimensional representation [10]).

$\mathfrak{so}(4, 3)$ = ($\mathfrak{su}(2) + \mathfrak{su}(2) + \mathfrak{su}(2)$) + [(Abelian subalgebra with semisimple elements, of order 3) + (nilpotent subalgebra of order 9, with order-5 commutative part)]
= $\mathfrak{so}(4, 2)$ + {same subspace as above}
= $\mathfrak{su}(2, 2)$ + {same subspace}

[as $\mathfrak{su}(2, 1)$ is real-irreducible with respect to the 6-dimensional representation of $\mathfrak{so}(4, 2)$], and therefore,

$\mathfrak{so}(4, 3)$ = ($\mathcal{P}$ + $\mathfrak{su}(2, 1)$ with order-3 intersection) + {same subspace}
= ($\mathcal{P}$ + $\mathfrak{g}_{2,2}$ with order-3 intersection).

Thus $\mathfrak{so}(4, 3) = U(\mathcal{P}, \mathfrak{g}_{2,2})$ with a 3-dimensional nilpotent intersection.

Thus, we cannot have nontrivial and nonbanal $U(\mathcal{P}, \mathfrak{X})$, with simple $\mathfrak{X}$, except perhaps with 40 simple L.a., namely $\mathfrak{g}_{2,2}$ and $\mathfrak{su}(2, 1)$, and 38 L.a. containing $\mathfrak{L}$ as subalgebra:

$$\mathfrak{L} = \mathfrak{so}(3, 1) = \mathfrak{sl}(2, \mathbf{C}) = \mathfrak{sp}(1, \mathbf{C}) = \mathfrak{so}(3, \mathbf{C});$$

$$\mathfrak{so}(3, 2) = \mathfrak{sp}(2, \mathbf{R}),\ \mathfrak{so}(4, 1) = \mathfrak{sp}(1, 1);$$

$$\mathfrak{sl}(4, \mathbf{R}) = \mathfrak{so}(3, 3),\ \mathfrak{su}^*(4) = \mathfrak{so}(5, 1),$$

$$\mathfrak{su}(2, 2) = \mathfrak{so}(4, 2),\ \mathfrak{su}(3, 1) = \mathfrak{so}^*(6);$$

$$\mathfrak{sl}(3, \mathbf{C});\ \mathfrak{so}(5, \mathbf{C}) = \mathfrak{sp}(2, \mathbf{C});$$

$$\mathfrak{so}(p, q)(1 \leq q \leq p, 7 \leq p + q \leq 10)$$

[including $\mathfrak{so}^*(8) = \mathfrak{so}(6, 2)$]; $\mathfrak{sp}(n, \mathbf{R})$ ($n = 3, 4, 5$); $\mathfrak{su}(3, 2)$, $\mathfrak{su}(4, 1)$; $\mathfrak{sl}(n, \mathbf{R})$ ($5 \leq n \leq 10$); and $\mathfrak{sl}(n, \mathbf{C})$ ($n = 4, 5$). This does *not* mean that, with each one, we can have other than trivial and banal unifications $U(\mathcal{P}, \mathfrak{X})$. However, with any other simple L.a., any $U(\mathcal{P}, \mathfrak{X})$ is either $\mathcal{P} \oplus \mathfrak{X}$, or, if $\mathcal{P} \subset \mathfrak{X}$, the banal $U(\mathcal{P}, \mathfrak{X}) = \mathfrak{X} \supset \mathcal{P}$ ($\supset \mathfrak{L}$ then), or else, if $\mathfrak{X}$ is of type a_1, the banal unification $U(\mathcal{P}, \mathfrak{X}) = \mathcal{P}$ (which is of no interest for our problem).

Now let $\mathfrak{X}$ be a semisimple L.a., $\mathfrak{X} = \mathfrak{X}_1 \oplus \cdots \oplus \mathfrak{X}_k$, where $\mathfrak{X}_j$ is a simple L.a. ($j = 1, \cdots, k$), and suppose that no $\mathfrak{X}_j$ is of type a_1, nor contains $\mathfrak{L}$, nor is $\mathfrak{su}(2, 1)$ or $\mathfrak{g}_{2,2}$[$U(\mathcal{P}, \mathfrak{X}_j)$ is then trivial for all j]. Set $U(\mathcal{P}, \mathfrak{X}) = \mathfrak{R} = \mathfrak{B} \cdot \mathfrak{R}_0$ ($\mathfrak{B}$ being Levi subalgebra). Then we have $\mathfrak{B} \approx \mathfrak{X}_1' \oplus \cdots \oplus \mathfrak{X}_{k'}'$, where $\mathfrak{X}_j'$ is simple. We cannot have $k' < k$ because we would then have either

$$\mathfrak{X}_1' \supset (\mathfrak{X}_1 \oplus \mathfrak{X}_2 \text{ and } \approx \mathfrak{L})$$

or

$$\mathfrak{X}_1' \supset \mathfrak{X}_1 \oplus \mathfrak{X}_2 \quad \text{and} \quad \mathfrak{X}_2' \supset (\mathfrak{X}_3 \text{ and } \approx \mathfrak{L}),$$

if we order the $\mathfrak{X}_i$'s and $\mathfrak{X}_i'$'s in a suitable way; and both hypotheses are excluded by dimension considerations, according to Dynkin's classification of semisimple subalgebras of simple L.a..[10] Thus *either* $k' = k+1$, $\mathfrak{X}_{k+1}' \approx \mathfrak{L}$, $\mathfrak{X}_i' = \mathfrak{X}_i$ $(j = 1, \cdots, k)$ (and thus, by our Lemma 1, $U(\mathcal{P}, \mathfrak{X}) \approx \mathcal{P} \oplus \mathfrak{X}$) *or* $k' = k$, $\mathfrak{X}_j' = \mathfrak{X}_j$ $(j > 1)$ and therefore $(\mathfrak{X}_1' \cdot \mathfrak{R}_0) = U(\mathcal{P}, \mathfrak{X}_1)$ and $U(\mathcal{P}, \mathfrak{X}) = (\oplus_{i>1} \mathfrak{X}_i) \oplus U(\mathcal{P}, \mathfrak{X}_1)$ from the same lemma. In a similar way, the results for $U(\mathcal{P}, \mathfrak{X}, \mathfrak{u}(1))$ are *extended to semisimple* L.a.

If now $\boldsymbol{X}$ be a *compact Lie group*, its L.a. $\mathfrak{X}$ is reductive, i.e., $\mathfrak{X} = \mathfrak{X}_1 \oplus \mathfrak{X}_2$, where $\mathfrak{X}_1$ is maximal semisimple and $\mathfrak{X}_2$ Abelian. Now, except the case in which $\mathfrak{X}_1$ has $\mathfrak{so}(3)$ as direct factor (in that case, a banal unification is possible), $U(\mathcal{P}, \mathfrak{X}_1) \approx \mathcal{P} \oplus \mathfrak{X}_1$. If $\mathfrak{X}_1 \approx \mathfrak{X}_1' \oplus \mathfrak{so}(3)$, we may have also the unification $U(\mathcal{P}, \mathfrak{X}_1') \approx \mathcal{P} \oplus \mathfrak{X}_1'$. Therefore, a Levi subalgebra $\mathfrak{B}$ of $U(\mathcal{P}, \mathfrak{X})$ necessarily contains $\mathfrak{L} \oplus \mathfrak{X}_1$ (or possibly $\mathfrak{L} \oplus \mathfrak{X}_1'$). By virtue of dimensionality, this must be, in fact, the total Levi subalgebra; therefore, the only possible unifications are

$$U(\mathcal{P}, \mathfrak{X}) \approx U(\mathcal{P}, \mathfrak{X}_2) \oplus \mathfrak{X}_1', \tag{3.12}$$

with $\mathfrak{X}_1' = \mathfrak{X}_1$ or possibly $\mathfrak{X}_1 \approx \mathfrak{X}_1' \oplus \mathfrak{so}(3)$, and with

$$U(\mathcal{P}, \mathfrak{X}_2) \approx \mathfrak{L} \cdot U(\mathcal{P}_0, \mathfrak{X}_2'), \tag{3.13}$$

where $\mathfrak{X}_2' \subseteq \mathfrak{X}_2$ and $\dim \mathfrak{X}_2 - \dim \mathfrak{X}_2' \leq 1$. Moreover, if we distinguish between compact and noncompact generators of the Abelian subalgebra of $U(\mathcal{P}, \mathfrak{X})$ which will come from $\mathfrak{X}_2$, we can have either $\mathcal{P} \oplus \mathfrak{X}$ or $\mathcal{P} \oplus \mathfrak{X}_1 \oplus \mathfrak{X}_2'$, with $\dim \mathfrak{X}_2 - \dim \mathfrak{X}_2' = 1$—or else, if $\mathfrak{X}_1' \oplus \mathfrak{so}(3) \approx \mathfrak{X}_1$, $\mathcal{P} \oplus \mathfrak{X}_1' \oplus \mathfrak{X}_2$.

Therefore, $g^{\mu\nu} p_\mu p_\nu$ is an invariant of any $\mathfrak{R} = U(\mathcal{P}, \mathfrak{X})$ for any L.a. $\mathfrak{X}$ of a compact Lie group $\boldsymbol{X}$ if we want to recover $\boldsymbol{X}$ by passing to a Lie group of Lie algebra $\mathfrak{R}$.

We sum up our results in two theorems.

Theorem I: *Let $\mathfrak{X}$ be a real semisimple Lie algebra.*

(i) *If $\mathfrak{X} \subset \mathfrak{L}$ [strictly, i.e., $\mathfrak{so}(3)$ or $\mathfrak{sl}(2, \mathbf{R})$], any $U(\mathcal{P}, \mathfrak{X})$ is either trivial $(\mathcal{P} \oplus \mathfrak{X})$ or banal $(\mathcal{P}$, uninteresting case for our purpose).*

(ii) *If $\mathfrak{X} = \mathfrak{su}(2, 1)$ [resp. $\mathfrak{X} = \mathfrak{g}_{2,2}$], one can have $\mathfrak{su}(2, 2) = U(\mathcal{P}, \mathfrak{su}(2, 1))$ [resp. $\mathfrak{so}(4, 3) = U(\mathcal{P}, \mathfrak{g}_{2,2})$] with intersection on the nilpotent noncommutative Lie algebra of order 3.*

(iii) *If no simple ideal of $\mathfrak{X}$ has a real representation in dimension ≤ 10, every $U(\mathcal{P}, \mathfrak{X})$ is either trivial $(\mathcal{P} \oplus \mathfrak{X})$ or banal $(\mathfrak{X}$, if $\mathfrak{X} \supset \mathcal{P})$.*

(iv) *If $\mathfrak{X}$ contains no subalgebra isomorphic to $\mathfrak{L}$, nor any simple ideal of the types considered in* (i) *and* (ii) *[$\mathfrak{so}(3)$, $\mathfrak{sl}(2, \mathbf{R})$; $\mathfrak{su}(2, 1)$, $\mathfrak{g}_{2,2}$], then every $U(\mathcal{P}, \mathfrak{X})$ is trivial $(\mathcal{P} \oplus \mathfrak{X})$.*

Corollary. In order to get $U(\mathcal{P}, \mathfrak{X}) \neq \mathcal{P} \oplus \mathfrak{X}$ and $\neq \mathcal{P}$, with semisimple $\mathfrak{X}$, it is necessary, if $\varphi(\mathcal{P}) \cap \psi(\mathfrak{X})$ is semisimple (in particular, $\{0\}$), that $\mathfrak{X}$ contain a subalgebra isomorphic to $\mathfrak{L}$, and, in any case, it is necessary that $\mathfrak{X}$ be non-compact.

If $\boldsymbol{X}$ is a compact Lie group, with Lie algebra $\mathfrak{X} = \mathfrak{X}_1 \oplus \mathfrak{X}_2$ ($\mathfrak{X}_1$ s.s. L.a.), every $U(\mathcal{P}, \mathfrak{X})$ is isomorphic to $U(\mathcal{P}, \mathfrak{X}_2) \oplus \mathfrak{X}_1'$, where $\mathfrak{X}_1' \subseteq \mathfrak{X}_1$ and $\mathfrak{X}_1'$ differs (possibly) from $\mathfrak{X}_1$ only by a direct factor $\mathfrak{so}(3)$ ("put in common" with $\mathcal{P}$). Moreover, the only topologically interesting unifications are $\mathcal{P} \oplus \mathfrak{X}'$, where $\mathfrak{X}' \subseteq \mathfrak{X}$ and may differ only by a direct factor $\mathfrak{u}(1)$ or (possibly) $\mathfrak{so}(3)$.

Theorem II: *Let $\mathfrak{X}$ be a real semisimple Lie algebra, containing no subalgebra isomorphic to $\mathfrak{L}$, nor any simple ideal of the types considered in* (i), (ii), *of Theorem I. Then every $U(\mathcal{P}, \mathfrak{X}, \mathfrak{u}(1))$ is isomorphic to $\mathfrak{X} \oplus U(\mathcal{P}, \mathfrak{u}(1))$, i.e., either $\mathcal{P} \oplus \mathfrak{X}$, or $\mathcal{P} \oplus \mathfrak{X} \oplus \mathfrak{u}(1)$, or $\mathfrak{X} \oplus$ der $(\mathcal{P})$. In the cases of $\mathfrak{so}(3)$ and $\mathfrak{sl}(2, \mathbf{R})$, there are also the banal unifications $\mathcal{P}$ and $\mathcal{P} \oplus \mathfrak{u}(1)$, and the semidirect products $\mathcal{P} \cdot \mathfrak{R}_0' = \mathfrak{L} \cdot (\mathcal{P}_0 \oplus \mathfrak{R}_0')$, with a 4-dimensional commutative ideal $\mathfrak{R}_0'$. In the cases of $\mathfrak{su}(2, 1)$ and of $\mathfrak{g}_{2,2}$, we have also the same result as in Theorem I* (ii), *with (or without) a possible direct factor $\mathfrak{u}(1)$ in addition.*

To conclude this section, we first give two *examples of unifications*, the first with $\mathfrak{L}$ different from the one already shown and the second with the de Sitter algebra $\mathfrak{so}(4, 1)$ (with which we get a nontrivial correction to $g^{\mu\nu} p_\mu p_\nu$). Next, we study, *a priori*, the general form of the unifications $U(\mathcal{P}, \mathfrak{X})$, for simple $\mathfrak{X}$.

Example 2. Let $\mathfrak{R}$ be a nontrivial unification $U(\mathcal{P}, \mathfrak{L}')$ with $\mathfrak{L}' \approx \mathfrak{L}$, $\mathfrak{B}$ a Levi subalgebra of unification: $6 \leq \dim \mathfrak{B} \leq \dim \mathfrak{R} \leq 16$, and $\mathfrak{B} \supset \approx \mathcal{P}$ if $\dim \mathfrak{B} > 12$. In this last category, the only possibility is $\mathfrak{B} = \mathfrak{R} = \mathfrak{so}(4, 2)$, but the structure of this algebra (cf. Table II.a in the Appendix) shows it is not a $U(\mathcal{P}, \mathfrak{L}')$ with order-1 intersection. Therefore, $\mathfrak{B} \approx \mathfrak{L}$; indeed, $\mathfrak{B}$ of type $\mathfrak{b}_2$ (noncompact real form) is excluded, because the inhomogeneous de Sitter Lie algebras [$\mathfrak{X} \cdot \mathfrak{R}_0$, with 5-dimensional $\mathfrak{R}_0$ and $\mathfrak{X} = \mathfrak{so}(3, 2)$ or $\mathfrak{so}(4, 1)$, the defining homomorphism being a 5 dimensional real representation of $\mathfrak{X}$] are $U(\mathcal{P}, \mathfrak{X}, \mathfrak{u}(1))$ with intersection on $\mathfrak{L}$, but are not $U(\mathcal{P}, \mathfrak{L}')$. Therefore $\mathfrak{R} = (\mathfrak{L} \cdot \mathfrak{R}_0)$, and, if we do not want the banal unification $\mathcal{P}$, we must have $\mathfrak{R} = \mathfrak{L}$. $(\mathfrak{R}_0 \oplus \mathfrak{Q}_0)$

with 4 or 6-dimensional commutative $\mathfrak{Q}_0$ (commutative, because if the defining homomorphism is a representation reducible by real transformations, we do not get a unification). As 6-dimensional real representation of $\mathfrak{L}$, we examined the adjoint, which is (in a suitable basis) "$I_2 \otimes D_3 + J_2 \otimes D_3$", where $I_2 = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$, $J_2 = \left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right)$, and D_3 is the adjoint representation of $\mathfrak{so}(3)$. The above expression means that an element $x \in \mathfrak{so}(3)$ is represented by $I_2 \otimes D_3(x)$, and $Jx \in J\mathfrak{so}(3)$ [J is a complex structure of $\mathfrak{L} \approx \mathfrak{so}(3) \otimes \mathbf{C}$ (see Ref. 10)] by $J_2 \otimes D_3(x)$. We have a similar result with "$D_3 \otimes J_2 + D_3 \otimes I_2$". We notice that the representation we denoted by D_4 is obtained in a similar way by considering $\mathfrak{L} \approx \mathfrak{sl}(2, \mathbf{R}) \otimes \mathbf{C}$ from a basic 2-dimensional real representation of $\mathfrak{sl}(2, \mathbf{R})$. Now, if we take $\mathfrak{Q}_0$ 4-dimensional, we cannot take the defining homomorphism of $(\mathfrak{L} \cdot \mathfrak{R}_0)$ as $D_N \oplus D_N$ because a computation shows that, in that case, the intersection is at least 3-dimensional, and we have thus only $U(\mathfrak{P}, \mathfrak{L}', \mathfrak{u}(1))$. But we can take $D_N \oplus D_4$ and

$$\varphi(\mathfrak{P}) = \mathfrak{L} \cdot \mathfrak{P}_0, \qquad \psi(\mathfrak{L}') = (I + \operatorname{ad} q_0)\mathfrak{L}$$

for $q_0 \in \mathfrak{Q}_0$. A not too difficult computation shows then that we have intersection on

$$\sum_k (\lambda_k a_k + \mu_k b_k) \in \mathfrak{L}$$

if a system of 4 homogeneous linear equations in the 6 (λ_k, μ_k) is verified, system which has, in general (for nonsingular q_0), a solution in, e.g., λ_k, μ_1 for fixed μ_2, μ_3 (the solution being never such that all the λ_k's or all the μ_k's are simultaneously 0), and that if (λ_k, μ_k) is a solution, $(\mu_k, -\lambda_k)$ also. Therefore, we have, in general, a $U(\mathfrak{P}, \mathfrak{L}')$ with a 2-dimensional commutative intersection.

Remark. Given $U(\mathfrak{P}, \mathfrak{L}')$, we have

$$U(\mathfrak{P}, \mathfrak{L}_1' \oplus \mathfrak{L}_2') = \mathfrak{L}_1' \oplus U(\mathfrak{P}, \mathfrak{L}_2') \qquad (3.14)$$

($\mathfrak{L}_1'$, $\mathfrak{L}_2'$ isomorphic to $\mathfrak{L}$). Thus we have all the possible unifications of $\mathfrak{P}$ and $\mathfrak{L}_1' \oplus \mathfrak{L}_2'$.

Example 3. Now, let $\mathfrak{X} = \mathfrak{so}(4, 1)$, take as basis the generators $(a_k, a_k'; b_k, c_1)$ of $\mathfrak{so}(4, 2)$, and form the semi-direct product $(\mathfrak{X} \cdot \mathfrak{R}_0)$, with 10-dimensional $\mathfrak{R}_0$, by the adjoint representation (in that basis). This representation splits on $\mathfrak{L} = (a_k; b_k)$ in $D_N \oplus \mathrm{ad}_{\mathfrak{L}}$; let us take as basis of $\mathfrak{R}_0$ the elements (p_k, p_4, q_k, q_k'), and let q_0 be a general linear combination of the q_k's and q_k''s. We may take

$$\psi(\mathfrak{X}) = \mathfrak{X}, \qquad \varphi(\mathfrak{P}) = (I + \operatorname{ad} q_0)(\mathfrak{L} \cdot \mathfrak{P}_0);$$

$\mathfrak{R}_0$ being commutative, one sees that, as a consequence of our treatment of the first example of

$$U(\mathfrak{P}, \mathfrak{L}, \mathfrak{u}(1) \oplus \mathfrak{u}(1)) = \mathfrak{L} \cdot \mathfrak{R}_0$$

(with the same $\mathfrak{R}_0$ as here), we have

$$\mathfrak{R} = \varphi(\mathfrak{P}) + \psi(\mathfrak{X}) + \psi'(\mathfrak{u}(1) \oplus \mathfrak{u}(1)).$$

Moreover, the adjoint representation of $\mathfrak{so}(4, 1)$ being "orthogonal" [more exactly, in $\mathfrak{so}(6, 4)$ realized by its basic representation], we have an invariant

$$C = g^{\mu\nu} p_\mu p_\nu + \sum_k (q_k^2 - q_k'^2) \qquad (3.15)$$

of that unification, and thus a nontrivial "correction" to $-g^{\mu\nu} p_\mu p_\nu$ (we choose $g_{kk} = \delta_{kk}$, $g_{44} = -1$).

Types of Unifications

Now let $\mathfrak{X}$ be a *simple* L.a. and set $U(\mathfrak{P}, \mathfrak{X}) = \mathfrak{R} = (\mathfrak{B} \cdot \mathfrak{R}_0)$, where $\mathfrak{R}_0$ is the (solvable) radical; we have $\mathfrak{R} = \psi(\mathfrak{X}) + \varphi(\mathfrak{P})$. There are then two possibilities: either $\mathfrak{B} = \psi(\mathfrak{X})$, or $\mathfrak{B} \supset \psi(\mathfrak{X})$ strictly, if $\mathfrak{B}$ is a Levi subalgebra containing $\psi(\mathfrak{X})$. We suppose $\mathfrak{R}$ nontrivial ($\neq \mathfrak{P} \oplus \mathfrak{X}$).

If, in the first case, moreover, $\mathfrak{X} \supset \approx \mathfrak{P}$, we have $\mathfrak{R} \approx (\mathfrak{X} \cdot \mathfrak{R}_0)$, where the semidirect product must clearly be nondirect. From dimension considerations (on the possible intersection of the supposed unification), $\mathfrak{R}_0$ must be either $\{0\}$ (banal case) or solvable and 6–10-dimensional. If $\mathfrak{X} \supsetneq \approx \mathfrak{P}$, we have $\mathfrak{R}_0 \supset \mathfrak{P}_0$, the semidirect (nondirect) product being defined by a representation containing D_N on a subalgebra $\approx \mathfrak{L}$ of $\mathfrak{X}$ ($\mathfrak{X} \supset \approx \mathfrak{L}$ for $\mathfrak{X}$ is here a Levi subalgebra of the unification).

Now if $\mathfrak{R} = (\mathfrak{B} \cdot \mathfrak{R}_0)$ with $\mathfrak{B} \supset \mathfrak{X}$ (strictly), and if $\mathfrak{B} \supsetneq \approx \mathfrak{P}$, $\mathfrak{R}_0 \supset \mathfrak{P}_0$, and thus $\dim \mathfrak{B} - \dim \mathfrak{X} \leq 6$; then the only possibility is [by dimensional considerations of the algebras, or of their representations, and by the structure of the inhomogeneous $\mathfrak{sl}(4, \mathbf{R})$ algebra, $\mathfrak{sl}(4, \mathbf{R}) \cdot \mathfrak{P}_0$ with 4-dimensional commutative $\mathfrak{P}_0$] the banal unification $\mathfrak{P}$, for $\mathfrak{X}$ of type $\mathfrak{a}_1$. If $\mathfrak{B} \supset \approx \mathfrak{P}$, $\dim \mathfrak{R}_0 \geq 5$ or $=0$, and

$$10 \geq \dim \mathfrak{R}_0 + (\dim \mathfrak{B} - \dim \mathfrak{X}) \geq 6,$$

the dimensions of subalgebras of $\mathfrak{P}$ (an intersection of two L.a. is a subalgebra of both), and the classification of simple L.a. (and of the dimensions of their basic representations) show that $\mathfrak{R}_0 = \{0\}$.

Therefore, there are only *three possible types of nontrivial and nonbanal* $U(\mathfrak{P}, \mathfrak{X}) = \mathfrak{R}$:

(*UI*) $\mathfrak{R} \approx (\mathfrak{X} \cdot \mathfrak{R}_0)$, $\mathfrak{X} \supset \approx \mathfrak{P}$, $\mathfrak{R}_0$ solvable of dimension 6–10.

(*UII*) $\mathfrak{R} \approx (\mathfrak{X} \cdot \mathfrak{R}_0)$, $\mathfrak{X} \supsetneq \approx \mathfrak{P}$, $\mathfrak{R}_0 \supset \mathfrak{P}_0$, the defining homomorphism giving D_N on a subalgebra $\approx \mathfrak{L}$ of $\mathfrak{X}$.

(*UIII*) $\mathfrak{R}$ simple $\supset \mathfrak{X}$ (strictly)—and of course then $\mathfrak{R} \supset \approx \mathfrak{P}$.

A priori (we must still verify in most cases that we have indeed a unification, and this may very well not be), (*UI*) can be the case of $\mathfrak{sl}(5, \mathbf{R})$ with a 10-dimensional representation; of $\mathfrak{sl}(n, \mathbf{R})$ $(6 \leq n \leq 10)$ with a basic representation; of $\mathfrak{sl}(n, \mathbf{C})$ $(n = 4, 5)$, $\mathfrak{sp}(5, \mathbf{R})$, $\mathfrak{so}(p, q)$ with $p \geq 4$, $q \geq 2$, $p + q \leq 10$; and $\mathfrak{su}(3, 2)$, with basic real representations. (*UII*) is surely the case of $\mathfrak{L}$, $\mathfrak{sl}(4, \mathbf{R})$, and perhaps other algebras. (*UIII*) is the case of the "exceptional" unifications, for $\mathfrak{su}(2, 1)$ and $\mathfrak{g}_{2,2}$, and perhaps of $\mathfrak{so}(p, q) = \mathfrak{X}$ [with $\mathfrak{R} = \mathfrak{so}(p + 1, q)$ or $\mathfrak{so}(p, q + 1)$, $p \geq q$, $p + q \leq 10$, and $p + 1 \geq 4$, $q \geq 2$ or $p \geq 4$, $q + 1 \geq 2$] or $\mathfrak{sl}(4, \mathbf{R}) = \mathfrak{X}$ [with $\mathfrak{R} = \mathfrak{sl}(5, \mathbf{R})$].

Note. This work was written before L. O'Raifeartaigh's structural considerations[12] were published. One may notice that these (however unwieldy their development) are essentially parallel to the study of types of unifications we made here [both being based on the Levi–Malcev theorem and the fact that $\mathfrak{P} = \mathfrak{L} \cdot \mathfrak{P}_0$, wherefrom if $\mathfrak{R} = (\mathfrak{B} \cdot \mathfrak{R}_0)$ is any L.a. containing $\mathfrak{P}$, either there is a Levi factor $\mathfrak{B}$ containing $\mathfrak{P}$, or $\mathfrak{P}_0 \subset \mathfrak{R}_0$]. As to the problem of the spectrum of the operator M representing $-g^{\mu\nu}p_\mu p_\nu$ in some representation of a Lie algebra $\mathfrak{R} \supset \mathfrak{P}$ in which it is self-adjoint, the claimed result of O'Raifeartaigh is not true,[13] though it must be said that it did lead to a public clarification of the question. In fact, a much weaker result could be proved by Jost[14]—and the (final version of the) proof shows that all the assumptions are needed in an essential way—namely that for irreducible unitary (continuous) representations of a connected (finite-order) Lie group $\mathbf{G} \supset \mathbf{P}$, the spectrum of M is a connected set (thus, while an isolated eigenvalue is the whole spectrum, eigenvalues with "continuous background" are not excluded). This does not hold as far as only Lie algebras are concerned (and we dealt mainly with Lie algebras in this paper), and counter-examples can be given.[13]

4. APPLICATION TO A MASS FORMULA

We have seen that, except for two "exceptional" cases, any interesting semisimple internal L.a. must contain $\approx \mathfrak{L}$. But $\mathfrak{L}$ itself is not enough, since we have only type (*UII*) unifications (as is easy to check), and then, the ideal being commutative, $g^{\mu\nu}p_\mu p_\nu$ is an invariant. On the other hand, we have some reasons to think[4] that the internal L.a. is a subalgebra of $\mathfrak{sl}(4, \mathbf{C})$. Now, the (phenomenological) classification of elementary particles according to some representations of the type-$\mathfrak{a}_2$ complex algebra [the compact real form of which is $\mathfrak{su}(3)$] seems until now, at least for baryons, in good accordance with the experimental data. For these (and other) reasons, we are particularly interested [besides the exceptional unification $\mathfrak{so}(4, 2) = U(\mathfrak{P}, \mathfrak{su}(2, 1))$, the implications of which we deal with more details elsewhere (our interest is this unification arose first during informal discussions with Dr. G. Rideau)], in the possible unification of $\mathfrak{P}$ with the algebra $\mathfrak{sl}(3, \mathbf{C}) = \mathfrak{su}(3) \otimes \mathbf{C}$ [which contains $\mathfrak{L}$ and is contained in $\mathfrak{sl}(4, \mathbf{C})$].

Let us therefore look for the unifications $U(\mathfrak{P}, \mathfrak{X})$ and $U(\mathfrak{P}, \mathfrak{X}, \mathfrak{u}(1))$, for $\mathfrak{X} \approx \mathfrak{sl}(3, \mathbf{C})$. If $\mathfrak{R}$ denotes a unification, we write, as usual, its Levi decomposition $\mathfrak{R} = \mathfrak{B} \cdot \mathfrak{R}_0$, where $\mathfrak{R}_0$ is the radical and $\mathfrak{B}$ a Levi subalgebra. If $\mathfrak{R}$ is not isomorphic to $\mathfrak{X}$ (which is of complex type $\mathfrak{a}_2 \oplus \mathfrak{a}_2$), we have at least $\mathfrak{R}$ of types $\mathfrak{a}_5$ or $\mathfrak{a}_3 \oplus \mathfrak{a}_3$, the dimension of which is greater than 27; since by extension of the base field, a unification remains a unification, we must have $\mathfrak{R} = \mathfrak{X} \cdot \mathfrak{R}_0$, where the semi-direct product gives $\mathfrak{P}$ with a subalgebra $\approx \mathfrak{L}$ in $\mathfrak{X}$ and a commutative subalgebra $\mathfrak{P}_0 \subset \mathfrak{R}_0$. Now, $\mathfrak{R}_0$ is of order ≤ 11; because of the dimensions of the real irreducible representations of $\mathfrak{L}$ in dimensions ≤ 7, and of the irreducible representations of $\mathfrak{X}$ in dimensions ≤ 11, we may *a priori* have $\mathfrak{R}_0$ of orders 8, 9, 10. Thus, our first step is to look for a real 8, 9, or 10, dimensional representation of $\mathfrak{X} = \mathfrak{sl}(3, \mathbf{C})$ containing D_N on a subalgebra $\approx \mathfrak{L}$. Then, we must prove that we indeed get a unification [and, in the case we consider, we see that this is a mere consequence of our study of $U(\mathfrak{P}, \mathfrak{L})$]. Once we have a suitable unification, we try to find a "correction" to $g^{\mu\nu}p_\mu p_\nu$, and then express it in terms of internal quantum numbers.

As we do not know the exact way in which the representations of $\mathfrak{X}$ split on the various subalgebras $\approx \mathfrak{L}(\mathfrak{so}(3, \mathbf{C})$ and $\mathfrak{sl}(2, \mathbf{C}))$, we start from some reduced representation of a certain subalgebra $\mathfrak{L}_e \subset \mathfrak{X}$, and try and build therefrom a representation of $\mathfrak{X}$, that is to say find 10 more matrices, representing generators of a supplementary subspace. Let us first consider the 10-dimensional representation $D_N \oplus \mathrm{ad}\ \mathfrak{L}$, on $\mathfrak{L}_e = \mathfrak{so}(3, \mathbf{C}) = (a_k; b_k)$ (in the basis of Table III.a—cf. Appendix). But if we then try to determinate the coefficients of the represent-

[12] L. O'Raifeartaigh, Phys. Rev. **139**, B1052 (1965); see also Phys. Rev. Letters **14**, 332, 575 (1965).
[13] M. Flato and D. Sternheimer, Phys. Rev. Letters **15**, 934 (1965); **16**, 1185 (1966).
[14] R. Jost [private communications; the final version is published in Helv. Phys. Acta. **39**, 369 (1966)].

ing matrices of $\mathfrak{L}_3 \approx \mathfrak{sl}(2, \mathbf{C}) \subset \mathfrak{X}$ (for instance), we arrive at contradictory conditions, and the same thing occurs if, instead of ad $\mathfrak{L}$, we take another 6-dimensional real representation of $\mathfrak{L}$, obtained from the canonical 3-dimensional complex realization as $\mathfrak{so}(3, \mathbf{C})$ by doubling the dimension. Again, the same thing occurs if we take $\mathfrak{L}_s = \mathfrak{L}_3$ [e.g., see the Appendix for this notation of a $\mathfrak{sl}(2, \mathbf{C})$ subalgebra of $\mathfrak{sl}(3, \mathbf{C})$] and try to extend it to $(a_k; b_k) = \mathfrak{so}(3, \mathbf{C}) \subset \mathfrak{X}$. If one seeks an eight-dimensional representation of $\mathfrak{X}$, decomposed on a subalgebra $\mathfrak{L}_s \approx \mathfrak{L}$ in $D_N \oplus D_4$, one arrives at the same impossibility, whether one takes $\mathfrak{L}_s = (a_k; b_k)$ or $\mathfrak{L}_s = \mathfrak{L}_3$ (or, which is equivalent, $\mathfrak{L}_1$ or $\mathfrak{L}_2$).

Our problem is mainly to get a real-irreducible (irreducible by real transformations) representation (with real coefficients, of course) of $\mathfrak{sl}(3, \mathbf{C}) = \mathfrak{X}$, and then to see if, on some $\mathfrak{L}_s \subset \mathfrak{X}$, it splits so that we get $D_N \oplus$ (another real representation of $\mathfrak{L}_s$). We do know that the D_N representation of $\mathfrak{L} \approx \mathfrak{sl}(2, \mathbf{C})$ is the Kronecker product of a (basic) 2-dimensional representation by its contragredient, and this suggests the consideration of the Kronecker product of a (basic) 3-dimensional representation of $\mathfrak{X}$ by its contragredient, which will contain D_N on a $\mathfrak{sl}(2, \mathbf{C})$ subalgebra of $\mathfrak{X}$ (i.e., a subalgebra like $\mathfrak{L}_1$, $\mathfrak{L}_3$, or $\mathfrak{L}_2$), and be real. If, moreover, among all complex-equivalent representations, we can select one that splits on a $\mathfrak{L} \subset \mathfrak{X}$ in a way so as to obtain a unification, our first (and most unwieldy) step is accomplished.

We therefore try and build a 9-dimensional representation of $\mathfrak{X}$, which, on $\mathfrak{L}_3$, splits into $D_N \oplus (-\tilde{D}_4) \oplus O$, where $-\tilde{D}_4$ is also a representation of $\mathfrak{L}$ (the $\sim$ standing for transposition), and O is the trivial representation in one dimension. For commodity in the calculation (and also in order to check the 10-dimensional case from another point of view), we try and prolong the 10-dimensional representation $D_N \oplus (-\tilde{D}_4) \oplus O \oplus O$ of $\mathfrak{L}_3$ to all $\mathfrak{X}$. In this way, we obtain a $U(\mathfrak{P}, \mathfrak{X}, \mathfrak{u}(1))$.

We take as representing matrices of the generators $(\frac{1}{2}c_3', \cdots, \frac{1}{2}b_3)$ of $\mathfrak{L}_3$ (cf. Appendix), those obtained in representing the elements $(a_1, \cdots, b_3)$ of $\mathfrak{L}$ by D_N and $(-\tilde{D}_4)$; we need 10 more basis elements. Let us then set

$$a_1 = \sum_{\alpha,\beta} \lambda_{\alpha\beta} E_{\alpha\beta} (\alpha, \beta = 1, 2, \cdots, 9, 0) \qquad (4.1)$$

($E_{\alpha\beta}$ is the canonical basis of the $\mathfrak{gl}$'s). We have

$$\tfrac{1}{2}a_3 = E_{21} - E_{12} + \tfrac{1}{2}(E_{65} - E_{56} + E_{87} - E_{78}).$$

If we write down the conditions given by $[a_3, a_1] = a_2$, and then by $[a_2, a_3] = a_1$, we get a_1 and a_2 with only 40 coefficients λ (instead of 100), in 10 sets of 4. Now we put

$$b_1 = \sum \lambda'_{\alpha\beta} E_{\alpha\beta}, \qquad (4.2)$$

and express $[a_3, b_1] = b_2$, $[b_2, a_3] = b_1$, to get b_1, b_2 with 40 coefficients λ' (same expression as above with the λ's). Moreover, $[a_1, b_3] = -b_2$, whence the λ''s in terms of the λ's. We then write $[a_1, a_2] = a_3$, from where we obtain relations in the λ's, and $[b_1, b_2] = -a_3$, whence (opposite) relations in the λ''s, with which we can make those in the λ's more precise. We further write down the relations between the λ's obtained from $[a_1, b_1] = 0$, $[b_1, a_2] = b_3$. As we construct a matrix Lie algebra, the Jacobi identity is automatically verified; using it, we find that all the commutation relations of the algebra $\mathfrak{L}_0 = \mathfrak{so}(3, \mathbf{C}) = (a_k; b_k)$ are verified. The conditions on the λ's we get are too long to be given here, even after simplification by use of vectorial considerations. We give them a much simpler form by supposing (this seems to be necessary, from the general look of the conditions) that the representing matrix of a_1 (a "compact generator", i.e., generating a compact one-parameter subgroup by the exponential mapping, in the given 3-dimensional realization, for instance) is skew-symmetric (therefore, that of a_2 is also a skew-symmetric matrix). We can now simplify the conditions by considering the commutators

$$b_2' = [a_1, b_3'] = [b_2, c_3'] = [b_3'', b_1],$$

and also by expressing them in a vectorial form. We obtain (for $a_1 = \sum \lambda_{\alpha\beta} E_{\alpha\beta}$ as above)

$$\lambda_{\alpha\beta} = -\lambda_{\beta\alpha}, \qquad \lambda_{26} = \lambda_{15}$$
$$= \cos\theta \cos\varphi = \lambda_{45} = \lambda_{37} = k'\lambda_{50} = kk'\lambda_{59},$$

where $k'^2 = \frac{1}{2}(1 + k^2)$ (k is const $\neq 0$), $0 \leq \theta \leq \frac{1}{2}\pi$, $0 \leq \varphi < 2\pi$, and 3 similar sets of equalities, the other coefficients being null. If now we write $[c_1', b_1''] = 2a_1$, we get $k^2 = 1$ and $\sin 2\varphi = 0$, or $\sin 2\varphi = 0$ and $\sin 2\theta = 0$. We now verify that (if $k^2 = 1$ in both cases) $[a_2, b_2''] = 2c_2'$, and $[a_2, b_2'] = 2c_2$, and that we get the desired results with the commutators of c_1' with b_1'', a_1, b_2''; of c_1 with a_1, b_1', b_1''; of c_2' with b_2'', a_2; of c_2 with a_2, b_2', b_2''; and of c_3 and c_3' with a_1, b_1, b_1', b_1'', a_2, b_2, b_2', b_2''. By use of the Jacobi identity, we check that all other commutation relations are verified (of course, one uses also the antisymmetry property of commutators).

All the matrices representing elements of $\mathfrak{X}$ are then of the form

$$\begin{bmatrix} A & kk'\alpha & k'\alpha \\ kk'\beta & h & kh \\ k'\beta & kh & h \end{bmatrix},$$

where A is an 8×8 matrix, $k^2 = k'^2 = 1$, α is an 8-component column vector and β an 8-component row vector, the representing matrices of the base elements being either symmetric or skew-symmetric. The matrix

$$V = \begin{bmatrix} I_8 & 0 \\ 0 & \frac{1}{\sqrt{2}} \begin{pmatrix} kk' & k' \\ k' & -kk' \end{pmatrix} \end{bmatrix} = V^{-1}$$

brings them to the form

$$\begin{bmatrix} A & \sqrt{2}\alpha & 0 \\ \sqrt{2}\beta & 2h & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and thus reduces the obtained representation to $D_9 \oplus O$, where D_9 is a *9-dimensional real* representation of $\mathfrak{X}$ (I_n denotes here the order-n unit matrix). This is an *irreducible representation*, because (after reduction) $2c_1 + c_3$ is represented by

$$A_1 = \begin{bmatrix} -2I_4 & 0 & 0 \\ 0 & I_4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \cos 2\varphi$$

(with $\cos^2 2\varphi = 1$), thus by a regular matrix, that cannot be reduced into a direct sum of an 8-dimensional matrix $\oplus$ 0 by a similitude $A_1 \to V_1^{-1} A_1 V_1$, (V_1 being the regular 9-dimensional matrix). However, D_9 is reducible when restricted to the maximal compact subalgebra $\mathfrak{su}(3)$ generated by (a_k, b'_k, c'_1, c'_3) $(k = 1, 2, 3)$, and only into an 8-dimensional real representation (equivalent to the adjoint) $\oplus$ O (O is the trivial 1-dimensional representation) as can be seen by an argument similar to the one we just used (the representing matrices of the above-mentioned base elements have a 9th row and column which is $\sqrt{2}$ times the 4th). Actually, we obtained here two (irreducible) representations D_9 and D'_9 (since θ is defined modulo π only, there are 2 possible choices of φ such that $\sin 2\varphi = 0$), and these coincide with $D_N \oplus (-\tilde{D}_4) \oplus O$ on $\mathfrak{L}_3$ and are related by $D_9(\mathfrak{L}_1) = D'_9(\mathfrak{L}_2)$ and $D_9(\mathfrak{L}_2) = -D'_9(\mathfrak{L}_1)$. Thus, they are essentially the same representation. In the Appendix, we have given D_9. More precisely (Table IV), we have given the commutators between $\mathfrak{X}$ and the ideal $\mathfrak{R}_0$ in the semidirect product $\mathfrak{R} = \mathfrak{X} \cdot \mathfrak{R}_0$ of defining homomorphism D_9, from where it is easy to get D_9. We have taken a basis (p_μ) $(\mu = 1, 2, 3, 4)$ of $\mathcal{P}_0 \subset \mathfrak{R}_0$, (q_μ) of $\mathfrak{Q}_0 \subset \mathfrak{R}_0$, and $r_0 = \sqrt{2} r'_0$ of a supplementary subspace; we see that, e.g.,

$$[a_1, p_1] = -q_1 \cos\theta - q_4 \sin\theta.$$

As we take the basis of $\mathfrak{R}_0$ in the ordering $(p_1, \cdots, p_4; q_1, \cdots, q_4; r_0)$, this shows that the matrix representing a_1 has a first column with 0 coefficients, except $-\cos\theta$ in the 5th row and $-\sin\theta$ in the 8th. Similarly, the relation

$$[a_1, q_4] = \sin\theta(p_1 + p_4 - r_0)$$

shows its 8th column is 0 except for $\sin\theta$ in the 1st and 4th rows and $-\sqrt{2}\sin\theta$ in the 9th. Here, we have a representation (and therefore a unification) depending on a parameter $\theta (0 \leq \theta \leq \frac{1}{2}\pi, \text{ mod } \pi)$, although we do not need this fact in the following.

It is then easy to check that *we have a unification*

$$\mathfrak{R} = U(\mathcal{P}, \mathfrak{X}, \mathfrak{u}(1)), \tag{4.3}$$

with a (commutative) 2-dimensional intersection between $\mathcal{P}$ and $\mathfrak{X}$. First (and this is because of the irreducibility of D_9) one can check, by an elementary computation [looking for commutation relations of $\mathfrak{R}_0$ such that $D_9(\mathfrak{X}) \subset \text{der}(\mathfrak{R}_0)$, i.e., $D_9(x) \in \text{der}(\mathfrak{R}_0)$. $\forall x \in \mathfrak{X}$; we express this on a basis] that $\mathfrak{R}_0$ must be commutative (and not only solvable). Next, we consider the algebra

$$\mathfrak{L}'_3 = (I + \text{ad } q_0)\mathfrak{L}_3,$$

where q_0 is a linear (real) combination of the q_μ's. For instance, we may take $q_0 = 2\alpha q_4$, where α is a (real) constant. Since $\mathfrak{R}_0$ is commutative, $[\mathfrak{L}'_3, \mathcal{P}_0] = [\mathfrak{L}_3, \mathcal{P}_0]$, i.e., $[l'_3, p_0] = [l_3, p_0]$ $\forall p_0 \in \mathcal{P}_0$, $l_3 \in \mathfrak{L}_3$, $l'_3 = (I + \text{ad } q_0) l_3$. Thus, we have

$$\varphi(\mathcal{P}) = (I + \text{ad } q_0)(\mathfrak{L}_3 \cdot \mathcal{P}_0) \approx \mathcal{P}. \tag{4.4}$$

But, one checks, exactly in the same way as for $U(\mathcal{P}, \mathfrak{L})$ with $D_N \oplus D_4$, that, in general (and, in particular, for $q_0 = 2\alpha q_4$), $\mathfrak{L}'_3 \cap \mathfrak{L}_3$ is a commutative 2-dimensional subalgebra. Therefore, $\mathfrak{X} \cap \varphi(\mathcal{P})$ is 2-dimensional (for "nondegenerate" q_0's), and $\mathfrak{X} \cdot \mathfrak{R}_0 = \mathfrak{R}$ is a unification $U(\mathcal{P}, \mathfrak{X}, \mathfrak{u}(1))$, the added $\mathfrak{u}(1)$ being generated by r_0, for instance.

Let us now pass to the second step, the search of a suitable *correction to* $g^{\mu\nu} p_\mu p_\nu$ in the above unification. $g^{\mu\nu} p_\mu p_\nu$ is a second-degree term in the enveloping algebra of the unification $\mathfrak{R}$; we would like to find an element in the center of that enveloping algebra which contains it. But the representation D_9 of $\mathfrak{sl}(3, \mathbf{C}) = \mathfrak{X}$ is not "orthogonal", i.e., $D_9(\mathfrak{X})$ is not contained in a 9-dimensional realization of a

real form of the complex algebra $\mathfrak{b}_4$. We know that[10] $\mathfrak{a}_2 \oplus \mathfrak{a}_2$ is a maximal regular semisimple algebra of $\mathfrak{b}_6$, and is also maximal (nonregular, nonsimple) in $\mathfrak{a}_8$ [as is $SL(3) \times SL(3)$ in $SL(9)$], but is not contained in $\mathfrak{b}_4$; this is *a fortiori* true when passing to real forms. As we see later (and can guess from Dynkin's remarks on invariants[10]), this is a source of difficulties, since we are looking for a second-degree invariant in the enveloping algebra of $\mathfrak{R}$.

Indeed, let

$$x, y, z \in \mathfrak{X}, \gamma_0, g_0, h_0 \in \mathfrak{R}_0,$$

and set

$$\gamma_0' = [\gamma_0, y], \quad \gamma_0'' = [\gamma_0, z], \quad g_0' = [x, g_0],$$
$$y' = [x, y], \quad h_0' = [x, h_0], \quad \text{and} \quad z' = [x, z].$$

Since $\mathfrak{R}_0$ is a commutative ideal in $\mathfrak{R}$, one sees immediately that

$$[x, (y + g_0)^2] = 2(x + g_0)y' + g_0'(y + g_0) + [y', g_0] + [y, g_0'] + [y', x],$$
$$[\gamma_0, (y + g_0)^2] = 2\gamma_0'(y + g_0) + [y, \gamma_0'],$$
$$[x, (y + g_0)(z + h_0)] = (y' + g_0')(z + h_0) + (y + g_0)(z' + h_0'),$$
$$[\gamma_0, (y + g_0)(z + h_0)] = (\gamma' z_0 + y\gamma_0'') + (\gamma_0' h_0 + g_0\gamma_0'').$$

[It is also possible to make these considerations more precise by writing

$$\mathfrak{R}_0 = \mathfrak{P}_0 + \mathfrak{Q}_0 + (r_0)$$

and

$$\mathfrak{X} = \mathfrak{su}(3) + J\mathfrak{su}(3) = (a_k, b_k'', c_1', c_3') + (b_k, b_k', c_1, c_3),$$

but we need no such considerations.]

If now we are looking for a second-degree expression

$$\sum (y + g_0)^2 + \sum (y + g_0)(z + h_0) \tag{4.5}$$

commuting with all $\mathfrak{X}$, we must have, for all $x \in \mathfrak{X}$,

$$[x, \sum y^2 + \sum yz] = 0, \tag{4.6}$$

and therefore $\sum y^2 + \sum yz$ will have to be a Casimir operator of $\mathfrak{X}$. If, moreover, we want an expression commuting with all $\mathfrak{R}$, it is not possible to take all the g_0's and h_0's as zero, and then it is necessary that

$$\sum \sum g_0^2 + \sum g_0 h_0] = 0 \tag{4.7}$$

(in any case, this is needed if we are looking for an invariant really containing $g^{\mu\nu}p_\mu p_\nu$), and it is sufficient if we take every y and z as zero.

So, we must find a second-degree expression in the elements of $\mathfrak{R}_0$ containing $-m^2 = g^{\mu\nu}p_\mu p_\nu$ and (if possible) commuting with all $\mathfrak{X}$. We know that $[m^2, \mathfrak{L}_3] = 0$, and see further that $[c_1', m^2] = 0$. We now write out a big table (not reproduced here) giving the commutators between the 16 base elements of $\mathfrak{X}$ and all 45 two-by-two products of base elements of $\mathfrak{R}_0$ (i.e., p_μ^2, q_μ^2, r_0^2, $p_\mu p_\nu$ and $q_\mu q_\nu$ for $\mu \neq \nu$, $p_\mu r_0$, $p_\mu q_\nu$, $q_\mu r_0$). In this table, we see that

$$\tfrac{1}{2}[a_1, g^{\mu\nu}p_\mu p_\nu + \delta^{\mu\nu}q_\mu q_\nu] = -q_1 \cos\theta(r_0 - 2p_4) - q_4 \sin\theta(r_0 - 2p_1) = [a_1, 2p_4 r_0]. \tag{4.8}$$

In the same way, we see that

$$g^{\mu\nu}p_\mu p_\nu + \delta^{\mu\nu}q_\mu q_\nu - 2p_4 r_0$$

commutes with all the generators of the maximal compact subalgebra of $\mathfrak{X}$, $\mathfrak{su}(3) = (a_k, b_k'', c_1', c_3')$, but not with the other generators of $\mathfrak{X}$. Moreover, it is enough to look at the commutators of $\mathfrak{L}_3$ and of a_1 with the 2-by-2 above-mentioned products in order to be convinced that no second-degree expression in $\mathfrak{R}_0$ can commute with all $\mathfrak{X}$. Therefore there exists, in the enveloping algebra of $\mathfrak{R}$, no (global) invariant of degree 2. On the other hand, the expressions

$$C_1 = p_4 + \tfrac{1}{2}r_0 \tag{4.9}$$

and

$$C_2 = p_1^2 + p_2^2 + p_3^2 + \tfrac{1}{3}(p_4 - r_0)^2 + q_1^2 + q_2^2 + q_3^2 + q_4^2 \tag{4.10}$$

and therefore also

$$-C = C_2 - \tfrac{4}{3}C_1^2 = \mathbf{p}^2 - p_4^2 + \mathbf{q}^2 + q_4^2 - 2p_4 r_0 \tag{4.11}$$

(where $\mathbf{p}^2 = \sum_1^3 p_k^2$, $\mathbf{q}^2 = \sum_1^3 q_k^2$) commute with the $\mathfrak{su}(3) = (a_k, b_k'', c_1', c_3')$ subalgebra, and of course also with the ideal $\mathfrak{R}_0$. An inspection of C_1 and C_2 shows (once more) that D_9 is reducible on $\mathfrak{su}(3)$ into an 8-dimensional ("orthogonal") representation, $\oplus O$. We have

$$2p_4 r_0 = 2C_1 r_0 - r_0^2.$$

Thus

$$m^2 = -g^{\mu\nu}p_\mu p_\nu = C - 2C_1 r_0 + \mathbf{q}^2 + q_4^2 + r_0^2, \tag{4.12}$$

where C and C_1 commute with $\mathfrak{su}(3)$ and $\mathfrak{R}_0$, and therefore with the subalgebra $\mathfrak{su}(3)\cdot\mathfrak{R}_0$ of $\mathfrak{R}$.

Remark. By using the inclusion $\mathfrak{a}_2 \oplus \mathfrak{a}_2 \subset \mathfrak{b}_6$, and the already mentioned fact concerning the inclusion $\mathfrak{a}_2 \subset \mathfrak{a}_3 = \mathfrak{b}_3$ (and its consequence on 6-dimensional representation of $\mathfrak{a}_2$), we can get a 12-dimensional "orthogonal" representation of $\mathfrak{X}$, which is real-irreducible—because it can be obtained from a 6-dimensional representation of $\mathfrak{su}(3)$ by doubling the dimension of the (complexified) 6-dimensional representation of $\mathfrak{X}$. With the corresponding inhomogeneous algebra, we have a quadratic invariant, but in order for it to give us what we want, it is necessary for there to be, among all the equivalent representations, one that, on a certain $\mathfrak{L}_s \subset \mathfrak{X}$, splits into $D_N \oplus$ (something else); and even then, we have to add at least two supplementary $\mathfrak{u}(1)$ in order to get a unification. In our case, we can also have a global invariant[10] but it will be of higher order (probably of order 9—a determinant—), and thus a most unwieldy expression, in which it is very difficult to find a mass formula.

We now express the "correction" in terms of internal quantum numbers[15]; but before, we show, by use of a "$\mathfrak{sl}(3, \mathbf{C})$ *contracted model*", why it can be considered that we obtained a correction term.

Let $u \in \mathfrak{R} = \mathfrak{X} \cdot \mathfrak{R}_0$ and set $u = a + b + r$, where

$$a \in \mathfrak{a} = \mathfrak{su}(3), \qquad b \in \mathfrak{b} = J\mathfrak{su}(3),$$

$\mathfrak{a} + \mathfrak{b}$ being a Cartan decomposition of $\mathfrak{sl}(3, \mathbf{C}) = \mathfrak{X}$, and $r \in \mathfrak{R}_0$. The commutation relations of $\mathfrak{X}$ (Table III) give us the following relations:

$$[\mathfrak{a}, \mathfrak{a}] = \mathfrak{a}, \quad [\mathfrak{a}, \mathfrak{b}] = \mathfrak{b},$$
$$[\mathfrak{a}, \mathfrak{R}_0] = (\mathfrak{P}_0 + \mathfrak{Q}_0) \subset \mathfrak{R}_0,$$
$$[\mathfrak{b}, \mathfrak{b}] = \mathfrak{a}, \quad [\mathfrak{b}, \mathfrak{R}_0] = \mathfrak{R}_0.$$

If we multiply the generators of $\mathfrak{b}$ and $\mathfrak{R}_0$ by $\rho \in \mathbf{R}$, and set $\mathfrak{b}' = \rho\mathfrak{b}$, $\mathfrak{R}_0' = \rho\mathfrak{R}_0'$, we get a L.a. $\mathfrak{R}(\rho)$, isomorphic to $\mathfrak{R} = \mathfrak{R}(1)$ for $\rho \neq 0$, the structure constants of which depend on ρ, and can be written as follows:

$$\mathfrak{R}(\rho) : [\mathfrak{a}, \mathfrak{a}] = \mathfrak{a}, \quad [\mathfrak{a}, \mathfrak{b}'] = \mathfrak{b}',$$
$$[\mathfrak{a}, \mathfrak{R}_0'] = (\mathfrak{P}_0' \oplus \mathfrak{Q}_0') \subset \mathfrak{R}_0',$$
$$[\mathfrak{b}', \mathfrak{b}'] = \rho^2\mathfrak{a}, \qquad [\mathfrak{b}', \mathfrak{R}_0'] = \rho\mathfrak{R}_0';$$

i.e., the structure constants of $\mathfrak{R}(\rho)$ are those of $\mathfrak{R}$, except for the last two types of commutators. If $\rho \to 0$, $\mathfrak{R}(\rho)$ has for a limit (in the sense defined, e.g., by Segal[16]) the *limit-algebra* $\mathfrak{R}(0) \approx \mathfrak{a} \cdot (\mathfrak{b}_0 \oplus \mathfrak{R}_0)$, semidirect product of $\mathfrak{su}(3)$ by an Abelian (8 + 9 = 17-dimensional) ideal, defined by the representation $\mathrm{ad}_{\mathfrak{a}} \oplus D_9$ (which, as we have seen, is reducible on $\mathfrak{su}(3)$).

We develop here some remarks on the physical meaning of the notion of limit-algebra, also called contracted algebra.[17] It is well known that, in physics, the commutation or noncommutation of quantities, as well as the commutatiors themselves, characterize the "level" in which we are working. For instance, at the "classical level", everything is commutative (only functions are considered); at the "quantum level", the p's (e.g.) commute, and also the q's, but $[p, q] \neq 0$. At a "finer" level, the p's (e.g.) do not necessarily commute. On the other hand, a very small change in structure constants (in commutation relations) do not change the level in which we are working, and besides, from the experimental point of view, a sufficiently small change in the structure constants cannot be observed. These considerations may give physical meaning to the notions of "neighboring" algebras (in the sense of Segal, i.e., with "neighboring" structure constants) and of limit-algebra, of which we have here shown an example. Moreover, it is not very surprising that in the limit we get what can be called an "extended $\mathfrak{su}(3)$ model", containing $\mathfrak{su}(3)$ and the squared mass-operator [and besides it happens, as seen later, that the operators used in the mass formula are composed of elements of the subalgebra $\mathfrak{su}(3) \cdot \mathfrak{R}_0$, of which C and C_1 are invariants], since we know that, phenomenologically, $\mathfrak{su}(3)$ has given, to a certain extent, suitable results.

In the limit-algebra $\mathfrak{R}(0)$, C and C_1 are global invariants. In $\mathfrak{R}(\rho)$, we have $[\mathfrak{b}', C_1'] \subset \rho\mathfrak{R}_0'$, and $[\mathfrak{b}', C'] \subset \rho\mathfrak{R}_0'\mathfrak{R}_0'$, where C', C_1' denote the elements corresponding to C, C_1 in $\mathfrak{R}_0'\mathfrak{R}_0'$ (set of linear combinations of elements $r'r_1'$, where r', $r_1' \in \mathfrak{R}_0'$) and $\mathfrak{R}_0'$. For very small ρ, $[\mathfrak{b}', C_1']$ and $[\mathfrak{b}', C']$ are very small compared to, e.g., $[\mathfrak{a}, m'^2]$ (where $m'^2 = -g^{\mu\nu}p_\mu'p_\nu'$), i.e.,

$$-2C_1'r_0' = -2\tilde{C}_1'r_0' + \epsilon_1(\rho),$$

and $C' = \tilde{C}' + \epsilon(\rho)$, where $\tilde{C}'$ and $\tilde{C}_1'$ commute with all $\mathfrak{R}(\rho)$ and $\epsilon(\rho)$ have commutators [with $\mathfrak{R}(\rho)$] very small compared to $[\mathfrak{a}, m'^2]$. In the L.a. $\mathfrak{R}(\rho)$, isomorphic to $\mathfrak{R}$ for $\rho \neq 0$, we thus have the formula

$$m'^2 = \tilde{C}' - 2\tilde{C}_1'r_0' + \mathfrak{q}'^2 + q_4'^2 + r_0'^2 + \epsilon(\rho) + \epsilon_1(\rho). \qquad (4.13)$$

[15] M. Flato and J. Sternheimer, Compt. Rend. **259**, 3455 (1964).

[16] I. E. Segal, Duke Math. J. **18**, 221 (1951).

[17] E. P. Wigner and E. Inonü, Proc. Natl. Acad. Sci. U. S. **39**, 510 (1953).

Therefore, Eq. (4.12) gives us (after expressing the correction term as a function of internal quantum numbers) a mass formula in our original sense (cf. Introduction), up to a perturbation which is as small as we want. Rigorously speaking, we consider that we do here a *contracted unification* of $\mathfrak{X}$, $\mathcal{P}$, and $\mathfrak{u}(1)$, which is also a unification of $\mathfrak{sl}(3, \mathbf{C})$ contracted to $\mathfrak{su}(3)\cdot\mathfrak{b}_0$, of a $\mathfrak{u}(1)$, and of $\mathcal{P}$ contracted to $\mathfrak{so}(3)\cdot(\mathfrak{B}_0 \oplus \mathcal{P}_0)$ where $\mathfrak{B}_0$ is an order-3-Abelian ideal. We further notice that, in these contractions, the commutation relations of the subalgebra $\mathfrak{su}(3)\cdot\mathfrak{R}_0$ (in which are contained the squared mass-operator, its correction, and the expression of that correction in terms of internal quantum numbers) are the same as in the unification $\mathfrak{R}$.

Now, set

$$\mathfrak{L}_3' = (I + 2 \text{ ad } \alpha q_4)\mathfrak{L}_3 = \exp(2\alpha \text{ ad } q_4)\mathfrak{L}_3,$$

and take $\mathcal{P} = \mathfrak{L}_3'\cdot\mathcal{P}_0$. Denote the generators of $\mathfrak{L}_3'$ by J_k, N_k, the J_k's being generators of the $\mathfrak{so}(3)$ in $\mathfrak{L}_3'$. Denote by

$$\mathbf{R} = (R_k) = (\tfrac{1}{2}c_3', \tfrac{1}{2}b_3'', \tfrac{1}{2}a_3)$$

the "rotation vector" of $\mathfrak{L}_3$ ($k = 1, 2, 3$):

$$[R_i, R_j] = \epsilon_{ijk}R_k,$$

and set $\mathbf{J} = (J_k)$, $\mathbf{q} = (q_1, q_2, q_3)$. From the definition of $\mathfrak{L}_3'$, we have then $\mathbf{J} = \mathbf{R} + \alpha\mathbf{q}$ and $N_i = \frac{1}{2}c_3 - \alpha q_4$, from where we deduce

$$\mathbf{q}^2 = (1/\alpha^2)(\mathbf{J} - \mathbf{R})^2$$
$$= (1/\alpha^2)(\mathbf{J}^2 + \mathbf{R}^2 - \mathbf{J}\cdot\mathbf{R} - \mathbf{R}\cdot\mathbf{J}), \qquad (4.14)$$

i.e.,

$$\mathbf{q}^2 = (1/\alpha^2)(\mathbf{J}^2 + \mathbf{R}^2 - 2\overline{\mathbf{J}\cdot\mathbf{R}}),$$

where $\overline{\mathbf{J}\cdot\mathbf{R}}$ stands for the symmetrized expression $\frac{1}{2}(\mathbf{J}\cdot\mathbf{R} + \mathbf{R}\cdot\mathbf{J})$. Moreover, we have $\alpha = \alpha(J)$, fixed for fixed J.

Now, we consider $\mathbf{R}$ represented linearly in an operator space $\mathcal{E}$; we have

$$\mathbf{R} = \mathcal{E} \otimes E, \qquad \mathbf{J} = \mathcal{E} \otimes E',$$

where E and E' are 3-dimensional real vector spaces. If (which is always possible) we suppose that a duality between those spaces is defined, we can define the symmetrized scalar product $\overline{\mathbf{J}\cdot\mathbf{R}} \in \mathcal{E}$. We also notice that the definition of $\mathcal{E} \otimes E$ is actually that of irreducible tensor operators (in the sense of Wigner–Racah).

If A is a matrix belonging to $SO(3)$ in its 3-dimensional (real) realization, since it conserves the scalar product of vectors, we have

$$\overline{(A\mathbf{J}\cdot A\mathbf{R})} = \overline{\mathbf{J}\cdot\mathbf{R}}.$$

However $A \in \text{Aut } (\mathfrak{so}(3))$ also, since it is easy to check that

$$\text{Aut } (\mathfrak{so}(3)) = \text{Int } (\mathfrak{so}(3)) \approx SO(3).$$

Thus, $[AR_i, AR_j] = \epsilon_{ijk}AR_k$, and similarly for $\mathbf{J}$. Moreover, it is always possible to choose A so that, when taking the eigenvalues of the corresponding operators (represented in $\mathcal{E}$), $A\mathbf{J} = (0, 0, AJ_3)$; and then $\overline{\mathbf{R}\cdot\mathbf{J}} = (AJ_3)(AR_3)$ (for given J).

It is necessary to stress one point, in regard to what we have done here, and in general, with the expression of a mass formula. The treatment done, concerning applications to finite-dimensional classification representations, is purely formal, because the spectra of the considered operators can be calculated only in infinite-dimensional representations. Finite-dimensional representations will therefore have to do with fields, or with a so-called "subquantum level".

Now, we do the identification $\mathbf{R} = \mathbf{I}$ (isospin), the product $2\overline{\mathbf{J}\cdot\mathbf{R}}$ defining $(PC)I_3 = AR_3$, and giving $2\lambda(J)(PC)I_3$ for fixed $\mathbf{J}$ in the formula. Here, following standard convention, PC stands for the particle–antiparticle passage. We can also identify

$$r_0 = \beta(PC)Q + \gamma(PC)I_3, \qquad (4.15)$$

which defines the electric charge Q [by means of the added $\mathfrak{u}(1)$], and $q_4^2 = -\beta'^2(\frac{1}{4}Y^2)$ (as q_4 is a difference of "noncompact generators"), or, if it seems more convenient,

$$q_4^2 = -\eta^2\mathbf{I}^2 - \beta'^2(\tfrac{1}{4}Y^2) \qquad (4.16)$$

(as $\mathbf{q}^2$ contained also an $\mathbf{I}^2$ term), where Y is the hypercharge, and β, γ, β', η, are "identification constants". Now if we apply the Gell-Mann–Nishijima formula [which relates the added $\mathfrak{u}(1)$ to the enveloping algebra of $\mathfrak{X}$], and put $\beta + \gamma = \delta$, we have the formula

$$m^2 = (C + \alpha^{-2}\mathbf{J}^2) - C_1\beta(PC)Y$$
$$+ (\alpha^{-2} - \eta^2)\mathbf{I}^2 - 2(C_1\delta + \alpha^{-2}\lambda(J))(PC)I_3$$
$$+ \beta\delta I_3 Y + \delta^2 I_3^2 + (\beta^2 - \beta'^2)\tfrac{1}{4}Y^2. \qquad (4.17)$$

We can remark here that the identification $\mathbf{R} = \mathbf{I}$ and $\mathbf{J} = \exp(2\alpha \text{ ad } q_4)\mathbf{R}$, i.e., the fact we get $\mathbf{J}$ from $\mathbf{I}$ by a special automorphism of the unification $\mathfrak{R}$ (which, in view of our identification of q_4^2, is closely related to the strangeness S) is fully compatible with the phase relation

$$(-1)^{2(I-J)} = (-1)^S, \qquad (4.18)$$

which we can obtain from the already mentioned relations $(-1)^{2I} = (-1)^Y$ and $(-1)^{2J} = (-1)^{B+L}$,

and which shows that, for a fixed S, I and J are not independent; this automorphism expresses a kind of "*J–I symmetry*".

Moreover, the presence of $\mathbf{J}^2$, of $\alpha = \alpha(J)$, and of $\lambda(J)$ in the formula hints at a kind of "*J^P assignment*" that we get by fixing J when going from representations of $\mathfrak{R}$ to that of the classification L.a. $\mathfrak{X}$ ($\mathbf{J}$ does not appear in the intersection $\mathfrak{P} \cap \mathfrak{X}$).

If $\beta^2 = \beta'^2$, we have no Y^2 term, and it can be shown that the experimental situation suggests such an identification. In the formula, we have two kinds of coefficients: C, C_1, α, $\lambda(J)$ depend *a priori* (by their origin) on the classification representation, while the others ($\beta-$ and $\beta'-$, δ, and possibly η) are "identification coefficients", which do not necessarily depend on the representation. We thus get the following *mass-formula*:

$$m^2 = m_0^2 - a(PC)Y + bI(I+1) + cYI_3 - d(PC)I_3 + eI_3^2 \qquad (4.19)$$

($+fY^2$ eventually), where m_0^2, a, b, d depend, *a priori*, on the representation, and c, e, (and f) do not. As it is easy to verify, the average of I_3^2 on an isomultiplet is

$$\langle I_3^2 \rangle = \tfrac{1}{3}I(I+1). \qquad (4.20)$$

Thus, identifying $\beta^2 = \beta'^2$, we get the *strong mass-formula*

$$\langle m^2 \rangle = m_0^2 - a(PC)Y + b'I(I+1), \qquad (4.21)$$

where $b' = b + \frac{1}{3}e$ (and is numerically very close to b, as is seen later), and the *electromagnetic correction*

$$cYI_3 - d(PC)I_3 + e[I_3^2 - \tfrac{1}{3}I(I+1)], \qquad (4.22)$$

the average of which on each isomultiplet is 0. It is interesting to notice here that, even with somewhat different identifications of generators appearing in the "raw" mass formula, the strong mass-splitting given by (4.21) will be similar.

For completeness we now give some indications on the *experimental verification* of the mass formula, following J. Sternheimer (Ref. 15, and private communications). It is seen that we can consider the (identification) coefficient of Y^2 to be 0, since the influence of such a term is less than half of the experimental margin (while that of the I_3^2 term is similar to the margin); m_0^2, a, b are much greater than c, d, e (which can be considered as constant when J varies), c and d being greater than e. This shows that the numerical contribution of the "invariants" C and C_1 in the formula (and that of α^{-1}) is greater than that of the "identification coefficients".

If we apply the strong mass-formula

$$\langle m^2 \rangle = m_0^2 - a(PC)Y + b'I(I+1) + fY^2 \qquad (4.23)$$

to the case of the basic $\frac{1}{2}^+$ octet, one finds $-f \simeq 0.001 \pm 0.0015$ BeV^2, and therefore, in that case, it is fully justified to identify $\beta^2 = \beta'^2$, so that $f = 0$. For the $\frac{3}{2}^+$ decuplet, we have phenomenologically $I = 1 + \frac{1}{2}Y$, and so

$$Y^2 = 4I(I+1) - 6Y - 8.$$

In that case, there will necessarily be a compatibility condition, which is written later, as a consequence of (4.23), and so nothing opposes the identification $f = 0$. We see that the formula (4.21) seems very well verified for the $\frac{5}{2}^+$ spin; besides, we see that it is possible to consider the identification coefficients (which appear in the electromagnetic correction) as constant as the spin varies. Therefore, the mass formula will be taken *without a Y^2 term*. For the I_3^2 term, we first notice that I_3^2 itself becomes appreciable only for $|I_3| = \frac{3}{2}$ or more. Next, we see that, in the case of the $\frac{1}{2}^+$ octet, the coefficient of I_3^2 is about 2×10^{-2} BeV^2, while $c \simeq 6 \times 10^{-3}$ BeV^2 and $d \simeq (8.5) \times 10^{-3}$ BeV^2, and that these values are also appropriate in the case of the $\frac{3}{2}^+$ spin (for other spins, there are yet no experimental informations on the electromagnetic correction), the coefficients a, b, and m_0^2 being much bigger (the smallest, b, is about 9×10^{-2} BeV^2, for these two spins). So, on one hand, the introduction of an I_3^2 term is not appreciable in the strong formula, and on the other hand there are now only two indications in favor of such a term in the mass formula.

Let us now denote by m_E the mass of the particle $E(J^P, I, Y, m_E)$, or the strong mass of the corresponding isomultiplet. In the case of the *basic $\frac{1}{2}^+$ octet*, we get from (4.21) the following compatibility relation:

$$4(m_N^2 + m_\Xi^2) = 3m_\Sigma^2 + 5m_\Lambda^2, \qquad (4.24)$$

which is verified with a precision of 0.05% for an experimental margin of 0.08%, while the Okubo–Gell-Mann formula shows a difference of 0.5% for the same margin. Moreover, we get the following electromagnetic compatibility relation (there are two such relations, if we neglect the I_3^2 term[15]):

$$(m_{\Xi^-}^2 - m_{\Xi^0}^2) + (m_n^2 - m_p^2) = (m_{\Sigma^-}^2 - m_{\Sigma^+}^2) \qquad (4.25)$$

from (4.19), which is verified with a precision of the same order than the well-known corresponding linear relation.

Obviously, the same formulas hold for antiparticles. In the case of this octet, the only experimental indication in favor of an I_3^2 term is the value of m_{Σ^-}, which is found by the most recent experiments to be about 1197 MeV (instead of 1196 as before).

In the case of spin $\frac{3}{2}^+$, with the quadruplet Δ, the triplet Y_{13}^*, the doublet Ξ_{13}^*, and the singlet Ω^-, we find a strong compatibility relation:

$$m_{\Omega^-}^2 + 3m_{Y_{13}^*}^2 = m_\Lambda^2 + 3m_{\Xi_{13}^*}^2, \qquad (4.26)$$

which is verified with a precision of 0.035% for an experimental margin of 0.18%. There are also 3 (or 4, if we neglect the I_3^2 term) electromagnetic compatibility relations, of which only one can be verified at present (and the experimental situation is not very clear):

$$(m_{\Xi_{13}^{*0}}^2 - m_{\Xi_{13}^{*-}}^2) + (m_{\Delta^+}^2 + -m_{\Delta^0}^2) = (m_{Y_{13}^{*+}}^2 - m_{Y_{13}^{*-}}^2). \qquad (4.27)$$

The most serious indication in favor of the I_3^2 term is given by recent values of mass differences of the Δ particles, for it seems now that $m_{\Delta^{++}}$ is very close, but slightly superior, to m_{Δ^+}.

For *spin* $\frac{5}{2}^+$, it seems now possible (as predicted in Ref. 15) that Y_1^* (?, 1, 0, 1650 $\pm$ 6) has a spin $\frac{5}{2}^+$, and that it is also the case of the particle Ξ_1^* (?, $\frac{1}{2}$, -1, 1816 ± 3) (also considered with that spin in Ref. 15). With $N_{15}^*(\frac{5}{2}^+, \frac{1}{2}, 1{,}1688)$ and $Y_{05}^*(\frac{5}{2}^+, 0, 0, 1815)$, we see that the compatibility relation [which is a consequence of (4.21)]

$$4(m_{N_{13}^*}^2 + m_{\Xi_1^*}^2) = 3m_{Y_1^*}^2 + 5m_{Y_0^*}^2 \qquad (4.28)$$

is verified with a precision of 0.12% for an experimental margin of 0.3%. We also notice that there exists at present no other formula that gives a suitable mass for Ξ_1^* (the Okubo mass formula gives 1865 MeV, and the other formula are worse).

For *mesons of negative parity*, we remark that our D_9 (real) representation of $\mathfrak{sl}(3, \mathbf{C})$ may permit to *classify the nonets* of presently well-established mesons (pseudo-scalars and vector mesons) with spins 0^- and 1^-. It is found phenomenologically[15] that it is convenient to replace in formula (4.21) the hypercharge Y by a quantum number n that takes integer values (maybe a kind of principal quantum number) and therefore to replace (4.21) by the formula

$$\langle m^2 \rangle = m_0^2 + an + bI(I+1). \qquad (4.29)$$

For spin 0^-, with η, K, π, and $(\eta 2\pi)$, we find, by elimination of m_0^2, a, b,

$$\frac{8m_K^2 - 3m_\pi^2 - 5m_\eta^2}{m_{(\eta 2\pi)}^2 - m_\eta^2} = \frac{8n_K - 3n_\pi - 5n_\eta}{n_{(\eta 2\pi)} - n_\eta} = \frac{2}{3} \qquad (4.30)$$

with a precision better than 0.1%. If we take (in accordance with the above relation) $n_\eta = 1$, $n_K = 2$, $n_\pi = 3$, $n_{(\eta 2\pi)} = 4$ [thus, n distinguishes between η and $(\eta 2\pi)$], we get[15] a compatibility relation that is verified with a precision of 0.08% for a tolerated margin of 0.15%.

In the case of spin 1^-, we take ρ, K^*, φ, and ω. Here, the experimental situation does not allow us to attribute, with certainty, values of n to the different particles; if we take for n the values 0, 2, 4, and 1, respectively, for the above-mentioned particles, we get a very well verified compatibility relation. In both cases, by attributing different values to n, we may predict the masses of some particles. [One can also introduce such a "principal" quantum number by "degeneracy-lifting" considerations in Riemannian manifolds.[18]]

In the case of negative parity for baryons (and positive for mesons), we cannot say, at present, anything in favor of, or against, the proposed formula.

CONCLUSION

We end this work by three remarks. The first (which we develop in more detail in other papers—see, e.g., Refs. 13 and 19) is that it must be worthwhile to consider the "exceptional" unification $\mathfrak{su}(2, 2) = U(\mathcal{P}, \mathfrak{su}(2, 1))$ [thus taking $\mathfrak{su}(2, 1)$ as the internal L.a.], because of (among other reasons) the physical meaning of the conformal group, the Lie algebra of which is $\mathfrak{su}(2, 2)$, and then, for instance, to try to apply [by a study of the invariants of $\mathfrak{su}(2, 2)$] our treatment of the squared mass-operator (or a similar one) in order to get a mass formula (or look directly for the spectrum of the mass-operator[13]).

In that case also, difficulties due to the problem of unitary representations of $SU(2, 2)$ (or of representation of its Lie algebra) will arise, and this also leads us to our last remark.

The second remark concerns finite "internal" groups: several attempts have been made in this direction. It is true that (non-connected) extensions of [resp., by] $\boldsymbol{P}$ by [resp., of] finite groups (or "uni-

[18] M. Flato, D. Sternheimer, J. Sternheimer, J. P. Vigier, and G. Wataghin, Nuovo Cimento **42**, 431 (1966).

[19] M. Flato, D. Sternheimer, and J. P. Vigier, Compt. Rend. **260**, 3869 (1965); D. Bohm, M. Flato, D. Sternheimer, and J. P. Vigier, Nuovo Cimento **38**, 1941 (1965).

fications" of these—in the sense of coset-products) may give a mass-spectrum (spectrum of the operator representing $-g^{\mu\nu}p_\mu p_\nu$), though no real mass formula. But, as will be shown, this is about the only feature of some interest to be expected from such a treatment, that on the other hand raises many difficulties which cannot be overcome [such as natural derivation of charge-independance (which is usually supposed *ad hoc* in this type of models, if we want the discrete "internal" group to play the role of invariance group), nonpossibility of getting pure mesonic states, etc.]. This being said, we may proceed freely to our last remark.

We have seen that, because of the noncompacity of the groups we consider, because of the nilpotency of the p_μ's in finite-dimensional representations, in order to give a nonformal sense to the mass formula, etc. . . , we must study infinite-dimensional (unitary or "local") representations (and these do not always give the wanted result concerning the eigenvalues of the operators appearing in the formula). We have thus introduced a classification principle, which enables us to give a formal sense to the (then experimentally well verified) mass formula, and to classify the particles as usual. Mathematically, this principle raises the problem of the connection between some infinite-dimensional representations (of the unification) and some finite-dimensional representations (of the internal symmetry). Physically, this principle too suggests the

TABLE I.a. $\mathfrak{sl}(4, \mathbf{R})$.

] \ [	a_1	a_2	a_3	b_1	b_2	b_3	a'_1	a'_2	a'_3	b'_1	b'_2	b'_3	d_1	d_2	d_3
a_1	0	a_3	$-a_2$	0	b_3	$-b_2$	0	a'_3	$-a'_2$	$-2(d_2-d_3)$	$-b'_3$	b'_2	0	b'_1	$-b'_1$
a_2	$-a_3$	0	a_1	$-b_3$	0	b_1	$-a'_3$	0	a'_1	b'_3	$-2(d_3-d_1)$	$-b'_1$	$-b'_2$	0	b'_2
a_3	a_2	$-a_1$	0	b_2	$-b_1$	0	a'_2	$-a'_1$	0	$-b'_2$	b'_1	$-2(d_1-d_2)$	b'_3	$-b'_3$	0
b_1	0	b_3	$-b_2$	0	$-a_3$	a_2	$2d_1$	b'_3	b'_2	0	a'_3	a'_2	$2a'_1$	a'_1	a'_1
b_2	$-b_3$	0	b_1	a_3	0	$-a_1$	b'_3	$2d_2$	b'_1	a'_3	0	a'_1	a'_2	$2a'_2$	a'_2
b_3	b_2	$-b_1$	0	$-a_2$	a_1	0	b'_2	b'_1	$2d_3$	a'_2	a'_1	0	a'_3	a'_3	$2a'_3$
a'_1	0	a'_3	$-a'_2$	$-2d_1$	$-b'_3$	$-b'_2$	0	a_3	$-a_2$	0	b_3	b_2	$2b_1$	b_1	b_1
a'_2	$-a'_3$	0	a'_1	$-b'_3$	$-2d_2$	$-b'_1$	$-a_3$	0	a_1	b_3	0	b_1	b_2	$2b_2$	b_2
a'_3	a'_2	$-a'_1$	0	$-b'_2$	$-b'_1$	$-2d_3$	a_2	$-a_1$	0	b_2	b_1	0	b_3	b_3	$2b_3$
b'_1	$2(d_2-d_3)$	$-b'_3$	b'_2	0	$-a'_3$	$-a'_2$	0	$-b_3$	$-b_2$	0	a_3	$-a_2$	0	a_1	$-a_1$
b'_2	b'_3	$2(d_3-d_1)$	$-b'_1$	$-a'_3$	0	$-a'_1$	$-b_3$	0	$-b_1$	$-a_3$	0	a_1	$-a_2$	0	a_2
b'_3	$-b'_2$	b'_1	$2(d_1-d_2)$	$-a'_2$	$-a'_1$	0	$-b_2$	$-b_1$	0	a_2	$-a_1$	0	a_3	$-a_3$	0
d_1	0	b'_2	$-b'_3$	$-2a'_1$	$-a'_2$	$-a'_3$	$-2b_1$	$-b_2$	$-b_3$	0	a_2	$-a_3$	0	0	0
d_2	$-b'_1$	0	b'_3	$-a'_1$	$-2a'_2$	$-a'_3$	$-b_1$	$-2b_2$	$-b_3$	$-a_1$	0	a_3	0	0	0
d_3	b'_1	$-b'_2$	0	$-a'_1$	$-a'_2$	$-2a'_3$	$-b_1$	$-b_2$	$-2b_3$	a_1	$-a_2$	0	0	0	0

TABLE I. b. $\mathfrak{sl}(4, \mathbf{R})$.

d_1	$-a_3+b'_3$	$a_2+b'_2$	$b_1-a'_1$
$a_3+b'_3$	d_2	b'_1-a_1	$b_2-a'_2$
$-a_2+b'_2$	b'_1+a_1	d_3	$b_3-a'_3$
$b_1+a'_1$	$b_2+a'_2$	$b_3+a'_3$	$-d_1$ $-d_2$ $-d_3$

0	$-\alpha_3$	α_2	γ_2	β_3	$-\beta'_2$
α_3	0	$-\alpha_1$	$-\beta'_3$	γ_3	β_1
$-\alpha_2$	α_1	0	β_2	$-\beta'_1$	γ_1
γ_2	$-\beta'_3$	β_2	0	$-\alpha'_3$	α'_2
β_3	γ_3	$-\beta'_1$	α'_3	0	$-\alpha'_1$
$-\beta'_2$	β_1	γ_1	$-\alpha'_2$	α'_1	0

existence of a possible connection between our quantum level and a so-called "subquantum level", or more traditionally, between states and fields.

APPENDIX. SOME LIE ALGEBRAS

We give here, in skew-symmetric tables, the commutation relations of three L.a.: $\mathfrak{sl}(4,\mathbf{R})$, $\mathfrak{su}(2,2)$, and $\mathfrak{sl}(3,\mathbf{C})$, and those of the unification $U(\mathcal{P}, \mathfrak{sl}(3,\mathbf{C}), \mathfrak{u}(1))$ we have constructed (the ideal $\mathcal{R}_0$ being there commutative). We also give the basic representations of these L.a., and indicate how the representations D_N and D_4 of $\mathfrak{L}$ appear in them. The tables are expressed according to a certain basis of the considered L.a., as are the basic representations, which are given in a condensed way by a matrix constructed as follows: if the element x is represented by a matrix, the coefficient of which is λ at a certain place, we write λx at this place; and if, in addition, y is represented by a matrix with μ at the same place, we write $\lambda x + \mu y$, etc. For a well-chosen basis (and a semisimple L.a.), we can get all the fundamental invariants (i.e., a basis of the polynomial ring which is the center of the enveloping algebra) as coefficients of the formal symmetrized characteristic polynomial of the matrix.

For $\mathfrak{sl}(4,\mathbf{R})$, (Table I.b) we give the basic four-

TABLE II. a. $\mathfrak{su}(2,2)$

[\ ]	a_1	a_2	a_3	b_1	b_2	b_3	a'_1	a'_2	a'_3	c_1	b'_1	b'_2	b'_3	c_2	c_3
a_1	0	a_3	$-a_2$	0	b_3	$-b_2$	0	a'_3	$-a'_2$	0	0	b'_3	$-b'_2$	0	0
a_2	$-a_3$	0	a_1	$-b_3$	0	b_1	$-a'_3$	0	a'_1	0	$-b'_3$	0	b'_1	0	0
a_3	a_2	$-a_1$	0	b_2	$-b_1$	0	a'_2	$-a'_1$	0	0	b'_2	$-b'_1$	0	0	0
b_1	0	b_3	$-b_2$	0	$-a_3$	a_2	c_1	0	0	a'_1	c_2	0	0	b'_1	0
b_2	$-b_3$	0	b_1	a_3	0	$-a_1$	0	c_1	0	a'_2	0	c_2	0	b'_2	0
b_3	b_2	$-b_1$	0	$-a_2$	a_1	0	0	0	c_1	a'_3	0	0	c_2	b'_3	0
a'_1	0	a'_3	$-a'_2$	$-c_1$	0	0	0	a_3	$-a_2$	b_1	c_3	0	0	0	$-b'_1$
a'_2	$-a'_3$	0	a'_1	0	$-c_1$	0	$-a_3$	0	a_1	b_2	0	c_3	0	0	$-b'_2$
a'_3	a'_2	$-a'_1$	0	0	0	$-c_1$	a_2	$-a_1$	0	b_3	0	0	c_3	0	$-b'_3$
c_1	0	0	0	$-a'_1$	$-a'_2$	$-a'_3$	$-b_1$	$-b_2$	$-b_3$	0	0	0	0	$-c_3$	$-c_2$
b'_1	0	b'_3	$-b'_2$	$-c_2$	0	0	$-c_3$	0	0	0	0	$-a_3$	a_2	$-b_1$	$-a'_1$
b'_2	$-b'_3$	0	b'_1	0	$-c_2$	0	0	$-c_3$	0	0	a_3	0	$-a_1$	$-b_2$	$-a'_2$
b'_3	b'_2	$-b'_1$	0	0	0	$-c_2$	0	0	$-c_3$	0	$-a_2$	a_1	0	$-b_3$	$-a'_3$
c_2	0	0	0	$-b'_1$	$-b'_2$	$-b'_3$	0	0	0	c_3	b_1	b_2	b_3	0	$-c_1$
c_3	0	0	0	0	0	0	b'_1	b'_2	b'_3	c_2	a'_1	a'_2	a'_3	c_1	0

TABLE II. b. $\mathfrak{su}(2,2)$

$ia_1-b_1+q_0$	a_2+ib_2 $+ia_3-b_3$	$2(-p_1+p_4)$	$2(ip_2-p_3)$
$-a_2-ib_2$ $+ia_3-b_3$	$-ia_1+b_1+q_0$	$-2(ip_2+p_3)$	$2(p_1+p_4)$
$-2(q_1+q_4)$	$2(iq_2-q_3)$	$ia_1+b_1-q_0$	a_2-ib_2 $+ia_3+b_3$
$-2(iq_2+q_3)$	$2(q_1-q_4)$	$-a_2+ib_2$ $+ia_3+b_3$	$-ia_1-b_1-q_0$

0	$-a'_1$	$-a'_2$	$-a'_3$	c_1	$-c_3$
a'_1	0	$-a_3$	a_2	b_1	b'_1
a'_2	a_3	0	$-a_1$	b_2	b'_2
a'_3	$-a_2$	a_1	0	b_3	b'_3
c_1	b_1	b_2	b_3	0	c_2
$-c_3$	b'_1	b'_2	b'_3	$-c_2$	0

TABLE III. a. $\mathfrak{sl}(3, C)$

	a_1	a_2	a_3	b_1	b_2	b_3	b_1'	b_2'	b_3'	b_1''	b_2''	b_3''	c_1	c_3	c_1'	c_3'	
a_1	0	a_3	$-a_2$	0	b_3	$-b_2$	$-2c_1$	$-b_3'$	b_2'	$-2c_1'$	$-b_3''$	b_2''	$2b_1'$	$-b_1'$	$2b_1''$	$-b_1''$	a_1
a_2	$-a_3$	0	a_1	$-b_3$	0	b_1	b_3'	$2(c_1+c_3)$	$-b_1'$	b_3''	$2(c_1'+c_3')$	$-b_1''$	$-b_2'$	$-b_2'$	$-b_2''$	$-b_2''$	a_2
a_3	a_2	$-a_1$	0	b_2	$-b_1$	0	$-b_2'$	b_1'	$-2c_3$	$-b_2''$	b_1''	$-2c_3'$	$-b_3'$	$2b_3'$	$-b_3''$	$2b_3''$	a_3
b_1	0	b_3	$-b_2$	0	$-a_3$	a_2	$-2c_1'$	$-b_3''$	b_2''	$2c_1$	b_3'	$-b_2'$	$2b_1''$	$-b_1''$	$-2b_1'$	b_1'	b_1
b_2	$-b_3$	0	b_1	a_3	0	$-a_1$	b_3''	$2(c_1'+c_3')$	$-b_1''$	$-b_3'$	$-2(c_1+c_3)$	b_1'	$-b_2''$	$-b_2''$	b_2'	b_2'	b_2
b_3	b_2	$-b_1$	0	$-a_2$	a_1	0	$-b_2''$	b_1''	$-2c_3'$	b_2'	$-b_1'$	$2c_3$	$-b_3''$	$2b_3''$	b_3'	$-2b_3'$	b_3
b_1'	$2c_1$	$-b_3'$	b_2'	$2c_1'$	$-b_3''$	b_2''	0	a_3	$-a_2$	0	b_3	$-b_2$	$2a_1$	$-a_1$	$2b_1$	$-b_1$	b_1'
b_2'	b_3'	$-2(c_1+c_3)$	$-b_1'$	b_3''	$-2(c_1'+c_3')$	$-b_1''$	$-a_3$	0	a_1	$-b_3$	0	b_1	$-a_2$	$-a_2$	$-b_2$	$-b_2$	b_2'
b_3'	$-b_2'$	b_1'	$2c_3$	$-b_2''$	b_1''	$2c_3'$	a_2	$-a_1$	0	b_2	$-b_1$	0	$-a_3$	$2a_3$	$-b_3$	$2b_3$	b_3'
b_1''	$2c_1'$	$-b_3''$	b_2''	$-2c_1$	b_3'	$-b_2'$	0	b_3	$-b_2$	0	$-a_3$	a_2	$2b_1$	$-b_1$	$-2a$	a_1	b_1''
b_2''	b_3''	$-2(c_1'+c_3')$	$-b_1'$	$-b_3'$	$2(c_1+c_3)$	b_1'	$-b_3$	0	b_1	a_3	0	$-a_1$	$-b_2$	$-b_2$	a_2	a_2	b_2''
b_3''	$-b_2''$	b_1''	$2c_3'$	b_2'	$-b_1'$	$-2c_3$	b_2	$-b_1$	0	$-a_2$	a_1	0	$-b_3$	$2b_3$	a_3	$-2a_3$	b_3''
c_1	$-2b_1''$	b_2'	b_3'	$-2b_1''$	b_2''	b_3''	$-2a_1$	a_2	a_3	$-2b_1$	b_2	b_3	0	0	0	0	c_1
c_3	b_1'	b_2'	$-2b_3'$	b_1''	b_2''	$-2b_3''$	a_1	a_2	$-2a_3$	b_1	b_2	$-2b_3$	0	0	0	0	c_3
c_1'	$-2b_1''$	b_2''	b_3''	$2b_1'$	$-b_2'$	$-b_3'$	$-2b_1$	b_2	b_3	$2a_1$	$-a_2$	$-a_3$	0	0	0	0	c_1'
c_3'	b_1''	b_2''	$-2b_3''$	$-b_1'$	$-b_2'$	$2b_3'$	b_1	b_2	$-2b_3$	$-a_1$	$-a_2$	$2a_3$	0	0	0	0	c_3'

TABLE III. b. $\mathfrak{sl}(3, C)$

$c_3 + ic_3'$	$-a_3 - ib_3$ $+b_3' + ib_3''$	$a_2 + ib_2$ $+b_2' + ib_2''$
$a_3 + ib_3$ $+b_3' + ib_3''$	$-c_3 - ic_3'$ $+c_1 + ic_1'$	$-a_1 - ib_1$ $+b_1' + ib_1''$
$-a_2 - ib_2$ $+b_2' + ib_2''$	$a_1 + ib_1$ $+b_1' + ib_1''$	$-c_1 - ic_1'$

and six-dimensional representations as $\mathfrak{sl}(4, R)$ and as $\mathfrak{so}(3, 3)$. The correspondance between them is

$$\alpha_k = \tfrac{1}{2}(a_k - a_k'), \qquad \alpha_k' = \tfrac{1}{2}(a_k + a_k'),$$

$$\beta_k = \tfrac{1}{2}(b_k - b_k'), \qquad \beta_k' = \tfrac{1}{2}(b_k + b_k') \ (k = 1, 2, 3),$$

and $\gamma_k = \tfrac{1}{2}(d_k + d_i - d_j)$ (ijk cyclic here).

The "compact generators" (those of a maximal compact subalgebra) are (a_k, a_k'), or (α_k, α_k'). In the four- and six-dimensional representations, respectively, we get $D_N(\mathfrak{L})$ and ad $\mathfrak{L}$ with the representing matrices of (a_k, b_k) (we have $a_k = \alpha_k + \alpha_k'$, $b_k = \beta_k + \beta_k'$).

We obtain a subalgebra isomorphic to $\mathfrak{L}$ with, e.g. [instead of $(a_k, b_k]$,

$$(\alpha_1, \alpha_2, \alpha_3; \beta_2', -\beta_1, -\gamma_1),$$

or

$$(\alpha_1, \alpha_2, \alpha_3; \gamma_2, -\beta_3', \beta_2),$$

$$(\alpha_1, \alpha_2, \alpha_3; \beta_3, \gamma_3, -\beta_1'),$$

$$(\alpha_1', \alpha_2', \alpha_3'; -\beta_3', \gamma_3, \beta_1),$$

$$(\alpha_1', \alpha_2', \alpha_3'; \gamma_2, \beta_3, -\beta_2'),$$

$$(\alpha_1', \alpha_2', \alpha_3'; \beta_2, -\beta_1', \gamma_1).$$

Any two of these, having the same compact subalgebra [(α_k), or (α_k')] generate a $\mathfrak{so}(3, 2)$ subalgebra with a 10th element (an α_i', or α_i). We get $D_4(\mathfrak{L})$ with $\mathfrak{L}$ written as $(\alpha_1', \alpha_2', \alpha_3'; \gamma_2, \beta_3, -\beta_2')$ in the *four*-dimensional representation. This representation is not equivalent to D_N.

For $\mathfrak{su}(2, 2) = \mathfrak{so}(4, 2)$, we have given here the basic four- and six-dimensional representations

TABLE IV. Commutation relations [$\mathfrak{sl}(3, \mathbf{C})$, ideal], with $C = \cos\theta$, $S = \sin\theta$.

] / [	p_1	p_2	p_3	p_4	q_1	q_2	q_3	q_4	r_0
a_1	$-q_1C-q_4S$	$-q_2C+q_3S$	$-q_2S-q_3C$	$-q_1C-q_4S$	$(p_1+p_4-r_0)C$	p_2C+p_3S	$-p_2S+p_3C$	$(p_1+p_4-r_0)S$	$2(q_1C+q_4S)$
a_2	q_2C-q_3S	$-q_1C-q_4S$	q_1S-q_4C	$-q_2C+q_3S$	p_2C-p_3S	$(-p_1+p_4-r_0)C$	$(p_1-p_4+r_0)S$	p_2S+p_3C	$2(q_2C-q_3S)$
a_3	$2p_2$	$-2p_1$	0	0	q_2	$-q_1$	q_4	$-q_3$	0
b_1	$-q_1S+q_4C$	$-q_2S-q_3C$	q_2C-q_3S	$-q_1S+q_4C$	$-(p_1+p_4+r_0)S$	$-p_2S+p_3C$	$-p_2C-p_3S$	$(p_1+p_4+r_0)C$	$2(-q_1S+q_4C)$
b_2	q_2S+q_3C	$-q_1S+q_4C$	$-q_1C-q_4S$	$-q_2S-q_3C$	$-p_2S-p_3C$	$(p_1-p_4-r_0)S$	$(p_1-p_4-r_0)C$	p_2C-p_3S	$2(-q_2S-q_3C)$
b_3	0	0	$-2p_4$	$-2p_3$	$-q_3$	$-q_4$	$-q_1$	$-q_2$	0
b'_1	q_1C+q_4S	q_2C-q_3S	q_2S+q_3C	q_1C+q_4S	$(p_1+p_4+r_0)C$	p_2C+p_3S	$-p_2S+p_3C$	$(p_1+p_4+r_0)S$	$2(q_1C+q_4S)$
b'_2	q_2C-q_3S	$-q_1C-q_4S$	q_1S-q_4C	$-q_2C+q_3S$	$-p_2C+p_3S$	$(p_1-p_4-r_0)C$	$(-p_1+p_4-r_0)S$	$-p_2S-p_3C$	$2(-q_2C+q_3S)$
b'_3	0	$2p_4$	0	$2p_2$	q_2	q_1	$-q_4$	$-q_3$	0
b''_1	q_1S-q_4C	q_2S+q_3C	$-q_2C+q_3S$	q_1S-q_4C	$(-p_1-p_4+r_0)S$	$-p_2S+p_3C$	$-p_2C-p_3S$	$(p_1+p_4-r_0)C$	$2(-q_1S+q_4C)$
b''_2	q_2S+q_3C	$-q_1S+q_4C$	$-q_1C-q_4S$	$-q_2S-q_3C$	p_2S+p_3C	$(-p_1+p_4-r_0)S$	$(-p_1+p_4-r_0)C$	$-p_2C+p_3S$	$2(q_2S+q_3C)$
b''_3	$-2p_3$	0	$2p_1$	0	$-q_3$	q_4	q_1	$-q_2$	0
c_1	$-(p_1+p_4)$	$-p_2$	$-p_3$	$-(p_1+p_4)$	0	q_2	q_3	0	$2r_0$
c_2	$2p_4$	0	0	$2p_1$	q_1	$-q_2$	$-q_3$	q_4	0
c'_1	0	$-p_3$	p_2	0	$-2q_4$	q_3	$-q_2$	$2q_1$	0
c'_2	0	$2p_3$	$-2p_2$	0	q_4	q_3	$-q_2$	$-q_1$	0

(Table II.b), with the basis that can be represented in four dimensions (in terms of the elements of $\mathfrak{sl}(4, \mathbf{R})$) by

$$a_1 = i\gamma_3 \quad (\text{or } i\gamma_3); \qquad a'_1 = -\alpha'_2 \quad \text{or} \quad -i\gamma_2;$$
$$a_2 = -\alpha'_3 \quad (\text{or } -\alpha'_3); \qquad a'_2 = i\beta_1 \quad \text{or} \quad -\alpha_3;$$
$$a_3 = i\beta'_3 \quad (\text{or } i\beta'_3); \qquad a'_3 = \alpha'_1 \quad \text{or} \quad -i\beta_3;$$
$$b_1 = \gamma_2 \quad \text{or} \quad -i\alpha'_2; \qquad b'_1 = \beta_2 \quad (\text{or } \beta_2);$$
$$b_2 = -i\alpha_3 \quad \text{or} \quad -\beta_1; \qquad b'_2 = i\alpha_1 \quad (\text{or } i\alpha_1);$$
$$b_3 = \beta_3 \quad \text{or} \quad i\alpha'_1; \qquad b'_3 = -\beta'_1 \quad (\text{or } -\beta'_1);$$
$$c_2 = \alpha_2 \quad \text{or} \quad i\gamma_1;$$
$$c_1 = \beta'_2 \quad \text{or} \quad -\beta'_2;$$
$$c_3 = \gamma_1 \quad \text{or} \quad -i\alpha_2.$$

If we set $p_k = a'_k + b'_k$, $p_4 = c_1 + c_2$, we get $\mathcal{P}$ with $(a_k; b_k; p_\mu)$. In the first way, above, the p_μ's are represented by triangular matrices; as far as $\mathcal{P}$ is concerned, we can multiply all of them by any complex number. The second way is the canonical[7] representation of $\mathfrak{su}(2, 2)$, and in it, $\mathfrak{su}(2, 1)$ is immediately seen as order-3-matrices. In Table II.b, we have given only the first way: the second can be constructed immediately (from Table I.b, for example); there we set $q_0 = c_3$ and $q_k = b'_k - a'_k$, $q_4 = c_2 - c_1$. (In order to get the invariants by the above-mentioned method, we have to "normalize" the nilpotent p_μ's and q_μ's by suppressing the factor 2 in front of them.) We get a $\mathfrak{so}(3, 2)$ subalgebra with $(a_k, c_2; b_k, b'_k)$ and a $\mathfrak{so}(4, 1)$ subalgebra with $(a_k, a'_k; b_k, c_1)$.

For $\mathfrak{sl}(3, \mathbf{C})$, we have taken the basis (a_k, b'_k, c_1, c_3) of $\mathfrak{sl}(3, \mathbf{R})$ and completed it into a basis

$$(a_k, b_k, b'_k, b''_k, c_1, c_3, c'_1, c'_3)$$

of $\mathfrak{sl}(3, \mathbf{C})$ (Tables III). We get $\mathfrak{su}(2, 1)$ subalgebras with the generators

$$(a_3, b''_3, c'_3, 2c'_1 + c'_3; b_1, b_2, b'_1, b'_2)$$

or with $(a_1, b''_1, c'_1, 2c'_3 + c'_1; b_2, b_3, b'_2, b'_3)$,

and a maximal compact subalgebra $\mathfrak{su}(3)$ with (a_k, b''_k, c'_1, c'_3). We realize a subalgebra $\approx \mathfrak{L}$ as $\mathfrak{so}(3, \mathbf{C})$ by $(a_k; b_k)$, and three others, as $\mathfrak{sl}(2, \mathbf{C})$, with the base elements $\frac{1}{2}(-a_k, b''_k, c'_k; b_k, b'_k, c_k)$, which we denoted by $\mathfrak{L}_k$ $(k = 1, 2, 3)$; here, we have set $-c_2 = c_1 + c_3$, and $-c'_2 = c'_1 + c'_3$. Thus, $\mathfrak{sl}(3, \mathbf{C})$ can be considered as "coupling", in a nontrivial way, three "relativistic rotators", i.e., three subalgebras isomorphic to $\mathfrak{L}$.

A Manifestly Covariant Description of Arbitrary Dynamical Variables in Relativistic Quantum Mechanics

GORDON N. FLEMING
Department of Physics, The Pennsylvania State University, University Park, Pennsylvania
(Received 14 April 1966)

The concepts of instantaneous observables and dynamical variables are analyzed and generalized to arbitrary spacelike hyperplanes. A formalism is developed which gives the basic equations of relativistic quantum mechanics for dynamical variables on arbitrary hyperplanes a manifestly covariant form. A covariant linear transformation on the Poincaré generators introduces the hyperplane generators which yield commutation relations displaying a clear separation of the kinematical and dynamical properties of dynamical variables. An axiomatic study of the center-of-mass position operator yields the uniqueness of the operator and completes the physical interpretation of the hyperplane generators. The Poincaré invariance and hyperplane independence of the scattering operator is related to asymptotic conservation laws in the hyperplane formalism, and finally, a nonlocal, hyperplane-dependent, field theory of free spinless particles is considered.

INTRODUCTION

THIS paper contains a geometrical analysis of the concept of a dynamical variable as it appears in the framework of the conventional foundations of relativistic quantum mechanics.[1] The analysis is presented in four stages. First, the dominant concepts of nonrelativistic quantum mechanics—observables and measurements associated with instants of time—are generalized to the relativistically invariant concepts of observables and measurements associated with arbitrary spacelike hyperplanes.[2] Second, the transformation equations, commutation relations, and Heisenberg equations of motion associated with the geometrical invariance principles[3] of relativistic quantum mechanics are written down and discussed for arbitrary hyperplane dynamical variables. A notational development is employed here to effect a cleancut separation between the dynamical equations of motion and the purely kinematical transformation properties of hyperplane variables in a manifestly covariant manner. I believe this is a novel result. Third, the presence of interactions, via the old-fashioned device of separating the Poincaré generators into "free" and "interaction" parts, is considered and the connection between symmetries and conservation laws is discussed. These very familiar topics are raised to display the novelties of the presentation which arise from the use of the covariant "hyperplane formalism," as it is called. In the course of this discussion, the uniqueness of the "center-of-mass" position operator is demonstrated.[4] The fourth and last stage takes up the question of a field-theoretic description of the evolution of noninteracting physical systems within the "hyperplane formalism." One is led quite naturally to the introduction of *nonlocal hyperplane dependent* field operators which, in the absence of interactions, provides a description of free particles completely equivalent to the familiar description in terms of *local* quantized free fields.[5] This equivalence seems to detract somewhat from the sanctified position of the microcausality assumption which has never enjoyed a secure physical foundation.[6]

Throughout this paper, no new physical postulates are invoked, and, in this sense, the treatment is entirely within the framework of conventional rel-

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939); Nuovo Cimento **3**, 517 (1956); R. Hagedorn, Nuovo Cimento Suppl. **12**, 73 (1959); A. Barut and A. S. Wightman, *ibid.* **14**, 81 (1959); E. Fabri, Nuovo Cimento **14**, 1130 (1959); A. S. Wightman, in *Dispersion Relations and Elementary Particles*, C. de Witt and R. Omnes, Eds. (John Wiley & Sons, Inc., New York, 1960), p. 159. For a concise, up-to-date discussion see R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and all That* (W. A. Benjamin, Inc., New York, 1964), Chap. 1. It is obvious that what I mean by the "conventional foundations of relativistic quantum mechanics" is the theory of Poincaré invariance in Hilbert space.

[2] The Heisenberg picture, in which the operators describe the evolution of the system, will be used throughout.

[3] A recent extensive treatment of the role of geometrical invariance principles has been given by R. M. F. Houtappel, H. Van Dam, and E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).

[4] The "center-of-mass" position operator as well as two other position operators which have received attention in the past have been discussed in a covariant way in G. N. Fleming, Phys. Rev. **137**, B188 (1964).

[5] See, for example, S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson and Company, Evanston, Illinois, 1961), Pt. 2, pp. 121–253.

[6] Microcausality is usually based on the assumption of the measurability of the fields. The classic discussion of this topic is given by N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat. fys. Medd. **12**, No. 8 (1933); Phys. Rev. **78**, 794 (1950). A careful statement of microcausality is given by R. Haag and B. Schroer, J. Math. Phys. **3**, 248 (1962). The measurability of many fields is undermined by superselection rules which were first considered by G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. **88**, 101 (1952).

ativistic quantum mechanics. On the other hand, the simplest *pseudo-local* interactions[7] which can be written down in terms of the nonlocal, hyperplane-dependent fields yield dynamical results that are fundamentally different from those arising out of local interactions of local fields. The detailed investigation of this exciting problem is under way and will be the subject of subsequent papers employing the hyperplane formalism developed here.

It must be realized by now that the analysis of the concept of observable presented here is in no sense complete. I will concentrate almost exclusively on the question of the transformation properties of arbitrary dynamical variables with an eye to *answering kinematical questions without invoking dynamical assumptions.* In particular, nothing is said about the perplexing problem of deciding which self-adjoint operators with complete sets of eigenstates are in fact "measurable" or how they are to be "measured."[8]

Before proceeding to the main discussion, I want to make clear the attitude adopted here regarding the notion of *instantaneous* measurements. This topic has been the subject of much discussion for many years,[9] and the belief that the instantaneous measurement is a useful theoretical idealization in the relativistic domain is currently suffering a marked decrease in popularity.

From the literature on the subject of time in quantum mechanics, one may discern four principle subdivisions. First there is the division between internal time variables and the external time variable, or parameter as it is frequently called. Second, associated with each of these subtopics there are questions of the existence of a theoretical lower limit to the duration required for a measurement to be made, or to the uncertainty accompanying any measurement of the time of an event. I am concerned here only with the external time variable which is the independent variable of the state vector in the Schrödinger picture or of the operators in the Heisenberg picture. It has been known for some time that one can associate Hermitian operators with the internal times of interest such as the time delay of scattering theory.[10] In such a circumstance the investigation of limitations on measurement can be carried out in the traditional vein of looking at commutators of the time operator with other observables. One does not have recourse to such techniques with the external parameter, and any conjecture of measurement limitations associated with it involves fundamental modifications of the general principles of quantum mechanics.

Limitations on the duration of measurements or on the precision of time measurements are two very different kinds of limitations and either could exist without the other. Thus a lower limit to the length of time required for a measurement does not *a priori* prohibit an arbitrarily precise determination of the instants that the measurement began and ended. Conversely, a lower limit on the uncertainty accompanying a time measurement does not *a priori* prohibit the expectation values of the terminal times of a measurement from being arbitrarily close together.

Any real measurement does, of course, require a finite time for its execution. If the measurement yields a result associated with the system of interest at the instant the interaction with the apparatus began then the measurement may be called postdictive[11] and can be used to test theoretical predictions of the evolution of probability distributions. If the measured result is characteristic of the quantum state at the instant the interaction with the apparatus ceases, then the measurement may be called predictive[11] and can be used to prepare quantum mechanical states. Measuring the momenta of particles by observing stopping tracks in emulsions

[7] By pseudo-local I mean that while the fields themselves do not obey the microcausality postulate the Hamiltonian density, nevertheless, involves fields at only one value of their space–time parameters.

[8] It has become increasingly evident in recent years (see Ref. 3, pp. 614, 625–627) that some of the most familiar and common operators pose grave problems for the relativistic quantum theory of measurement. Thus the projection operators for position observables do not commute even when they project onto mutually space-like regions [see Ref. 14 and G. N. Fleming, Phys. Rev. **139**, B963 (1965)].[3] The possibility that the hyperplane formalism may at least permit an invariant characterization of the result of measuring two noncommuting mutually spacelike projectors is presently under investigation.

[9] N. Bohr, Die Naturwissenschaftern 251 (1928); L. Landau and R. Peirl, Z. Physik **69**, 56 (1931); L. Mandelstamm and I. Tamm, Zh. Techn. Fiz. **9**, 249 (1945); V. Fock and N. Krylov, *ibid.* **11**, 112 (1947); H. L. Armstrong, Am. J. Phys. **22**, 195 (1947); L. Landau and E. Lifschitz, *Quantum Mechanics-Nonrelativistic Theory* (Pergamon Press, Inc., New York, 1959), pp. 150–153; W. Heisenberg, *Physics and Philosophy* (Allen and Unwin, London, 1959), Chap. 3; A. Messiah, *Quantum Mechanics* (North-Holland Publishing Company, Amsterdam, 1961), Vol. I, pp. 137, 319; Y. Aharonov and D. Bohm, Phys. Rev. **122**, 1649 (1961); H. Paul, Ann. Physik **9**, 252 (1962); M. L. Goldberger and K. M. Watson, Phys. Rev. **127**, 2284 (1962); E. Frick and F. Engelmann, Z. Physik **175**, 271 (1963); **178**, 551 (1964); Y. Aharonov, P. G. Bergmann, and I. L. Lebowitz, Phys. Rev. **134**, B1410 (1964); Y. Aharonov and D. Bohm, *ibid.* **134**, B1417 (1964); W. C. Davidon and H. Ekstein, J. Math. Phys. **5**, 1588 (1964); H. P. Stapp, Phys. Rev. **139**, B257 (1965); B. Rankin, J. Math. Phys. **6**, 1057 (1965); R. Omnes, Phys. Rev. **140**, B1474 (1965).

[10] F. T. Smith, Phys. Rev. **118**, 349 (1960); **130**, 394 (1963); **131**, 2803 (1963). See, also, T. T. Gien, J. Math. Phys. **6**, 671 (1965).

[11] This is not the terminology of L. Landau and E. Lifschitz (Ref. 9, p. 5).

is a post-dictive measurement. Using accelerators and electromagnetic selecting and focusing devices to prepare monoenergetic particles is a predictive measurement. The location of a particle via its passage through a geiger tube connected to an amplifying circuit is both post-dictive and predictive. There is no evidence, at present, that the time intervals required for these measurements or the uncertainties in the times that they begin and end cannot be arbitrarily decreased by technological innovations.

This situation could be modified. An increased understanding of the influence of cosmological structure on both classical apparatus and quantum systems could yield fundamental limitations on time measurements,[12] while the discovery of an upper limit to the mass spectrum of fundamental particles could yield limitations on the duration of any measurement.[13] Such notions are highly speculative now and can be pursued only with great difficulty. On the other hand, taking the naive position that instantaneous measurements at precisely determined instants are theoretical possibilities and pursuing this premise to the conclusions demanded by consistency with special relativity should be enlightening. This is all the more likely if it is accomplished independently of specific dynamical postulates such as field theory or analytic S-matrix theory. This, then, defines the attitude I adopt here regarding measurements and observables defined at an instant.

In Sec. (1) the transformation rules for classical dynamical variables are discussed in a general way indicating the dependence of the solution on the *pointlike* character of the variables. Quantum mechanical variables, which are not pointlike, are considered in Sec. (2) and the simplicity of the transformation problem under the Galilean group is displayed. The translation of the description of quantum mechanical measurements from one inertial frame to another in the relativistic domain is taken up in Sec. (3). In Sec. (4) the hyperplane solution of the translation and transformation problems for arbitrary relativistic observables is described and briefly criticized. In Sec. (5) the transformation equations for arbitrary dynamical variables under the Poincaré group are written down and this is followed in Sec. (6) by the introduction of the hyperplane generators which simplify the transformation equations and display a clean separation of the kinematical and dynamical aspects of the transformation equations. In Sec. (7) the center-of-mass position operator is defined axiomatically and proved unique.[14] It is employed here to demonstrate a kinematical interpretation of the commutation relations between the free and interacting parts of the hyperplane generators. It also establishes the connection between the two dynamical problems posed by the hyperplane formalism. A connection is needed to maintain equivalence with the conventional formulation in which the determination of the time dependence of the dynamical variables is the only dynamical problem. A brief discussion of the connection between the Poincaré invariance of the scattering operator and the conservation of hyperplane observables occupies Sec. (8). Finally, in Sec. (9) the hyperplane-dependent field operator for a free scalar particle is introduced providing a nonlocal field theory of such particles completely equivalent to the traditional local field theory. The hyperplane field operator is essentially the creation operator for the hyperplane generalization of a Newton–Wigner position eigenstate.[14]

THE GEOMETRICAL TRANSFORMATION OF DYNAMICAL VARIABLES

(1) Dynamical Variables in Classical Physics

In classical physics we learn the fertility of describing arbitrary physical systems in terms of material particles having no extension and fields defined over the points of space at instants of time. This conception of the nature of physical systems combined with the assumption of the arbitrarily precise, simultaneous, measurability of arbitrary sets of dynamical variables simplified the theory of measurement in classical physics to the point of triviality. In particular, the relations between the results of measurements of one and the same fundamental quantity performed by observers in two physically equivalent reference frames is very simply expressed in terms of the transformation properties of the points of space and instants of time themselves.

Thus all fundamental quantities refer to a point of the space–time manifold whether they are field-like quantities evaluated at the point of interest or particle variables evaluated at a point on the

[12] H. Salecker and E. P. Wigner, Phys. Rev. **109**, 571 (1958); C. A. Mead, *ibid.* **135**, B849 (1964).

[13] One may expect the time required for a light signal to traverse the Compton wavelength of the most massive fundamental particle to be a natural lower limit to the duration of any physical process.

[14] The corresponding proof for the familiar Newton–Wigner position operator is given by T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949). See, also, A. S. Wightman, *ibid.* **34**, 845 (1962); A. Galindo, "On the Uniqueness of the Position Operator for Relativistic Elementary Systems" (CERN Preprint, 1964).

trajectory or worldline of the particle. In either case one finds sets of quantities which, under homogeneous transformations of the space and time coordinates, transform among themselves with a transformation rule having the form,

$$A'_\alpha(P) = S^\beta_\alpha(T)A_\beta(P). \tag{1.1}$$

In this equation the $A_\alpha(\alpha = 1, \cdots, n)$ comprise the set of fundamental variables, P denotes the space–time point of interest, T the transformation of coordinates, and $S^\beta_\alpha(T)$ the dimensionless transformation coefficients for the A_α (the summation convention on β is implied). If the A_α are fieldlike quantities, then on the right-hand side of (1.1), P is replaced by (x_i, t) and on the left-hand side by (x'_i, t'); the coordinates of the point in question as determined from the unprimed and primed reference frames, respectively. If the A_α are particle variables and the particles in the system are labeled by an index i, then on the right we replace P by (i, t) and on the left by (i, t'), since the spatial coordinates cannot be known until the equations of motion are solved.

The transformations T form a group and the coefficients $S^\beta_\alpha(T)$ a matrix representation of the group in the sense that

$$S^\beta_\alpha(T_2T_1) = S^\gamma_\alpha(T_2)S^\beta_\gamma(T_1), \tag{1.2}$$

where T_2T_1 is the transformation resulting from the successive application of T_1 and then T_2. Consequently, the problem of determining the possible transformations of dynamical quantities is related to the problem of determining the matrix representations of the group of coordinate transformations between equivalent reference frames.[15]

These considerations hold equally in the relativistic and nonrelativistic cases, and that is the source of the ease with which one can make the transition from Newtonian to classical relativistic mechanics. The coordinate transformation rules are different in the two cases, but the way in which the transformation properties of fundamental quantities depend on the coordinate transformations is the same. In both cases the values of quantities *referring to one and the same space–time point* are related by the matrix elements of a representation of the transformation group.

I have belabored these well-known matters in order to emphasize the contrast with quantum mechanics.

[15] These common terms of contemporary theoretical physics are defined, among other places, in E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).

(2) Nonrelativistic Quantum Mechanics

In nonrelativistic quantum mechanics, the concept of a material particle retains a fundamental status. The notion that such a particle has no extension in space, however, loses most of its usefulness if not its meaning. Thus, there are particle variables the measurement of which requires one to forego any precise knowledge of the location of the particle. Furthermore, the fact that the momentum of the particle is just such a variable, leading to the Heisenberg uncertainty relations, means that the concept of the trajectory or world-line of a particle plays only a limiting statistical role in quantum mechanics. Nor is this situation to be attributed to mere ignorance on our part of where the particle is at any given time. We learn, very early, how the superposition principle and the interference of probabilities dooms to failure any attempt to retain such a semi-classical picture of reality.[16] One is rather forced to give up the notion that quantum mechanical particles *have* definite positions, momenta, or spins, etc., at any given time. Instead, one can *transform the particle* into a state in which, momentarily at least, a particular variable has a definite value, by performing a predictive measurement of that variable. What a quantum mechanical particle does have at any given instant is an association to a ray of a Hilbert space and a set of Hermitian operators defined on the Hilbert space, called observables, which describe a statistical relation of the possible results of measurements performed on the particle to other measurements performed on other particles prepared in an identical manner. The word particle is used to describe the system because of certain similarities between the measurements that can be performed on it and those that can be performed on classical particles. Thus it is possible to momentarily localize the system in space by what is *therefore called* a position measurement.

The only particle observables in nonrelativistic quantum mechanics which refer to even finite bounded regions R of space at a definite time are the so-called projection operators $\Pi(R)$, having possible values unity or zero corresponding to the system being confined or excluded from the region R immediately after the measurement of $\Pi(R)$. Geiger counters, occupying the region R and connected to an amplifying circuit only momentarily, at the time t, provide a crude approximation to an apparatus for measuring $\Pi_t(R)$. That the position observable itself refers collectively to all of space at

[16] For a modern proof of the absence of hidden variables see J. M. Jauch and C. Piron, Helv. Phys. Acta 36, 827 (1963).

a given instant is demonstrated mathematically by the familiar equation,

$$q_i(t) \equiv \int_{-\infty}^{\infty} x_i \, d\Pi_t(\mathbf{x}' : x_i' \leq x_i), \qquad (2.1)$$

for the Cartesian components of the position operator in terms of the projection operators for regions containing all points $\mathbf{x}'$ for which $x_i' \leq x_i$, where x_i is given.[17] The same fact is demonstrated physically by the necessity of using a battery of Geiger counters extending throughout space and sensitized simultaneously at the time t if one is to be sure of getting *some* result from a measurement of $\mathbf{q}(t)$ employing Geiger counters.

Now the homogeneous transformations of the Galilean group leave the time invariant. Furthermore, all transformations of the Galilean group leave time intervals invariant. Consequently, a measurement performed at a definite time in one inertial frame looks like a measurement performed at a definite time when viewed from another inertial frame independently of whether the measurement refers to space, pointwise, collectively or not at all. It is therefore easy and natural to fit the quantum mechanical concept of measurement into the space–time structure of nonrelativistic physics.[18]

(3) Relativistic Invariance in Quantum Mechanics

Such is not the case for relativistic physics where the coordinate transformation group is the inhomogeneous Lorentz group, or Poincaré group $\mathcal{P}$, which does not, in general, leave time intervals invariant. If a measurement at a definite time, referring to a finite region of space, or all of space, or not referring to space at all, is carried out in one Lorentz frame of reference, then the result cannot be uniquely associated with *any* definite time in another Lorentz frame moving relatively to the first. Therefore, describing the measurements of physical observables as occurring at a definite time is not a suitable procedure for formulating the transformation properties of the results of measurements.

Two famous ways exist for solving, or rather bypassing, this problem. The first way imposes the restriction that all the observables of relativistic quantum mechanical systems can be expressed in terms of a fundamental set of field operators defined over the manifold of space–time points.[19] The transformation properties of the fields can be described simply in the same way as for the classical case, and the transformation properties for arbitrary observables are implicitly contained in the prescription for constructing them out of the fields. The fields themselves do not have to be observables although that was usually assumed in the early days of quantum field theory. In general, the observables of such a formalism are integrals over all space, at some time, of a fieldlike integrand which is some local function of the basic fields. If one places integrals of such local functions over arbitrary spacelike surfaces on the same footing as the instantaneous spatial integrals, a manifestly covariant formalism results.[20] The transformation problem for arbitrary observables is then essentially solved in an explicit manner because the instantaneous observables transform into functionals on flat spacelike surfaces.

Unfortunately, this covariant formalism is simple and elegant only if all spacelike surfaces are treated equivalently, curved as well as flat. Also this solution of the *kinematical* problem of the transformation of observables depends crucially on the *dynamical* assumption of a local field theoretic description of the evolution of physical systems.

The second way of handling the problem, rather than imposing a restriction, relaxes a conventional assumption, namely, the assumption made in the introduction to this paper, that measurements can, in principle, refer to instants of time. One envisions a theory with observables referring to arbitrary bounded regions of space–time, extended in the time direction as well as through space.[21] Such regions are invariant entities and the transformation rules can be easily formulated. It is very hard, however, to decide what the physically important observables would be in such a theory. Presumably something *like* position and momentum measurements at more or less definite times exists. If so, how does one generalize these observables to finite time intervals and what are the restrictions on the smallness of the time intervals, if any? In general, the ambiguities inherent in this approach to the theory of measurement have left it essentially undeveloped except for the efforts of the axiomatic field theorists who, having already adopted the field theoretic approach, find it mathematically convenient to work with

[17] A. S. Wightman (see Ref. 14).

[18] The classic discussion of the nonrelativistic quantum theory of measurement is J. von Newmann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton University Press, Princeton, New Jersey, 1955). A recent discussion of current problems in the subject is given by J. M. Jauch, Helv. Phys. Acta 37, 293 (1964).

[19] The detailed development of the consequences of this point of view, quantum field theory, is given in many books. One of the most comprehensive is S. S. Schweber (see Ref. 5).

[20] S. Tomonaga, Progr. Theoret. Phys. (Kyoto) 1, 27 (1946); J. Schwinger, Phys. Rev. 82, 914 (1951).

[21] R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1965).

certain sets of bounded four-dimensional observables which form von Neumann rings of bounded operators.[22]

A third approach to the problem of measurement is that presented by axiomatic or analytic S-matrix theory.[23] In that theory the notion of measurements at finite times is replaced by the notion of measurements of conserved quantities at infinite times. The Poincaré group leaves infinite times infinite, and so the S-matrix theorists' conception of fundamental measurements is relativistically invariant. Also, the most interesting dynamical properties of fundamental particles are strictly conserved (in systems of several particles) only in the asymptotically infinite past and future. On the other hand, all real measurements are performed at finite times, and it is understood that the infinite time interval of S-matrix theory is a convenient idealization of the extreme ratio between the time intervals separating the initial and final measurements of a scattering experiment and the time intervals during which the interaction of the participating particles takes place. Very interesting efforts to formulate a bona fide description of macroscopic finite time and space intervals within the formalism of S-matrix theory have been made but the ideas involved must at present be regarded as very tentative.[24]

Field theory, then, provides the only extant solution to the problem of the transformation of observables in relativistic quantum mechanics. It does so, furthermore, only when functionals on arbitrary spacelike surfaces are given equal billing with instantaneous observables. Otherwise, the demand for relativistic invariance is satisfied only in the sense that a prescription exists, having the same form in each inertial frame, for constructing any given observable out of the field operators which do have manifestly covariant transformation properties. One does achieve, thereby, an equivalent description of physical systems in each inertial frame but the explicit transformation rules for most of the interesting observables are undetermined. To evaluate such a situation requires a clear understanding of the precise demands which the principle of special relativity makes on arbitrary physical theories.

A very clear and careful discussion of this problem occurs in a paper by Wigner written in 1955.[25] In that paper Wigner attributes to R. Haag the following postulates as being included in any invariance principle one may impose on a physical theory:

"(a) It should be possible to translate a complete description of a physical system from one coordinate system into every equivalent coordinate system.

(b) That the translation of a dynamically possible description be again dynamically possible. Expressed in a somewhat more simple language: a succession of events which appears possible to one observer should appear possible also to any other observer.

(c) That the criteria for the dynamical possibility of complete descriptions be identical for equivalent observers."

Wigner emphasizes that the concept of instantaneous observables referring to extended regions of space does not allow, in any obvious way, the translation required in (a) to be made since the t = const hyperplanes of space–time are different for different observers.

Clearly, if field functionals over arbitrary spacelike hyperplanes are admitted as observables, then the translation can be made since instantaneous hyperplanes transform into spacelike hyperplanes under any element of the Poincaré group. It is undesirable, however, that the introduction of hyperplane observables, which facilitate the solution of the translation problem, should occur within the context of the dynamical scheme of quantum field theory.

In the next section, hyperplane observables are introduced in a completely general way without any commitment to a particular dynamical scheme.

(4) Observables on Hyperplanes

Consider an arrangement of apparatus distributed throughout some region of space and used to measure in a predictive way some dynamical variable of the system at the time t. If the measurement is nearly instantaneous, the various pieces of apparatus are observed to be open to interaction with the physical system of interest simultaneously at the time t. What does the same measurement look like from an inertial frame moving relative to the original frame? If the various pieces of apparatus are located, roughly speaking, at the points $\mathbf{x}_i$ in the original frame, then the times, in the new frame, at which the pieces are activated are

$$t_i' = [t - (\mathbf{v}/c^2)\cdot\mathbf{x}_i](1 - v^2/c^2)^{-\frac{1}{2}}, \qquad (4.1)$$

[22] R. Haag and B. Schroer (see Ref. 6). See Ref. 21, also.

[23] H. P. Stapp, Phys. Rev. **125**, 2139 (1962); **139**, B257 (1965); D. Olive, *ibid.* **135**, B745 (1964); J. Gunson and J. G. Taylor, *ibid.* **119**, 112 (1960); J. R. Taylor, *ibid.* **140**, B187 (1965); **142**, 1236 (1966); J. Math. Phys. **7**, 181 (1966).

[24] M. L. Goldberger and K. M. Watson, Phys. Rev. **127**, 2284 (1962); M. Froissart, M. L. Goldberger, and K. M. Watson, *ibid.* **131**, 2820 (1963); R. Omnes, *ibid.* **140**, B1474 (1965).

[25] E. P. Wigner, Nuovo Cimento **3**, 517 (1956).

while the locations of the pieces at the moments of activation are

$$\mathbf{x}_i'^{\parallel} = (\mathbf{x}_i^{\parallel} - \mathbf{v}t)(1 - v^2/c^2)^{-\frac{1}{2}}, \tag{4.2}$$

$$\mathbf{x}_i'^{\perp} = \mathbf{x}_i^{\perp}. \tag{4.3}$$

In these equations $\mathbf{v}$ is the velocity of the origin of the "moving" frame relative to the original frame while the superscripts $\parallel$ and $\perp$ denote the projections of the vectors parallel and perpendicular to $\mathbf{v}$, respectively. The essential point is that in the primed frame the various pieces of apparatus are activated at different times and the measurement does not appear to be an instantaneous one at all.

The number of pieces of more or less localizable apparatus required in a measurement will differ greatly from one measurement to the next. Hence it is desirable to obtain a compact way of describing the possible combinations of position and time coordinates in the primed frame which can result from apparatus used in instantaneous measurements in the original frame. From Eqs. (4.1)–(4.3), it follows that

$$(1 - v^2/c^2)^{-\frac{1}{2}}ct_i' - (1 - v^2/c^2)^{-\frac{1}{2}}(\mathbf{v}/c)\cdot\mathbf{x}_i' = ct \tag{4.4}$$

or, introducing the notation,

$$\mathbf{n} = (1 - v^2/c^2)^{-\frac{1}{2}}\mathbf{v}/c, \tag{4.5}$$

$$\eta_0 = (1 - v^2/c^2)^{-\frac{1}{2}} = (1 + \mathbf{n}^2)^{\frac{1}{2}}, \tag{4.6}$$

we obtain

$$\eta_0 x_0' - \mathbf{n}\cdot\mathbf{x}' = ct \tag{4.7}$$

as the equation relating the space–time coordinates in the primed frame of apparatus used in an instantaneous measurement in the unprimed frame. For fixed t, (4.7) is the equation of a two-dimensional plane which sweeps through the space of the primed frame with a normal velocity of

$$u = (\eta_0/|\mathbf{n}|)c > c. \tag{4.8}$$

Each piece of apparatus is activated as the plane sweeps over it. Since the plane is moving with a velocity greater than light, any two space–time points on the moving plane are separated by a spacelike interval. Consequently, the three-dimensional manifold of space–time points generated by the moving plane is called a space-like hyperplane.[26]

According to the Haag–Wigner postulate (a), the physical equivalence of the two inertial frames requires that the result of the measurement being considered is a legitimate constituent of a complete description of the physical system as observed from the primed frame. Furthermore, since the original reference frame could be any inertial frame, it follows that the activation sequence of the pieces of apparatus is constrained only to satisfy an equation of the form

$$\eta^\mu x_\mu' = \tau, \tag{4.9}$$

where,

$$\eta^\mu\eta_\mu = 1; \qquad \eta_0 \geq 1. \tag{4.10}$$

Any such measurement appears instantaneous in *some* inertial frame, and hence the primed observer must regard *all* such measurements on an equal footing. The measurements cannot all be associated with a definite time t', but each one can be associated with a definite spacelike hyperplane (η, τ). Finally, since the primed observer is also arbitrary, one is led to conclude that an appropriate concept of measurement in the relativistic domain and consequently of observables in relativistic quantum mechanics is the concept of measurements and observables associated with, or defined on, arbitrary spacelike hyperplanes.

An idealized example of the kind of measurement discussed above is provided by the measurement of the position $\mathbf{x}(t)$ of a particle via a battery of small Geiger counters occupying the region R within which the particle is already known to be confined. By a prearranged setting of an elaborate clock mechanism, the amplification circuits for the Geiger counters are closed simultaneously for a short time interval and the position is measured.

In a moving frame, however, the circuits appear to be closed in a very rapid sequence, the closing time for any given circuit depending upon the spatial location of its associated Geiger tube. The closing sequence proceeds among the circuits more rapidly than any signal could propagate and so it is clear to the primed observer that the sequence is a result of a prearranged setting of the whole battery of apparatus. Notice also that in the primed frame the time coordinate of the result of the measurement can vary with repetitions of the measurement on identically prepared systems. Thus the primed observer sees a probability distribution in the time component of the position four-vector as well as the space components. All of this leads the primed observer to associate the measurement with a four-vector Hermitian operator $x_\mu(\eta; \tau)$,[4] where (η, τ) determine the spacelike hyperplane along which the circuit closing sequence occurs.

[26] J. L. Synge, *Relativity, The Special Theory* (North-Holland Publishing Company, Amsterdam, 1956), Appendix A.

The concept of a spacelike hyperplane is an invariant one and when, in the next section, I proceed to write down transformation equations for arbitrary hyperplane observables, it will be observables referring to one and the same hyperplane that are related by the equations. In other words, one is relating the descriptions by two observers of one physical quantity. The parameters (η, τ) which describe the single hyperplane are different for the two observers and under the Poincaré transformation

$$x'_\mu = A_\mu{}^\nu x_\nu + A_\mu; \tag{4.11}$$

the hyperplane parameters transform according to

$$\eta'_\mu = A_\mu{}^\nu \eta_\nu, \tag{4.12}$$

$$\tau' = \tau + \eta^{\mu'} A_\mu. \tag{4.13}$$

Note that η_μ is changed only by homogeneous Lorentz transformations while τ is changed only by translations. I call η_μ the *timelike normal unit vector* or simply normal vector for the hyperplane and τ the *"invariant" parameter* of the hyperplane.

The reader may wish to object to the earlier discussion on the ground that in the primed frame the measuring apparatus is moving, and it is reasonable to require that the apparatus defining an observable in a given frame be at rest in that frame. Whether such a requirement is reasonable or not, it seems likely that the same results could, in fact, be obtained in the primed frame with the use of stationary apparatus. To the extent that the Geiger counters in the cited example yield results dependent on the relative velocities of the counters and the particle one cannot regard the counters as providing an ideal measurement of the position of the particle. In general, an ideal hyperplane measurement employs apparatus which yields results dependent on where the apparatus is at the time of activation but independent of the motion of the apparatus. It is assumed that in principle such apparatus exists. This assumption is already a tacit one in nonrelativistic quantum mechanics where observers in distinct Galilean inertial frames are expected to be able, in principle, to duplicate each others' measurements with apparatus stationary in their respective frames.

A second objection that can be raised is that, after all, the original problem of relating the instantaneous observables in two different inertial frames has not been solved. *This is correct,* but what has been made clear, I believe, is that this original problem has nothing to do with relativistic invariance. Instantaneous observables in two inertial frames refer in general to different sets of hyperplanes. The theory of relativity requires only the direct translation of the results of measurements of one and the same observable. The problem of relating instantaneous observables in two frames reduces to the problem of relating instantaneous observables in one frame to hyperplane observables in the same frame, but on hyperplanes which appear instantaneous in the second frame. This problem is essentially dynamical since it requires the calculation of

$$\partial A(\eta, \tau)/\partial \eta_\mu$$

for a hyperplane observable $A(\eta, \tau)$.

Finally, one may regard the example of the idealized position measurement and even the more general discussion preceding it to be too specific or unrealistic to permit the extension, of the conclusions reached, to arbitrary observables. However, the entire problem of learning how to translate the description of observables from one frame to another appears to be that of finding an *invariant geometrical construct*[26] to which the observables refer. The *only* invariant geometrical construct which can be associated with arbitrary observables defined in one frame at a definite time, and without any further *a priori* reference to geometrical entities, is the set of all those space–time points which have the given time coordinate in the given frame, i.e., a spacelike hyperplane. Furthermore, the set of all spacelike hyperplanes is the smallest set of geometrical entities containing the instantaneous hyperplanes of a given inertial frame and being invariant under the active interpretation of the Poincaré group. In this sense the generalization of the concept of observable that is proposed here is the simplest one which is relativistically invariant.

To the reader who has found these lengthy discussions of frequently elementary topics tiring, my apologies. My only justification is the desire to leave no stone unturned in the attempt to make the general concept of hyperplane observables physically clear.

THE HYPERPLANE FORMALISM

(5) The Transformation Equations for Hyperplane Operators

The discussion is simplified if I temporarily refer only to the expectation values of the hyperplane operators. The transformation properties of the quantum mechanical operators are completely determined by the transformation properties of their expectation values in arbitrary states. Furthermore, the discussion is applicable to classical physics as

well, when expressed in terms of expectation values.

Let $A(\eta, \tau)$ then denote the expectation value of a single dynamical variable on the hyperplane characterized by (η, τ) in the original (unprimed) reference frame. Under an infinitesimal homogeneous transformation,

$$x'_\mu = x_\mu + \delta\omega_\mu{}^\nu x_\nu, \qquad \delta\omega_\mu{}^\nu = -\delta\omega_\nu{}^\mu, \tag{5.1}$$

the description of the same hyperplane changes to (η', τ), where

$$\eta'_\mu \equiv \eta_\mu + \delta\eta_\mu = \eta_\mu + \delta\omega_\mu{}^\nu \eta_\nu \tag{5.2}$$

and the expectation value becomes

$$A'(\eta', \tau) = A(\eta, \tau) + H^{\mu\nu}(\eta, \tau)\, \delta\omega_{\mu\nu}. \tag{5.3}$$

Since the infinitesimal coefficients $\delta\omega_{\mu\nu}$ are dimensionless, the dimensions of $H^{\mu\nu}$ and A must be the same. This fact motivates one to look for sets of quantities,

$$A_\alpha(\eta, \tau) \quad (\alpha = 1, \cdots, n),$$

such that

$$A(\eta, \tau) = A_1(\eta, \tau)$$

and

$$A'_\alpha(\eta', \tau) = A_\alpha(\eta, \tau) + H_\alpha{}^{\mu\nu}(\eta, \tau)\, \delta\omega_{\mu\nu}, \tag{5.4}$$

where

$$H_\alpha{}^{\mu\nu}(\eta, \tau) = C_\alpha{}^{\mu\nu\beta} A_\beta(\eta, \tau), \tag{5.5}$$

the $C_\alpha{}^{\mu\nu\beta}$ being dimensionless numerical coefficients. Such sets of quantities, called tensors,[27] are relatively easy to find and permit the formulation of equations which are form invariant under Lorentz transformations. For finite homogeneous transformations

$$x'_\mu = a_\mu{}^\nu x_\nu,$$

they satisfy

$$A'_\alpha(\eta', \tau) = S_\alpha{}^\beta(a) A_\beta(\eta, \tau), \tag{5.6}$$

where the $S_\alpha{}^\beta$ are the matrix elements of a finite-dimensional linear representation of the homogeneous Lorentz group. The infinitesimal elements of the representation are

$$S_\alpha{}^\beta(g + \delta\omega) = \delta_\alpha{}^\beta + C_\alpha{}^{\mu\nu\beta}\, \delta\omega_{\mu\nu}. \tag{5.7}$$

Under infinitesimal translations

$$x'_\mu = x_\mu + \delta a_\mu, \tag{5.8}$$

the tensors satisfy

$$A'_\alpha(\eta, \tau') = A_\alpha(\eta, \tau) + T_\alpha{}^\mu(\eta, \tau)\, \delta a_\mu, \tag{5.9}$$

where

$$\tau' = \tau + \eta^\mu\, \delta a_\mu. \tag{5.10}$$

In this case the infinitesimals δa_μ have the dimension of length, and so the $T_\alpha{}^\mu$ must be regarded as essentially new quantities. They are, of course, new tensors. Thus from

$$x''_\mu = a_\mu{}^\nu(x_\nu + \delta b_\nu) = x'_\mu + a_\mu{}^\nu\, \delta b_\nu,$$

we have

$$\begin{aligned} A''_\alpha &= S_\alpha{}^\beta(a)(A_\beta + T_\beta{}^\nu\, \delta b_\nu) \\ &= S_\alpha{}^\beta(a) A_\beta + T_\alpha{}^{\mu'} a_\mu{}^\nu\, \delta b_\nu. \end{aligned}$$

Consequently,

$$T_\alpha{}^{\mu'} a_\mu{}^\nu\, \delta b_\nu = S_\alpha{}^\beta(a) T_\beta{}^\nu\, \delta b_\nu.$$

Since δb_ν is arbitrary and

$$a_\mu{}^\nu a^\lambda{}_\nu = g^\lambda{}_\mu,$$

this yields

$$T_\alpha{}^{\lambda'} = S_\alpha{}^\beta(a) a^\lambda{}_\nu T_\beta{}^\nu, \tag{5.11}$$

and the $T_\alpha{}^\mu$ are therefore tensors. The transformation of the A_α's under finite translations can, in general, be very complicated, and the matter is not pursued further here. In cases of practical physical interest, the result is usually simple.

From the foregoing equations one can now obtain general expressions for the commutation relations of arbitrary operators $A_\alpha(\eta, \tau)$ with the generators $M_{\mu\nu}$ and P_μ of the unitary representations of the Poincaré group[28] in the quantum mechanical Hilbert space. I work exclusively in the Heisenberg picture in which the concepts of observables and measurement are foremost. The Heisenberg picture is defined by two conditions:

(i) The operators corresponding to dynamical variables carry the dependence on the hyperplane parameters (η, τ) and therefore describe the evolution of the physical system.

(ii) Under a Poincaré transformation $\{a_\mu{}^\nu, a_\mu\}$ of the coordinate frame of reference the state vectors and operators describing the system transform according to

$$|\psi\rangle \to U(\omega_\mu{}^\nu; a_\mu)\, |\psi\rangle \equiv |\psi'\rangle, \tag{5.12}$$

$$A_\alpha(\eta, \tau) \to A_\alpha(\eta', \tau'), \tag{5.13}$$

where

$$U(\omega_\mu{}^\nu; a_\mu) \equiv \exp\left[(i/\hbar) P^\lambda a_\lambda\right] \exp\left[-(i/\hbar) M^{\lambda\rho} \tfrac{1}{2} \omega_{\lambda\rho}\right] \tag{5.14}$$

[27] The restriction to tensors holds, of course, only for observables. For arbitrary dynamical variables on hyperplanes, spinors must also be admitted. This does not modify the equations in any way.

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

is the unitary operator which induces the transformation in Hilbert space.

For an arbitrary state $|\psi\rangle$ and infinitesimal $a_\mu \equiv \delta b_\mu$, it then follows that

$$\langle\psi'|\, A_\alpha(\eta', \tau')\, |\psi'\rangle = S_\alpha{}^\beta(g+\omega)\langle\psi|\, A_\beta(\eta, \tau)\, |\psi\rangle + S_\alpha{}^\beta(g+\omega)(g^\nu{}_\mu + \omega^\nu{}_\mu)\langle\psi|\, T_\beta{}^\mu(\eta, \tau)\, |\psi\rangle\, \delta b_\nu. \quad (5.15)$$

The arbitrariness of $|\psi\rangle$ permits (5.15) to be written as an operator equation,[29]

$$U(\omega; \delta b)^{-1} A_\alpha(\eta', \tau') U(\omega; \delta b) = S_\alpha{}^\beta(g+\omega)\{A_\beta(\eta, \tau) + (g_\mu{}^\nu + \omega^\nu{}_\mu) T_\beta{}^\mu(\eta, \tau)\, \delta b_\nu\}, \quad (5.16)$$

to first order in δb_ν.

Letting $\omega_{\mu\nu} = 0$ and writing

$$U(0; \delta b) = \exp[(i/\hbar) P^\lambda\, \delta b_\lambda] \simeq I + (i/\hbar) P^\lambda\, \delta b_\lambda \quad (5.17)$$

yields

$$A_\alpha(\eta, \tau + \eta\, \delta b) - A_\alpha(\eta, \tau) \equiv \frac{\partial A_\alpha(\eta, \tau)}{\partial \tau}\, \eta^\lambda\, \delta b_\lambda = \frac{i}{\hbar}[P^\lambda, A_\alpha(\eta, \tau)]\, \delta b_\lambda + T_\alpha{}^\lambda(\eta, \tau)\, \delta b_\lambda$$

or

$$[P^\lambda, A_\alpha(\eta, \tau)] = i\hbar\left\{T_\alpha{}^\lambda(\eta, \tau) - \eta^\lambda \frac{\partial A_\alpha(\eta, \tau)}{\partial \tau}\right\}. \quad (5.18)$$

On the other hand, setting $\delta b_\nu = 0$ and making $\omega_\mu{}^\nu = \delta\omega_\mu{}^\nu$ infinitesimal yields

$$U(\delta\omega; 0) \simeq I - (i/\hbar) M^{\lambda\rho} \tfrac{1}{2} \delta\omega_{\lambda\rho} \quad (5.19)$$

and

$$A_\alpha(\eta + \eta\, \delta\omega, \tau) - A_\alpha(\eta, \tau) = -\frac{i}{\hbar}[M^{\lambda\rho}, A_\alpha(\eta, \tau)]\, \frac{\delta\omega_{\lambda\rho}}{2} + C_\alpha{}^{\lambda\rho\beta} A_\beta(\eta, \tau)\, \delta\omega_{\lambda\rho}. \quad (5.20)$$

Since the $\delta\omega_{\lambda\rho}$ are antisymmetric in λ and ρ, it follows that the variation $\delta\eta_\mu = \delta\omega_{\mu\lambda}\eta\lambda$ automatically satisfies the constraint

$$\eta^\mu\, \delta\eta_\mu = 0 \quad (5.21)$$

imposed by the fixed norm of η_μ. Consequently,

$$A_\alpha(\eta + \eta\, \delta\omega, \tau) - A_\alpha(\eta, \tau) = [\partial A_\alpha(\eta, \tau)/\partial\eta_\lambda]\, \delta\omega_{\lambda\rho}\eta\rho \quad (5.22)$$

may be used in (5.20). The $\delta\omega_{\lambda\rho}$, however, cannot be regarded as independent variations until the coefficients of $\delta\omega_{\lambda\rho}$ are made manifestly antisymmetric in λ and ρ. Once that is done, the infinitesimal parameters may be removed to yield the commutation relations

$$[M^{\lambda\rho}, A_\alpha(\eta, \tau)] = i\hbar\left\{(C_\alpha{}^{\rho\lambda\beta} - C_\alpha{}^{\lambda\rho\beta}) A_\beta(\eta, \tau) + \left(\eta^\rho \frac{\partial A_\alpha(\eta, \tau)}{\partial\eta_\lambda} - \eta^\lambda \frac{\partial A_\alpha(\eta, \tau)}{\partial\eta_\rho}\right)\right\}. \quad (5.23)$$

The commutation relation for translations (5.18) is the generalization to the hyperplane formalism of the conventional Heisenberg equations of motion, and the sense in which it describes the dynamical evolution of the system is well known. Not so familiar, although emphasized in the past by Dirac,[30] is the fact that the commutation relations with the homogeneous generators of the Poincaré group may also be complicated by dynamical considerations and, conversely, provide part of the dynamical description of the system. Thus in the conventional "instant" form[30] of relativistic quantum mechanics, the generators $M_{ij}(i, j = 1, 2, 3)$ refer to spatial rotations at a given time, and their commutators with instantaneous operators have the same form whether or not interactions are present. The commutators with the M_{i0}, however, which describe the results of pure Lorentz transformations depend crucially on the presence and nature of interactions. The latter commutation relations are as hard to "solve," in the presence of interactions, as the Heisenberg equations of motion. In the hyperplane formalism, this complication is displayed by the appearance, on the right-hand side of (5.23), of the partial derivatives,

$$\partial A_\alpha/\partial\eta_\lambda.$$

It is precisely these derivatives that determine the relationship in a *single* reference frame, between corresponding dynamical variables defined on differently oriented hyperplanes, and it is in this sense that the question of relating *instantaneous* dynamical variables in two *different* frames is primarily, if not entirely, a problem in dynamics.

(6) The Hyperplane Generators and Dynamical Equations

In 1949 Dirac[30] presented a rather general investigation of several forms that could be used for the description of relativistic dynamics. Two of the

[29] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Fredrick Ungar Publishing Company, New York, 1955), pp. 200–202.

[30] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949); **34**, 592 (1962).

three forms discussed by Dirac are, in fact, special cases occurring within the hyperplane formalism. Thus Dirac's *instant form* is the conventional description in terms of instantaneous hyperplanes with

$$\eta_\mu = (1, 0).$$

The *front form* of relativistic dynamics, however, is also a (limiting) case of the hyperplane formalism. In the front form dynamical variables are described on a two-dimensional plane surface, or front, moving with the velocity of light in the direction of the normal to the surface. The space–time points lying on such a moving front comprise a limit for spacelike hyperplanes with timelike normal unit vectors approaching the limit,

$$\eta_\mu \equiv (\eta_0, \mathbf{n}) \underset{\lambda\to\infty}{\to} [(1 + \lambda^2)^{\frac{1}{2}}, \lambda\hat{n}], \tag{6.1}$$

for a front moving in the direction of the spatial unit vector $\hat{n}$. The *point form* of dynamics described by Dirac has no simple relation to the hyperplane formalism.

Dirac showed how the structure of the *ten fundamental quantities*, the Poincaré generators, varied in complexity with the choice of a form of dynamical description. Thus in the instant form the "spatial translators" P_i and "rotators" J_i are simple combinations of the basic variables of the system, whether or not interactions are present. The "time translator" P_0 and the "accelerators" M_{i0} become complicated in the presence of interactions and describe the dynamical evolution of the system. In the front form with the front moving along the positive 3-axis, the simple generators are

$$P_i, \quad (i = 1, 2), \qquad P_0 - P_3, \qquad M_{12},$$
$$M_{i0} - M_{i3}, \quad (i = 1, 2), \qquad M_{30};$$

and the complicated ones are

$$P_0 + P_3, \qquad M_{i0} + M_{i3}, \quad (i = 1, 2).$$

The fact that there are seven simple generators and only three complicated ones in the front form as opposed to the instant form in which there are six and four, respectively, reflects the singular limiting nature of the front form. Had the front been moving with a speed greater than that of light, it would have connected space–time points on a bona fide spacelike hyperplane with finite η_μ and M_{30} would then also be complicated. The invariance of the speed of light simplifies M_{30} for Dirac's front form.

Now in both forms of dynamics, those generators which leave the hyperplanes of interest invariant are simple while those generators which change the hyperplane become complicated in the presence of interactions. One may ask whether a similar separation into simple and complicated generators of the Poincaré group may be carried out for the arbitrary spacelike hyperplanes of the hyperplane formalism. If such can be done, then the construction of the generators for specific systems should be simplified, as should the problem of solving the dynamical equations once the construction is carried out.

I begin by giving a name to the quantity

$$\eta^\mu P_\mu \equiv H. \tag{6.2}$$

This operator is an invariant for a given hyperplane, and it reduces to P_0 for the reference frame in which the hyperplane (η, τ) appears instantaneous. It occurs frequently in the subsequent discussion and plays a role analogous to both the mass and the Hamiltonian of nonrelativistic quantum mechanics. On the basis of this analogy and the relativistic relation

$$E = Mc^2$$

for energies defined at a definite time, I call H/c the *invariant hyperplane mass* of the system or simply the mass. It is, of course, to be distinguished from the *rest mass* $c^{-1}(P^\mu P_\mu)^{\frac{1}{2}}$, to which it is equal only for instantaneous hyperplanes in the rest frame. It is clear from the definition of H that it acts as the generator of translations in the direction of η_μ, normal to the hyperplane on which it is defined. Consequently, H does not leave the hyperplane invariant and can be expected to be complicated in the presence of interactions.

The translators which leave the hyperplanes (η, τ) invariant and therefore are not modified by interactions are the

$$K_\mu \equiv P_\mu - \eta_\mu H. \tag{6.3}$$

The constraint

$$\eta^\mu K_\mu = 0 \tag{6.4}$$

satisfied by K_μ is an expression of the fact that K_μ leaves the hyperplane invariant and also maintains the number of independent translators at four.

The generators of homogeneous transformations $M_{\mu\nu}$ may be separated into simple and complicated parts by noting that the simple parts must reduce to the rotators J_i in those frames where $\eta_\mu = (1, \mathbf{0})$. Such generators are provided by the constrained four-vector

$$J_\mu \equiv -\tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}M^{\alpha\beta}\eta^\gamma, \tag{6.5}$$

$$\eta^\mu J_\mu = 0, \tag{6.6}$$

which leaves the hyperplane, on which it is defined, invariant. The remaining generators are suggested by the identity

$$M_{\mu\nu} \equiv \epsilon_{\mu\nu\alpha\beta} J^\alpha \eta^\beta + M_{\mu\lambda}\eta^\lambda \eta_\nu - M_{\nu\lambda}\eta^\lambda \eta_\mu, \tag{6.7}$$

and I define

$$N_\mu \equiv M_{\mu\lambda}\eta^\lambda, \tag{6.8}$$

where again

$$\eta^\mu N_\mu = 0. \tag{6.9}$$

It is to be expected that N_μ does not leave the hyperplane invariant but reorients it (changes η_μ) and is therefore complicated by interactions.

The conventional generators P_μ, $M_{\mu\nu}$ satisfy the familiar commutation relations

$$[P_\mu, P_\nu] = 0, \quad [M_{\mu\nu}, P_\lambda] = i\hbar\{g_{\nu\lambda}P_\mu - g_{\mu\lambda}P_\nu\},$$

and

$$[M_{\mu\nu}, M_{\lambda\rho}] = i\hbar\{g_{\nu\lambda}M_{\rho\mu} - g_{\nu\rho}M_{\lambda\mu} + g_{\mu\rho}M_{\lambda\nu} - g_{\mu\lambda}M_{\rho\nu}\}.$$

From these relations and the definitions (6.2, 3, 5, 8), it is straightforward to derive the equations

$$[K_\mu, K_\nu] = 0, \tag{6.10}$$

$$[K_\mu, H] = 0, \tag{6.11}$$

$$[J_\mu, H] = 0, \tag{6.12}$$

$$[J_\mu, K_\nu] = i\hbar\epsilon_{\mu\nu\alpha\beta}K^\alpha \eta^\beta, \tag{6.13}$$

$$[J_\mu, J_\nu] = i\hbar\epsilon_{\mu\nu\alpha\beta}J^\alpha \eta^\beta, \tag{6.14}$$

$$[J_\mu, N_\nu] = i\hbar\epsilon_{\mu\nu\alpha\beta}N^\alpha \eta^\beta, \tag{6.15}$$

$$[N_\mu, N_\nu] = -i\hbar\epsilon_{\mu\nu\alpha\beta}J^\alpha \eta^\beta, \tag{6.16}$$

$$[K_\mu, N_\nu] = i\hbar(g_{\mu\nu} - \eta_\mu\eta_\nu)H, \tag{6.17}$$

$$[N_\mu, H] = i\hbar K_\mu. \tag{6.18}$$

The formal similarity between these relations and those for the instant form generators $\mathbf{P}$, $H = P_0$, $\mathbf{J}$, and $\mathbf{N}$ is striking. It is not, however, surprising since these *hyperplane generators* reduce to the instant generators on instantaneous hyperplanes.

By taking the commutators of the hyperplane generators with arbitrary hyperplane operators, one verifies that the desired separation into simple and complicated generators has indeed been effected. Applying the definitions of the hyperplane generators to (5.18) and (5.23) yields

$$[H, A_\alpha] = i\hbar\{T_\alpha{}^\lambda \eta_\lambda - \partial A_\alpha/\partial\tau\}, \tag{6.19}$$

$$[K_\mu, A_\alpha] = i\hbar(g_{\mu\nu} - \eta_\mu\eta_\nu)T_\alpha{}^\nu, \tag{6.20}$$

$$[J_\mu, A_\alpha] = i\hbar\epsilon_{\mu\rho\lambda\gamma}C_\alpha{}^{\rho\lambda\beta}A_\beta\eta^\gamma, \tag{6.21}$$

and

$$[N^\mu, A_\alpha] = i\hbar\{(C_\alpha{}^{\rho\mu\beta} - C_\alpha{}^{\mu\rho\beta})\eta_\rho A_\beta + (g^{\mu\nu} - \eta^\mu\eta^\nu)\,\partial A_\alpha/\partial\eta^\nu\}. \tag{6.22}$$

The way A_α behaves under small translations (determined by the $T_\alpha{}^\mu$), and the way A_α behaves under "rotations" (determined by the $C_\alpha{}^{\mu\nu\beta}$) is not dependent on the presence or nature of interactions. Consequently, (6.20) and (6.21) express the kinematical properties of A_α alone while the dynamical evolution of the system is accounted for by (6.19) and (6.22).

The derivatives $\partial A_\alpha/\partial\eta_\nu$, which appear frequently in the preceding discussion, have not been unambiguously defined. The difficulty is a consequence of the interdependence of the η_μ by virtue of the constraint equation (4.10) or (5.21). Thus, the defining equation for hyperplane derivatives is

$$\delta A_\alpha \equiv (\partial A_\alpha/\partial\tau)\,\delta\tau + (\partial A_\alpha/\partial\eta_\nu)\,\delta\eta_\nu, \tag{6.23}$$

and from (5.21) it is clear that any term proportional to η^ν may be added to $\partial A_\alpha/\partial\eta_\nu$ without influencing (6.23). To determine this derivative uniquely, I impose the constraint

$$\eta_\nu(\partial A_\alpha/\partial\eta_\nu) = 0. \tag{6.24}$$

The recipe for calculating the derivative is then:

(i) calculate the derivative as though all the η_ν *were* independent;

(ii) project out, via the tensor $(g_{\mu\nu} - \eta_\mu\eta_\nu)$, the part of the derivative orthogonal to η_ν.

The consistency of (i) and (ii) is guaranteed by the idempotent property of the tensor $(g_{\mu\nu} - \eta_\mu\eta_\nu)$. Thus

$$(g_\mu{}^\lambda - \eta_\mu\eta^\lambda)(g_{\lambda\nu} - \eta_\lambda\eta_\nu) = g_{\mu\nu} - \eta_\mu\eta_\nu. \tag{6.25}$$

The hyperplane translators provide a simple example of the difficulty one may encounter if the η_ν are regarded as independent when taking the derivative. If such were the case then

$$\partial H/\partial\eta_\nu = P^\nu = K^\nu + \eta^\nu H,$$

and

$$\partial K_\mu/\partial\eta_\nu = -g_\mu{}^\nu H - \eta_\mu(K^\nu + \eta^\nu H).$$

Consequently,

$$0 = \partial\eta^\lambda K_\lambda/\partial\eta_\nu = K^\nu + \eta^\lambda(\partial K_\lambda/\partial\eta_\nu) = K^\nu - \eta^\nu H - K^\nu - \eta^\nu H = -2\eta^\nu H,$$

which is absurd. Employing the recipe yields

$$\partial\eta_\mu/\partial\eta_\nu = (g_\mu{}^\nu - \eta_\mu\eta^\nu), \tag{6.26}$$

$$\partial H/\partial\eta_\nu = K^\nu, \tag{6.27}$$

$$\partial K_\mu/\partial\eta_\nu = -(g_\mu{}^\nu - \eta_\mu\eta^\nu)H - \eta_\mu K^\nu, \tag{6.28}$$

and

$$\frac{\partial\eta^\lambda K_\lambda}{\partial\eta_\nu} = (g^{\nu\lambda} - \eta^\nu\eta^\lambda)K_\lambda - \eta^\lambda(g_\lambda{}^\nu - \eta^\nu\eta_\lambda)H - K^\nu \equiv 0. \tag{6.29}$$

Finally, notice that the constraint (6.24) enables one to simplify (6.22) slightly to

$$[N^\mu, A_\alpha] = i\hbar\{(C_\alpha{}^{\rho\mu\beta} - C_\alpha{}^{\mu\rho\beta})\eta_\rho A_\beta + \partial A_\alpha/\partial\eta_\mu\}. \tag{6.30}$$

(7) The Center of Mass and Interactions

In the previous section a transformation of the conventional Poincaré generators was found leading to the hyperplane generators. These latter effect a clean separation between those commutation relations which display the purely kinematical behavior of dynamical variables under Poincaré transformations, which leave the hyperplane parameters unchanged, and those which include contributions from the dynamical evolution of the system. It is physically clear that the dynamical problems posed by the derivatives $\partial A_\alpha/\partial\tau$ and $\partial A_\alpha/\partial\eta_\nu$ appearing in (6.19) and (6.30), respectively, cannot be entirely unrelated. In the first place, if they were unrelated, the hyperplane formalism would contain much more physical structure than conventional relativistic quantum mechanics. The hyperplane formalism, however, is nothing more than a reformulation of the general principles of relativistic quantum mechanics which has the merit of being manifestly covariant. In the second place, a small change in the hyperplane parameter η_ν has a much greater effect on those portions of the hyperplane far removed from the point $x_\mu{}^{(0)} = \eta_\mu\tau$ than it does on those portions near the point (see Fig. 1). Consequently, the observables associated with a physical system more or less confined to a region far removed from $x_\mu{}^{(0)} = \eta_\mu\tau$ suffer a comparatively large change from the small variation $\delta\eta_\nu$. This change can, to some extent, be counterbalanced by a corresponding change in $\delta\tau$. This change is clearly larger the greater the distance of the system from $x_\mu{}^{(0)}$. It would be very desirable to have a precise statement of this relationship between dependence on η_ν and τ, and it is obvious that some description of the "position" of the system

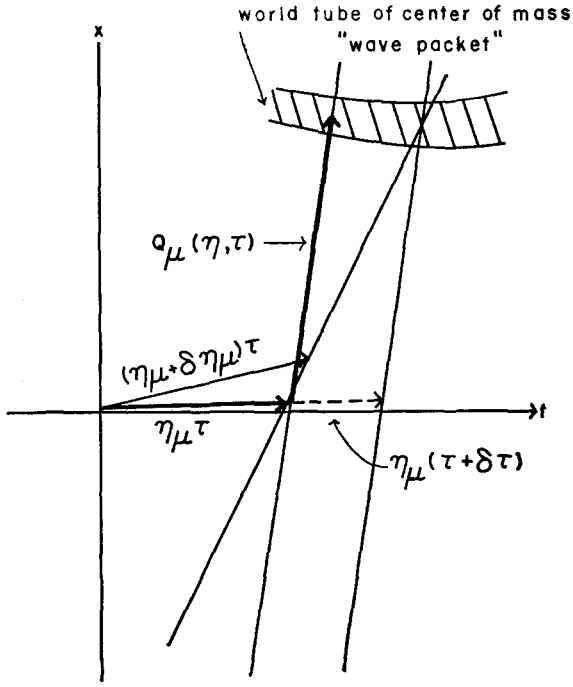


FIG. 1. Space–time diagram displaying the geometrical relation between $\delta\eta_\mu$, $\delta\tau$, and $Q_\mu(\eta, \tau)$. As Q_μ gets smaller, so does $\delta\tau$ for fixed $\delta\eta_\mu$.

relative to $x_\mu{}^{(0)}$ is needed to obtain the desired relationship. In this connection it proves convenient to introduce the notion of the center of the mass, or, more precisely, the center of the invariant hyperplane mass H/c.

Consider the manifold of state vectors describing the possible states of a single massive stable particle. This vector space is an irreducible representation space for the Poincaré group,[28] i.e., every vector in the space can be reached from any given vector in the space (up to a numerical factor) via a Poincaré transformation. From this property it follows that any single particle observable defined in the space must be expressible in terms of the Poincaré generators and multiples of the identity operator. Consequently, the possibility of introducing the concept of a center-of-mass position observable depends on the possibility of constructing a Hermitian operator function $Q_\mu(\eta, \tau)$ of the Poincaré generators which satisfies *all* the requirements one may legitimately impose on a center-of-mass position operator. Once the expression for Q_μ as a function of the generators has been obtained for a single stable particle, the same expression is used for an arbitrary closed system. This is required since the center-of-mass operator, which describes the system as a whole, must always have the same commutators with the generators for the entire system.

Let $|\Phi_1\rangle$ and $|\Phi_2\rangle$ denote arbitrary state vectors from the state spaces of two single particle systems. If these states are normalizable then the incoming direct product state,

$$|\Phi_1, \Phi_2 \text{ in}\rangle = |\Phi_1\rangle \overset{\text{in}}{\otimes} |\Phi_2\rangle, \tag{7.1}$$

is well defined.[31] This state describes a system which in the asymptotically infinite past consists of two noninteracting particles in the single particle states $|\Phi_1\rangle$ and $|\Phi_2\rangle$. The asymptotically free nature of the system requires that, for any Poincaré transformation corresponding to the unitary operator U,

$$U\,|\Phi_1, \Phi_2 \text{ in}\rangle = \{U\,|\Phi_1\rangle\} \overset{\text{in}}{\otimes} \{U\,|\Phi_2\rangle\} \tag{7.2}$$

holds.[32] This equation in turn demands that the generators satisfy

$$P_\mu\,|\Phi_1, \Phi_2 \text{ in}\rangle = \{P_\mu\,|\Phi_1\rangle\} \overset{\text{in}}{\otimes} |\Phi_2\rangle + |\Phi_1\rangle \overset{\text{in}}{\otimes} \{P_\mu\,|\Phi_2\rangle\} \tag{7.3}$$

and

$$M_{\mu\nu}\,|\Phi_1, \Phi_2 \text{ in}\rangle = \{M_{\mu\nu}\,|\Phi_1\rangle\} \overset{\text{in}}{\otimes} |\Phi_2\rangle + |\Phi_1\rangle \overset{\text{in}}{\otimes} \{M_{\mu\nu}\,|\Phi_2\rangle\}. \tag{7.4}$$

In other words, the hyperplane generators H, K_μ, J_μ, and N_μ are additive for noninteracting composite systems. The physical interpretation of H, K_μ, and J_μ as hyperplane mass, momentum, and angular momentum leads one to expect this additive property for these operators. These considerations are important since the first condition to be demanded of the center-of-mass position operator is that the product of H and Q_μ also be additive on the state (7.1). Specifically, using the notation

$$A{:}B \equiv \tfrac{1}{2}(AB + BA), \tag{7.5}$$

the additive property is

$$H{:}Q_\mu\,|\Phi_1, \Phi_2 \text{ in}\rangle = \{H{:}Q_\mu\,|\Phi_1\rangle\} \overset{\text{in}}{\otimes} |\Phi_2\rangle + |\Phi_1\rangle \overset{\text{in}}{\otimes} \{H{:}Q_\mu\,|\Phi_2\rangle\}. \tag{7.6}$$

It follows immediately from (7.6) that $H{:}Q_\mu$ must be a *linear* function of the hyperplane generators

$$H{:}Q_\mu = a_\mu H + b_\mu{}^\nu K_\nu + C_\mu{}^\nu J_\nu + d_\mu{}^\nu N_\nu. \tag{7.7}$$

The additive postulate (7.6) contains the essence of the concept of a center-of-mass position operator and is an obvious generalization of the nonrelativistic relation in which H is replaced by the total mass of the system.

Under an arbitrary translation of the coordinate system *along the hyperplane*, the operator Q_μ changes by the addition of the translation vector. Thus

$$\exp[-(i/\hbar)K^\lambda\,\delta a_\lambda]Q_\mu \exp[(i/\hbar)K^\lambda\,\delta a_\lambda] = Q_\mu + (g_\mu{}^\lambda - \eta_\mu\eta^\lambda)\,\delta a_\lambda. \tag{7.8}$$

Hence,

$$[K_\mu, Q_\nu] = i\hbar(g_{\mu\nu} - \eta_\mu\eta_\nu). \tag{7.9}$$

Taking the commutator of K_ν with (7.7) and substituting (6.10), (6.11), (6.13), (6.17), and (7.9) into the resulting equation yields

$$H(g_{\mu\nu} - \eta_\mu\eta_\nu) = C_\mu{}^\lambda\epsilon_{\nu\lambda\alpha\beta}K^\alpha\eta^\beta + d_\mu{}^\lambda(g_{\lambda\nu} - \eta_\lambda\eta_\nu)H.$$

The operators appearing in the foregoing equation are independent, except for the constraint (6.4), and therefore the coefficients must satisfy

$$(g_\mu{}^\lambda - d_\mu{}^\lambda)(g_{\lambda\nu} - \eta_\lambda\eta_\nu) = 0$$

and

$$C_\mu{}^\lambda\epsilon_{\nu\lambda\alpha\beta}\eta^\beta = 0.$$

The general solution of these equations are

$$d_\mu{}^\lambda = g_\mu{}^\lambda + h_\mu\eta^\lambda \tag{7.10}$$

and

$$C_\mu{}^\lambda = 1_\mu\eta^\lambda, \tag{7.11}$$

where h_μ and 1_μ are arbitrary. Substituting (7.10) and (7.11) back into (7.7) yields

$$H{:}Q_\mu = a_\mu H + b_\mu{}^\nu K_\nu + N_\mu. \tag{7.12}$$

Under rotations of the coordinate system *within the hyperplane*, Q_μ must behave as a hyperplane four-vector, i.e.,

$$[J_\nu, Q_\mu] = i\hbar\epsilon_{\nu\mu\alpha\beta}Q^\alpha\eta^\beta. \tag{7.13}$$

Taking the commutator of (7.12) with J_ν and again making the appropriate substitutions yields first

$$\epsilon_{\nu\mu\alpha\beta}H{:}Q^\alpha\eta^\beta = b_\mu{}^\lambda\epsilon_{\nu\lambda\alpha\beta}K^\alpha\eta^\beta + \epsilon_{\nu\mu\alpha\beta}N^\alpha\eta^\beta, \tag{7.14}$$

and then, from substitution of (7.12) into the left-hand side of (7.14),

$$\epsilon_{\nu\mu\alpha\beta}a^\alpha\eta^\beta H + \epsilon_{\nu\mu\alpha\beta}b^{\alpha\lambda}\eta^\beta K_\lambda = b_\mu{}^\lambda\epsilon_{\nu\lambda\alpha\beta}\eta^\beta K^\alpha. \tag{7.15}$$

By virtue of the constraint (6.4) no generality is lost in choosing $b_\mu{}^\nu$ to satisfy $b_\mu{}^\nu\eta_\nu = 0$. With this choice (7.15) yields

$$\epsilon_{\nu\mu\alpha\beta}a^\alpha\eta^\beta = 0$$

[31] R. Haag, Kgl. Danske. Videnskab. Selskab, Mat. fys. Medd. **29**, No. 12 (1955); Nuovo Cimento Suppl. **14**, 131 (1959).

[32] See, for example, R. F. Streater and A. S. Wightman (Ref. 1, pp. 24–27).

and

$$\epsilon_{\nu\tau\alpha\beta}\eta^{\beta}\{g^{\tau}{}_{\mu}b^{\alpha}{}_{\lambda} - g_{\lambda}{}^{\alpha}b_{\mu}{}^{\tau}\} = 0.$$

The solution of the first equation is

$$a^{\alpha} = h\eta^{\alpha}, \tag{7.16}$$

but the second one is more involved. Employing identities similar to the pair (6.5) and (6.7) for antisymmetric tensors, yields

$$g^{\tau}{}_{\mu}b^{\alpha}{}_{\lambda} - g^{\alpha}{}_{\lambda}b_{\mu}{}^{\tau} + g^{\tau}{}_{\lambda}b_{\mu}{}^{\alpha} - g^{\alpha}{}_{\mu}b^{\tau}{}_{\lambda}$$
$$= g^{\tau}{}_{\mu}b^{\nu}{}_{\lambda}\eta_{\nu}\eta^{\alpha} - \eta_{\lambda}\eta^{\alpha}b_{\mu}{}^{\tau} - \eta_{\mu}\eta^{\alpha}b^{\tau}{}_{\lambda}$$
$$- g^{\alpha}{}_{\mu}b^{\nu}{}_{\lambda}\eta_{\nu}\eta^{\tau} + \eta_{\lambda}\eta^{\tau}b_{\mu}{}^{\alpha} + \eta_{\mu}\eta^{\tau}b^{\alpha}{}_{\lambda}.$$

Contracting on τ and μ, one obtains

$$4b^{\alpha}{}_{\lambda} - g^{\alpha}{}_{\lambda}b + b_{\lambda}{}^{\alpha} - b^{\alpha}{}_{\lambda}$$
$$= 2\eta^{\alpha}b^{\nu}{}_{\lambda}\eta_{\nu} - \eta_{\lambda}\eta^{\alpha}b + \eta_{\lambda}b_{\nu}{}^{\alpha}\eta^{\nu} + b^{\alpha}{}_{\lambda},$$

where

$$b \equiv b_{\mu}{}^{\mu}.$$

Hence,

$$2b^{\alpha}_{\mu\lambda} + b_{\lambda}{}^{\alpha} = (g^{\alpha}{}_{\lambda} - \eta^{\alpha}\eta_{\lambda})b$$
$$+ 2\eta^{\alpha}\eta_{\nu}b^{\nu}{}_{\lambda} + \eta_{\lambda}\eta^{\nu}b_{\nu}{}^{\alpha}. \tag{7.17}$$

To go further in the determination of a_{μ} and $b_{\mu}{}^{\nu}$, it is necessary to invoke the constraint

$$\eta^{\mu}Q_{\mu} = 0, \tag{7.18}$$

which is a consequence of the physical interpretation of Q_{μ} as describing the position of the center of mass relative to the point $x_{\mu}{}^{(0)} = \eta_{\mu}\tau$. The space–time interval being described *lies in the hyperplane*, and therefore (7.18) must hold. Applying this to (7.12) yields

$$\eta^{\mu}a_{\mu} = 0 \quad \text{and} \quad \eta^{\mu}b_{\mu}{}^{\nu} = 0.$$

These results in turn reduce (7.16) to

$$a^{\alpha} = 0,$$

and (7.17) to

$$b_{\mu}{}^{\nu} = (\tfrac{1}{3}b)(g_{\mu}{}^{\nu} - \eta_{\mu}\eta^{\nu}) \equiv (g_{\mu}{}^{\nu} - \eta_{\mu}\eta^{\nu})\sigma.$$

The expression (7.12), then, can now be replaced by

$$H{:}Q_{\mu} = \sigma K_{\mu} + N_{\mu}. \tag{7.19}$$

Finally, σ can be determined up to an additive constant by the requirement that the total hyperplane momentum K_{μ} be contained in the center-of-mass motion, i.e.,

$$H{:}\partial Q_{\mu}/\partial\tau = K_{\mu}. \tag{7.20}$$

Since H, K_{μ}, and N_{μ} are independent of τ, this yields

$$\sigma = \tau + \sigma_0.$$

The constant σ_0 cannot be determined. It can be changed by a unitary transformation

$$e^{(i/\hbar)H\Delta\sigma_0},$$

which corresponds on the one hand to a change in the origin of the τ parameter, and on the other hand to the fact that the physical interpretation of N_{μ} has not yet been fixed. The *choice* $\sigma_0 = 0$ determines N_{μ} as the moment of the hyperplane mass at the "hyperplane time" $\tau = 0$. Therefore,

$$H{:}Q_{\mu} = \tau K_{\mu} + N_{\mu}. \tag{7.21}$$

From (7.9), (7.20) and (6.19), (6.20), it follows that

$$[Q_{\mu}, H] = i\hbar K_{\mu}/H. \tag{7.22}$$

This last commutator can be used to solve (7.21) for Q_{μ}, the result being,

$$Q_{\mu} = \{\tau K_{\mu} + N_{\mu}\}{:}H^{-1}. \tag{7.23}$$

The postulates (7.6), (7.9), (7.13), (7.18), and (7.20) along with the choice $\sigma_0 = 0$ have uniquely determined the center-of-mass position operator. If there are other, independent postulates that must be imposed on physical grounds, it is unlikely that (7.23) will satisfy them. In such a case one would have to conclude that a physically acceptable center-of-mass position operator does not exist. Further on in this section I discuss some of the alternative postulates that have been adopted by other authors writing about relativistic position operators. For the moment, however, I return to the original motivation for introducing the center-of-mass position operator; the desire to find relations between the hyperplane derivatives of arbitrary dynamical variables.

Taking the commutator of (7.21) with A_{α} and using (6.19), (6.20), and (6.30) yields

$$\frac{\partial A_{\alpha}}{\partial\eta^{\mu}} = \frac{H}{i\hbar}{:}[Q_{\mu}, A_{\alpha}] + (C_{\alpha\mu\rho}{}^{\beta} - C_{\alpha\rho\mu}{}^{\beta})\eta^{\rho}A_{\beta}$$
$$- \tau(g_{\mu\nu} - \eta_{\mu}\eta_{\nu})T_{\alpha}{}^{\nu} + Q_{\mu}{:}\left\{T_{\alpha}{}^{\nu}\eta_{\nu} - \frac{\partial A_{\alpha}}{\partial\tau}\right\}. \tag{7.24}$$

It appears, then, that the kinematical transformation properties of A_{α} and the commutator of A_{α} and Q_{μ} completely determine the derivatives $\partial A_{\alpha}/\partial\eta^{\mu}$ once the derivative $\partial A_{\alpha}/\partial\tau$ is given. Note, in passing, that (7.24) is consistent with

$$\eta^{\mu}\,\partial A_{\alpha}/\partial\eta^{\mu} = 0.$$

If interaction terms are added to the hyperplane

generators H and N_μ, then the demand that the free generators and the total generators satisfy the same commutator algebra among themselves leads to commutation relations between the interaction terms and the free generators.[30] In the conventional formulations of relativistic quantum mechanics, the physical interpretation of those commutation relations which are linear in the interaction terms has been clear. Accordingly, they have been the commutation relations which were easy to satisfy when constructing interaction terms. The physical interpretation of the nonlinear relations, however, has been obscure and the relations difficult to satisfy.[33] The center-of-mass position operator seems to provide easy access to a natural physical interpretation of the nonlinear relations, as will now be demonstrated.

The relationship between the full- and free-hyperplane generators is assumed to be

$$K_\mu = K_\mu{}^{(0)}, \tag{7.25}$$

$$J_\mu = J_\mu{}^{(0)}, \tag{7.26}$$

$$H = H^{(0)} + V, \tag{7.27}$$

and

$$N_\mu = N_\mu{}^{(0)} + U_\mu, \tag{7.28}$$

where the superscript (0) denotes the free generators. Since (6.10)–(6.18) must hold for both the free and full generators, it follows that the "potentials" V and U_μ must satisfy

$$[K_\mu{}^{(0)}, V] = 0, \tag{7.29}$$

$$[J_\mu{}^{(0)}, V] = 0, \tag{7.30}$$

$$[K_\mu{}^{(0)}, U_\nu] = i\hbar(g_{\mu\nu} - \eta_\mu\eta_\nu)V, \tag{7.31}$$

$$[J_\mu{}^{(0)}, U_\nu] = i\hbar\epsilon_{\mu\nu\alpha\beta}U^\alpha\eta^\beta, \tag{7.32}$$

$$[N_\mu{}^{(0)}, V] + [U_\mu, H^{(0)}] + [U_\mu, V] = 0, \tag{7.33}$$

$$[N_\mu{}^{(0)}, U_\nu] + [U_\mu, N_\nu{}^{(0)}] + [U_\mu, U_\nu] = 0. \tag{7.34}$$

From (7.25), (7.26), Eqs. (7.29)–(7.32) are simply statements of the hyperplane translational and rotational properties of V and U_ν. V is invariant under such transformations while U_ν is a vector which under translation is modified by

$$\exp\ [-(i/\hbar)K^\lambda a_\lambda]U_\mu \exp\ [(i/\hbar)K^\lambda a_\lambda] = U_\mu + (g_\mu{}^\lambda - \eta_\mu\eta^\lambda)a_\lambda V. \tag{7.35}$$

As they stand, the meaning of the remaining relations (7.33), (7.34) is not so clear. From (7.21), however,

$$N_\mu{}^{(0)} = H^{(0)}{:}Q_\mu{}^{(0)} - \tau K_\mu{}^{(0)} \tag{7.36}$$

and

$$U_\mu = H{:}Q_\mu - H^{(0)}{:}Q_\mu{}^{(0)} = V{:}Q_\mu + H^{(0)}{:}\Delta Q_\mu, \tag{7.37}$$

where

$$\Delta Q_\mu \equiv Q_\mu - Q_\mu{}^{(0)}. \tag{7.38}$$

Substituting these equations into (7.33) yields after several algebraic manipulations

$$[H{:}Q_\mu, H] - [H^{(0)}{:}Q_\mu{}^{(0)}, H^{(0)}] = 0. \tag{7.39}$$

Therefore if (7.22) is given for the free system, (7.39), in conjunction with (7.25), demands that (7.22) also hold for the interacting system. In other words, (7.33) is the requirement, albeit in disguised form, that *the center-of-mass motion continue to carry the total linear momentum when the interactions are turned on.* If (7.34) is rewritten as

$$[N_\mu, N_\nu] - [N_\mu{}^{(0)}, N_\nu{}^{(0)}] = 0$$

and (7.21), (7.36) substituted into the left-hand operator of each commutator, then use of (7.31) yields

$$[H{:}Q_\mu, N_\nu] - [H^{(0)}{:}Q_\mu{}^{(0)}, N_\nu{}^{(0)}] = i\hbar(g_{\mu\nu} - \eta_\mu\eta_\nu)V. \tag{7.40}$$

This equation describes the change in the behavior of the center-of-mass motion, or the moment of the hyperplane mass distribution, under pure Lorentz transformations when the interaction is turned on. A closer analogy to the interpretation of (7.39), however, is provided by looking at the hyperplane derivatives. Thus (7.39) may be written

$$\frac{\partial H{:}Q_\mu}{\partial\tau} = \left[\frac{\partial H^{(0)}{:}Q_\mu{}^{(0)}}{\partial\tau}\right]^{(0)}, \tag{7.41}$$

while (7.40) becomes

$$\frac{\partial H{:}Q_\mu}{\partial\eta^\nu} = \left[\frac{\partial H^{(0)}{:}Q_\mu{}^{(0)}}{\partial\eta^\nu}\right]^{(0)} - (g_{\mu\nu} - \eta_\mu\eta_\nu)V - \eta_\mu U_\nu. \tag{7.42}$$

Finally, it is interesting to use the center-of-mass position operator to effect a partition of the hyperplane angular momentum J_μ into an "orbital" contribution from the center-of-mass motion and the remainder term representing the internal angular momentum relative to the center of mass. The partition is effected by

$$J_\mu = -\epsilon_{\mu\alpha\beta\gamma}Q^\alpha{:}K^\beta\eta^\gamma + \Sigma_\mu \tag{7.43}$$

in close analogy with the nonrelativistic expression.

[33] L. L. Foldy [Phys. Rev. 122, 275 (1961)] has analyzed a perturbative solution of these nonlinear relations for a relativistic system of a fixed number of particles in the conventional "instant" formalism.

The internal angular momentum Σ_μ is translationally invariant,

$$[K_\mu, \Sigma_\nu] = [H, \Sigma_\nu] = 0, \tag{7.44}$$

and the further partition

$$\Sigma_\mu \equiv S_\mu^{\parallel} + (Mc/H)S_\mu^{\perp}, \tag{7.45}$$

where

$$Mc \equiv |(P^\mu P_\mu)^{\frac{1}{2}}|, \tag{7.46}$$

$$S_\mu^{\parallel} \equiv K_\mu K^\lambda \Sigma_\lambda (K^\rho K_\rho)^{-1}, \tag{7.47}$$

and

$$S_\mu^{\perp} \equiv (H/Mc)[\Sigma_\mu - K_\mu K^\lambda \Sigma_\lambda (K^\rho K_\rho)^{-1}] \tag{7.48}$$

yields the translationally invariant *spin vector* S_μ satisfying[34]

$$[S_\mu, S_\nu] = i\hbar\epsilon_{\mu\nu\alpha\beta}S^\alpha \eta^\beta. \tag{7.49}$$

Many people have studied position variables for relativistic systems in the past. Although the position variables treated have often been called the "center of mass" of the physical system, they have rarely satisfied the additive postulate (7.6) or any simple generalization of it. On the other hand, the analogus of (7.9), (7.13), and (7.20) have been almost universally invoked. Since the properties assumed here uniquely determine the center of mass, it is necessary to justify the exclusion of the more conventional assumptions which have replaced the additive postulate in other works.

By far the most frequent assumption is that of locality,[14,35] i.e.,

$$[Q_\mu, Q_\nu] = 0. \tag{7.50}$$

This assumption can be formulated in classical physics via Poisson brackets, and it shares there, with quantum mechanics, the advantage of great mathematical convenience. Its physical significance in quantum mechanics is, however, much greater than in classical physics owing to the quantum theory of measurement. Thus in classical physics (7.50) permits the use of the Q_μ as generalized coordinates, while in quantum mechanics it permits the Q_μ to possess a complete set of simultaneous eigenstates. In quantum mechanics the existence of a complete set of simultaneous eigenstates for Hermitian operators means that the operators can all be measured precisely *on the same hyperplane*, whereas such measurability is assumed in classical physics regardless of whether (7.50) is satisfied.

The center-of-mass position operator is not local. A straightforward but tedious calculation based on (7.23) yields

$$[Q_\mu, Q_\nu] = -(i\hbar/H^2)\epsilon_{\mu\nu\alpha\beta}[S^{\alpha\parallel} + (Mc/H)S^{\alpha\perp}]\eta^\beta. \tag{7.51}$$

This result *does not* prohibit one, categorically, from performing precise measurements of all the components of the center of mass on a given hyperplane. It does, however, restrict such results to measurements performed on a class of states which, for a spinning particle, contains contributions from arbitrarily high hyperplane mass or energy. On the other hand, (7.51) does demand that the statistical distributions of eigenvalues of any two components of Q_μ in a state of finite expectation value for H are constrained by a lower limit on the product of their rms deviations in the usual way. The lower limit, furthermore, depends on the expectation value of H.

It seems to me that none of these considerations clash with the *essential* characteristics of a quantum mechanical position operator. The essential characteristics being certain transformation properties under translations and rotations and the description of the average motion of the system by the expectation value of the operator.

If one does employ the locality postulate instead of the additive property in the search for a position operator, then it is well known that the result is unique for elementary systems.[14] In the hyperplane formalism the physical interpretation of the local position operator $\tilde{Q}_\mu$ is most strongly suggested by the equation

$$J_\mu = -\epsilon_{\mu\alpha\beta\gamma}\tilde{Q}^\alpha : K^\beta \eta^\gamma + S_\mu. \tag{7.52}$$

The local operator $\tilde{Q}_\mu$, however, does not possess any simple combinatorial property even for noninteracting composite systems.

More recently, a postulate concerning the behavior under pure Lorentz transformations has been invoked as an essential characteristic of a relativistic position variable.[36–38] This postulate, which has

[34] The Casimir operators of the Poincaré group are $H^2 + K_\mu K^\mu$ and $(H^2 + K_\mu K^\mu)\, S_\nu S^\nu$ and for a massive irreducible representation of "spin" s, $S_\nu S^\nu = -\hbar^2 s(s+1)$. In general S_ν does not exist for massless representations.

[35] G. N. Fleming (see Refs. 4 and 8) contains further references on this topic.

[36] D. G. Currie, J. Math. Phys. 4, 1470 (1963); "A New Class of Invariance Transformations in Classical Hamiltonian Particle Dynamics," Northeastern University preprint (1965); Phys. Rev. 142, 817 (1966); D. G. Currie and E. J. Saletan, J. Math. Phys. 7, 967 (1966). I wish to thank Professor Currie for sending me copies of his papers prior to publication.

[37] P. Havas and J. Plebanski, Bull. Am. Phys. Soc. 5, 433 (1961); H. Van Dam and E. P. Wigner, Phys. Rev. 138, B1576 (1965); 142, 138 (1966). These papers have pioneered a reinvestigation of the notion that particle interactions must proceed via "signals" with speeds not exceeding that of light.

[38] E. H. Kener, J. Math. Phys. 6, 1218 (1965); R. N. Hill, "Instantaneous Action-at-a-Distance In Classical Relativistic Mechanics," University of Delaware preprint (1966). I wish to thank Professor Hill for sending me a copy of his paper prior to publication.

been discussed most carefully by Currie and called by him the *strong world-line condition*, is the requirement that the position vectors $\mathbf{x}(t)$ and $\mathbf{x}'(t')$, describing the position from two inertial frames, trace out the same absolute world line. In quantum mechanics the demand is made on the expectation values of the position operator. Neither the center-of-mass operator nor the local operator $\tilde{Q}_\mu$ satisfy the strong world-line condition.

In the instant form of relativistic mechanics the strong world-line condition can *only* be formulated as a comparison of instantaneous position coordinates determined in two inertial frames. The existence of frame dependent world lines, as such, seems very unnatural, and in this circumstance the strong world-line condition is very compelling. The hyperplane formalism, however, provides an alternative interpretation. The distinct world lines for *instantaneous positions in various frames* are identical to the distinct world lines for *hyperplane position variables* on *variously oriented* (i.e., different η_μ values) *hyperplanes* in *one frame* of reference. Such a manifold of distinct world lines in one frame indicates that the physical system in question has a finite extension. Since, in classical physics the position vectors associated with extended systems are the result of some averaging process carried out over the system, it should not be surprising that the world line traced out by the position vector should depend on the orientation of the hyperplanes over which the averaging process is carried out. Therefore, it is only when one can regard the physical system as consisting of a *point particle* with no spatial extension that the strong world-line condition can be justified *a priori*.

That is, of course, precisely the kind of system that Currie and others have been addressing themselves to in the classical case and their conclusions can hardly be challenged there. The application of the world-line condition to quantum mechanical particles, however, can not be so easily justified. Material particles are *defined* in fundamentally different ways in relativistic classical and quantum physics. In the classical case they are defined as systems without spatial extension, this being the primary property distinguishing them from fields, the other basic ingrediant of classical physical systems. As such, a theorem of Moller prevents classical point particles from having any spin.[39] The elementary particles of quantum mechanics, on the other hand, are defined in terms of a rather abstract property of the linear manifold of their physical states,[14,28,40] and this definition accomodates the notion of spinning particles very naturally. Hence, the spatial extension of elementary particles in relativistic quantum mechanics is a problem to be investigated and not decided via *a priori* assumptions.

I have discussed elsewhere, in more detail, the case for the concept of the spatially extended elementary particle.[8]

(8) The Poincaré Invariance of the Scattering Operator and Conservation Laws

It is very problematical at present whether the same dynamical variables that are convenient to use in a complete description of systems of free particles are useful or even meaningful in the description of a relativistic quantum mechanical system undergoing interactions.[41] A strong conviction among many physicists that this is not the case has resulted in the so-called scattering operator assuming a prominent position in the theoretical analysis of relativistic particle interactions.[42] Without dwelling on the justification of the conviction or the present status of S-matrix theory compared to field theory, I would like to describe briefly here the manner in which the scattering operator enters into the hyperplane formalism and the relation between the Poincaré invariance of the scattering operator and conservation laws in the formalism.

Let the symbol α stand for a complete description of an instantaneous state of a system of free particles in the conventional formulation of relativistic quantum mechanics. The Heisenberg picture state vector

$$|\alpha, t\rangle$$

then describes a system which *at the time t is in the instantaneous state α*. Thus, although a Heisenberg state vector refers to the entire evolution of a physical system, the characterization of that evolution in terms referring to a complete set of instantaneous dynamical variables requires the specification of

[39] C. Moller, Commun. Dublin Inst. Adv. Studies **A** No. 5 (1949); see, also, *The Theory of Relativity* (Clarendon Press, Oxford, England, 1952), p. 173.

[40] Two interesting discussions of the physical meaning and mathematical definition of single particle states in relativistic quantum mechanics is given by B. Schroer, Acta Phys. Austriaca 17, 72 (1963); H. Ekstein, Commun. Math. Phys. 1, 6 (1965). I find it hard to evaluate Ekstein's no-interaction theorem because the physical interpretation of his basic postulate (see p. 10) is not nearly as clear as the strong world-line condition. Nevertheless, although direct-interaction particle theories are not my primary concern here, it seems to me that Ekstein's postulate is at variance with the notions that have been advanced here. Thus Ekstein seems to deny the possibility of dynamical content in the derivative $\partial A(\eta, \tau)/\partial\eta^\mu$.

[41] G. F. Chew, Physics 1, 77 (1964).

[42] The literature is voluminous. See, for example, Phys. Rev. 140, 7AB, Sec. 65, AB28 (1965).

the time at which the characterization applies. Now, since an arbitrary spacelike hyperplane appears instantaneous in some frame, the descriptions α can easily be generalized to refer to arbitrary hyperplanes, and I use the same symbol to denote the generalized description. Thus the meaning of the symbol

$$|\alpha;\ \eta,\ \tau\rangle$$

as a Heisenberg state vector for a free system is clear.

In order to characterize Heisenberg state vectors for fully interacting systems, it is customary to make the *dynamical* assumption that in the asymptotically infinite future and past *timelike* directions the system approaches free behavior.[31] Making the assumption here one can write

$$|\alpha;\ \eta,\ \tau(\text{in})\rangle$$

as the symbol for that Heisenberg state which, in the infinite *past* timelike directions, approaches a free state which *if there were no interactions* would assume the hyperplane state α on the hyperplane $(\eta,\ \tau)$. Similarly,

$$|\alpha;\ \eta,\ \tau(\text{out})\rangle$$

denotes the Heisenberg state that approaches, in the infinite timelike *future*, a free state which *in the absence of interactions* would assume the hyperplane state α on the hyperplane $(\eta,\ \tau)$.

The superposition principle applied to the free hyperplane states α then tells one that the (in) and (out) states are related by a linear transformation $S(\eta_2,\ \tau_2;\ \eta_1,\ \tau_1)$, independent of α, such that

$$|\alpha;\ \eta_2,\ \tau_2(\text{in})\rangle \equiv S(\eta_2,\ \tau_2;\ \eta_1,\ \tau_1)\ |\alpha;\ \eta_1\tau_1(\text{out})\rangle. \tag{8.1}$$

On the other hand, relativistic invariance demands that if, under the proper inhomogeneous transformation 1,

$$U(1)\ |\alpha;\ \eta,\ \tau(\text{in})\rangle = |\alpha';\ \eta',\ \tau'(\text{in})\rangle, \tag{8.2}$$

then

$$U(1)\ |\alpha;\ \eta,\ \tau(\text{out})\rangle = |\alpha';\ \eta',\ \tau'(\text{out})\rangle, \tag{8.3}$$

where the primes denote the description of the hyperplane state from the transformed frame. From the preceding equation it follows that

$$U(1)^{-1}S(\eta_2',\ \tau_2';\ \eta_1',\ \tau_1')U(1) = S(\eta_2,\ \tau_2;\ \eta_1,\ \tau_1). \tag{8.4}$$

This equation cannot immediately lead to commutation relations with the hyperplane generators because the hyperplane generators require a single value of η_μ for their definition. To get around this, note that (8.2), (8.3) lead to

$$i\hbar(\partial/\partial\eta\mu)\ |\alpha;\ \eta,\ \tau(\text{in, out})\rangle = N_\mu\ |\alpha;\ \eta,\ \tau(\text{in, out})\rangle, \tag{8.5}$$

so that

$$|\alpha;\ \eta_2,\ \tau(\text{in, out})\rangle = P\left\{\exp\left(-\frac{i}{\hbar}\int_{\eta_1}^{\eta_2} N_\mu(\eta)\ d\eta^\mu\right)\right\} \times |\alpha;\ \eta_1,\ \tau(\text{in, out})\rangle, \tag{8.6}$$

where the P symbol plays the role of an ordering operator in the multiple integrals of the expansion of the exponent. The ordering is somewhat more involved here than in the conventional time ordering of the Dyson P-symbol since the ordering parameter is a four-vector. The problem can, of course, be reduced to the usual one by introducing a continuous one-parameter family $\eta^\mu(\rho)$ of timelike unit four-vectors such that

$$\eta^\mu(0) = \eta_1{}^\mu, \qquad \eta^\mu(1) = \eta_2{}^\mu.$$

The ordering operator then refers to ρ in the usual way. Writing the ordered exponential operator in (8.6) as

$$T(\eta_2;\ \eta_1)$$

and ignoring here the mathematical problem of the path independence in η^μ space of $T(\eta_2;\ \eta_1)$, yields with (8.1)

$$S(\eta_2,\ \tau_2;\ \eta_1,\ \tau_1) = T(\eta_2;\ \eta_1)S(\eta_1,\ \tau_2;\ \eta_1\tau_1). \tag{8.7}$$

From (8.4) it follows that the commutation relations of the hyperplane generators with

$$S(\eta_1,\ \tau_2;\ \eta_1,\ \tau_1) \equiv S(\eta_1;\ \tau_2,\ \tau_1) \tag{8.8}$$

are

$$[K_\mu,\ S] = 0, \tag{8.9}$$

$$[J_\mu,\ S] = 0, \tag{8.10}$$

$$[H,\ S] = -i\hbar\{(\partial S/\partial\tau_1) + (\partial S/\partial\tau_2)\}, \tag{8.11}$$

$$[N_\mu,\ S] = i\hbar(\partial S/\partial\eta^\mu). \tag{8.12}$$

The first two equations assert the asymptotic conservation of hyperplane momentum and angular momentum, a consequence of the translational and rotational invariance of S. The third equation can be simplified by using (8.2), (8.3) again in the form

$$-i\hbar(\partial/\partial\tau)\ |\alpha;\ \eta,\ \tau(\text{in, out})\rangle = H\ |\alpha;\ \eta,\ \tau(\text{in, out})\rangle, \tag{8.13}$$

which eventually leads to

$$S(\eta;\tau_2,\tau_1) = e^{(i/\hbar)H(\tau_2-\tau_1)}S(\eta;0) \equiv e^{(i/\hbar)H(\tau_2-\tau_1)}S(\eta). \quad (8.14)$$

Substituting (8.14) into (8.11) yields

$$[H(\eta), S(\eta)] = 0, \quad (8.15)$$

and in (8.12) leads to

$$[N_\mu(\eta), S(\eta)] = i\hbar\,\partial S(\eta)/\partial\eta^\mu, \quad (8.16)$$

where

$$[N_\mu, e^{(i/\hbar)H(\tau_2-\tau_1)}] = i\hbar(\partial/\partial\eta^\mu)e^{(i/\hbar)H(\tau_2-\tau_1)} = -K_\mu(\tau_2-\tau_1)e^{(i/\hbar)H(\tau_2-\tau_1)}$$

has been used. Equation (8.15) asserts the asymptotic conservation of the hyperplane mass, and its derivation demonstrates the dependence of this result on the τ-translational invariance of $S(\eta;\tau_2,\tau_1)$.

If (7.21) is used to rewrite N_μ in (8.16), then (8.9), (8.10), (8.15) allow (8.16) to be replaced by

$$H{:}[Q_\mu, S] = i\hbar(\partial S/\partial\eta^\mu), \quad (8.17)$$

and the asymptotic conservation of the center-of-mass motion is related to the hyperplane orientation independence of S. The proof of this independence involves the assumption that *any* (in) or (out) state may be described in terms of free hyperplane states on *any* hyperplane. In other words, if

$$|\alpha;\eta,\tau(\text{in, out})\rangle \equiv |\psi\rangle,$$

then for any (η',τ') there exists β such that

$$|\beta;\eta',\tau'(\text{in, out})\rangle = |\psi\rangle.$$

This postulate is related to the aspect of relativistic invariance which demands that if one observer can describe a system in terms of (η,τ) hyperplane states than a second observer, whose (η,τ) hyperplanes are (η',τ') hyperplanes for the first observer, must also be able to employ that mode of description. This plus the translatability of a given description yields the postulate. The postulate in turn yields

$$|\alpha;\eta,\tau(\text{in})\rangle = |\beta;\eta',\tau'(\text{in})\rangle$$

and

$$|\alpha;\eta,\tau(\text{out})\rangle = |\beta;\eta',\tau'(\text{out})\rangle.$$

But since

$$S(\eta)\,|\alpha;\eta,\tau(\text{out})\rangle = |\alpha;\eta,\tau(\text{in})\rangle = |\beta;\eta',\tau'(\text{in})\rangle = S(\eta')\,|\beta;\eta',\tau'(\text{out})\rangle,$$

it must be that

$$S(\eta) = S(\eta') \quad (8.18)$$

and

$$[Q_\mu, S] = 0. \quad (8.19)$$

Finally, since the norm of a free hyperplane state does not depend on when that state is approached, it must be true that

$$\langle\alpha;\eta,\tau(\text{in})\,|\alpha;\eta,\tau(\text{in})\rangle = \langle\alpha;\eta,\tau(\text{out})\,|\alpha;\eta,\tau(\text{out})\rangle$$

for arbitrary α. Hence

$$S^+S = I = SS^+ \quad (8.20)$$

and the *scattering operator* S is unitary.

(9) Hyperplane Field Theory of Spinless Particles

In this section I would like to discuss a field-theoretic mode of description which appears quite natural in the hyperplane formalism. The description differs in several respects from what is to be called *conventional local quantum field theory* or c.l.q.f.t.[5] Whether the differences can be significant for the formulation of dynamical theories is a question that is not pursued at length here. Accordingly, I restrict the discussion to the hyperplane field-theoretic description of free spinless particles. Within that limited domain the major difference is that the hyperplane field operators do not satisfy the microcausality condition which is a basic property of c.l.q.f.t.[6] Nevertheless, the physical content of the description of free spinless particles is identical with that provided by c.l.q.f.t.

To ensure the truth of the last statement, I begin by considering the creation and destruction operators for momentum eigenstates of such particles

$$\tilde\Phi^+(\mathbf{p},t);\quad \tilde\Phi(\mathbf{p},t),$$

where the time dependence (trivial in the case of free particles) is included to emphasize the essentially noncovariant character of these operators. Whether interactions are present or not, these operators satisfy the commutation relations[43]

$$[\mathbf{P}, \tilde\Phi(\mathbf{p},t)] = -\mathbf{p}\tilde\Phi(\mathbf{p},t) \quad (9.1)$$

and

$$[H, \tilde\Phi(\mathbf{p},t)] = -i\hbar(\partial/\partial t)\tilde\Phi(\mathbf{p},t), \quad (9.2)$$

where $\mathbf{P}$ and H are the conventional total momentum and energy operators. In the absence of interactions, (9.2) is equivalent to

$$[H, \tilde\Phi(\mathbf{p},t)] = -c(\mathbf{p}^2 + m^2c^2)^{\frac{1}{2}}\tilde\Phi(\mathbf{p},t), \quad (9.3)$$

[43] If interactions are present, of course, then the mass spectrum of the state $\tilde\Phi^+(p,t)\,|0\rangle$ is not that of a single stable particle. See H. Lehmann, Nuovo Cimento 11, 342 (1954); P. J. Peebles, Phys. Rev. 128, 1412 (1962).

so that

$$\tilde{\Phi}(\mathbf{p}, t) = \tilde{\Phi}(\mathbf{p}) \exp [-(i/\hbar)(\mathbf{p}^2 + m^2c^2)^{\frac{1}{2}}ct]. \quad (9.4)$$

In accordance with the philosophy behind the hyperplane formalism, the instantaneous destruction operator $\tilde{\Phi}(\mathbf{p}, t)$ must be replaced by a destruction operator associated with a given hyperplane (η, τ). Furthermore, the description of the momentum $\mathbf{p}$ that is removed from the state by the action of $\tilde{\Phi}(\mathbf{p}, t)$ must be replaced by a four-vector variable k_μ denoting the hyperplane momentum removed from the state. The resulting hyperplane destruction operator may be written

$$\tilde{\Phi}(k; \eta, \tau).$$

The fact that the hyperplane momentum operator satisfies the constraint equation (6.4) places an awkward limitation on the domain of definition of the variable k_μ. It is more convenient for calculations to enlarge the domain of definition to the entire four-dimensional k-space and impose the constraint equation

$$\eta^\mu(\partial/\partial k^\mu)\tilde{\Phi}(k; \eta, \tau) = 0 \quad (9.5)$$

on the destruction operator. The ambiguity thereby introduced into the result of applying $\tilde{\Phi}$ to a state vector is removed by the commutation relations

$$[K_\mu, \tilde{\Phi}(k; \eta, \tau)] = -q_\mu\tilde{\Phi}(k; \eta, \tau), \quad (9.6)$$

$$[H, \tilde{\Phi}(k; \eta, \tau)] = -i\hbar(\partial/\partial\tau)\tilde{\Phi}(k; \eta, \tau)$$
$$= -(m^2c^2 - q^2)^{\frac{1}{2}}\tilde{\Phi}(k; \eta, \tau), \quad (9.7)$$

where

$$q_\mu \equiv k_\mu - \eta_\mu(\eta k). \quad (9.8)$$

The limit

$$\lim_{\substack{\eta_0 \to 1 \\ \tau \to ct}} \tilde{\Phi}(k; \eta, \tau) = \tilde{\Phi}(\mathbf{k}, t) \quad (9.9)$$

is consistent with these definitions.

In c.l.q.f.t. the space–time field operators would now be introduced via the three-dimensional instantaneous Fourier transform

$$\tilde{\Phi}(\mathbf{x}, t) = (2\pi\hbar)^{-\frac{3}{2}} \int \frac{d^3p}{2p_0} \{\tilde{\Phi}(\mathbf{p}, t)e^{(i/\hbar)\mathbf{p}\cdot\mathbf{x}}$$
$$+ \tilde{\Phi}^+(\mathbf{p}, t)e^{(i/\hbar)\mathbf{p}\cdot\mathbf{x}}\}, \quad (9.10)$$

and relativistic invariance of the theory is guaranteed by the requirement that $\Phi(\mathbf{x}, t)$ transform like a scalar

$$\Phi'(\mathbf{x}', t') = \Phi(\mathbf{x}, t). \quad (9.11)$$

This requirement (9.11) then implicitly determines the transformation properties of $\tilde{\Phi}(\mathbf{p}, t)$. In particular, when no interactions are present, the invariance of the differential element $d^3p/2p_0$ plus (9.4) yields the scalar transformation rule

$$\tilde{\Phi}(\mathbf{p}')' = \tilde{\Phi}(\mathbf{p}),$$

which usually provokes the manifestly covariant notation,

$$\tilde{\Phi}(\mathbf{p}) \equiv \tilde{\Phi}(p_\mu); \quad p_0 = (\mathbf{p}^2 + m^2c^2)^{\frac{1}{2}}.$$

In the hyperplane formalism, $\tilde{\Phi}(k; \eta, \tau)$ is defined at the outset to satisfy

$$\tilde{\Phi}'(k'; \eta', \tau') = \tilde{\Phi}(k; \eta, \tau), \quad (9.12)$$

and the transformation to a space–time field operator can be effected via the generalization (from nonrelativistic quantum field theory)

$$d^3p \to d^4k\ \delta(\eta k) \quad (9.13)$$

rather than

$$d^3p \to (mc/p_0)\ d^3p$$

as in c.l.q.f.t. Using (9.13) one obtains

$$\Phi(x; \eta, \tau) = (2\pi\hbar)^{-\frac{3}{2}} \int d^4k\ \delta(\eta k)$$
$$\times \{\tilde{\Phi}(k; \eta, \tau)e^{-(i/\hbar)kx} + \tilde{\Phi}^+(k; \eta, \tau)e^{(i/\hbar)kx}\}, \quad (9.14)$$

and the resulting constraint

$$\eta^\mu(\partial/\partial x^\mu)\Phi(x; \eta, \tau) = 0 \quad (9.15)$$

keeps the number of independent variables from proliferating unjustifiably.

Combining (9.6), (9.7), and (9.14) yields the Heisenberg equations of motion

$$[K_\mu, \Phi(x; \eta, \tau)] = -i\hbar(\partial/\partial x^\mu)\Phi(x; \eta, \tau), \quad (9.16)$$

$$[H, \Phi(x; \eta, \tau)] = -i\hbar(\partial/\partial\tau)\Phi(x; \eta, \tau). \quad (9.17)$$

Furthermore, the scalar transformation property

$$\Phi'(x'; \eta', \tau') = \Phi(x; \eta, \tau), \quad (9.18)$$

which follows from (9.12) and (9.14), yields the commutators

$$[J_\mu, \Phi(x; \eta, \tau)]$$
$$= -i\hbar\epsilon_{\mu\alpha\beta\gamma}x^\alpha(\partial/\partial x_\beta)\Phi(x; \eta, \tau)\eta^\gamma \quad (9.19)$$

and

$$[N_\mu, \Phi(x; \eta, \tau)]$$
$$= i\hbar\{(\eta x)(\partial/\partial x^\mu) + (\partial/\partial\eta^\mu)\}\Phi(x; \eta, \tau). \quad (9.20)$$

Assuming the *equal hyperplane* commutation relations, for free fields,[44]

$$[\tilde{\Phi}(k'; \eta, \tau), \tilde{\Phi}(k; \eta, \tau)] = 0, \tag{9.21}$$

$$\delta(\eta k - \eta k')[\tilde{\Phi}(k'; \eta, \tau), \tilde{\Phi}^{+}(k; \eta, \tau)] = \delta^4(k - k') \tag{9.22}$$

yield

$$[\Phi^{(-)}(x'; \eta, \tau'), \Phi^{(-)}(x; \eta, \tau)] = 0, \tag{9.23}$$

$$[\Phi^{(-)}(x'; \eta, \tau'), \Phi^{(+)}(x; \eta, \tau)] = C(x - x', \tau - \tau'; \eta), \tag{9.24}$$

where

$$\Phi^{(-)}(x; \eta, \tau) = (2\pi\hbar)^{-\frac{3}{2}} \int d^4k\ \delta(\eta k)\tilde{\Phi}(k; \eta, \tau)e^{-(i/\hbar)kx}, \tag{9.25}$$

$$\Phi^{(+)}(x; \eta, \tau) = \{\Phi^{(-)}(x; \eta, \tau)\}^{+}, \tag{9.26}$$

and

$$C(x - x'; \tau - \tau'; \eta) = (2\pi\hbar)^{-3} \int d^4k\ \delta(\eta k)e^{(i/\hbar)[h(\tau - \tau') + k(x - x')]}, \qquad h = (m^2c^2 - k^2)^{\frac{1}{2}}. \tag{9.27}$$

The function C is the space–time propagator for the free field in the sense that

$$\int d^4x'\ \delta(\eta x')C(x - x', \tau - \tau'; \eta)\Phi^{(+)}(x'; \eta, \tau') = \Phi^{(+)}(x; \eta, \tau). \tag{9.28}$$

This function displays the nonlocality of the field since

$$C(x, \tau; \eta) \not\equiv 0 \tag{9.29}$$

for

$$\tau^2 + [x - \eta(\eta x)]^2 < 0. \tag{9.30}$$

In fact, the state

$$\Phi^{(+)}(x; \eta, \tau)\ |0\rangle = |x; \eta, \tau\rangle \tag{9.31}$$

is a Newton–Wigner position eigenstate generalized to the hyperplane (η, τ), and the nonlocal behavior of these states is well known.[14] The position operator appearing in (7.51) satisfies

$$\tilde{Q}_\mu(\eta, \tau)\ |x; \eta, \tau\rangle = [x_\mu - \eta_\mu(\eta x)]\ |x; \eta, \tau\rangle. \tag{9.32}$$

Thus the field $\Phi(x; \eta, \tau)$ has a more direct physical interpretation than the c.l.q.f.t. operator $\Phi(x)$ since the former is a linear combination of creation and annihilation operators for position eigenstates.

The commutation relations (9.16), (9.17), and (9.19) are easily "solved" by choosing[45]

$$K_\mu = \int d^4x\ \delta(\eta x)\mathcal{K}_\mu(x; \eta, \tau), \tag{9.33}$$

$$J_\mu = \int d^4x\ \delta(\eta x)\epsilon_{\mu\alpha\beta\gamma}x^\alpha\mathcal{K}^\beta(x; \eta, \tau)\eta^\gamma, \tag{9.34}$$

and

$$H = \int d^4x\ \delta(\eta x)\mathcal{H}(x; \eta, \tau), \tag{9.35}$$

where

$$\mathcal{K}_\mu(x; \eta, \tau) = \tfrac{1}{2}i\hbar\Phi^{(+)}(x; \eta, \tau) \times (\partial^{\leftrightarrow}/\partial x^\mu)\Phi^{(-)}(x; \eta, \tau) \tag{9.36}$$

and

$$\mathcal{H}(x; \eta, \tau) = \tfrac{1}{2}i\hbar\Phi^{(+)}(x; \eta, \tau) \times (\partial^{\leftrightarrow}/\partial \tau)\Phi^{(-)}(x; \eta, \tau). \tag{9.37}$$

These hyperplane generator densities satisfy, as a consequence of the field equation,

$$\left[\frac{\partial^2}{\partial\tau^2} + \frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x_\mu} + \frac{m^2c^2}{\hbar^2}\right]\Phi(x; \eta, \tau) = 0, \tag{9.38}$$

the analog of the local conservation equation for the stress–energy–momentum tensor of c.l.q.f.t.

$$\partial\mathcal{H}/\partial\tau = -\partial\mathcal{K}_\mu/\partial x_\mu. \tag{9.39}$$

Accordingly, it is tempting to write

$$N_\mu = \int d^4x\ \delta(\eta x)\{x_\mu\mathcal{H}(x; \eta, \tau) - \tau\mathcal{K}_\mu(x; \eta, \tau)\}, \tag{9.40}$$

pursuing the analogy further. The justification of this last equation (which is valid) would take me too far afield of my desire to present a short survey of the elements of the hyperplane field theory of free spinless particles. Suffice it to say that the recon-

[44] This set of fundamental commutation relations is equivalent to the canonical quantization scheme only in the case of free fields. Furthermore, even for free fields the *space–time field*, $\Phi(x; \eta, \tau)$, does not satisfy canonical commutation relations such as $\delta(\eta x - \eta x')\ [\Phi(x'; \eta, \tau), (\partial/\partial\tau)\Phi(x; \eta, \tau)] = i\hbar\ \delta^4(x - x')$. This is probably the greatest drawback to the use of such fields in the description of interacting systems, since one does not have a canonical scheme for equal hyperplane commutation relations to invoke. On the other hand, such canonical schemes have almost dropped out of the picture in the modern approach to quantum field theory. Their residue, in the form of the microcausality postulate, would be replaced here by the equal hyperplane commutation relation $[\Phi(x'; \eta, \tau), \Phi(x; \eta, \tau)] = 0$ for $x_\mu - \eta_\mu(\eta x) \neq x'_\mu - \eta_\mu(\eta x')$.

[45] Equations (9.33)–(9.35) also "solve" the commutation relations (6.10)–(6.14).

ciliation of (9.40) with (9.20) involves the application of (7.24) to the hyperplane field operator

$$\frac{\partial\Phi}{\partial\eta^\mu} = \frac{H}{i\hbar}:[Q_\mu, \Phi] - \tau\frac{\partial\Phi}{\partial x^\mu} - Q_\mu\cdot\frac{\partial\Phi}{\partial\tau}. \qquad (9.41)$$

The fact that this section began with the generalization to arbitrary hyperplanes of the familiar momentum space creation and destruction operators for spinless particles guarantees that the resulting hyperplane field theory has the same physical content as c.l.q.f.t. of free particles. If interactions are introduced, however, the choice of a local function of the field operators for the current operator j appearing in

$$\left(\frac{\partial^2}{\partial\tau^2} + \frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x_\mu} + \frac{m^2c^2}{\hbar^2}\right)\Phi(x;\eta,\tau) = j(x;\eta,\tau) \qquad (9.42)$$

does not yield dynamical results equivalent to those obtained from assuming local interactions in c.l.q.f.t.[46] In short, for a scalar field, the dynamics of

$$j(x;\eta,\tau) = \lambda\Phi^2(x;\eta,\tau) \qquad (9.43)$$

may be expected to be very different from that of

$$j(x) = \lambda\Phi^2(x)$$

in c.l.q.f.t. Such differences are to be studied elsewhere.

SUMMARY AND CONCLUSION

I have presented a formulation of some of the basic concepts and principles of conventional relativistic quantum mechanics in a manifestly covariant manner. The underlying physical idea of the entire development is that only by treating observables or dynamical variables on arbitrary spacelike hyperplanes equivalently can one achieve full consistency with the requirements of special relativity. The resulting hyperplane formalism is a minimal generalization of the conventional "instant" formalism, since the class of spacelike hyperplanes has no proper subclass which is invariant under the Poincaré group. The introduction of the hyperplane generators clarifies the relation between the dynamical and kinematical properties of dynamical variables in the relativistic domain. The axiomatic study of the center-of-mass position operator crystalizes the physical interpretation of the generator N_μ as well as shedding light on some of the commutation relations between the free and interacting parts of the generators. The asymptotic conservation laws associated with the Poincaré invariance of the scattering operator was considered and led to the relation between the law of center-of-mass motion and the independence of S on η_μ. Finally, a nonlocal hyperplane-dependent field theory of free spinless particles was displayed which suggests interesting avenues of research in the study of dynamical models of interacting particles.

In the near future I hope to present a thorough treatment of the hyperplane field theory of free particles of arbitrary spin as well as considering some very perplexing problems in the (really nonexistent) relativistic quantum theory of measurement. More distant are studies of a fully covariant model field theory with persistent single-particle states,[47] covariant helicity amplitudes,[48] and the formulation of "phenomenological" causality[49] and the proof of dispersion relations within the framework of hyperplane field theory.

ACKNOWLEDGMENTS

I have at various times benefited from enlightening conversations with Professor B. Schroer and Professor T. F. Jordan, at the University of Pittsburgh, and Professor E. H. Kerner and Professor R. N. Hill, of the University of Delaware, on topics related to the subject of this paper.

[46] The nonlocal character of the hyperplane field operators should not be taken as an implication that the theories obtained from (9.42), (9.43) say, would be more convergent than conventional field theories. The free field propagator (9.27) still has singularities on the light cone, and the absence of the $(1/p_0)$ factor in the definition (9.14) of $\Phi(x;\eta,\tau)$ gives the space–time field an even larger contribution from the high energy region than in c.l.q.f.t.

[47] M. Dresden and P. B. Kahn, Rev. Mod. Phys. **34**, 401 (1962).

[48] Here I am referring to matrix elements of the scattering operator between eigenstates of the hyperplane invariant operator $K^\mu S_\mu/|K^\lambda K_\lambda|^{1/2}$.

[49] E. P. Wigner, in *Dispersion Relations and their Connection with Causality* (Academic Press Inc., New York, 1964), pp. 40–67.

The Asymptotic Theory of Cerenkov Radiation*

R. HANDELSMAN†‡ AND R. M. LEWIS
Courant Institute of Mathematical Sciences, New York University, New York City
(Received 21 April 1966)

Recently developed methods of asymptotic analysis are applied to the problem of Cerenkov radiation. The mathematical description of this physical phenomenon is given by the integro-differential system of equations for the electromagnetic field in a dispersive medium. The parameter λ is introduced into these equations, where λ is a characteristic frequency of the medium. It is for large λ that the asymptotic expansion of the electromagnetic field is sought. Isotropic, uniaxial crystalline and gyrotropic media are treated in detailed. The source function which appears in the field equations is taken to be quite general, e.g., it may be used to represent the current associated with any moving "multipole" source. By applying the method of stationary phase to an integral representation of the solution, the leading term of the asymptotic expansion of the electromagnetic field is obtained. More precisely, a parametric representation of the expansion is found in which certain space–time curves called "rays" play a key role. An expression for the total energy of the radiation is then determined.

INTRODUCTION

IN some of the previously published works on Cerenkov radiation, including the original work of Frank and Tamm,[1] asymptotic methods have been used to obtain results. These methods, however, have not been employed in a systematic way. For the most part, they involve the use of known asymptotic expansions of special functions that appear in the analysis. In this paper, we present a complete and systematic asymptotic theory which enables us to obtain approximate solutions to a wide variety of problems that arise in the study of Cerenkov radiation.

In this paper, the problem of Cerenkov radiation in infinite homogeneous media is considered. Isotropic, uniaxial crystalline, and gyrotropic media are treated in detail. In future papers inhomogeneous media and interface problems are to be treated. It is found that the results of the present paper are essential in the asymptotic analysis of these more difficult problems.

Our mathematical description of Cerenkov radiation is based on the time-dependent form of Maxwell's equations for "dispersive media." In this case, the constituitive equation takes the form of a convolution integral. Thus, we are led to consider the asymptotic solution of an integro-differential system of equations. The methods to be employed have been developed[2] for a more general system of equations. What distinguishes Cerenkov radiation from other electromagnetic phenomena is the type of source function that appears in the field equations. As is well known, Cerenkov radiation can only occur when the source is moving. It is shown in Sec. 2 that the type of source function considered in this paper is quite general. In fact, we find that it may be used to represent the current associated with any multipole source moving along an arbitrary trajectory. Moreover, the source function is allowed to have an "oscillatory factor" so that the Cerenkov–Doppler effect may be treated.

As is pointed out in Sec. 2, the large expansion parameter λ is a characteristic frequency of the medium. The true meaning of our asymptotic expansion is better understood when an equivalent dimensionless parameter λ_0 is found. If dimensionless variables are introduced throughout, we find that $\lambda_0 = \lambda a/c$, where "a" is a characteristic dimension of the problem and c is the speed of light in free space. We can conclude that the correct interpretation of our expansion is that it is valid for $1 \ll \lambda_0$. Since, in this paper, we deal with infinite homogeneous media, "a" can be taken to be the distance from the source trajectory to the point in space at which the solution is to be obtained.

From results given in Ref. 2, an integral representation of the Fourier type is obtained for the electromagnetic field which is valid for both isotropic and anisotropic media. By applying the method of stationary phase in several dimensions to this integral, parametric representations for the leading terms of the asymptotic expansions of the fields are found. In these representations, certain

* This research was supported by the Office of Naval Research under Contract NONR 285(48). Reproduction in whole or in part is permitted for any purpose of the United States Government.

† This paper is based on part of R. Handelsman's Doctoral thesis, *The Asymptotic Theory of Cerenkov Radiation*, submitted to New York University, (1965).

‡ Present address: Division of Applied Mathematics, Brown University, Providence, Rhode Island.

[1] I. M. Frank and Ig. Tamm, Compt. rend. acad. sci. U.R.S.S. **14**, 109 (1937).

[2] R. M. Lewis, Arch. Ratl. Mech. Anal. **20**, 3 (1965).

straight lines called "rays" play a central role. Once the fields have been determined, an expression for what we call the "average asymptotic energy" is obtained. We emphasize the fact that all our results are valid for a wide class of moving sources. Whenever possible, we specialize these results to sources which have been treated by other authors for the purpose of comparison. In all cases, the comparison yields perfect agreement.

CERENKOV RADIATION IN INFINITE HOMOGENEOUS MEDIA

1. The Electromagnetic Equations for Dispersive Media and the Fourier Integral Representation of the Solution

In Gaussian units, Maxwell's equations take the form

$$\mathbf{D}_t - c\nabla\times\mathbf{H} = -4\pi\mathbf{J}, \quad \mathbf{B}_t + c\nabla\times\mathbf{E} = 0, \tag{1.1}$$

$$\nabla\cdot\mathbf{D} = 4\pi\rho, \quad \nabla\cdot\mathbf{B} = 0. \tag{1.2}$$

Here, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{J}$ are t-dependent 3-vectors and $\mathbf{X} = (x_1, x_2, x_3)$. The source functions $\rho(t, \mathbf{X})$ and $\mathbf{J}(t, \mathbf{X})$ must satisfy the continuity equation

$$\rho_t + \nabla\cdot\mathbf{J} = 0. \tag{1.3}$$

It follows from (1) and (3) that

$$\partial/\partial t(\nabla\cdot\mathbf{D} - 4\pi\rho) = 0 \quad \text{and} \quad \partial/\partial t(\nabla\cdot\mathbf{B}) = 0. \tag{1.4}$$

Therefore, if Eqs. (2) are satisfied at any time t, they are satisfied for all time.

We assume that the source and fields are identically zero for $t < 0$, i.e.,

$$\rho(t, \mathbf{X}) \equiv \mathbf{J}(t, \mathbf{X}) \equiv \mathbf{H}(t, \mathbf{X}) \equiv \mathbf{E}(t, \mathbf{X}) \equiv \mathbf{B}(t, \mathbf{X}) \equiv \mathbf{D}(t, \mathbf{X}) \equiv 0; \qquad t < 0. \tag{1.5}$$

Eq. (1.5) implies that Eqs. (1.2) are satisfied for $t < 0$ and hence are satisfied for all t.

Dispersive media are characterized by the fact that $\mathbf{D}(t, \mathbf{X})$ does not depend on the value of $\mathbf{E}(t, \mathbf{X})$ at time t alone as in nondispersive media, but rather depends on the values of $\mathbf{E}(t, \mathbf{X})$ at earlier times as well (see p. 248 of Ref. 3). There is a similar dependence of $\mathbf{B}(t, \mathbf{X})$ on $\mathbf{H}(t, \mathbf{X})$. These relationships are most simply expressed by the constituitive equation

$$\mathbf{v}(t, \mathbf{X}) = \int_0^\infty \mathfrak{F}(\tau)\mathbf{u}(t - \tau, \mathbf{X})\, d\tau. \tag{1.6}$$

In (1.6), we have introduced the column vectors $\mathbf{u}$ and $\mathbf{v}$ having 6 components defined by

$$\mathbf{u} = [\mathbf{E}, \mathbf{H}] = (E_1, E_2, E_3, H_1, H_2, H_3),$$
$$\mathbf{v} = [\mathbf{D}, \mathbf{B}] = (D_1, D_2, D_3, B_1, B_2, B_3). \tag{1.7}$$

[In what follows other column vectors $\mathbf{a}$ with 6 components are defined by an ordered pair of two 3-vectors. Thus $\mathbf{a} = [\mathbf{A}, \mathbf{B}] = (A_1, A_2, A_3, B_1, B_2, B_3)$.] $\mathfrak{F}(t)$ is a 6×6 matrix which is a real function of time. In matrix block notation $\mathfrak{F}(t)$ takes the form

$$\mathfrak{F}(t) = \begin{bmatrix} F_1(t) & 0 \\ 0 & F_2(t) \end{bmatrix}. \tag{1.8}$$

Here, $F_1(t)$ and $F_2(t)$ are real 3×3 matrices. Furthermore, we assume that the *causality condition*

$$\mathfrak{F}(t) \equiv 0 \quad \text{for} \quad t < 0 \tag{1.9}$$

is satisfied.

We now define the matrix $\hat{\mathcal{E}}(\hat{\omega})$ to be the Fourier transform of $\mathfrak{F}(t)$. That is,

$$\mathfrak{F}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\hat{\omega}t}\hat{\mathcal{E}}(\hat{\omega})\, d\hat{\omega}. \tag{1.10}$$

In Gaussian units, $\hat{\mathcal{E}}(\hat{\omega})$ is dimensionless and therefore must be a function of the dimensionless variable $\omega = \hat{\omega}/\lambda$, where λ is a characteristic frequency of the medium. (For example, in an isotropic plasma we may take λ to be the plasma frequency.) As a result, we may write

$$\hat{\mathcal{E}}(\hat{\omega}) = \mathcal{E}(\omega). \tag{1.11}$$

Then, setting $\hat{\omega} = \lambda\omega$ in (1.10), we obtain

$$\mathfrak{F}(t) = \frac{\lambda}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda\omega t}\mathcal{E}(\omega)\, d\omega. \tag{1.12}$$

We assume that the matrix $\mathcal{E}(\omega)$ is Hermitian for real ω and independent of λ. Physically, this implies we are neglecting the dissipative effects of the medium. The effects of "weak dissipation" is to be treated in a subsequent paper.

It is convenient to write (1.1) in matrix form. We introduce the antisymmetric matrix $(\mathbf{Z})$ corresponding to any 3-vector $\mathbf{Z}$, given by

$$(\mathbf{Z}) = \begin{bmatrix} 0 & -z_3 & z_2 \\ z_3 & 0 & -z_1 \\ -z_2 & z_1 & 0 \end{bmatrix}. \tag{1.13}$$

Then, if $\mathbf{W}$ is an arbitrary 3-vector, $(\mathbf{Z})\mathbf{W} = \mathbf{Z}\times\mathbf{W}$. We also define the three 6×6 matrices, A^1, A^2, A^3 by

$$\sum_{r=1}^{3} k_r A^r = \begin{bmatrix} 0 & -c(\mathbf{K}) \\ c(\mathbf{K}) & 0 \end{bmatrix}, \tag{1.14}$$

where $\mathbf{K} = (k_1, k_2, k_3)$. The matrix A^1 is obtained by setting $k_1 = 1$ and $k_2 = k_3 = 0$ in (1.14). A^2 and A^3 are determined in a similar manner. We note

[3] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Inc., New York, 1960).

that the matrices A^r are real and symmetric. Using (1.14), (1.1) can be written in the compact form

$$\partial \mathbf{v}/\partial t + A^r(\partial \mathbf{u}/\partial x_r) = \mathbf{f}. \tag{1.15}$$

(The summation convention with respect to repeated indices from 1 to 3 is used.) Here, $\mathbf{f}$ is the column vector with components, $-4\pi(J_1, J_2, J_3, 0, 0, 0)$. Equation (1.5), when written in terms of $\mathbf{u}(t, \mathbf{X})$ and $\mathbf{v}(t, \mathbf{X})$ implies the "initial conditions,"

$$\mathbf{u}(t, \mathbf{X}) \equiv \mathbf{v}(t, \mathbf{X}) \equiv 0 \quad \text{for} \quad t < 0. \tag{1.16}$$

An integral representation for the solution of the integro-differential system (1.6), (1.15) with "initial conditions" (1.16), can be obtained using Fourier transform techniques. This has been done in Ref. 2. In order to describe the result obtained there, it is necessary to give some preliminary definitions. We define the inner product of any two column vectors $\mathbf{x}$ and $\mathbf{y}$ having 6 components by

$$(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{6} x_j \bar{y}_j, \tag{1.17}$$

where the bar denotes complex conjugate. Furthermore, we introduce the Hermitian matrix

$$G(\omega, \mathbf{K}) = k_r A^r - \omega \mathcal{E}(\omega) \tag{1.18}$$

and the corresponding *dispersion relation*,

$$\det G(\omega, \mathbf{K}) = 0. \tag{1.19}$$

We denote the real solutions of the dispersion relation by $\omega = h(\mathbf{K})$ [We should index the functions $h(\mathbf{K})$. However, not doing so simplifies the notation and need not cause any confusion.] and the linearly independent null eigenvectors of $G(h, \mathbf{K})$ by $\mathbf{r}^j(\mathbf{K})$; $j = 1, \cdots, q$. That is,

$$G(h, \mathbf{K})\mathbf{r}^j = 0; \qquad j = 1, 2, \cdots, q. \tag{1.20}$$

Thus, q is the "nullity" of the singular matrix $G(h, \mathbf{K})$. (We see that for isotropic media $q = 2$, while for anisotropic media $q = 1$.) The vectors $\mathbf{r}^j(\mathbf{K})$, which of course depend on the root $\omega = h(\mathbf{K})$ of (1.19), are orthonormalized by the condition

$$(\mathbf{r}^i, A^0[\omega]\mathbf{r}^j) = \delta_{ij}; \qquad i, j = 1, \cdots, q. \tag{1.21}$$

Here, δ_{ij} is the Kronecker delta. The Hermitian matrix $A^0[\omega] = d/d\omega[\omega\mathcal{E}(\omega)]$ is assumed to be positive definite and therefore it is always possible to accomplish the orthonormalization.

An integral representation of the solution $\mathbf{u}$ is given by Eq. (4.3.6) of Ref. 2, which yields

$$\mathbf{u}(t, \mathbf{X}) = \left(\frac{\lambda}{2\pi}\right)^3 \int_{-\infty}^{\infty} d\mathbf{K} \int_{-\infty}^{\infty} d\mathbf{Z} \int_0^t d\tau \times \sum_{\omega = h} \exp\{i\lambda[k_r(x_r - z_r) - \omega(t - \tau)]\} \times \sum_{j=1}^{q} (\mathbf{f}[\tau, \mathbf{Z}], \mathbf{r}^j)\mathbf{r}^j; \qquad t > 0. \tag{1.22}$$

Equation (1.22) is the exact integral representation of $\mathbf{u}$ only if all the roots of (1.19) are real. If any roots lie in the lower half of the ω plane, (1.22) is an asymptotic equation for $\lambda \to \infty$. It is assumed that no roots lie in the upper half-plane or, that if any do, their contributions to the solution can be neglected.

Until now we have said nothing about the nature of the source function $\mathbf{f}(t, \mathbf{X})$. Our concern here is with the moving sources. The current $\mathbf{J}(t, \mathbf{X})$ corresponding to a particle with charge e moving along the trajectory $\mathbf{X} = \mathbf{Y}(t)$ is given by

$$\begin{aligned}\mathbf{J}(t, \mathbf{X}) &= e\dot{\mathbf{Y}}(t)\ \delta[\mathbf{X} - \mathbf{Y}(t)] \\ &= e\dot{\mathbf{Y}}(t)\ \delta[x_1 - y_1(t)]\ \delta[x_2 - y_2(t)] \\ &\quad \times \delta[x_3 - y_3(t)].\end{aligned} \tag{1.23}$$

For greater generality, we consider source functions of the form

$$\mathbf{f}(t, \mathbf{X}; \lambda) = \lambda^d \mathbf{g}\{t, \lambda[\mathbf{X} - \mathbf{Y}(t)]\}. \tag{1.24}$$

Here d is a real number and $\mathbf{g}(t, \mathbf{X})$ is taken to be real and to have, for each value of t, compact support in $\mathbf{X}$. [$\mathbf{f}(\mathbf{X})$ has "compact support" if it vanishes outside a bounded region called the "support" of f.] We note that for large λ the support of $\mathbf{f}(t, \mathbf{X}; \lambda)$ shrinks to the point $\mathbf{X} = \mathbf{Y}(t)$. Therefore, (1.24) can be used to represent a source which is nonzero only in a small neighborhood of the moving point $\mathbf{X} = \mathbf{Y}(t)$. Remembering that $\mathbf{J}(t, \mathbf{X})$ represents the first three components of $\mathbf{f}(t, \mathbf{X})$, it is easy to see that (1.23) is a special case of (1.24). In fact, using the relation $\lambda\delta(\lambda x) = \delta(x)$, (1.23) becomes

$$\begin{aligned}\mathbf{J}(t, \mathbf{X}) &= e\dot{\mathbf{Y}}(t)\lambda^3\ \delta\{\lambda[x_1 - y_1(t)]\} \\ &\quad \times \delta\{\lambda[x_2 - y_2(t)]\}\ \delta\{\lambda[x_3 - y_3(t)]\}.\end{aligned} \tag{1.25}$$

In a similar manner, we can show that (1.24) includes all moving "multipole" sources. That is, all sources whose corresponding current terms $\mathbf{J}$ are given by linear combinations of partial derivatives of the three-dimensional δ-function.

A further generalization of the moving source is obtained by introducing an *oscillatory factor* $\cos[\lambda q(t)]$. Then, $\mathbf{f}$ has the form

$$\mathbf{f}(t, \mathbf{X}; \lambda) = \lambda^d \mathbf{g}\{t, \lambda[\mathbf{X} - \mathbf{Y}(t)]\} \cos[\lambda q(t)]. \tag{1.26}$$

Such a source describes an oscillatory current nonzero only in a small neighborhood of the moving point $\mathbf{X} = \mathbf{Y}(t)$. Equation (1.26) can also be used to express the current associated with a particle of fixed charge, moving with a small oscillation superimposed on a smooth trajectory. For such a source, (1.23) holds with

$$\mathbf{Y}(t) = \mathbf{Y}_0(t) + [\mathbf{M}(t)/\lambda] \sin [\lambda q(t)], \qquad \dot{\mathbf{Y}}(t) = \dot{\mathbf{Y}}_0(t) + \mathbf{M}(t)\dot{q}(t) \cos [\lambda q(t)] + O(1/\lambda). \tag{1.27}$$

Since

$$\delta[\mathbf{X} - \mathbf{Y}(t)] = \delta[\mathbf{X} - \mathbf{Y}_0(t)] + O(1/\lambda), \tag{1.28}$$

(1.23) becomes

$$\mathbf{J}(t, \mathbf{X}; \lambda) = e\{\dot{\mathbf{Y}}_0(t) + \mathbf{M}(t)\dot{q}(t) \cos [\lambda q(t)]\} \times \delta[\mathbf{X} - \mathbf{Y}_0(t)] + O(1/\lambda). \tag{1.29}$$

Equation (1.29) shows that the current is asymptotic to a sum of two terms, one of which has an oscillatory factor, as in (1.26), while the other does not.

If in Eq. (1.22), we take $\mathbf{f}$ to be given by (1.26) and make the change of variables, $q_\nu = \lambda[z_\nu - y_\nu(\tau)]$, we obtain

$$\mathbf{u}(t, \mathbf{X}) = \frac{\lambda^d}{(2\pi)^3} \int_{-\infty}^{\infty} d\mathbf{K} \int_0^t d\tau \times \sum_{\omega = h} \exp \{i\lambda[k_\nu(x_\nu - y_\nu) - \omega(t - \tau)]\} \times \sum_{j=1}^{q} (\hat{\mathbf{g}}[\tau, \mathbf{K}], \mathbf{r}^j)\mathbf{r}^j \cos [\lambda q(t)], \tag{1.30}$$

where

$$\hat{\mathbf{g}}(\tau, \mathbf{K}) = \int_{-\infty}^{\infty} \exp (-ik_\nu q_\nu)\mathbf{g}(\tau, \mathbf{Q})\, d\mathbf{Q}; \qquad \mathbf{Q} = (q_1, q_2, q_3). \tag{1.31}$$

It is important to note that the function $\hat{\mathbf{g}}(\tau, \mathbf{K})$ is independent of λ. We use this fact when the asymptotic expansion of (1.30) for large λ is obtained.

2. The Asymptotic Expansion for Isotropic Media

Isotropic media are characterized by the fact that the matrix $\mathcal{E}(\omega)$ has the form

$$\mathcal{E}(\omega) = \begin{bmatrix} \epsilon(\omega)I & 0 \\ 0 & \mu(\omega)I \end{bmatrix}. \tag{2.1}$$

Here $\epsilon(\omega)$ and $\mu(\omega)$ are scalars and I is the 3×3 unit matrix. It is readily seen from Eqs. (2.1) and (1.18) that the dispersion relation (1.19) implies

$$k = m(\omega), \tag{2.2}$$

where

$$k = (k_\nu k_\nu)^{\frac{1}{2}} = |\mathbf{K}|, \qquad m(\omega) = |\omega|/c[\epsilon(\omega)\mu(\omega)]^{\frac{1}{2}} = |\omega|/c[n(\omega)]. \tag{2.3}$$

In (2.3), $n(\omega) = [\epsilon(\omega)\mu(\omega)]^{\frac{1}{2}}$ is the index of refraction of the medium. We let $\omega = h(k)$ represent those real-valued functions for which

$$k = m[h(k)]. \tag{2.4}$$

We now introduce $\mathbf{A} = (\alpha_1, \alpha_2, \alpha_3)$, the unit vector in the direction of $\mathbf{K}$. Thus,

$$\mathbf{K} = k\mathbf{A}. \tag{2.5}$$

Furthermore, we let $\hat{\mathbf{N}}$ and $\hat{\mathbf{B}}$ be any two real unit vectors such that $\hat{\mathbf{N}}$, $\hat{\mathbf{B}}$ and sgn $[m']\mathbf{A}$ form a right-handed orthonormal set. The null vectors $\mathbf{r}^j$; $j = 1, \cdots, q$, defined by Eqs. (1.20) and (1.21) can be easily determined. We find that q equals two, and that

$$\mathbf{r}^1 = [(\zeta/\epsilon)^{\frac{1}{2}}\hat{\mathbf{B}}, (\zeta/\mu)^{\frac{1}{2}}\hat{\mathbf{N}}], \tag{2.6}$$

$$\mathbf{r}^2 = [(\zeta/\epsilon)^{\frac{1}{2}}\hat{\mathbf{N}}, -(\zeta/\mu)^{\frac{1}{2}}\hat{\mathbf{B}}]. \tag{2.7}$$

($\hat{\mathbf{N}}$ and $\hat{\mathbf{B}}$ are determined up to a rotation which leaves all our results invariant.) The function $\zeta(\omega)$ appearing in (6) and (7) is determined by the orthonormality condition (1.21), which yields

$$\zeta(\omega) = \frac{\epsilon(\omega)\mu(\omega)}{\epsilon(d/d\omega)(\omega\mu) + \mu(d/d\omega)(\omega\epsilon)}. \tag{2.8}$$

It can be seen from the definition of $\mathcal{E}(\omega)$ given in Sec. 1 that $\mathcal{E}(-\omega) = \overline{\mathcal{E}(\omega)}$. Therefore, $\epsilon(\omega)$ and $\mu(\omega)$ are even functions of ω. Equation (2.3) then shows that $m(\omega)$ is even in ω. Thus, if $\omega = h(k)$ is a root of the dispersion relation, $\omega = -h(k)$ is also a root. We now define the functions $\mathbf{u}_+(t, \mathbf{X})$ and $\mathbf{u}_-(t, \mathbf{X})$ by the right side of (1.30), except that for $\mathbf{u}_\pm(t, \mathbf{X})$ we sum only over the (positive/negative) roots $\omega = h(k)$. Then, assuming we may neglect the contribution from the zero root (see Ref. 2, Sec. 7.1),

$$\mathbf{u}(t, \mathbf{X}) = \mathbf{u}_+(t, \mathbf{X}) + \mathbf{u}_-(t, \mathbf{X}). \tag{2.9}$$

Equation (1.18) shows that $G(-h, -\mathbf{K}) = -G(h, \mathbf{K})$. Since $\mathbf{g}$ is a real quantity, it follows from (1.31) that $\hat{\mathbf{g}}(\tau, -\mathbf{K}) = \overline{\hat{\mathbf{g}}(\tau, \mathbf{K})}$. Using these two facts, it can be shown [by replacing $\mathbf{K}$ by $-\mathbf{K}$ and $h(k)$ by $-h(k)$ in the representation of $u_+(t, \mathbf{X})$] that $\overline{\mathbf{u}_+(t, \mathbf{X})} = \mathbf{u}_-(t, \mathbf{X})$. It therefore follows from (2.9) that $\mathbf{u}(t, \mathbf{X})$ is real and is given by

$$\mathbf{u}(t, \mathbf{X}) = \mathbf{u}_+ + \bar{\mathbf{u}}_+ = 2 \text{ Re } [\mathbf{u}_+(t, \mathbf{X})]. \tag{2.10}$$

As a result, we may replace $\cos [\lambda q(\tau)]$ by $\exp \{i\lambda q(\tau)\}$

in (1.30), if the real part of the integral is taken. That is,

$$\mathbf{u}(t, \mathbf{X}) = \frac{\lambda^d}{(2\pi)^3} \operatorname{Re}\left\{\int_{-\infty}^{\infty} d\mathbf{K} \int_0^t d\tau \sum_{\omega=h} \exp[i\lambda\phi] \times \sum_{i=1}^{2} (\hat{\mathbf{g}}[\tau, \mathbf{K}], \mathbf{r}^i[\mathbf{K}])\mathbf{r}^i[\mathbf{K}]\right\}, \quad (2.11)$$

where

$$\phi = k_\nu[x_\nu - y_\nu(\tau)] - (t - \tau)h(k) + q(\tau). \quad (2.12)$$

We point out that when representation (2.11) is used Eq. (2.10) holds only if $q(\tau) \equiv 0$.

We now apply the method of stationary phase in several dimensions (see Appendix II of Ref. 4 for a description of the method of stationary phase in several dimensions) to obtain the asymptotic expansion (for $\lambda \to \infty$) of the integral (2.11). The "phase function" ϕ, corresponding to a particular root $\omega = h(k)$, is given by Eq. (2.12). ϕ is stationary at those points $(\tau, \mathbf{K})$, such that

$$\partial\phi/\partial k_\nu = x_\nu - y_\nu(\tau) - (t - \tau)\,\partial h/\partial k_\nu = 0; \quad \nu = 1, 2, 3 \quad (2.13)$$

and

$$\partial\phi/\partial\tau = -k_\nu\dot{y}_\nu(\tau) + h(k) + \dot{q}(\tau) = 0. \quad (2.14)$$

We introduce $\mathbf{G} = (g_1, g_2, g_3)$, the group velocity vector and $g = (g_\nu g_\nu)^{\frac{1}{2}}$ the group speed, where

$$g_\nu = \partial h/\partial k_\nu = h'(k)(k_\nu/k) = h'(k)\alpha_\nu; \quad \nu = 1, 2, 3. \quad (2.15)$$

Then, if $\mathbf{T}$ is the unit vector in the direction of $\mathbf{G}$,

$$\mathbf{G} = g\mathbf{T}, \quad (2.16)$$

where

$$g = [h'(k)] \quad \text{and} \quad \mathbf{T} = \operatorname{sgn}[h'(k)]\mathbf{A}. \quad (2.17)$$

We see from (2.17) that $\mathbf{G}$ is either in the direction of $\mathbf{K}$ or $-\mathbf{K}$. The ambiguity is resolved by the sign of $h'(k)$. Equations (2.13), (2.14), and (2.15) show that, at the stationary points,

$$\mathbf{X} = \mathbf{Y}(\tau) + (t - \tau)\mathbf{G}; \quad t \geq \tau \quad (2.18)$$

and

$$[\mathbf{K}\cdot\dot{\mathbf{Y}}(\tau)] = h(k) + \dot{q}(\tau). \quad (2.19)$$

If $v(\tau) = |\dot{\mathbf{Y}}(\tau)|$ is the source speed and θ is the angle between $\mathbf{G}$ and $\dot{\mathbf{Y}}(\tau)$, Eqs. (2.17) and (2.19) yield

$$\cos\theta = \operatorname{sgn}[h'(k)]\{[h(k) + \dot{q}(\tau)]/kv(\tau)\}. \quad (2.20)$$

When $\dot{q}(\tau) = 0$, (2.20) becomes the well-known "Cerenkov condition" and θ becomes the "Cerenkov angle." When $\dot{q}(\tau) \neq 0$, (2.20) is usually referred to as the "Cerenkov–Doppler condition."

For fixed $(t, \mathbf{X})$, the stationary points $(\tau, \mathbf{K})$ must be obtained by solving (2.18) and (2.19) which are, in general, transcendental equations. To avoid this difficulty, we take a different but equivalent point of view. We consider Eq. (2.18) as defining, for fixed $(\tau, \mathbf{K})$, a locus in $(t, \mathbf{X})$ space. This locus is a straight line which we call a *ray*. In (2.18), the quantity τ represents the time of emission of the ray from the source trajectory. Since Eq. (2.19) is a relation between the four quantities τ, k_ν, only three of them are independent. Therefore, as τ and $\mathbf{K}$ vary, Eq. (2.18) represents a three-parameter family of rays. Zero, one, or more rays may pass through a given point $(t, \mathbf{X})$. The asymptotic expansion of $\mathbf{u}$ at this point, as determined by the method of stationary phase, is obtained by summing the contributions corresponding to those rays.

The value s of the phase at the stationary point is determined by Eqs. (2.12), (2.15), and (2.18). They yield

$$s = [kh'(k) - h(k)] + q(\tau). \quad (2.21)$$

Then, using formula (12) in Appendix II of Ref. 4, we obtain

$$\mathbf{u}(t, \mathbf{X}) \sim \frac{\lambda^{d-2}}{2\pi} \operatorname{Re}\left[\sum_{\omega=h} \sum_{i=1}^{2} (|\det(\phi_{\nu\beta})|)^{-\frac{1}{2}} \times (\hat{\mathbf{g}}, \mathbf{r}^i)\mathbf{r}^i \exp\left\{i\lambda[(t - \tau) \times (kh' - \omega) + q(\tau)] + \frac{\pi i}{4}\operatorname{sig}(\phi_{\nu\beta})\right\}\right], \quad (2.22)$$

as the asymptotic expansion of (2.11). For each ray (2.18) which passes through $(t, \mathbf{X})$, the appropriate term in the sum (2.22) represents the corresponding contribution to the expansion. In (2.22), $(\phi_{\nu\beta})$ is the matrix of second derivatives of ϕ with respect to τ and k_ν, and sig $(\phi_{\nu\beta})$ denotes the signature of this matrix. By choosing a coordinate system in which $\mathbf{A} = (0, 0, 1)$, the elements of $(\phi_{\nu\beta})$ simplify greatly. In this system

$$(\phi_{\nu\beta}) = \begin{bmatrix} -(t-\tau)h'' & 0 & 0 & (h'-\dot{y}_1) \\ 0 & -(t-\tau)\dfrac{h'}{k} & 0 & -\dot{y}_2 \\ 0 & 0 & -(t-\tau)\dfrac{h'}{k} & -\dot{y}_3 \\ (h'-\dot{y}_1) & -\dot{y}_2 & -\dot{y}_3 & (\ddot{q}-k\ddot{y}_1) \end{bmatrix}. \quad (2.23)$$

[4] R. M. Lewis, *Asymptotic Methods for the Solution of Dispersive Hyperbolic Equations*, Research Rept. EM-197 New York University, (1964).

The fact that $\mathbf{A} = (1, 0, 0)$ implies

$$\dot{y}_1(\tau) = (\ddot{\mathbf{Y}}\cdot\mathbf{A}) = \operatorname{sgn}[h'(k)]v(\tau)\cos\theta, \qquad \dot{y}_2^2 + \dot{y}_3^2 = v^2(\tau)\sin^2\theta \tag{2.24}$$

and

$$\ddot{y}_1(\tau) = (\dddot{\mathbf{Y}}\cdot\mathbf{A}). \tag{2.25}$$

Using (2.24) and (2.25), we obtain

$$\det(\phi_{\nu\beta}) = \left[(t-\tau)\frac{h'}{k}\right]^2\left\{(t-\tau)[k(\dddot{\mathbf{Y}}\cdot\mathbf{A}) - \ddot{q}]h'' - \frac{h''}{k}v^2\sin^2\theta - [|h'| - v\cos\theta]^2\right\}. \tag{2.26}$$

We note that (2.26) has been expressed in a form independent of our special choice of the coordinate system.

Equations (2.18), (2.20), (2.22), and (2.26) yield a parametric representation of the first term of the asymptotic expansion of $\mathbf{u}(t, \mathbf{X})$, which is valid only when $\det(\phi_{\nu\beta}) \neq 0$. Those points $(t, \mathbf{X})$, corresponding to parameter values for which $\det(\phi_{\nu\beta}) = 0$, form a locus (called a "caustic") in $(t, \mathbf{X})$ space. To obtain the asymptotic expansion of $\mathbf{u}(t, \mathbf{X})$ valid at caustic points, a more detailed analysis is required. An interesting effect may occur when $\ddot{\mathbf{Y}} \equiv q(t) \equiv 0$. In this case, there may be a caustic surface similar to a "Mach cone" along which the solution $\mathbf{u}$ is more intense. This cone proceeds into the medium with the source. The reader is referred to Appendix II for an analysis of this effect.

It was pointed out above that only three of the four quantities k_ν, τ are independent. Actually, three functions of k_ν and τ may be selected as independent parameters. To determine the most convenient set of independent parameters, we introduce along the trajectory $\mathbf{X} = \mathbf{Y}(\tau)$, the orthonormal set $\mathbf{T}^*, \mathbf{N}^*, \mathbf{B}^*$ consisting of the tangent, principal normal, and binormal vectors. If the trajectory is a straight line, $\mathbf{N}^*$ and $\mathbf{B}^*$ can be any two unit vectors such that $\mathbf{T}, \mathbf{N}^*, \mathbf{B}^*$ form an orthonormal set and $\mathbf{B}^* = \mathbf{T}^*\cdot\mathbf{N}^*$. We now define γ to be the angle which the projection of $\mathbf{T}$ into the $\mathbf{N}^*, \mathbf{B}^*$ plane makes with $\mathbf{N}^*$, as measured in a counterclockwise direction from $\mathbf{N}^*$. Then, we have

$$\mathbf{T} = \operatorname{sgn}[h'(k)]\mathbf{A} = \cos\theta\mathbf{T}^* + \sin\theta\cos\gamma\mathbf{N}^* + \sin\theta\sin\gamma\mathbf{B}^*. \tag{2.27}$$

Equations (2.5) and (2.27) show that $\mathbf{K}$ depends on θ, τ, k, and γ. We see from (2.20) that θ can be eliminated in favor of k and τ. Equation (2.3) shows that k in turn may be eliminated in favor of ω. (It is advantageous to make this replacement because $m(\omega)$ is a single-valued function, whereas its inverse is, in general, multiple-valued.) We see therefore that τ, ω, and γ may always be selected as independent parameters.

For any root $\omega = h(k)$, we have $\omega = h[m(\omega)]$. Implicit differentiation of this relation with respect to ω yields

$$h'(k) = \frac{1}{m'(\omega)} \quad \text{and} \quad h''(k) = \frac{-m''(\omega)}{[m'(\omega)]^3}. \tag{2.28}$$

Inserting Eqs. (2.2) and (2.28) into (2.20), we obtain

$$\cos\theta = \operatorname{sgn}[m'(\omega)]\left[\frac{\omega + \dot{q}(\tau)}{v(\tau)m(\omega)}\right]. \tag{2.29}$$

The ray equations (2.18) are now given by

$$\mathbf{X} = \mathbf{Y}(\tau) + [(t-\tau)/|m'(\omega)|]\mathbf{T}(\tau, \omega, \gamma), \tag{2.30}$$

where $\mathbf{T}$ is expressed in terms of τ, ω, and γ through (2.27) and (2.29).

The parameters $\mathbf{P} = (\tau, \omega, \gamma)$ lie in a certain *parameter space* $\mathcal{P}$ which is determined below. For any fixed value of t, Eq. (2.30) defines a transformation from $\mathbf{P}$-space to $\mathbf{X}$-space. The Jacobian $j(t; \mathbf{P})$ of this *ray transformation* is defined by

$$j(t; \mathbf{P}) = \partial(x_1, x_2, x_3)/\partial(\tau, \omega, \gamma). \tag{2.31}$$

Equation (AI22) in Appendix I yields the relation

$$\left|\frac{j(t; \mathbf{P})}{\det(\phi_{\nu\beta})}\right| = \frac{m(\omega)\,|m'(\omega)|}{v(\tau)}, \tag{2.32}$$

where, as seen from Eqs. (2.26) and (2.28),

$$\det(\phi_{\nu\beta}) = \left(\frac{t-\tau}{mm'}\right)^2\left\{(t-\tau)[(\dddot{\mathbf{Y}}\cdot\mathbf{A})m - \ddot{q}]\frac{m''}{(m')^3} + \left[1 - \left(\frac{\omega+\dot{q}}{vm}\right)^2\right]\frac{mm''}{(m')^2} - \left(\frac{1}{m'} - \frac{\omega+\dot{q}}{vm}\right)^2\right\}. \tag{2.33}$$

Furthermore, (2.28) and (2.32) show that Eq. (2.22) can be written in the form

$$\mathbf{u} \sim \operatorname{Re}\left[\mathbf{z}(t; \mathbf{P})\exp[i\lambda s(t; \mathbf{P})]\right] = \operatorname{Re}\left[\frac{\mathbf{d}(\mathbf{P})}{|j(t; \mathbf{P})|^{\frac{1}{2}}}\exp[i\lambda s(t; \mathbf{P})]\right], \tag{2.34}$$

where

$$s(t; \mathbf{P}) = (t-\tau)(m/m' - \omega) + q(\tau) \tag{2.35}$$

and

$$\mathbf{d}(\mathbf{P}) = |j(t; \mathbf{P})|^{\frac{1}{2}}\,\mathbf{z}(t; \mathbf{P}) = \frac{\lambda^{d-2}}{2\pi}\left[\frac{m\,|m'|}{v}\right]^{\frac{1}{2}} \times \exp\left[\frac{\pi i}{4}\operatorname{sig}(\phi_{\nu\beta})\right]\sum_{i=1}^{2}(\hat{\mathbf{g}}, \mathbf{r}^i)\mathbf{r}^i. \tag{2.36}$$

In (2.36), $\hat{\mathbf{g}}(\tau, \mathbf{K})$ and $\mathbf{r}^i(\mathbf{K})$ are expressed in terms of $\mathbf{P}$ through Eqs. (2.2), (2.5), (2.27), and (2.29). We note that no sum over the roots of the dispersion relation appears in (2.34). This simplification is a direct consequence of the single-valuedness of $m(\omega)$.

It remains to determine the domain of the independent parameters $\mathbf{P}$ in our parametric representation (2.30), (2.34). We claim that this domain is defined by the inequalities

$$0 \leq \tau, \tag{2.37}$$

$$1 \leq \{v(\tau)m(\omega)/[\omega + \dot{q}(\tau)]\}^2, \tag{2.38}$$

$$0 \leq \gamma < 2\pi. \tag{2.39}$$

Condition (2.37) is obvious. The requirement that, for real $\mathbf{K}$, Eq. (2.19) be satisfied implies that θ is a real angle. Therefore, condition (2.38) follows immediately from (2.29). If we restrict θ to lie between 0 and π, we can allow γ to vary between 0 and 2π. For fixed $\tau \geq 0$, each value of ω satisfying (2.38), defines a hyper-cone of rays in space–time. The hyper-cone is generated as γ varies from 0 to 2π.

In the special case $q(\tau) \equiv 0$, we make use of relation (2.10). Equation (2.34) still holds except now we take

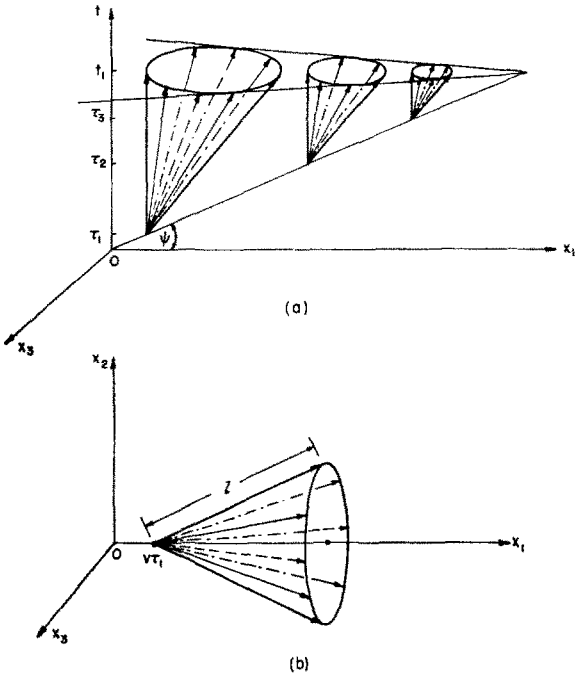


FIG. 1(a). A space–time diagram of the rays. For each of the three times of emission τ_1, τ_2, and τ_3, those rays which lie in the x_1, x_3 plane are depicted. (b). The projection into space of the hyper-cone of rays in space–time corresponding to the values $\tau = \tau_1$ and $\omega = \omega_1$. [Note that Figs. 1(a) and 1(b) are not drawn to the same scale.]

$$\mathbf{d}(\mathbf{P}) = \frac{\lambda^{d-2}}{\pi}(m\,|m'|/v)^{\frac{1}{2}} \times \exp\left[\frac{\pi i}{4}\operatorname{sig}(\phi_{,\beta})\right] \sum_{i=1}^{2} (\hat{\mathbf{g}}, \mathbf{r}^i)\mathbf{r}^i \tag{2.40}$$

and add the restriction $0 < \omega$ to the definition of the domain of $\mathbf{P}$. If we set $\beta(\tau) = v(\tau)/c$ and note Eq. (2.3), condition (2.38) becomes

$$\beta^2(\tau)n^2(\omega) \geq 1. \tag{2.41}$$

It follows from (2.41) that, for given values of τ and ω, the source speed $v(\tau)$ must be greater than the corresponding phase speed $c/n(\omega)$, for Cerenkov radiation to occur.

To conclude this section, we discuss two ray diagrams. We consider first a space–time picture of the rays. For simplicity we assume that $q(\tau) \equiv 0$ and that the source velocity is uniform in the direction of the positive x_1 axis. Because of the azimuthal symmetry of the problem, only those rays which lie in one plane passing through the source trajectory need be considered. Let that plane be the x_1, x_3 plane. (That is, we consider only those rays which correspond either to $\gamma = \frac{1}{2}\pi$ or to $\gamma = \frac{3}{2}\pi$.)

In Fig. 1(a) the vertical axis is the time axis. The source trajectory is a straight line in the (x_1, t) plane passing through the origin and making an angle ψ with the x_1 axis. It is easy to see that $\psi = \cot^{-1}v$, where v is the source speed. In general, for fixed time of emission τ, there is a conoidal surface of rays generated as ω varies in the range $\beta n(\omega) \geq 1$. A similar surface is produced at each point along the source trajectory. Of course the actual configuration in any given problem depends on the function $m(\omega)$ associated with that problem. Here, for definiteness, we set $\mu(\omega) \equiv 1$ and take $\epsilon(\omega)$ to be of the form

$$\epsilon(\omega) = 1 + p^2/(r^2 - \omega^2). \tag{2.42}$$

[Theoretical derivations of the dielectric permeability for isotropic media with a single "resonance frequency" given in Ref. 5 lead to functions $\epsilon(\omega)$ of this form.] Then, if $\beta^2p^2 \leq (1 - \beta^2)r^2$, Cerenkov radiation can occur only in the range

$$[r^2 - p^2\beta^2/(1 - \beta^2)]^{\frac{1}{2}} \leq \omega \leq r. \tag{2.43}$$

Using the functions $\epsilon(\omega)$ and $\mu(\omega)$ defined above, we have depicted in Fig. 1(a) three ray surfaces corresponding to three distinct times of emission. For each time of emission τ, these surfaces are depicted for $\tau \leq t \leq t_1$. (The reader is referred to Appendix

[5] A. Sommerfeld, *"Optics" in Lectures on Theoretical Physics* (Academic Press Inc., New York, 1964), Vol. IV.

II for a discussion of the significance of the v-shaped figure appearing in this diagram.)

The projection into X-space of the hyper-cone of rays in space–time, which corresponds to the values $\tau = \tau_1$ and $\omega = \omega_1$, is represented in Fig. 1(b). Actually, only that portion of the projection defined by $\tau_1 \leq t \leq t_1$ is shown. The two rays which lie in the (x_1, x_3) plane in this diagram also appear in Fig. 1(a). These rays are represented by dashed lines in both diagrams. The vertex angle Θ_1 of the cone in Fig. 1(b) is defined by

$$\Theta_1 = 2 \cos^{-1} [\beta n(\omega_1)]^{-1}. \tag{2.44}$$

The length l of the generators of the cone is determined from the ray Eqs. (2.30). We see that

$$l = (t_1 - \tau_1)/m'(\omega_1). \tag{2.45}$$

Associated with each ω lying in the range (2.43), there is an analogous cone defined by Eqs. (2.44) and (2.45) with ω_1 replaced by ω. Moreover, it can be shown that, as ω varies throughout this range, these cones fill out a 3-dimensional region in X-space.

3. The Energy of Cerenkov Radiation for Isotropic Media

We define the "average asymptotic energy density" $w(t, \mathbf{X})$ by

$$w(t, \mathbf{X}) = [16\pi\, |j(t; \mathbf{P})|]^{-1}(\mathbf{d}[\mathbf{P}], A^0\, \mathbf{d}[\mathbf{P}]), \tag{3.1}$$

where $|j(t; \mathbf{P})|$ is obtained from Eqs. (2.32) and (2.33), and $\mathbf{d}[\mathbf{P}]$ is given by (2.36). In (3.1), the following convention is used. The right side of (1) is summed over all rays which pass through the point $(t, \mathbf{X})$; [i.e., over all values of $\mathbf{P} = (\tau, \omega, \gamma)$ which, for fixed $(t, \mathbf{X})$, satisfy (2.30)]. It is shown in Ref. 2, Appendix B that

$$w_t + \nabla \cdot \langle \mathbf{S} \rangle = 0, \tag{3.2}$$

where $\langle \mathbf{S} \rangle$ is the average over a small time interval of the "Poynting vector," $\mathbf{S} = (c/4\pi)\, \mathbf{E} \cdot \mathbf{H}$. Equation (3.2) is the well-known equation of energy conservation and justifies the designation of $w(t, \mathbf{X})$ as the average asymptotic energy density.

Let $W(\tau_1, t)$ be the total energy, measured at time $t > \tau_1$, radiated from the source as it traverses the portion of trajectory defined by $0 \leq \tau \leq \tau_1$. Then,

$$W(\tau_1, t) = \int w(t, \mathbf{X})\, d\mathbf{X} = \frac{1}{16\pi} \int \frac{(\mathbf{d}, A^0\mathbf{d})}{|j|}\, d\mathbf{X}. \tag{3.3}$$

The integrand in the last term of (3.3) is to be summed over all values of $\mathbf{P}$ such that the corresponding ray passes through $(t, \mathbf{X})$ and $0 \leq \tau \leq \tau_1$. Alternatively, this sum can be replaced by a single integral over an appropriate domain in a "multiple X-space," consisting of κ replicas of physical X-space. Here κ is the maximum number of rays that pass through any point $\mathbf{X}$ at time t. The ray-transformation (2.30) maps the parameter space $\mathcal{P}$ in a 1–1 manner on this multiple $\mathbf{X}$ space (whereas the transformation is multiple valued for physical $\mathbf{X}$ space). The change of variables from $\mathbf{X}$ to $\mathbf{P}$ leads to the simple result,

$$W(\tau_1, t) = W(\tau_1) = \frac{1}{16\pi} \int_{\mathcal{P}_1} (\mathbf{d}, A^0\mathbf{d})\, d\mathbf{P}. \tag{3.4}$$

Here, $d\mathbf{P} = d\tau d\omega d\gamma$ and $\mathcal{P}_1$ is the domain in parameter space defined by the inequalities (2.37)–(2.39) with the additional restriction $\tau \leq \tau_1$. We note that (3.4) is independent of t, an expression of the fact that energy is conserved.

From Eq. (2.36) and the orthonormality condition (1.21), we obtain

$$(\mathbf{d}, A^0\mathbf{d}) = \frac{\lambda^{2(d-2)}}{4\pi^2} \frac{m(\omega)\, |m'(\omega)|}{v(\tau)} \sum_{i=1}^{2} |(\hat{\mathbf{g}}, \mathbf{r}^i)|^2. \tag{3.5}$$

By inserting (3.5) into (3.4) and noting the definition of $\mathcal{P}_1$ given above, we have

$$W(\tau_1) = \frac{\lambda^{2(d-2)}}{64\pi^3} \int_0^{2\pi} d\gamma \int_0^{\tau_1} d\tau \times \int_{((\omega+\delta)/s m)^2 \leq 1} d\omega\, \frac{m(\omega)\, |m'(\omega)|}{v(\tau)} \sum_{i=1}^{2} |(\hat{\mathbf{g}}, \mathbf{r}^i)|^2. \tag{3.6}$$

In the case $q(\tau) = 0$, $\mathcal{P}_1$ is defined by inequalities (2.37)–(2.39) with the additional restrictions $\tau \leq \tau_1$ and $0 < \omega$. Equations (1.21), (2.40), and (3.4) show that

$$W(\tau_1) = \frac{\lambda^{2(d-2)}}{16\pi^3} \int_0^{2\pi} d\gamma \int_0^{\tau_1} d\tau \times \int_{\substack{\beta^2(\tau) n^2(\omega) \geq 1 \\ \omega > 0}} d\omega\, \frac{m(\omega)\, |m'(\omega)|}{v(\tau)} \sum_{i=1}^{2} |(\hat{\mathbf{g}}, \mathbf{r}^i)|^2. \tag{3.7}$$

We observe that $j(t; \mathbf{P})$ does not appear in the integrals (3.6) and (3.7), whereas it does appear in the expression for the fields. Equations (2.32) and (2.33) show that $j(t; \mathbf{P})$ is a very complicated expression. Therefore, we can conclude that the degree of computational difficulty required to obtain the energy results is significantly less than that required to obtain the fields.

4. Examples for Isotropic Media

In this section we consider separately the Cerenkov radiation due to a uniformly moving charged particle and due to a uniformly moving dipole. In both cases the function $q(\tau)$ is taken to be zero. Since the source velocity vector is constant in

these problems, we are free to select our spatial axes so that the positive x_1 axis coincides with the trajectory. Thus, if $\hat{\mathbf{X}}_\nu$ is the unit vector in the direction of the positive x_ν axis ($\nu = 1, 2, 3$), we may set

$$\mathbf{Y}(\tau) = v\tau\hat{\mathbf{X}}_1, \quad \mathbf{N}^* = \hat{\mathbf{X}}_2, \quad \text{and} \quad \mathbf{B}^* = \hat{\mathbf{X}}_3. \tag{4.1}$$

Furthermore, we may choose $\hat{\mathbf{B}}$ to be perpendicular to $\hat{\mathbf{X}}_1$ as well as to $\mathbf{T}$. Then, Eqs. (4.1) and (2.27) show that we may take

$$\mathbf{A} = \operatorname{sgn}[m'(\omega)](\cos\theta, \sin\theta\cos\gamma, \sin\theta\sin\gamma), \tag{4.2}$$

$$\hat{\mathbf{B}} = (0, \sin\gamma, -\cos\gamma), \tag{4.3}$$

and

$$\hat{\mathbf{N}} = (\sin\theta, -\cos\theta\cos\gamma, -\cos\theta\sin\gamma), \tag{4.4}$$

where

$$\cos\theta = \operatorname{sgn}[m'(\omega)]/\beta(\tau)n(\omega), \qquad \omega > 0. \tag{4.5}$$

The null vectors $\mathbf{r}^1$ and $\mathbf{r}^2$ are determined by inserting (4.3) and (4.4) into Eqs. (2.6) and (2.7) respectively.

A. Uniformly Moving Charged Particle

For this source, the first three components of $\lambda^d\mathbf{g}\{t, \lambda[\mathbf{X} - \mathbf{Y}(t)]\}$, aside from a factor of -4π, are given by Eq. (1.25). We see that $d = 3$ and

$$\hat{\mathbf{g}}(\tau, \mathbf{Q}) = -4\pi ev\,\delta(q_1)\,\delta(q_2)\,\delta(q_3)[\hat{\mathbf{X}}_1, 0]. \tag{4.6}$$

Equation (1.31) then yields

$$\hat{\mathbf{g}}(\tau, \mathbf{K}) = -4\pi ev[\hat{\mathbf{X}}_1, 0] \tag{4.7}$$

and therefore

$$(\hat{\mathbf{g}}, \mathbf{r}^1) = 0 \quad \text{and} \quad (\hat{\mathbf{g}}, \mathbf{r}^2) = -4\pi ev(\zeta/\epsilon)^{\frac{1}{2}}\sin\theta. \tag{4.8}$$

The asymptotic expansion of $\mathbf{u}$ is given parametrically by (2.34), (2.40), and (2.30). These equations along with (4.8) yield

$$\mathbf{u}(t, \mathbf{X}) = [\mathbf{E}, \mathbf{H}] \sim (-4e\lambda)\left[\frac{v\zeta(\omega)m(\omega)\,|m'(\omega)|}{\epsilon(\omega)\,|j(t;\mathbf{P})|}\right]^{\frac{1}{2}} \times \sin\theta\cos[\lambda s + \tfrac{1}{4}\pi\operatorname{sig}(\phi_{\nu\beta})]\mathbf{r}^2 \tag{4.9}$$

and

$$\mathbf{X} = \mathbf{Y}(\tau) + [(t - \tau)/m'(\omega)]\mathbf{A}. \tag{4.10}$$

In (4.9), $j(t;\mathbf{P})$ and s are obtained by setting $q(\tau) = 0$ in (2.33) and (2.35) respectively. The signature of the matrix $(\phi_{\nu\beta})$ can be determined from (2.23).

We define $\Omega(\mathbf{X})$ to be the plane which passes through the x_1 axis and the point $(\mathbf{X})$. For any ray through $(\mathbf{X})$, the vector $\hat{\mathbf{N}}$ lies in $\Omega(\mathbf{X})$ and the vector $\hat{\mathbf{B}}$ is normal to $\Omega(\mathbf{X})$. Thus, Eqs. (2.7) and (4.9) show that whenever the fields are nonzero at $(\mathbf{X})$, $\mathbf{E}(t, X)$ lies in $\Omega(\mathbf{X})$ and $\mathbf{H}(t, \mathbf{X})$ is normal to $\Omega(\mathbf{X})$.

The energy $W(\tau_1)$ is given by Eq. (3.7). It can be shown from (2.3) and (2.8) that, for $\omega > 0$,

$$\frac{m(\omega)\,|m'(\omega)|}{\epsilon(\omega)}\,\zeta(\omega) = \frac{\omega\,|\mu(\omega)|}{2c^2}. \tag{4.11}$$

We recall that ω is related to the frequency $\hat{\omega}$ by $\omega = \hat{\omega}/\lambda$. Then, inserting Eqs. (4.8) and (4.11) into (3.7) and using (1.11), we obtain, after performing the τ and γ integrations,

$$W(\tau_1) = \frac{e^2v\tau_1}{c^2}\int_{\substack{\beta^2\hat{\epsilon}\hat{\mu}\geq 1\\ \hat{\omega}>0}} \hat{\omega}\,|\hat{\mu}(\hat{\omega})|\left(1 - \frac{1}{\beta^2\hat{\epsilon}\hat{\mu}}\right)d\hat{\omega}. \tag{4.12}$$

If $\hat{\mu}(\hat{\omega}) \equiv 1$, (4.12) agrees with the result obtained by Tamm[6]. We point out here that Tamm's result is exact whereas (4.12) is asymptotic. The agreement can be explained by the following argument. We note that (4.12) holds for all time $t > \tau_1$, and in particular for $t = \infty$. Equation (4.10) shows that at $t = \infty$, $x_2^2 + x_3^2 = \infty$. Therefore, in $\lambda_0 = \lambda a/c$, $a = \infty$ and, in that case, we expect our asymptotic expansion to be exact.

B. Uniformly Moving Dipole

For this source the current $\mathbf{J}(t, \mathbf{X})$ is given by

$$\mathbf{J}(t, \mathbf{X}) = -[c\nabla\times\mathbf{M}(t, \mathbf{X}) + (\partial/\partial t)\mathbf{F}(t, \mathbf{X})]. \tag{4.13}$$

Here, $\mathbf{M}$ is the magnetic moment vector and $\mathbf{F}$ is the electric moment vector. We consider the special case of a moving magnetic dipole, where

$$\mathbf{M}(t, \mathbf{X}) = m_0\hat{\mathbf{X}}_2\,\delta(x_1 - vt)\,\delta(x_2)\,\delta(x_3). \tag{4.14}$$

The magnetic moment induces an electric moment as seen in the rest frame of the observer, given by

$$\mathbf{F}(t, \mathbf{X}) = (m_0/c)\dot{\mathbf{Y}}\times\hat{\mathbf{X}}_2\,\delta[\mathbf{X} - \mathbf{Y}(t)] = m_0\beta\hat{\mathbf{X}}_3\,\delta(x_1 - vt)\,\delta(x_2)\,\delta(x_3). \tag{4.15}$$

(See Ginzburg and Eidman[7] for a discussion of this source.) It can be shown from Eqs. (4.13)–(4.15) that, in the representation of the source by the function $\lambda^d\mathbf{g}\{t, \lambda[\mathbf{X} - \mathbf{Y}(t)]\}$, we should set $d = 4$ and

$$\mathbf{g}(\tau, \mathbf{Q}) = 4\pi cm_0\{\delta'(q_1)\,\delta(q_2)\,\delta(q_3)(1 - \beta^2)[\hat{\mathbf{X}}_3, 0] - \delta'(q_3)\,\delta(q_1)\,\delta(q_2)[\hat{\mathbf{X}}_1, 0]\}. \tag{4.16}$$

Equation (1.31) then yields

$$\hat{\mathbf{g}}(\tau, \mathbf{K}) = 4\pi im_0c \times\{k_1(1 - \beta^2)[\hat{\mathbf{X}}_3, 0] - k_3[\hat{\mathbf{X}}_1, 0]\}. \tag{4.17}$$

[6] Ig. Tamm, Zh. Teor. Fiz. S.S.S.R. 1, 439 (1939).

[7] V. L. Ginzburg and V. Ia. Eidman, Zh. Eksperim. i Teor. Fiz. 35, 1508 (1958) [English transl.: Soviet Phys.—JETP 8, 1055 (1959)].

The quantities k_1 and k_3 are expressed in terms of τ, ω, γ through Eqs. (2.3), (4.2), and (4.5). We obtain

$$k_1 = \omega/v \quad \text{and} \quad k_3 = [\omega n(\omega)/c] \sin\gamma \times (1 - 1/\beta^2 n^2)^{\frac{1}{2}} \operatorname{sgn}[m'(\omega)]. \tag{4.18}$$

The null vectors $\mathbf{r}^1$ and $\mathbf{r}^2$ are as in Sec. 4.A and therefore

$$(\hat{\mathbf{g}}, \mathbf{r}^1) = 4\pi i m_0 \omega(\zeta/\epsilon)^{\frac{1}{2}}(\beta - 1/\beta)\cos\gamma$$

and (4.19)

$$(\hat{\mathbf{g}}, \mathbf{r}^2) = 4\pi i m_0 \omega(\zeta/\epsilon)^{\frac{1}{2}}(1/n - n) \times \sin\gamma \operatorname{sgn}[m'(\omega)].$$

Thus, we obtain

$$\mathbf{u}(t, \mathbf{X}) \sim 4m_0\lambda^2\left[\frac{m(\omega)\,|m'(\omega)|\,\zeta(\omega)}{\epsilon(\omega)\,|j(t;\mathbf{P})|\,v}\right]^{\frac{1}{2}}\omega \times \{(\beta - 1/\beta)\cos\gamma\,\mathbf{r}^1 + \operatorname{sgn}[m'(\omega)] \times (1/n - n)\sin\gamma\mathbf{r}^2\} \sin[\lambda s + \tfrac{1}{4}\pi \operatorname{sig}(\phi_{\nu\beta})]. \tag{4.20}$$

The rays are given by Eq. (4.10).

The energy $W(\tau_1)$ is obtained by inserting Eqs. (4.11) and (4.19) into (3.6). After setting $\omega = \hat{\omega}/\lambda$ and performing the τ and γ integrations, the result is

$$W(\tau_1) = \frac{m_0^2\tau_1}{2c^2v}\int_{\substack{\beta^2\hat{n}^2\geq 1\\ \hat{\omega}>0}} \hat{\omega}^3\,|\hat{\mu}(\hat{\omega})| \times \left[\left(\hat{n} - \frac{1}{\hat{n}}\right)^2 + \left(\beta - \frac{1}{\beta}\right)^2\right] d\hat{\omega}. \tag{4.21}$$

If we set $\hat{\mu}(\hat{\omega}) \equiv 1$, (4.21) agrees with the result obtained by Ginzburg and Eidman in Ref. 8.

5. The Asymptotic Expansion for Crystalline Media

In this section we assume that the medium is a uniaxial crystal. The crystal axis is taken to be the x_3 axis. If we set the magnetic permeability $\mu(\omega)$ equal to 1, the matrix $\mathcal{E}(\omega)$ takes the form

$$\mathcal{E}(\omega) = \begin{bmatrix}\mathcal{E}_1(\omega) & 0\\ 0 & I\end{bmatrix}; \quad \mathcal{E}_1(\omega) = \begin{bmatrix}\epsilon_1(\omega) & 0 & 0\\ 0 & \epsilon_1(\omega) & 0\\ 0 & 0 & \epsilon_2(\omega)\end{bmatrix}. \tag{5.1}$$

In this case, the dispersion relation (1.19) yields the two equations

$$k_0 = m_0(\omega) = (|\omega|/c)[\epsilon_1(\omega)]^{\frac{1}{2}}, \tag{5.2}$$

$$k_e = m_e(\omega, k_3) = \left[\frac{\omega^2\epsilon_2}{c^2} + k_{3e}^2\left(1 - \frac{\epsilon_2}{\epsilon_1}\right)\right]^{\frac{1}{2}}. \tag{5.3}$$

The subscripts o and e are used because the solutions corresponding to k_0 and k_e are commonly referred to as "ordinary" and "extraordinary" respectively. In (5.3), $k_{3e} = k_e\alpha_3$.

If $|\alpha_3| \neq 1$, there exists one null-eigenvector of the dispersion matrix G corresponding to each of the quantities k_0 and k_e. We denote these vectors by

$$\mathbf{r}_0 = [\mathbf{R}_{10}, \mathbf{R}_{20}] \quad \text{and} \quad \mathbf{r}_e = [\mathbf{R}_{1e}, \mathbf{R}_{2e}]. \tag{5.4}$$

If we set $n_{0,e} = ck_{0,e}/|\omega|$, it can be shown that the 3-vectors $\mathbf{R}_{\nu 0}$, $\mathbf{R}_{\nu e}$ take the form

$$\mathbf{R}_{10} = \zeta_0^{\frac{1}{2}}\mathbf{A}\times\hat{\mathbf{X}}_3, \qquad \mathbf{R}_{20} = \zeta_0^{\frac{1}{2}}n_0\mathbf{A}\times(\mathbf{A}\times\hat{\mathbf{X}}_3); \tag{5.5}$$

$$\mathbf{R}_{1e} = \zeta_e^{\frac{1}{2}}(n_e/\epsilon_1)[\mathbf{A}\times(\mathbf{A}\times\hat{\mathbf{X}}_3) + (1 - \epsilon_1/\epsilon_2)(1 - \alpha_3^2)\hat{\mathbf{X}}_3],$$

$$\mathbf{R}_{2e} = \zeta_e^{\frac{1}{2}}\hat{\mathbf{X}}_3\times\mathbf{A}. \tag{5.6}$$

The quantities ζ_0 and ζ_e are determined by condition (1.21) which yields

$$\zeta_0^{-1} = (1 - \alpha_3^2)\left[\frac{d}{d\omega}(\omega\epsilon_1) + \epsilon_1\right], \tag{5.7}$$

$$\zeta_e^{-1} = (1 - \alpha_3^2)\left\{1 + n_e^2\left[\frac{\alpha_3^2}{\epsilon_1^2}\frac{d}{d\omega}(\omega\epsilon_1) + \frac{(1 - \alpha_3^2)}{\epsilon_2^2}\frac{d}{d\omega}(\omega\epsilon_2)\right]\right\}. \tag{5.8}$$

If $|\alpha_3| = 1$, the expressions for $\mathbf{r}_0$ and $\mathbf{r}_e$ given above are not valid. In this case $k_0 = k_e$ and there exist two linearly independent null-vectors of G in the forms

$$\mathbf{r}^1 = [\hat{\mathbf{N}}, \hat{\mathbf{B}}], \quad \mathbf{r}^2 = [\hat{\mathbf{B}}, -\hat{\mathbf{N}}], \tag{5.9}$$

where $\hat{\mathbf{N}}\cdot\hat{\mathbf{B}} = \hat{\mathbf{N}}\cdot\hat{\mathbf{X}}_3 = \hat{\mathbf{B}}\cdot\hat{\mathbf{X}}_3 = 0$. The vectors $\hat{\mathbf{N}}$ and $\hat{\mathbf{B}}$ are determined only up to an arbitrary rotation about the $x_3^{\wedge}$ axis. When $|\alpha_3| = 1$, we are not able to carry out the stationary phase evaluation of the integral representation of $\mathbf{u}$. Therefore, when we obtain the asymptotic expansion of $\mathbf{u}$, we must exclude this case from consideration. Thus, all rays corresponding to parameter values for which $\mathbf{A} = (0, 0, \pm 1)$ will be omitted. The energy results, however, will not be affected because, at any time t, the solutions corresponding to the omitted rays contribute negligibly to the radiated energy.

The integral representation of the solution is given by Eq. (1.30) when the source function $\mathbf{f}$ is given by (1.26). Arguments similar to those used for isotropic media show that the solution is real. Let $\omega = h_0(k)$ represent those real functions for which $k = m_0[h_0]$, and let $\omega = h_e(k, k_3)$ represent those real functions for which $k = m_e[h_e, k_3]$. Then, it can be shown that the integral representation takes the form

$$\mathbf{u}(t, \mathbf{X}) = \mathbf{u}_0(t, \mathbf{X}) + \mathbf{u}_e(t, \mathbf{X}), \tag{5.10}$$

where

$$\mathbf{u}_{0,\epsilon} = \frac{\lambda}{(2\pi)^3} \operatorname{Re}\left[\int_{-\infty}^{\infty} d\mathbf{K} \int_0^t d\tau \times \sum_{\omega=h_{0,\epsilon}} \exp\left[i\lambda\phi_{0,\epsilon}\right](\hat{\mathbf{g}}, \mathbf{r}_{0,\epsilon})\mathbf{r}_{0,\epsilon}\right]. \quad (5.11)$$

Here, the functions $\phi_{0,\epsilon}$ are given by

$$\phi_{0,\epsilon} = k_\nu[x_\nu - y_\nu(\tau)] - (t - \tau)h_{0,\epsilon} + q(\tau). \quad (5.12)$$

In the special case $q(\tau) \equiv 0$, the relation $\mathbf{u} = 2 \operatorname{Re}[\mathbf{u}_+]$ holds, $\mathbf{u}_+$ is obtained by restricting ω to be positive in (5.11).

The computational difficulty involved in carrying out the stationary phase evaluation of (5.11) is significantly reduced if we take the source trajectory to be a straight line. Thus, we set

$$\mathbf{Y}(\tau) = y(\tau)\mathbf{T}^*, \quad (5.13)$$

where

$$\mathbf{T}^* = (\xi_1, 0, \xi_2); \quad (\xi_1^2 + \xi_2^2) = 1. \quad (5.14)$$

There is no loss of generality in the assumption $\mathbf{Y}(\tau)$ lies in the x_1, x_3 plane. This can always be accomplished by a rotation of the spatial axes about the crystal axis. Our results are invariant under such a rotation.

Let us consider first the asymptotic expansion of the integral representation of $\mathbf{u}_0(t, \mathbf{X})$. The stationary phase analysis of this term is essentially the same as the analysis given in Sec. 2 for isotropic media. Therefore, only the results will be given here. For anisotropic media, it is more convenient to define θ to be the angle between the vectors $\mathbf{A}$ and $\mathbf{T}^*$, rather than, as in the case of isotropic media, the angle between $\mathbf{T}$ and $\mathbf{T}^*$. We have

$$\mathbf{K}_0 = m_0(\omega)\mathbf{A}_0. \quad (5.15)$$

We may select,

$$\mathbf{N}^* = (\xi_2, 0, -\xi_1) \quad \text{and} \quad \mathbf{B}^* = (0, 1, 0). \quad (5.16)$$

Then, if γ is the angle formed by the projection of $\mathbf{A}_0$ into the $(\mathbf{N}^*, \mathbf{B}^*)$ plane and $\mathbf{N}^*$ (as measured in a counterclockwise direction from $\mathbf{N}^*$), we obtain

$$\mathbf{A}_0 = \cos\theta_0\mathbf{T}^* + \sin\theta_0 \cos\gamma\mathbf{N}^* + \sin\theta_0 \sin\gamma\mathbf{B}^*. \quad (5.17)$$

The Cerenkov–Doppler relation yields

$$\cos\theta_0 = [[\omega + \dot{q}(\tau)]/m_0(\omega)v(\tau)], \quad v(\tau) = |\dot{\mathbf{Y}}(\tau)|. \quad (5.18)$$

The parametric representation of the first term of the asymptotic expansion of $\mathbf{u}_0(t,\mathbf{X})$ is given by

$$\mathbf{u}_0 \sim \operatorname{Re}\left[\frac{\mathbf{d}_0(\mathbf{P})}{[j_0(t;\mathbf{P})]^{\frac{1}{2}}} \exp\left[i\lambda s_0(t;\mathbf{P})\right]\right]; \quad (5.19)$$

$$\mathbf{d}_0(\mathbf{P}) = \frac{\lambda^{d-2}}{2\pi}(m_0\,|m_0'|/v)^{\frac{1}{2}} \times \exp\left[\tfrac{1}{4}\pi i \operatorname{sig}(\phi^0_{\nu\beta})\right](\hat{\mathbf{g}}, \mathbf{r}_0)\mathbf{r}_0$$

and

$$\mathbf{X}_0 = \mathbf{Y}(\tau) + [(t - \tau)/m_0']\mathbf{A}_0; \quad t \geq \tau. \quad (5.20)$$

In (5.19), $s_0(t;\mathbf{P})$ is given by Eq. (2.35) and $j_0(t;\mathbf{P})$, the Jacobian of the ray transformation (5.20), is obtained from Eqs. (2.32) and (2.33). Of course $m(\omega)$ is to be replaced by $m_0(\omega)$ in these equations. The matrix $(\phi^0_{\nu\beta})$ is the matrix of second derivatives of the phase function ϕ_0 with respect to k_ν and τ. Sig $(\phi^0_{\nu\beta})$ denotes the signature of this matrix.

We denote by $\mathcal{P}_0$ the domain of the independent parameters $\mathbf{P} = (\tau, \omega, \gamma)$. It is defined by inequalities (2.37)–(2.39) with $m(\omega)$ replaced by $m_0(\omega)$. When $q(\tau) \equiv 0$, we make use of the relation $\mathbf{u}_0 = 2 \operatorname{Re}[\mathbf{u}_{0+}]$. In this case, (5.19) holds with $\mathbf{d}_0$ multiplied by a factor of 2, and the restriction $0 < \omega$ is added to the definition of $\mathcal{P}_0$.

We now consider the asymptotic analysis of $\mathbf{u}_\epsilon$. We set

$$\mathbf{K}_\epsilon = m_\epsilon(\omega, k_3)\mathbf{A}_\epsilon, \quad (5.21)$$

where

$$\mathbf{A}_\epsilon = \cos\theta_\epsilon\mathbf{T}^* + \sin\theta_\epsilon \cos\gamma\mathbf{N}^* + \sin\theta_\epsilon \sin\gamma\mathbf{B}^*. \quad (5.22)$$

The stationary condition $\partial\phi_\epsilon/\partial\tau = 0$, yields the Cerenkov–Doppler relation

$$\cos\theta_\epsilon = \left[\frac{\omega + \dot{q}(\tau)}{v(\tau)m_\epsilon(\omega, k_3)}\right] = \frac{\eta(\tau, \omega)}{m_\epsilon(\omega, k_3)}; \quad \eta(\tau, \omega) = \frac{\omega + \dot{q}(\tau)}{v(\tau)}. \quad (5.23)$$

The stationary conditions $\partial\phi_\epsilon/\partial k_\nu = 0$; $\nu = 1, 2, 3$, yield the equations for the rays, which, noting the relations

$$\frac{\partial h_\epsilon}{\partial k} = \left(\frac{\partial m_\epsilon}{\partial\omega}\right)^{-1} \quad \text{and} \quad \frac{\partial h_\epsilon}{\partial k_3} = -\left(\frac{\partial m_\epsilon}{\partial k_3}\right)\left(\frac{\partial m_\epsilon}{\partial\omega}\right)^{-1} = \left(\frac{\partial m_\epsilon}{\partial\omega}\right)^{-1}\alpha_{3\epsilon}\left(\frac{\epsilon_2}{\epsilon_1} - 1\right), \quad (5.24)$$

become

$$\mathbf{X}_\epsilon - \mathbf{Y}(\tau) = (t - \tau)\mathbf{G}_\epsilon = \frac{(t - \tau)}{(m_\epsilon)_\omega}\left[\mathbf{A}_\epsilon + \alpha_{3\epsilon}\left(\frac{\epsilon_2}{\epsilon_1} - 1\right)\hat{\mathbf{X}}_3\right]; \quad t \geq \tau. \quad (5.25)$$

Equation (5.25) shows that, in this case, the group velocity vector $\mathbf{G}_\epsilon$ is not in the direction of the wave vector $\mathbf{K}_\epsilon$. This is a characteristic of anisotropic media.

Before completing the stationary-phase evaluation, we show how to express the vector $\mathbf{A}_\epsilon$ in terms of $\mathbf{P} = (\tau, \omega, \gamma)$. It follows from (5.22) that, to accomplish this, we need only express θ_ϵ in terms of $\mathbf{P}$. Equations (5.16) and (5.22) yield

$$\alpha_{3\epsilon} = \xi_2 \cos\theta_\epsilon - \xi_1 \sin\theta_\epsilon \cos\gamma. \tag{5.26}$$

Equations (5.3), (5.23), and (5.26) can be used to determine $\tan\theta_\epsilon$. The result is

$$\begin{aligned}\tan\theta_\epsilon = [1 - (1 - \epsilon_2/\epsilon_1)\xi_1^2\cos^2\gamma]^{-1} \\ \times \Bigg(\left(\frac{\epsilon_2}{\epsilon_1} - 1\right)\xi_1\xi_2\cos\gamma \\ \pm \Bigg\{\left[1 - \left(1 - \frac{\epsilon_2}{\epsilon_1}\right)\xi_1^2\cos^2\gamma\right] \\ \times \left[\frac{\omega^2\epsilon_2}{c^2\eta^2} - 1 + \xi_2^2\left(1 - \frac{\epsilon_2}{\epsilon_1}\right)\right] \\ + \xi_1^2\xi_2^2\left(1 - \frac{\epsilon_2}{\epsilon_1}\right)^2\cos^2\gamma\Bigg\}\Bigg).\end{aligned} \tag{5.27}$$

In the treatment of isotropic media, we restricted θ to lie between 0 and π and allowed γ to vary between 0 and 2π. It is more convenient here to restrict γ to lie between 0 and π and allow θ_ϵ to lie between 0 and 2π. For a given value of $\mathbf{P}$, Eq. (5.27) yields two values of $\tan\theta_\epsilon$. We see from (5.23) that $\operatorname{sgn}[\cos\theta_\epsilon] = \operatorname{sgn}[\omega + \dot{q}(\tau)]$. Thus, for a given value of $\mathbf{P}$, two angles which we call $\theta_\epsilon^\pm$ are defined. We introduce the following notation. Whenever the symbol $[\ ]_+$ appears, the quantity within the brackets is to be evaluated with θ_ϵ set equal to θ_ϵ^+. The symbol $[\ ]_-$ is defined in a similar way.

Once θ_ϵ is determined m_ϵ is expressed in terms of $\mathbf{P}$ through Eq. (5.23). We see from (5.23) and (5.27) that τ is introduced into m_ϵ only through the quantity η. Therefore, we may write

$$m_\epsilon(\omega, k_3) = m_{1\epsilon}(\eta, \omega, \gamma). \tag{5.28}$$

Equation (AI.37) of Appendix I shows that

$$\left|\frac{j_\epsilon(t;\mathbf{P})}{\det(\phi^\epsilon_{\gamma\beta})}\right| = \frac{m_{1\epsilon}\,|(m_{1\epsilon})_\omega|}{v(\tau)}. \tag{5.29}$$

Here, $(\phi^\epsilon_{\gamma\beta})$ is the matrix of second derivatives of ϕ_ϵ with respect to k_γ and τ, and $j_\epsilon(t;\mathbf{P}) = \det(d\mathbf{X}_\epsilon/d\mathbf{P})$ is the Jacobian of the ray transformation (5.25). The stationary-phase formula yields

$$\mathbf{u}_\epsilon(t, \mathbf{X}) \sim \sum_\pm \operatorname{Re}\left[\frac{\mathbf{d}_\epsilon(\mathbf{P})}{|j_\epsilon(t;\mathbf{P})|^{\frac{1}{2}}} \times \exp[i\lambda s_\epsilon(t;\mathbf{P})]\right]_\pm, \tag{5.30}$$

where

$$\mathbf{d}_\epsilon(\mathbf{P}) = \frac{\lambda^{d-2}}{2\pi}\left[\frac{m_{1\epsilon}\,|(m_{1\epsilon})_\omega|}{v}\right]^{\frac{1}{2}} \times \exp[\tfrac{1}{4}\pi i \operatorname{sig}(\phi^\epsilon_{\gamma\beta})](\hat{\mathbf{g}}, \mathbf{r}_\epsilon)\mathbf{r}_\epsilon \tag{5.31}$$

and

$$s_\epsilon(t;\mathbf{P}) = \frac{m_\epsilon}{(m_\epsilon)_\omega}\left[1 + \alpha_{3\epsilon}^2\left(\frac{\epsilon_2}{\epsilon_1} - 1\right) - \omega\right] \times (t - \tau) + q(\tau). \tag{5.32}$$

The computations required to determine $j_\epsilon(t;\mathbf{P})$ are difficult and the resulting expression is complicated. Because the Jacobian itself is not needed for the energy results, we omit its calculation here.

We denote by $\mathcal{P}_\epsilon$ the domain of the independent parameters $\mathbf{P}$. This domain is defined by the inequalities

$$0 \le \tau, \tag{5.33}$$

$$0 \le \gamma < \pi, \tag{5.34}$$

$$\begin{aligned}0 \le \{\xi_1^2\xi_2^2(1 - \epsilon_2/\epsilon_1)^2\cos^2\gamma \\ + [1 - \xi_1^2\cos^2\gamma(1 - \epsilon_2/\epsilon_1)] \\ \times [\omega^2\epsilon_2 v^2/c^2(\omega + \dot{q})^2 - 1 + \xi_2^2(1 - \epsilon_2/\epsilon_1)]\}.\end{aligned} \tag{5.35}$$

The requirement that the angles $\theta_\epsilon^\pm$ be real is expressed by condition (5.35). When $q(\tau) \equiv 0$, we make use of the relation $\mathbf{u}_\epsilon = 2\operatorname{Re}[\mathbf{u}_{\epsilon+}]$. Equation (5.30) holds with $\mathbf{d}_\epsilon$ multiplied by a factor of 2 and the restriction $0 < \omega$ is added to the definition of $\mathcal{P}_\epsilon$.

We remind the reader that we must exclude from consideration all rays corresponding to values of $\mathbf{P}$ for which $\mathbf{A}_{0\epsilon} = (0, 0, \pm 1)$. Equation (5.26) shows that $|\alpha_{3\epsilon}| = 1$ only when $\gamma = 0$, $\sin\theta_\epsilon = \pm\xi_1$, $\cos\theta_\epsilon = \mp\xi_2$. ($\gamma = \pi$ does not lie in the domain $\mathcal{P}_\epsilon$.)

6. The Energy of Cerenkov Radiation for Crystalline Media

In this section an expression for $W(\tau_1)$, the energy radiated from the source as it traverses the portion of trajectory defined by $0 \le \tau \le \tau_1$, is obtained. The derivation of this expression is similar to the derivation of the energy results given in Sec. 3 for isotropic media and therefore most of the computations are omitted here. It can be shown that

$$W(\tau_1) = W_0(\tau_1) + W_\epsilon(\tau_1), \tag{6.1}$$

where

$$W_0(\tau_1) = \frac{1}{16\pi}\int_{\mathcal{P}_{1_0}} (\mathbf{d}_0, A^0\mathbf{d}_0)\, d\mathbf{P} \tag{6.2}$$

and

$$W_\epsilon(\tau_1) = \frac{1}{16\pi}\sum_{\pm}\int_{\mathcal{P}_{1_\epsilon}} [(\mathbf{d}_\epsilon, A^0\mathbf{d}_\epsilon)]_\pm\, d\mathbf{P}. \tag{6.3}$$

Here, the domains of integration $\mathcal{P}_{1_0}$ and $\mathcal{P}_{1_\epsilon}$ are respectively the domains $\mathcal{P}_0$ and $\mathcal{P}_\epsilon$ defined in Sec. 5, with the added restriction $\tau \leq \tau_1$.

When the expression for $\mathbf{d}_0(\mathbf{P})$ given by (5.19) is inserted into Eq. (6.2), we obtain

$$W_0(\tau_1) = \frac{\lambda^{2(d-2)}}{64\pi^3}\int_{\mathcal{P}_{1_0}} \frac{m_0(\omega)\,|m_0'(\omega)|}{v(\tau)}\,|(\hat{\mathbf{g}}, \mathbf{r}_0)|^2\, d\mathbf{P}. \tag{6.4}$$

Similarly, inserting (5.31) into (6.3), we have

$$W_\epsilon(\tau_1) = \frac{\lambda^{2(d-2)}}{64\pi^3} \times \sum_{\pm}\int_{\mathcal{P}_{1_\epsilon}} \left[\frac{m_{1_\epsilon}\,|(m_{1_\epsilon})_\omega|}{v}\,|(\hat{\mathbf{g}}, \mathbf{r}_\epsilon)|^2\right]_\pm d\mathbf{P}. \tag{6.5}$$

If $q(\tau) \equiv 0$, the domains $\mathcal{P}_{1_0}$ and $\mathcal{P}_{1_\epsilon}$ are further restricted by the condition $\omega > 0$ and the right sides of Eqs. (6.4) and (6.5) are multiplied by a factor of 4.

7. Example for Crystalline Media

In this section we apply the results of the previous two sections to the case of a moving charged particle. For such a source $q(\tau) = 0$, $d = 3$, and the function $\hat{\mathbf{g}}(\tau, \mathbf{K})$ is given by

$$\hat{\mathbf{g}}(\tau, \mathbf{K}) = -4\pi e v(\tau)[\mathbf{T}^*, 0]; \quad \mathbf{T}^* = (\xi_1, 0, \xi_2). \tag{7.1}$$

From Eqs. (7.1), (5.5), and (5.6) we obtain

$$(\hat{\mathbf{g}}, \mathbf{r}_0) = -4\pi e v \zeta_0^{\frac{1}{2}}\xi_1\alpha_{20} \tag{7.2}$$

and

$$(\hat{\mathbf{g}}, \mathbf{r}_\epsilon) = -4\pi e v \zeta_\epsilon^{\frac{1}{2}} n_\epsilon\left[\frac{\alpha_{1_\epsilon}\alpha_{3_\epsilon}}{\epsilon_1}\xi_1 + \frac{(\alpha_{3_\epsilon}^2 - 1)}{\epsilon_2}\xi_2\right]. \tag{7.3}$$

Equations (5.14) and (5.16) show that, for both $\mathbf{A}_0$ and $\mathbf{A}_\epsilon$,

$$\alpha_1 = \xi_1\cos\theta + \xi_2\sin\theta\cos\gamma, \tag{7.4}$$

$$\alpha_2 = \sin\theta\sin\gamma, \tag{7.5}$$

$$\alpha_3 = \xi_2\cos\theta - \xi_1\sin\theta\cos\gamma. \tag{7.6}$$

Equations (5.2) and (5.18) yield

$$\cos\theta_0 = \{\beta(\tau)[\epsilon_1(\omega)]^{\frac{1}{2}}\}^{-1}; \qquad \omega > 0. \tag{7.7}$$

The values of $\tan\theta_\epsilon^\pm$ are obtained by setting $q(\tau)$ equal to zero in (5.27).

The asymptotic expansions of the fields $\mathbf{u}_0$ and $\mathbf{u}_\epsilon$ are determined by inserting Eqs. (7.2) and (7.3) into (5.19) and (5.31), respectively. The energies W_0 and W_ϵ are given by (6.4) and (6.5) respectively. It can be shown that

$$m_0\,|m_0'|\,\zeta_0 = \omega/2c^2(1 - \alpha_{30}^2) \tag{7.8}$$

and

$$m_{1_\epsilon}\,|(m_{1_\epsilon})_\omega|\,\zeta_\epsilon = \omega[2c^2(1 - \alpha_{3_\epsilon}^2)]^{-1} \times |\epsilon_2[1 + \cos\gamma\xi_1(1 - \epsilon_2/\epsilon_1)(\cot\theta_\epsilon\xi_2 - \xi_1\cos\gamma)]^{-1}|. \tag{7.9}$$

Inserting (7.2) and (7.3) into Eqs. (6.4) and (6.5) and noting relations (7.8) and (7.9), we obtain

$$W_0(\tau_1) = \frac{\lambda^2 e^2\xi_1^2}{2c^2\pi}\int_{\mathcal{P}_{1_0}} \frac{v\omega\alpha_{20}^2}{(1 - \alpha_{30}^2)}\, d\mathbf{P} \tag{7.10}$$

and

$$W_\epsilon(\tau_1) = \frac{\lambda^2 e^2}{2c^2\pi}\sum_{\pm}\int_{\mathcal{P}_{1_\epsilon}} \left[\frac{\omega v n_\epsilon^2\,|\epsilon_2|\,[(\alpha_{1_\epsilon}\alpha_{3_\epsilon}/\epsilon_1)\xi_1 + (\xi_2/\epsilon_2)(\alpha_{3_\epsilon}^2 - 1)]^2}{(1 - \alpha_{3_\epsilon}^2)\,|1 + (1 - \epsilon_2/\epsilon_1)\xi_1\cos\gamma(\cot\theta_\epsilon\xi_2 - \xi_1\cos\gamma)|}\right]_\pm d\mathbf{P}. \tag{7.11}$$

The expressions for W_0 and W_ϵ simplify when the particle's motion is uniform and is either parallel or perpendicular to the crystal axis.

A. *Parallel Motion* ($\xi_1 = 0$)

We denote the energies of the radiation in this case by $W_0^\parallel$ and $W_\epsilon^\parallel$. Equation (7.10) yields

$$W_0^\parallel(\tau_1) = 0. \tag{7.12}$$

From Eq. (5.27), we obtain

$$\tan\theta_\epsilon^\pm = \pm[(\epsilon_2/\epsilon_1)(\beta^2\epsilon_1 - 1)]^{\frac{1}{2}}. \tag{7.13}$$

It can be shown that

$$W_\epsilon^\parallel(\tau_1) = \frac{v\tau_1 e^2}{c^2}\int_{\substack{(\hat{\epsilon}_\epsilon/\hat{\epsilon}_1)(\beta^2\hat{\epsilon}_1 - 1) \geq 0 \\ \hat{\omega} > 0}} \hat{\omega}\,|1 - 1/\beta^2\hat{\epsilon}_1|\, d\hat{\omega}. \tag{7.14}$$

In (7.14), we have reintroduced the frequency $\hat{\omega}$ through the relation $\hat{\omega} = \lambda\omega$.

B. *Perpendicular Motion* ($\xi_1 = 1$)

Here, we denote the energies by $W_0^\perp$ and $W_\epsilon^\perp$. Inserting Eqs. (7.5), (7.6), and (7.7) into (7.10), we obtain

$$W_0^{\perp}(\tau_1) = \frac{v\tau_1 e^2}{2\pi c^2}\int_0^{2\pi} d\gamma \times \int_{\substack{\beta^2\hat{\epsilon}_1 \geq 1 \\ \hat{\omega}>0}} d\hat{\omega}\,\frac{\hat{\omega}(\beta^2\hat{\epsilon}_1 - 1)\sin^2\gamma}{(\cos^2\gamma + \beta^2\hat{\epsilon}_1\sin^2\gamma)}. \quad (7.15)$$

Equation (5.27) yields

$$\tan\theta_e^{\pm} = \pm\left[\frac{\epsilon_1(\beta^2\epsilon_2 - 1)}{\epsilon_1\sin^2\gamma + \epsilon_2\cos^2\gamma}\right]^{\frac{1}{2}}. \quad (7.16)$$

It can then be shown that

$$W_e^{\perp}(\tau_1) = \frac{e^2 v\tau_1}{\pi c^2} \times \iint \frac{\hat{\omega}\cos^2\gamma\,|(\beta^2\hat{\epsilon}_2 - 1)|\,d\hat{\omega}\,d\gamma}{\beta^2\,|\sin^2\gamma\hat{\epsilon}_1 + \hat{\epsilon}_2\cos^2\gamma|\,|\hat{\epsilon}_1\beta^2\sin^2\gamma + \cos^2\gamma|}. \quad (7.17)$$

The domain of integration in (7.17) is defined by

$$0 \leq \frac{\hat{\epsilon}_1(\beta^2\hat{\epsilon}_2 - 1)}{\hat{\epsilon}_1\sin^2\gamma + \hat{\epsilon}_2\cos^2\gamma}; \quad 0 \leq \gamma < \pi; \quad 0 < \omega.$$

In this section, we have obtained an expression for the energy of the Cerenkov radiation which occurs when a charged particle moves along a straight line trajectory in a uniaxial crystal. We have seen that this energy depends on the angle that the source trajectory makes with the crystal axis. In the special cases of uniform parallel and perpendicular motion, the results are given by Eqs. (7.12), (7.14), (7.15), and (7.17). These results agree with those given in Ref. 8. (The azimuthal angle ϕ in Ref. 8 is the complement of our azimuthal angle γ.)

8. The Asymptotic Expansion for Optically Active Media

In this section we assume that the matrix $\mathcal{E}(\omega)$ has the form

$$\mathcal{E}(\omega) = \begin{bmatrix} \mathcal{E}_1(\omega) & 0 \\ 0 & I \end{bmatrix}; \quad \mathcal{E}_1(\omega) = \begin{bmatrix} \epsilon_1 & -i\epsilon_g & 0 \\ i\epsilon_g & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_2 \end{bmatrix}. \quad (8.1)$$

The fact that $\mathcal{E}(\omega)$ is Hermitian implies that the quantity $\epsilon_g(\omega)$, called the "gyration parameter," is real. It follows from the definition of $\mathcal{E}(\omega)$ given in Sec. 1 that $\mathcal{E}(-\omega) = \overline{\mathcal{E}(\omega)}$, and therefore $\epsilon_g(-\omega) = -\epsilon_g(\omega)$. A matrix $\mathcal{E}(\omega)$ of this form represents what is called an optically active or gyrotropic medium with the optic axis the x_3 axis.

The dispersion relation (1.19) yields the equation

$$[\omega^2\epsilon_1 - c^2k^2]\left[\omega^2\epsilon_2 - c^2k^2 + c^2k_3^2\left(1 - \frac{\epsilon_2}{\epsilon_1}\right)\right] + \frac{\omega^2\epsilon_g^2}{\epsilon_1}[c^2(k^2 - k_3^2) - \omega^2\epsilon_2] = 0. \quad (8.2)$$

Equation (8.2) can be solved for k^2. The result is

$$k_{0,e}^2 = m_{0,e}^2(\omega, k_3) = \frac{\omega^2}{2c^2}\left(\left[\epsilon_1 + \epsilon_2 - \frac{\epsilon_g^2}{\epsilon_1} + \frac{c^2k_3^2}{\omega^2}\left(1 - \frac{\epsilon_2}{\epsilon_1}\right)\right] \pm \left\{\left[\frac{\epsilon_g^2}{\epsilon_1} - \epsilon_1 - \epsilon_2 - \frac{c^2k_3^2}{\omega^2}\left(1 - \frac{\epsilon_2}{\epsilon_1}\right)\right]^2 - 4\left[\epsilon_1\epsilon_2 + (\epsilon_1 - \epsilon_2)\frac{c^2k_3^2}{\omega^2} - \frac{\epsilon_g^2}{\epsilon_1\omega^2}(c^2k_3^2 + \omega^2\epsilon_2)\right]\right\}^{\frac{1}{2}}\right). \quad (8.3)$$

We denote by $\mathbf{r}_{0,e} = [\mathbf{R}_{10,e}, \mathbf{R}_{20,e}]$ the null-eigenvectors of the dispersion matrix G corresponding to the quantities $k_{0,e}$. If we set $n_{0,e} = ck_{0,e}/|\omega|$ and assume $|\alpha_3| \neq 1$, the vectors $\mathbf{R}_{\nu 0,e}$ take the form

$$\mathbf{R}_{10,e} = \zeta_{0,e}^{\frac{1}{2}}[(\epsilon_1 - n_{0,e}^2)\mathbf{A} + \nu_{0,e}\alpha_3\hat{\mathbf{X}}_3 + i\epsilon_g\mathbf{A}\times\hat{\mathbf{X}}_3], \quad (8.4)$$

$$\mathbf{R}_{20,e} = n_{0,e}\zeta_{0,e}^{\frac{1}{2}}(\nu_{0,e}\alpha_3\mathbf{A}\times\hat{\mathbf{X}}_3 + i\epsilon_g\mathbf{A}\times\mathbf{A}\times\hat{\mathbf{X}}_3). \quad (8.5)$$

Here,

$$\nu_{0,e} = \left[\frac{(\epsilon_1 - \epsilon_2)(\epsilon_1 - n_{0,e}^2) - \epsilon_g^2}{\epsilon_2 - n_{0,e}^2}\right]. \quad (8.6)$$

The normalization factors $\zeta_{0,e}$ are determined by condition (1.21) and are given by

$$\zeta_{0,e}^{-1} = (1 - \alpha_3^2)\{(\omega\epsilon_1)_\omega[\epsilon_g^2 + (\epsilon_1 - n_{0,e}^2)^2] [(\omega\epsilon_2)_\omega\alpha_3^2(1 - \alpha_3^2)n_{0,e}^4\nu_{0,e}^2]\epsilon_2^{-2} - 2(\omega\epsilon_g)_\omega\epsilon_g(\epsilon_1 - n_{0,e}^2) + n_{0,e}^2(\alpha_3^2\nu_{0,e}^2 + \epsilon_g^2)\}. \quad (8.7)$$

When $|\alpha_3| = 1$, $n_{0,e}^2 = \epsilon_1 \pm \epsilon_g$ and the vectors $\mathbf{r}_{0,e}$ take the forms,

$$\mathbf{r}_0 = \{2[(\omega\epsilon_1)_\omega + \omega(\epsilon_g)_\omega + \epsilon_1]\}^{-\frac{1}{2}} \times \{[\hat{\mathbf{X}}_1, (\epsilon_1 + \epsilon_g)^{\frac{1}{2}}\hat{\mathbf{X}}_2] + i[\hat{\mathbf{X}}_2, -(\epsilon_1 + \epsilon_g)^{\frac{1}{2}}\hat{\mathbf{X}}_1]\} \quad (8.8)$$

and

$$\mathbf{r}_e = \{2[(\omega\epsilon_1)_\omega - \omega(\epsilon_g)_\omega + \epsilon_1]\}^{-\frac{1}{2}} \times \{[\hat{\mathbf{X}}_2, -(\epsilon_1 - \epsilon_g)^{\frac{1}{2}}\hat{\mathbf{X}}_1] + i[\hat{\mathbf{X}}_1, (\epsilon_1 - \epsilon_g)^{\frac{1}{2}}\hat{\mathbf{X}}_2]\}. \quad (8.9)$$

In this section, we take the source function $\mathbf{f}(t, \mathbf{X})$ to be of the form (1.26), where the source trajectory is given by Eqs. (5.13) and (5.14). Let $\omega = h_{0,e}(k, k_3)$ represent those real functions for which $k_{0,e} = m(h_{0,e}, k_3)$. It can be shown that

$$\mathbf{u}(t, \mathbf{X}) = \mathbf{u}_0(t, \mathbf{X}) + \mathbf{u}_e(t, \mathbf{X}), \quad (8.10)$$

[8] J. V. Jelly, *Cerenkov Radiation and Its Applications* (Pergamon Press, Inc., New York, 1959).

where the integral representations of $\mathbf{u}_{0,\epsilon}$ are given by Eqs. (5.10) and (5.11). Of course the null-vectors $\mathbf{r}_{0,\epsilon}$ and functions $\omega = h_{0,\epsilon}$ corresponding to gyrotropic media must be inserted into these equations.

The parametric representations of the first terms of the asymptotic expansions of $\mathbf{u}_0$ and $\mathbf{u}_\epsilon$ are given by

$$\mathbf{u}_{0,\epsilon}(t, \mathbf{X}) \sim \text{Re } [[|j_{0,\epsilon}(t; \mathbf{P})|]^{-\frac{1}{2}}\mathbf{d}_{0,\epsilon}(\mathbf{P}) \times \exp [i\lambda s_{0,\epsilon}(t; \mathbf{P})]], \quad \mathbf{P} = (\tau, \omega, \gamma); \tag{8.11}$$

$$\mathbf{d}_{0,\epsilon}(\mathbf{P}) = \frac{\lambda^{d-2}}{2\pi}[m_{10,\epsilon}\,|(m_{10,\epsilon})_\omega|/v]^{\frac{1}{2}} \times \exp [\tfrac{1}{4}\pi i \operatorname{sig}(\phi^{0,\epsilon}_{\nu\beta})](\hat{\mathbf{g}}, \mathbf{r}_{0,\epsilon})\mathbf{r}_{0,\epsilon}, \tag{8.12}$$

and

$$\mathbf{X}_{0,\epsilon} - \mathbf{Y}(\tau) = (t - \tau)\mathbf{G}_{0,\epsilon} = (t - \tau)/(m_{0,\epsilon})_\omega \times [\mathbf{A}_{0,\epsilon} - (m_{0,\epsilon})_{k_3}\hat{\mathbf{X}}_3], \quad t \geq \tau. \tag{8.13}$$

In (8.11),

$$s_{0,\epsilon}(t; \mathbf{P}) = \left[\frac{m_{0,\epsilon}}{(m_{0,\epsilon})_\omega} - \frac{(m_{0,\epsilon})_{k_3}}{(m_{0,\epsilon})_\omega}\alpha_{30,\epsilon} - \omega\right] \times (t - \tau) + q(\tau), \tag{8.14}$$

$j_{0,\epsilon} = \det (d\mathbf{X}_{0,\epsilon}/d\mathbf{P})$ and $(\phi^{0,\epsilon}_{\nu\beta})$ is the matrix of second derivatives of $\phi_{0,\epsilon}$ [defined by Eq. (5.12)] with respect to k_ν and τ.

We may write

$$\mathbf{K}_{0,\epsilon} = m_{0,\epsilon}\mathbf{A}_{0,\epsilon}; \quad \mathbf{A}_{0,\epsilon} = \cos\theta_{0,\epsilon}\mathbf{T}^* + \sin\theta_{0,\epsilon}\cos\gamma\mathbf{N}^* + \sin\theta_{0,\epsilon}\sin\gamma\mathbf{B}^*, \tag{8.15}$$

where $\mathbf{T}^*$, $\mathbf{N}^*$, and $\mathbf{B}^*$ are defined by Eqs. (5.14) and (5.16). The Cerenkov–Doppler relation yields

$$\cos\theta_{0,\epsilon} = \eta/m_{0,\epsilon}; \quad \eta(\tau, \omega) = [\omega + \dot{q}(\tau)]/v(\tau). \tag{8.16}$$

It can be shown that τ is introduced into $m_{0,\epsilon}$ only through the function $\eta(\tau, \omega)$, and therefore we may set $m_{0,\epsilon}(\omega, k_3) = m_{10,\epsilon}(\eta, \omega, \gamma)$. This defines the square-root term in Eq. (8.12).

All quantities that appear in Eqs. (8.11)–(8.14) must be expressed in terms of the independent parameters $\mathbf{P}$. To accomplish this, we need only obtain the functions $m_{10,\epsilon}$. Once this is done, $\cos\theta_{0,\epsilon}$ is determined through the Cerenkov–Doppler relation, and hence $\mathbf{K}_{0,\epsilon}$ can be expressed in terms of $\mathbf{P}$. From Eqs. (8.15), (5.14), and (5.16), we obtain

$$\alpha_{30,\epsilon} = \cos\theta_{0,\epsilon}\xi_2 - \sin\theta_{0,\epsilon}\cos\xi_1. \tag{8.17}$$

Equations (8.2), (8.16), and (8.17) can, in principle, be used to determine the quantities $m_{10,\epsilon}$. This, however, necessitates the solving of an algebraic equation of fourth degree. In the special cases of parallel motion ($\xi_1 = 0$) and perpendicular motion ($\xi_1 = 1$), the equation reduces to a biquadratic which can be easily solved. We therefore restrict our considerations to these two cases.

A. Parallel Motion

It follows from Eqs. (8.16) and (8.17) that in this case $k_{30,\epsilon} = \eta(\tau, \omega)$. By inserting this relation into Eq. (8.3), the quantities $m_{10,\epsilon}$ are immediately obtained. The Cerenkov–Doppler relation (8.16) then yields the values of $\cos\theta^{\|}_{0,\epsilon}$. (Here, the superscript $\|$ denotes parallel motion.) If $\epsilon_1 \equiv \epsilon_2 \equiv \epsilon$, (8.3) yields the simple result,

$$m^2_{10,\epsilon} = \frac{\omega^2}{2\epsilon c^2}\left[2\epsilon^2 - \epsilon_g^2 \pm \epsilon_g\left(\epsilon_g^2 + \frac{4\epsilon c^2\eta^2}{\omega^2}\right)^{\frac{1}{2}}\right]. \tag{8.18}$$

We restrict $\theta^{\|}_{0,\epsilon}$ to lie between 0 and π, and allow γ to vary between 0 and 2π. The domains $\mathcal{P}^{\|}_{0,\epsilon}$ of the independent parameters are defined by the inequalities

$$0 \leq \tau, \tag{8.19}$$

$$0 \leq \cos^2\theta^{\|}_{0,\epsilon} \leq 1, \tag{8.20}$$

$$0 \leq \gamma < 2\pi. \tag{8.21}$$

B. Perpendicular Motion

Here, it is more convenient to solve for $\tan^2\theta^{\perp}_{0,\epsilon}$. It can be shown that $\tan^2\theta^{\perp}_{0,\epsilon}$ are the roots of the biquadratic equation

$$\left(\sin^2\gamma + \frac{\epsilon_2}{\epsilon_1}\cos^2\gamma\right)\tan^4\theta + \left[\frac{\omega^2\epsilon_g^2\sin^2\gamma}{c^2\eta^2\epsilon_1} - \left(\frac{\omega^2\epsilon_2}{c^2\eta^2} - 1\right) - \left(\frac{\omega^2\epsilon_1}{c^2\eta^2} - 1\right)\left(\sin^2\gamma + \frac{\epsilon_2}{\epsilon_1}\cos^2\gamma\right)\right]\tan^2\theta + \left[\left(\frac{\omega^2\epsilon_2}{c^2\eta^2} - 1\right)\left(\frac{\omega^2\epsilon_1}{c^2\eta^2} - \frac{\epsilon_g^2\omega^2}{\epsilon_1 c^2\eta^2} - 1\right)\right] = 0. \tag{8.22}$$

When $\epsilon_1 \equiv \epsilon_2 \equiv \epsilon$, (8.22) yields

$$\tan^2\theta^{\perp}_{0,\epsilon} = \left(\frac{\omega^2\epsilon}{c^2\eta^2} - 1\right) - \frac{\omega^2\epsilon_g^2\sin^2\gamma}{2\epsilon c^2\eta^2} \pm \frac{\omega\epsilon_g}{2c\eta\epsilon} \times \left[\frac{\omega^2\epsilon_g^2\sin^4\gamma}{c^2\eta^2} + 4\epsilon\left(\frac{\omega^2\epsilon}{c^2\eta^2} - 1\right)\cos^2\gamma\right]^{\frac{1}{2}}. \tag{8.23}$$

The domains $\mathcal{P}^{\perp}_{0,\epsilon}$ are defined by inequalities (8.19)–(8.21) with $\theta^{\|}_{0,\epsilon}$ replaced by $\theta^{\perp}_{0,\epsilon}$.

If $q(\tau) \equiv 0$, it can be shown that the relations $\mathbf{u}_{0,\epsilon} = 2\,\text{Re}\,[\mathbf{u}_{0,\epsilon+}]$ hold, where $\mathbf{u}_{0,\epsilon+}$ are obtained by restricting ω to be positive. Therefore, in this case, we add the restriction $0 < \omega$ to the definitions

of $\mathcal{P}^{\parallel}_{0,\epsilon}$ and $\mathcal{P}^{\perp}_{0,\epsilon}$, and multiply the corresponding quantities $\mathbf{d}_{0,\epsilon}$ by a factor of 2.

9. The Energy of Cerenkov Radiation for Optically Active Media

It can be shown that $W(\tau_1)$, the energy radiated from the source as it traverses the portion of trajectory defined by $0 \leq \tau \leq \tau_1$, is given by

$$W(\tau_1) = W_0(\tau_1) + W_\epsilon(\tau_1), \tag{9.1}$$

where

$$W_{0,\epsilon}(\tau_1) = \frac{\lambda^{2(d-2)}}{64\pi^3} \int_{\mathcal{P}_{10,\epsilon}} \frac{m_{10,\epsilon}\,|(m_{10,\epsilon})_\omega|}{v} \times |(\hat{\mathbf{g}}, \mathbf{r}_{0,\epsilon})|^2\, d\mathbf{P}. \tag{9.2}$$

Here, $\mathcal{P}_{10,\epsilon}$ are the domains of the independent parameters with the added restriction $\tau \leq \tau_1$.

Furthermore, the relation

$$m_{10,\epsilon}\,|(m_{10,\epsilon})_\omega|\,\zeta_{0,\epsilon} = \frac{\omega[\epsilon_2 - n^2_{0,\epsilon} + n^2_{0,\epsilon}(\alpha_{30,\epsilon})^2(1 - \epsilon_2/\epsilon_1)]}{2c^2\epsilon_g^2[1 - (\alpha_{30,\epsilon})^2]\mathfrak{L}_{0,\epsilon}} \tag{9.3}$$

holds, where

$$\begin{aligned}\mathfrak{L}_{0,\epsilon} = \{&\epsilon_1 + \epsilon_2 - n^2_{0,\epsilon}(\alpha_{30,\epsilon})^2 \\ &\times (1 - \epsilon_2/\epsilon_1) - 2n^2_{0,\epsilon} - \epsilon_g^2/\epsilon_1 \\ &+ \xi_1 \cos\gamma[(1 - \epsilon_2/\epsilon_1)(\epsilon_1 - \eta^2_{0,\epsilon}) - \epsilon_g^2/\epsilon_1] \\ &\times (\xi_2 \cot\theta_{0,\epsilon} - \xi_1 \cos\gamma)\}.\end{aligned} \tag{9.4}$$

We now apply the results given above to obtain the energy of the Cerenkov radiation which occurs when a charged particle moves uniformly through a gyrotropic medium. We restrict our considerations to the cases $\xi_1 = 0$ and $\xi_1 = 1$. Moreover, we assume that $\epsilon_1 \equiv \epsilon_2 \equiv \epsilon$. For the source under consideration $q(\tau) = 0$, $d = 3$ and $\hat{\mathbf{g}}$ is given by Eq. (7.1). The inner products $(\hat{\mathbf{g}}, \mathbf{r}_{0,\epsilon})$ are obtained from Eqs. (7.1), (8.4), and (8.5). Then, by using (9.3) and (9.4), it can be shown that

$$W^{\parallel}_{0,\epsilon}(\tau_1) = \frac{e^2 v \tau_1}{2c^2} \int_{\substack{0 \leq \cos^2\theta_{0,\epsilon} \leq 1 \\ 0 < \hat{\omega}}} \hat{\omega} \left|\left(1 - \frac{1}{\beta^2\hat{\epsilon}}\right) \times \left[1 \pm \frac{\beta\hat{\epsilon}_g\,|\beta^2\hat{\epsilon} + 1|}{(1 - \beta^2\hat{\epsilon})(4\hat{\epsilon} + \beta^2\hat{\epsilon}_g)^{\frac{1}{2}}}\right]\right| d\hat{\omega} \tag{9.5}$$

and

$$W^{\perp}_{0,\epsilon}(\tau_1) = \frac{e^2 v \tau_1}{4\pi c^2} \int_{\substack{0 \leq \cos^2\theta^{\perp}_{0,\epsilon} \leq 1 \\ 0 < \hat{\omega}}} \hat{\omega} \left|\left(1 - \frac{1}{\beta^2\hat{\epsilon}}\right) \times \left\{1 \pm \frac{\beta\hat{\epsilon}_g \sin^2\gamma}{[4\hat{\epsilon}(\hat{\epsilon}\beta^2 - 1)\cos^2\gamma + \beta^2\hat{\epsilon}_g^2 \sin^4\gamma]^{\frac{1}{2}}}\right\}\right| d\hat{\omega}\, d\gamma. \tag{9.6}$$

In this section we have obtained a general expression for the energy of the radiation which occurs in a gyrotropic medium. The results given by Eqs. (9.5) and (9.6) for the special case treated above agree with those given in Ref. 8.

APPENDIX I. DETERMINATION OF $|j(t\,\mathbf{P})/\det(\phi_{\nu\beta})|$

In this section, we compare the two functions $j(t; \mathbf{P})$ and $\det(\phi_{\nu\beta})$ for both isotropic and anisotropic media. Here, $j(t; \mathbf{P})$ is the Jacobian of the ray transformation, and $(\phi_{\nu\beta})$ is the matrix of second derivatives of the phase function

$$\phi = k_\nu[x_\nu - y_\nu(\tau)] - (t - \tau)h(\mathbf{K}) + q(\tau) \tag{AI.1}$$

with respect to the k_ν's and τ. The elements of $(\phi_{\nu\beta})$ are

$$\begin{aligned}\frac{\partial^2\phi}{\partial k_\nu\,\partial k_\beta} &= -(t - \tau)\frac{\partial^2 h}{\partial k_\nu\,\partial k_\beta}, \\ \frac{\partial^2\phi}{\partial k_\nu\,\partial\tau} &= \left[\frac{\partial h}{\partial k_\nu} - \dot{y}_\nu\right], \\ \frac{\partial^2\phi}{\partial\tau^2} &= -[(\mathbf{K}\cdot\ddot{\mathbf{Y}}) - \ddot{q}].\end{aligned} \tag{AI.2}$$

The ray equations and the "Cerenkov condition" are obtained by setting $\partial\phi/\partial k_\nu = 0$ and $\partial\phi/\partial\tau = 0$, respectively. Thus

$$x_\nu = y_\nu(\tau) + (t - \tau)\,\partial h/\partial k_\nu, \quad \nu = 1, 2, 3 \tag{AI.3}$$

and

$$k_\nu\dot{y}_\nu = h + \dot{q}. \tag{AI.4}$$

Differentiating Eqs. (AI.3) and (AI.4) with respect to k_β, we obtain

$$\frac{\partial x_\nu}{\partial k_\beta} = (t - \tau)\frac{\partial^2 h}{\partial k_\nu\,\partial k_\beta} + \left[\dot{y}_\nu - \frac{\partial h}{\partial k_\nu}\right]\frac{\partial\tau}{\partial k_\beta} \tag{AI.5}$$

and

$$\frac{\partial\tau}{\partial k_\beta} = \frac{[\partial h/\partial k_\beta - \dot{y}_\nu]}{[(\mathbf{K}\cdot\ddot{\mathbf{Y}}) - \ddot{q}]}, \quad [(\mathbf{K}\cdot\ddot{\mathbf{Y}}) - \ddot{q}] \neq 0. \tag{AI.6}$$

Equations (AI.2), (AI.5), and (AI.6) show that

$$\frac{\partial x_\nu}{\partial k_\beta} = -\frac{\partial^2\phi}{\partial k_\nu\,\partial k_\beta} + \frac{\partial^2\phi}{\partial k_\nu\,\partial\tau}\frac{\partial^2\phi}{\partial k_\beta\,\partial\tau}\Big/\frac{\partial^2\phi}{\partial\tau^2}. \tag{AI.7}$$

We may write $j(t; \mathbf{P})$ as the product

$$j(t; \mathbf{P}) = j_1(t; \mathbf{P})j_K(\mathbf{P}), \tag{AI.8}$$

where

$$j_1(t; \mathbf{P}) = \frac{\partial(x_1, x_2, x_3)}{\partial(k_1, k_2, k_3)}$$

and

$$j_K(\mathbf{P}) = \frac{\partial(k_1, k_2, k_3)}{\partial(\tau, \omega, \gamma)}. \tag{AI.9}$$

It can be easily shown from (AI.7) that simple determinant operations yield

$$j_1(t; \mathbf{P}) = -(\partial^2\phi/\partial\tau^2)^{-1} \det(\phi_{\nu\beta}). \quad \text{(AI.10)}$$

Therefore, to obtain the ratio of $j(t; \mathbf{P})$ to $\det(\phi_{\nu\beta})$, we need only determine $j_K(\mathbf{P})$.

A. Isotropic Media

In this case $k = m(\omega)$ and Eqs. (2.5) and (2.27) yield

$$\mathbf{K} = m(\omega)\mathbf{A} = \operatorname{sgn}[m'(\omega)]m(\omega)(\cos\theta\mathbf{T}^* + \sin\theta\cos\gamma\mathbf{N}^* + \sin\theta\sin\gamma\mathbf{B}^*), \quad \text{(AI.11)}$$

where

$$\cos\theta = \operatorname{sgn}[m'(\omega)]\left[\frac{\omega + \dot{q}(\tau)}{m(\omega)v(\tau)}\right]. \quad \text{(AI.12)}$$

The Jacobian $j_K(\mathbf{P})$ can be expanded as a scalar triple product. That is,

$$j_K(\mathbf{P}) = \mathbf{K}_\gamma\cdot(\mathbf{K}_\omega \times \mathbf{K}_\tau). \quad \text{(AI.13)}$$

Equation (AI.11) yields

$$\mathbf{K}_\gamma = m(\omega)\mathbf{A}_\gamma = m(\omega)(\mathbf{T}^* \times \mathbf{A}), \quad \text{(AI.14)}$$

$$\mathbf{K}_\omega = m'(\omega)\mathbf{A} + m(\omega)\mathbf{A}_\omega, \quad \text{(AI.15)}$$

$$\mathbf{K}_\tau = m(\omega)\mathbf{A}_\tau. \quad \text{(AI.16)}$$

Inserting (AI.14) into (AI.13), we obtain

$$j_K = m(\omega)[(\mathbf{T}^* \times \mathbf{A})\cdot(\mathbf{K}_\tau \times \mathbf{K}_\omega)]$$
$$= m(\omega)[(\mathbf{T}^*\cdot\mathbf{K}_\tau)(\mathbf{A}\cdot\mathbf{K}_\omega) - (\mathbf{T}^*\cdot\mathbf{K}_\omega)(\mathbf{A}\cdot\mathbf{K}_\tau)]. \quad \text{(AI.17)}$$

To obtain $(\mathbf{T}^*\cdot\mathbf{K}_\tau)$, we differentiate (AI.4) with respect to τ. This yields

$$(\mathbf{T}^*\cdot\mathbf{K}_\tau) = -[(\mathbf{K}\cdot\ddot{\mathbf{Y}}) - \ddot{q}]/v. \quad \text{(AI.18)}$$

It is easily seen from Eq. (AI.11) that $(\mathbf{A}\cdot\mathbf{A}_\omega) = 0$. Furthermore, if we make use of the Frênet formulas (see Ref. 2, p. 87), we find that $\mathbf{A}\cdot\mathbf{A}_\tau = 0$. Thus,

$$(\mathbf{A}\cdot\mathbf{K}_\omega) = m'(\omega), \quad \text{(AI.19)}$$

$$(\mathbf{A}\cdot\mathbf{K}_\tau) = 0. \quad \text{(AI.20)}$$

Inserting Eqs. (AI.18)–(AI.20) into (AI.17), we obtain

$$j_K(\mathbf{P}) = -\frac{m(\omega)m'(\omega)}{v(\tau)}[(\mathbf{K}\cdot\ddot{\mathbf{Y}}) - \ddot{q}]. \quad \text{(AI.21)}$$

Finally, from Eqs. (AI.8), (AI.10), and (AI.21) we have

$$\left|\frac{j(t; \mathbf{P})}{\det(\phi_{\nu\beta})}\right| = \frac{m(\omega)\,|m'(\omega)|}{v(\tau)}. \quad \text{(AI.22)}$$

B. Anisotropic Media

We assume here that the source trajectory is a straight line so that $\mathbf{T}^*$, $\mathbf{N}^*$, and $\mathbf{B}^*$ may be taken independent of τ. In anisotropic media $k = m(\tau, \omega, \gamma)$. Furthermore, it is convenient to define θ as the angle between $\mathbf{A}$ and $\mathbf{T}^*$. Therefore, Eqs. (AI.11) and (AI.12) hold with $m(\omega)$ replaced by $m(\tau, \omega, \gamma)$ and $\operatorname{sgn}[m'(\omega)]$ set equal to 1. We then have

$$\mathbf{K}_\gamma = m_\gamma\mathbf{A} + m\mathbf{A}_\gamma = m_\gamma\mathbf{A} + m(\mathbf{T}^* \times \mathbf{A}), \quad \text{(AI.23)}$$

$$\mathbf{K}_\tau = m_\tau\mathbf{A} + m\mathbf{A}_\tau, \quad \text{(AI.24)}$$

$$\mathbf{K}_\omega = m_\omega\mathbf{A} + m\mathbf{A}_\omega. \quad \text{(AI.25)}$$

Inserting (AI.23) into (AI.13), we obtain

$$j_K = m_\gamma\mathbf{A}\cdot(\mathbf{K}_\tau \times \mathbf{K}_\omega) + m(\mathbf{T}^* \times \mathbf{A})\cdot(\mathbf{K}_\tau \times \mathbf{K}_\omega). \quad \text{(AI.26)}$$

Because $\mathbf{T}^*$, $\mathbf{N}^*$, and $\mathbf{B}^*$ are independent of τ, $\mathbf{A}_\omega$ and $\mathbf{A}_\tau$ are in the same direction. Equations (AI.24) and (AI.25) then show that $\mathbf{A}\cdot(\mathbf{K}_\omega \times \mathbf{K}_\tau) = 0$ and therefore (AI.17) holds with $m(\omega)$ replaced by $m(\tau, \omega, \gamma)$. Furthermore, Eqs. (AI.11), (AI.24), and (AI.25) yield the relations

$$(\mathbf{A}\cdot\mathbf{K}_\tau) = m_\tau, \quad \text{(AI.27)}$$

$$(\mathbf{A}\cdot\mathbf{K}_\omega) = m_\omega, \quad \text{(AI.28)}$$

$$(\mathbf{T}^*\cdot\mathbf{K}_\omega) = (m\cos\theta)_\omega, \quad \text{(AI.29)}$$

$$(\mathbf{T}^*\cdot\mathbf{K}_\tau) = (m\cos\theta)_\tau. \quad \text{(AI.30)}$$

We define the quantity $\eta(\tau, \omega) = m\cos\theta = [(\omega + \dot{q})/v]$. Using (AI.12), Eqs. (AI.29) and (AI.30) become

$$(\mathbf{T}^*\cdot\mathbf{K}_\omega) = \eta_\omega = 1/v \quad \text{(AI.31)}$$

and

$$(\mathbf{T}^*\cdot\mathbf{K}_\tau) = \eta_\tau = [\ddot{q} - (\ddot{\mathbf{Y}}\cdot\mathbf{K})]/v. \quad \text{(AI.32)}$$

We note that τ is introduced into $m(\tau, \omega, \gamma)$ through the quantity $\eta(\tau, \omega)$. Therefore, we may write

$$m(\tau, \omega, \gamma) = m_1(\eta, \omega, \gamma). \quad \text{(AI.33)}$$

From (AI.33), we have

$$m_\omega = (m_1)_\eta\eta_\omega + (m_1)_\omega = (m_1)_\eta/v + (m_1)_\omega \quad \text{(AI.34)}$$

and

$$m_\tau = (m_1)_\eta\eta_\tau = (m_1)_\eta[\ddot{q} - (\ddot{\mathbf{Y}}\cdot\mathbf{K})]^{-1}. \quad \text{(AI.35)}$$

Inserting (AI.27), (AI.28), (AI.31), and (AI.32) into (AI.17), we obtain

$$j_K = [m_1(m_1)_\omega v^{-1}][\ddot{q} - (\ddot{\mathbf{Y}}\cdot\mathbf{K})]. \quad \text{(AI.36)}$$

Finally, Eqs. (AI.8), (AI.10), and (AI.36) yield

$$\left|\frac{j(t;\mathbf{P})}{\det(\phi_{\nu\beta})}\right| = \frac{m_1\,|(m_1)_\omega|}{v}. \qquad \text{(AI.37)}$$

APPENDIX II. CAUSTICS IN ISOTROPIC MEDIA

Those points $(t, \mathbf{X})$, corresponding [through Eq. (2.30)] to parameter values $\mathbf{P} = (\tau, \omega, \gamma)$ for which $j(t; \mathbf{P})$ (the Jacobian of the ray transformation) vanishes, form a locus in space–time called a caustic. It is of interest to discuss the geometry of these caustics, because, as we shall see, the fields at caustic points are more intense (higher order in λ) than the fields at ordinary points.

We assume that $q(\tau) \equiv 0$ and that $\mathbf{Y}(\tau) = v\tau\hat{\mathbf{X}}_1$, where v is a constant. Equations (1.32) and (2.32) yield

$$j(t;\mathbf{P}) = \frac{(t-\tau)^2}{vmm'}\left[\frac{m''m}{(m')^2}v^2\left(1-\frac{\omega^2}{m^2v^2}\right) - \left(\frac{1}{m'}-\frac{\omega}{m}\right)^2\right]. \qquad \text{(AII.1)}$$

We see that $j(t; \mathbf{P})$ vanishes when $t = \tau$, an expression of the fact that the source trajectory itself is a caustic line. Our concern here is with caustics away from the trajectory. The factor within the brackets in Eq. (AII.1) is a function of ω only. Suppose there exists a value of ω, say ω_0, for which this factor vanishes, and such that $\beta n(\omega_0) \geq 1$. Then, any ray corresponding to parameter values $\mathbf{P} = (\tau, \omega_0, \gamma)$ lies on a caustic surface for all time $t > \tau$.

To investigate the geometry of this caustic surface, we must consider the space–time picture of the rays. Figure 1(a) is a typical space–time diagram restricted to two spatial dimensions. We see from this diagram that there is a v-shaped envelope of the ray surfaces formed whose vertex lies on the source trajectory. It can be shown that this envelope is generated by precisely those rays which correspond to $\omega = \omega_0$ and is therefore the caustic under consideration.

To obtain the space picture of the caustic at time t_1, we need only determine the intersection of the plane $t = t_1$ and the envelope. It is easy to see that the intersection is the boundary of the v-shaped figure which appears in Fig. 1(a). The third spatial dimension of the caustic is obtained by rotating this v-shaped boundary about the x_1 axis. The result is the surface of the cone depicted in Fig. 2. This

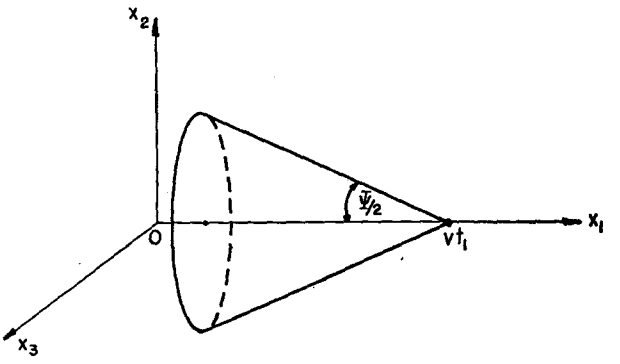


FIG. 2. Space picture of caustic surface at time $t = t_1$.

cone proceeds into the medium in the x_1 direction and with the source speed v. It can be shown that the vertex angle Φ is given by

$$\Phi = 2\tan^{-1}\{[v^2m^2(\omega_0)/\omega_0^2 - 1]^{-\frac{1}{2}} \times [1 - v^2m(\omega_0)m'(\omega_0)/\omega_0]\}. \qquad \text{(AII.2)}$$

To obtain the asymptotic expansion valid at the caustic, a more detailed stationary phase analysis than that given in Sec. 2 is required. The result is given parametrically by

$$\mathbf{u}(t,\mathbf{X}) \sim \frac{\lambda^{d-11/6}\Gamma(\frac{1}{3})m(m')^{3/2}}{(2\pi)^{3/2}(t-\tau)^{5/6}(3\,|m''|)^{1/2}(|d|)^{1/3}} \times \sum_{i=1}^{2}(\hat{\mathbf{g}},\mathbf{r}^i)\mathbf{r}^i\cos\left[\lambda\left(\frac{m}{m'}-\omega\right)\times(t-\tau)+\tfrac{1}{4}\pi(\operatorname{sgn} m''-2)\right] \qquad \text{(AII.3)}$$

and

$$\mathbf{X} = \mathbf{Y}(\tau) + (t-\tau)\mathbf{A}/m', \qquad \text{(AII.4)}$$

where

$$d = \frac{m'}{m''}\left(\frac{m'\omega}{m}-1\right)\left[\frac{m'''m'}{b(m'')^2}\left(\frac{m'\omega}{m}-1\right)^2 + \tfrac{3}{4}v^2(m')^2\left(1-\frac{\omega^2}{m^2v^2}\right)\right]. \qquad \text{(AII.5)}$$

In Eqs. (AII.3)–(AII.5), ω is to be set equal to ω_0.

Comparison of (AII.3) with Eq. (2.36) shows that the solution at the caustic is indeed of higher order in λ than the solution away from the caustic. In fact, we see that at caustic points the solution behaves like $\lambda^{d-11/6}$, whereas at ordinary points the solution behaves like λ^{d-2}.

Complex Temperatures and Phase Transitions*

GERALD L. JONES
Department of Physics, University of Notre Dame, Notre Dame, Indiana
(Received 17 March 1966)

The thermodynamic limit is considered for complex temperatures, and a picture of a phase transition, similar to the Yang–Lee picture, is proposed. For certain cases a representation of the partition function as an infinite product is obtained. Some simple models are considered.

I. INTRODUCTION

IN this paper we wish to propose a picture of phase transitions, which is similar to the Yang–Lee[1,2] picture except that we deal with complex temperatures rather than with complex fugacity. The reasons one might want to consider complex temperatures are analogous to those for complex fugacity. One expects the free energy of a thermodynamic system to be a nonanalytic function of the temperature in the neighborhood of a point of phase transition. This is particularly noticeable in what is apparently a logarithmic singularity in the specific heat of fluids as a function of temperature in the neighborhood of the critical point.[3] Since the free energy is an analytic function of the temperature for finite systems, this singularity must be connected with the thermodynamic limit. The hope is that one can form a simple picture of the way that the singularities occur, and more optimistically, something about their analytic properties, by considering the thermodynamic limit for complex temperatures. To this end we have essentially duplicated part of the Yang–Lee theory for complex temperatures. In Sec. II we prove, for a general class of systems, a theorem analogous to the second Yang–Lee theorem.[1] We treat both the canonical and the grand canonical ensemble, and in the grand canonical case both the temperature and the fugacity are complex numbers. Although we do not consider quantum systems, these results may follow by similar arguments for such systems, provided one makes allowances for the singular behavior of the free-particle partition function for bosons. In Sec. III we consider some applications to simple systems.

The results of this paper will also be used in a following paper, where the thermodynamic equivalence of a rather general class of ensembles is considered.

* This work was supported in part by the United States Atomic Energy Commission.
[1] C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404 (1952).
[2] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[3] M. E. Fisher, Phys. Rev. **136**, A1599 (1964).

II. SOME GENERAL RESULTS

We consider classical N-particle systems with Hamiltonian

$$H = \sum_{i=1}^{N} p_i^2/2m + u_N(\mathbf{r}_1 \cdots \mathbf{r}_N). \tag{1}$$

Define, for any complex z in the open left half-plane,

$$Z(z, N, \Omega) = \frac{\Lambda^{-3N}(z)}{N!} \int_\Omega e^{zu_N(\mathbf{r}_1 \cdots \mathbf{r}_N)}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \tag{2}$$

where we assume u_N bounded from below so that the integrand is finite for Re $z < 0$.

$$\Lambda(z) = (-zh^2/2\pi m)^{\frac{1}{2}}; \tag{3}$$

Ω is the spatial domain over which each $\mathbf{r}_i$ varies. We want to put sufficient restrictions on the system so that the thermodynamic limit exists for z real and negative. This problem has been treated recently and comprehensively by Ruelle[4] and Fisher.[5] One goes to the thermodynamic limit by choosing a sequence of numbers N_i and a sufficiently regular sequence of domains Ω_i with volume V_i such that $N_i \to \infty$, $V_i \to \infty$, $N_i V_i^{-1} \to 1/v$. With sufficient conditions on $u_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ and on the sequence of domains one can show that

$$\lim_{i\to\infty} N_i^{-1} \ln Z(z, N_i, \Omega_i) = f(z, v) \tag{4}$$

for z real and negative. Since v will be fixed in this paper, we omit it from now on. It is also known that $f(z)$ is a continuous convex downward function on the negative real axis. This ensures that the specific heat is positive whenever it exists.

For the exact conditions on the $u_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ and on the sequence Ω_i, we refer to Refs. 4 and 5. We state, however, one condition on u_N which we need. There exists a $K > 0$ such that

$$u_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \geq -NK \tag{5}$$

for all $\mathbf{r}_1 \cdots \mathbf{r}_N$ and all N. Such potentials are called stable. In addition we need sufficient continuity on the part of $u_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ to ensure that (2) is an

[4] D. Ruelle, Helv. Phys. Acta **36**, 183 (1963).
[5] M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 410 (1964).

analytic function of z in the open left half-plane. A sufficient condition, although not the most general, which seems to include most u_N of interest is that u_N be of the form

$$u_N(\mathbf{r}_1 \cdots \mathbf{r}_N) = R_N(\mathbf{r}_1 \cdots \mathbf{r}_N) + S_N(\mathbf{r}_1 \cdots \mathbf{r}_N), \quad (6)$$

where

$$R_N(\mathbf{r}_1 \cdots \mathbf{r}_N) = \begin{cases} +\infty & \text{if } (\mathbf{r}_1 \cdots \mathbf{r}_N) \in A_N, \\ 0 & \text{otherwise,} \end{cases}$$

where A_N is some measurable subset of R^{3N} (the space of $\mathbf{r}_1 \cdots \mathbf{r}_N$). We require $S_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ to be continuous, but not necessarily finite. Roughly, R_N contains the hard-core parts of the interaction and S_N contains everything else. We prove in Appendix I that, under this assumption, $Z_i(z) = Z(z, N_i, \Omega_i)$ is analytic in z for Re $z < 0$. Clearly, $Z_i(z) > 0$ for z real and negative, hence $N_i^{-1} \ln Z_i(z)$ is analytic in some neighborhood of the negative real axis, and therefore the free energy per particle is an analytic function of the (real) temperature for any finite system. The main results of this section are theorems which are analogous to the second Yang–Lee theorem.[1]

Theorem I: Let R be any bounded region of the open left-half z plane, and suppose R contains a segment of the real axis and no zeros of $Z_i(z)$ for any j. Then $\lim_{i\to\infty} N_i^{-1} \ln Z_i(z) = f(z)$ exists everywhere in R and $f(z)$ is analytic.

Proof: The proof of this theorem depends, as usual, on an application of the Vitali convergence theorem which, in one of its forms, asserts that, if a uniformly bounded sequence of functions, each of which is analytic in a region, converges to a limit on a set of points which has an accumulation point in the region, then the sequence converges uniformly everywhere in the region and the limit function is analytic.[6]

We first consider the upper bound on $Z_i(z)$ generated by the stability condition (5). If Re $z < 0$ we have

$$\begin{aligned} |Z_i(z)| &= \frac{|\Lambda(z)|^{-3N_i}}{N_i!} \left| \int_{\Omega_i} e^{zu_N(\mathbf{r}_1 \cdots \mathbf{r}_N)}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N \right| \\ &\leq \frac{|\Lambda(z)|^{-3N_i}}{N_i!} \int_{\Omega_i} \left| e^{zu_N(\mathbf{r}_1 \cdots \mathbf{r}_N)} \right| d\mathbf{r}_1 \cdots d\mathbf{r}_N \\ &= \frac{|\Lambda(z)|^{-3N_i}}{N!} \int_{\Omega_i} e^{(\mathrm{Re}\, z)u_N(\mathbf{r}_1 \cdots \mathbf{r}_N)}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N \\ &\leq \frac{|\Lambda(z)|^{-3N_i}}{N_i!} V_i^{N_i} e^{-(\mathrm{Re}\, z)N_i K}. \end{aligned} \quad (7)$$

[6] C. Carathéodory, *Theory of Functions* (Chelsea Publishing Company, New York, 1954), Vol. I.

Now, to get a uniformly bounded sequence, we define

$$T_i(z) = (Z_i(z) N_i! / V_i^{N_i})^{1/N_i}. \quad (8)$$

Since $Z_i(z) \neq 0$ in the region R by assumption, $T_i(z)$ can be chosen to be analytic everywhere in R for all j. From (7) we get a uniform upper bound on $T_i(z)$ of

$$|T_i(z)| \leq |\Lambda(z)|^{-1} e^{-\mathrm{Re}\, zK}, \quad (9)$$

which is finite for any bounded region of the open left-half z plane. We now show that $T_i(z)$ converges when z is real and negative. For any z in R, we can take the logarithm of (8), since $T_i(z) \neq 0$; hence

$$\ln T_i(z) = N_i^{-1} \ln Z_i(z) + N_i^{-1} \ln \frac{N_i!}{V_i^{N_i}}. \quad (10)$$

By Stirling's approximation,

$$N! = A(N) N^{N+\frac{1}{2}} e^{-N}, \quad \text{where} \quad e^{11/12} < A(N) < e$$

for all N. Hence,

$$\begin{aligned} N_i^{-1} \ln N_i! V_i^{-N_i} &= N_i^{-1} \ln A(N_i) \\ &\quad + (\tfrac{1}{2}) N_i^{-1} \ln N_i - 1 + \ln N_i V_i^{-1}. \end{aligned}$$

Then

$$\begin{aligned} \lim_{i\to\infty} N_i^{-1} \ln N_i! V_i^{-N_i} &= \ln (1/v) - 1 = -\ln v - 1. \end{aligned} \quad (11)$$

Therefore, if the limits in (10) exist, we must have

$$\lim_{i\to\infty} \ln T_i(z) = \lim_{i\to\infty} N_i^{-1} \ln Z_i(z) - \ln v - 1. \quad (12)$$

But the limit on the right is known to exist for z real and negative, hence, so must the limit on the left. By the continuity of the exponential function, then $\lim_{i\to\infty} T_i(z)$ exists for z real and negative. If we set $z = -\beta$, $0 < \beta < \infty$, and put $\lim_{i\to\infty} T_i(-\beta) = t(-\beta)$, then from (12) and (4) we have

$$t(-\beta) = v^{-1} e^{f(-\beta)-1}. \quad (13)$$

Now, by the Vitali theorem,

$$\lim_{i\to\infty} T_i(z) = t(z) \quad (14)$$

everywhere in R and $t(z)$ is analytic in R. Furthermore, by a theorem due to Hurwitz,[6] since $T_i(z) \neq 0$ in R for all j, we must have either $t(z) \neq 0$ for any z in R or $t(z)$ identically zero in R. Hence, by continuity of the logarithm $\lim_{i\to\infty} \ln T_i(z) = \ln t(z)$ everywhere in R. From (12)

$$\lim_{i\to\infty} N_i^{-1} \ln Z_i(z) = f(z) = \ln t(z) + \ln v + 1 \quad (15)$$

for all z in R. The convergence is uniform and the limit $f(z)$ is analytic in R.

The zeros of $Z_j(z)$ must be a set of isolated points, otherwise $Z_j(z)$ would vanish everywhere. The hope is, of course, that these zeros lie on definite curves in the z plane, and become more dense on these curves as $j \to \infty$. One hopes these curves cross the negative real axis at only a few points, or not at all, depending on the specific volume and the nature of the system. If this is the case, then $f(z)$ is analytic on the negative real axis, except at the points of crossing which are then the points of phase transition. We mean by a phase transition here a singularity in the temperature behavior of the free energy rather than a singularity in the density behavior of the free energy. Since the specific heat is proportional to the second temperature derivative of the free energy, there can be no singularity in the specific heat except the neighborhood of the zeros of $Z_j(z)$. This is certainly a possible picture of the way things happen, however, it is not proven in general. The above theorem in itself does not preclude the possibility of the zeros of $Z_j(z)$ becoming everywhere dense in the open left-hand plane, in which case it would be impossible to say anything about the analytic properties of $f(z)$ from the theorem. It seems difficult to deduce anything in general about the zeros of $Z_j(z)$ from general properties of $u_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ and barring this one is reduced to examining particular systems which may be simple. We take this up in the next section.

We now examine what can be said for the grand canonical ensemble. Set

$$Q_j(z, w) = \sum_{n=0}^{\infty} Z(z, n, \Omega_j)[\Lambda^3(z)w]^n, \quad (16)$$

where z and w are complex and Re $z < 0$. We want to prove that the series represents a function, which is separately analytic in z and w for all w and for all z such that Re $z < 0$. Let us look at the seqeuence of partial sums (for fixed j)

$$S_m(z, w) = \sum_{n=0}^{m} Z(z, n, \Omega_j)[\Lambda^3(z)w]^n. \quad (17)$$

We show that this sequence converges uniformly in any bounded region of the four-dimensional (z, w) space for which Re $z < 0$. Let A be such a region and take any $\epsilon > 0$. Then from the upper bound (7) we have

$$\begin{aligned} |S_{m+p}(z, w) - S_m(z, w)| &= |\sum_{n=m+1}^{m+p} Z(z, n, \Omega_j)[\Lambda^3(z)w]^n| \\ &\leq \sum_{n=m+1}^{m+p} \frac{V_j^n |w|^n e^{-(\mathrm{Re}\, z)nK}}{n!}. \end{aligned} \quad (18)$$

Let

$$\alpha = \max_{(z,w)\in A} |w| V_j e^{-(\mathrm{Re}\, z)K}.$$

Clearly $\alpha < \infty$ if A is a bounded region. Now choose m so large that

$$\sum_{n=m+1}^{n=m+p} \alpha^n/n! < \epsilon \quad \text{for all} \quad p.$$

This can be done because series $\sum_{n=0}^{\infty} \alpha^n/n!$ converges. Then, for all $(z, w) \in A$, and for all p

$$|S_{m+p}(z, w) - S_m(z, w)| < \epsilon. \quad (19)$$

By the Cauchy criterion $S_m(z, w)$ converges uniformly in both variables, in the region A. But, holding either variable fixed, $S_m(z, w)$ is analytic in the other; hence, the limit $Q_j(z, w)$ is analytic in each variable separately so long as Re$z < 0$. By Hartog's theorem, $Q_j(z, w)$ is analytic in both variables. It is clear from (18) that

$$|Q_j(z, w)| \leq \exp [V_j |w| e^{-(\mathrm{Re}\, z)K}]. \quad (20)$$

Theorem II: Let w be a fixed real positive number and let R be any bounded region of the open left-half z plane, containing a segment of the real axis and no zeros of $Q_j(z, w)$ for any j. Then $\lim_{j\to\infty} V_j^{-1} \ln Q_j(z, w) = \pi(z, w)$ exists everywhere in R and the limit is an analytic function of z in R.

Proof: Set $T_j(z, w) = [Q_j(z, w)]^{V_j^{-1}}$. $T_j(z, w)$ is a nonzero analytic function of z in R for all j. By (20) the sequence is uniformly bounded,

$$|T_j(z, w)| \leq \exp [|w| e^{-(\mathrm{Re}\, z)w}]. \quad (21)$$

Furthermore,

$$\lim_{j\to\infty} \ln T_j(z, w) = \lim_{j\to\infty} V_j^{-1} \ln Q_j(z, w) = \pi(z, w) \quad (22)$$

whenever either of these limits exist. But it is proven in Refs. 1, 4, and 5 that the limit on the right exists when z is real and negative. Now, using the Vitali theorem, and using an argument entirely analogous to the one used for the canonical ensemble, we extend the convergence to all of R and obtain the analyticity of the limit function.

The second Yang–Lee Theorem[1] states that, if z is real and negative and R' is a bounded region of the complex w plane containing a segment of the positive real axis and no zeros of $Q_j(z, w)$ for any j, then $\lim_{j\to\infty} V_j^{-1} \ln Q_j(z, w) = \pi(z, w)$ exists everywhere in R', and the limit is an analytic function of w in R'. We could give a proof of this theorem by exactly the same method used in the previous theorem. It should be pointed out that the method used in these proofs does not depend on representing the partition function as a finite or infinite product. The last two theorems can be combined to give the following.

Theorem III: Let R be a bounded region of the space of two complex variables (z, w), such that $\mathrm{Re}\, z < 0$ in R. Suppose R contains no zeros of $Q_j(z, w)$ for any j, and that R contains a region of the two-dimensional subspace, z real and negative, w real and positive. Then $\lim_{j\to\infty} V_j^{-1} \ln Q_j(z, w) = \pi(z, w)$ exists everywhere in R and $\pi(z, w)$ is analytic in each variable separately in R.

One might now hope the following picture of a phase transition is valid. The set of zeros of $Q_j(z, w)$ is a set of two-dimensional surfaces in the four-dimensional space. As $j \to \infty$ the surfaces coalesce to form a three-dimensional surface which may intersect the plane, z real, w real, in a curve. This curve then gives the values of z and w for which the phase transitions occur (of course we must restrict $\mathrm{Re}\, z < 0$ as always). Again it seems difficult to prove this kind of behavior in general.

We close this section with some results for a more restricted class of potentials. Let us associate with each potential of the form (6) another potential obtained by changing the sign of the nonhard core part,

$$\tilde{u}_N(\mathbf{r}_1 \cdots \mathbf{r}_N) = R_N(\mathbf{r}_1 \cdots \mathbf{r}_N) - S_N(\mathbf{r}_1 \cdots \mathbf{r}_N). \quad (23)$$

We now restrict ourselves to those sets of potentials such that both u_N and $\tilde{u}_N$ satisfy sufficient conditions for the existence of a thermodynamic limit. In particular, there exists a $K > 0$ such that for all N and $\mathbf{r}_1 \cdots \mathbf{r}_N$,

$$\begin{aligned} u_N(\mathbf{r}_1 \cdots \mathbf{r}_N) &\geq -NK, \\ \tilde{u}_N(\mathbf{r}_1 \cdots \mathbf{r}_N) &\geq -NK. \end{aligned} \quad (24)$$

For example, any u_N which is a sum of two-body potentials with hard cores and bounded, finite-range tails will fall into this class. It is convenient to deal only with the configuration integral so we define

$$\begin{aligned} W(z, N, \Omega) &= Z(z, N, \Omega)\Lambda(z)^{3N} \\ &= 1/N! \int_\Omega e^{zu_N}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \end{aligned} \quad (25)$$

$$\tilde{W}(z, N, \Omega) = 1/N! \int_\Omega e^{z\tilde{u}_N}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N.$$

Now, on the set A_N, $R_N(\mathbf{r}_1 \cdots \mathbf{r}_N) = +\infty$, and for $\mathrm{Re} z < 0$, the integrands in (25) are zero, hence we can integrate only over Δ, the set of points $(\mathbf{r}_1 \cdots \mathbf{r}_N)$ such that each $\mathbf{r}_i \in \Omega$ but $(\mathbf{r}_1 \cdots \mathbf{r}_N) \notin A_N$. We can write

$$\begin{aligned} W(z, N, \Omega) &= 1/N! \int_\Delta e^{zS_N}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \\ \tilde{W}(z, N, \Omega) &= 1/N! \int_\Delta e^{-zS_N}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N. \end{aligned} \quad (26)$$

Both of these integrals are analytic functions of z for $\mathrm{Re}\, z < 0$. Since

$$W(z, N, \Omega) = \tilde{W}(-z, N, \Omega), \quad (27)$$

it is clear that both W and $\tilde{W}$ are entire functions, one of which is, in fact, just the reflection of the other through the origin. For such potentials, then we have $W(z, N, \Omega)$ analytic everywhere. From (7) and (27) we get the upper bound.

$$|W(z, N, \Omega)| \leq \frac{V(\Omega)^N}{N!} e^{|\mathrm{Re}\, z|NK} \quad (28)$$

for all z. This bound implies that the order of the entire function $W(z, N, \Omega)$ is at most one and hence the following factorization.[6]

$$W(z, N, \Omega) = e^{a+bz} \prod_{i=1}^{\infty} (1 - z/z_i)e^{z/z_i} \quad (29)$$

or

$$W(z, N, \Omega) = \prod_{i=1}^{\infty} (1 - z/z_i), \quad (30)$$

depending on whether the order is one or less than one. A similar discussion does not hold for the grand canonical ensemble. From (16),

$$\begin{aligned} Q(z, w, \Omega) &= \textstyle\sum_{n=0}^{\infty} W(z, n, \Omega)w^n, \\ \tilde{Q}(z, w, \Omega) &= \textstyle\sum_{n=0}^{\infty} \tilde{W}(z, n, \Omega)w^n. \end{aligned} \quad (31)$$

From (27)

$$Q(z, w, \Omega) = \tilde{Q}(-z, w, \Omega). \quad (32)$$

From (20) and (32)

$$|Q(z, w, \Omega)| \leq \exp V_i\, |w|\, e^{|\mathrm{Re}\, z|K}. \quad (33)$$

Now (32) implies Q is an entire function of z, but the bound (33) is not strong enough to give any information about the order of Q.

III. EXAMPLES

Since one cannot, at the present time, verify the proposed picture of a phase transition, it would be of some interest to verify it for at least some model. The natural thing to try is the two-dimensional Ising (lattice-gas) model, since it has been solved exactly in the thermodynamic limit and is known to exhibit a phase transition. Unfortunately, it seems difficult to find the zeros of the partition function

for this model, and it has not yet been accomplished. As a start in this direction, we can consider the one-dimensional Ising model. The theorems we have proven do not apply to an Ising model but one can carry through corresponding proofs for this model. The exact solution of the one-dimensional model, with nearest-neighbor interactions is known to be of the form[7]

$$Q(z, H, N) = \lambda_+^N(z, H) + \lambda_-^N(z, H), \qquad (34)$$

where z is the complex temperature ($z = -1/kT$ for z real and negative) and H is the external magnetic field

$$\lambda_\pm(z, H) = e^{-\frac{1}{2}zJ}\cosh(mHz) \pm [e^{-zJ}\sinh^2(mHz) + e^{zJ}]^{\frac{1}{2}}. \qquad (35)$$

The energy of the system is of the form

$$E = -\tfrac{1}{2}J\sum_{(i,j)}\mu_i\mu_j - mH\sum_i U_i, \qquad (36)$$

where the sum $\langle i, j\rangle$ goes over nearest neighbors and μ_i takes on the values ± 1. If we are to look for zeros of $Q(z, H, N)$ we must set $\lambda_+ = (-1)^{1/N}\lambda_-$. Using (35) we obtain the following transcendental equation:

$$\tanh(mHz) + \frac{e^{2Jz}}{\cosh^2(mHz)} = -\tan^2\pi(k + \tfrac{1}{2})/N, \qquad (37)$$

where $k = 0, 1 \cdots N - 1$. It is clear that the left side of (37) is greater than zero for all real z and H, hence there will never be any roots near the real axis and therefore no phase transition. For real H and complex z, there are roots of (37) and they become dense on rather complicated curves in the complex z plane. For the case $H = 0$, the curves are easy to find. We have

$$e^{2Jz} = -\tan^2[\pi(k + \tfrac{1}{2})/N]$$

or

$$z_- = (2J)^{-1}\ln\tan^2[\pi(k + \tfrac{1}{2})/N] \pm i\frac{\pi(2m+1)}{2J}, \qquad (38)$$

where m is any integer. The zeros become dense on lines parallel to the real z axis and displaced by $\pi(2m + 1)/3J$. For $H \neq 0$ Eq. (37) can, in general, only be solved approximately.

Two other systems which have simple analytic structure in the z plane are the free gas and the gas of hard spheres. For both of these systems, the configuration integral is independent of z, hence, all of the z dependence of the partition function is in $\Lambda(z)$, and these systems can exhibit no singular behavior as a function of temperature.

Note added in proof: The use of complex temperatures has been previously proposed by M. E. Fisher.[8] In this article Fisher finds the zeros of the two-dimensional Ising model and verifies the proposed picture of the phase transition.

ACKNOWLEDGMENT

The author would like to thank Dr. W. McGlinn for a useful suggestion on the proof of the theorems.

APPENDIX

We want to show, under assumption (6), that

$$C(z) = \int_\Omega e^{zu_N(\mathbf{r}_1\cdots\mathbf{r}_N)}\,d\mathbf{r}_1\cdots d\mathbf{r}_N$$

is an analytic function of z for $\mathrm{Re}\ z < 0$. We set $z = +\beta + i\alpha$, where $\beta < 0$. Referring to (6), let Δ be the set of points $(\mathbf{r}_1\cdots\mathbf{r}_N)$ such that each $\mathbf{r}_i \in \Omega$ but $(\mathbf{r}_1\cdots\mathbf{r}_N) \notin A_N$. Using the decomposition (6) and the fact that $e^{zu_N} = 0$ on A_N, we can write

$$C(z) = C_r(\beta, \alpha) + iC_i(\beta, \alpha), \qquad (A1)$$

where

$$C_r(\beta, \alpha) = \int_\Delta e^{\beta S_N(\mathbf{r}_1\cdots\mathbf{r}_N)} \times \cos\alpha S_N(\mathbf{r}_1\cdots\mathbf{r}_N)\,d\mathbf{r}_1\cdots d\mathbf{r}_N,$$

$$C_i(\beta, \alpha) = \int_\Delta e^{\beta S_N(\mathbf{r}_1\cdots\mathbf{r}_N)} \times \sin\alpha S_N(\mathbf{r}_1\cdots\mathbf{r}_N)\,d\mathbf{r}_1\cdots d\mathbf{r}_N$$

when we have split C into its real and imaginary parts. We want to show the α and β derivatives of C_r and C_i exist, and satisfy the Cauchy–Reimann conditions. To do this we need a standard theorem on differentiating under an integral sign. We state it in a form suitable for this problem.[9]

Let x be a point in the n-dimensional Euclidian space R^n, and let y belong to R^1. Let B be a closed region of R^n and I a closed interval in R^1. Let $f(x, y)$ be defined for $x \in B$ and $y \in I$ and such that $(\partial f/\partial y)(x, y)$ exists there, and that this de-

[7] H. A. Kramers and G. H. Wannier, Phys. Rev. 60, 252 (1941).

[8] M. E. Fisher, in *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, Colorado, 1965), Vol. VII C.

[9] R. Courant, *Differential and Integral Calculus* (Interscience Publishers, Inc., New York, 1949).

rivative is continuous function of the $n + 1$ variables (x, y) in $B \times I$. Then

$$\frac{\partial}{\partial y}\int_B f(x, y)\,dx = \int_B \frac{\partial f}{\partial y}(x, y)\,dx$$

for all $y \in I$.

Let us apply this theorem to $C_r(\beta, \alpha)$ and consider the β derivative. Then $y = \beta$, $x = (\mathbf{r}_1 \cdots \mathbf{r}_N)$ and

$$f(x, y) = e^{\nu S_N(x)} \cos \alpha S_N(x).$$

We take I to be any closed interval on the negative real axis which does not include the origin. Now if x is such that $S_N(x) \neq +\infty$ then

$$\frac{\partial f}{\partial y}(x, y) = S_N(x) e^{\nu S_N(x)} \cos \alpha S_N(x).$$

If $S_N(x) = +\infty$ then $(\partial f/\partial y)(x, y) = 0$ for all $y \in I$. Now using the continuity of $S_N(x)$ in x, it is easy to show that $(\partial f/\partial y)(x, y)$ is continuous in x for any fixed $y \in I$, including points for which $S_N(x) = +\infty$. For any fixed x, it is clear that $(\partial f/\partial y)(x, y)$ is continuous everywhere in I. However, continuity in x and y separately does not imply continuity in (x, y). For this we need the following theorem.

Theorem: Let $g(x, y)$ be continuous in y at (x_0, y_0), and continuous in x at x_0 uniformly in y, for all y in some neighborhood of y_0. Then $g(x, y)$ is continuous at (x_0, y_0).

Proof: Given $\epsilon > 0$ we find $\delta > 0$ such that $|y - y_0| < \delta$ implies $|f(x_0, y) - f(x_0, y_0)| < \frac{1}{2}\epsilon$ and also $|y - y_0| < \delta$ and $|x - x_0| < \delta$ implies $|f(x, y) - f(x_0, y)| < \frac{1}{2}\epsilon$. Now, let $|(x_0, y_0) - (x, y)| < \delta$. Then $|x - x_0| < \delta$ and $|y - y_0| < \delta$ hence

$$|f(x, y) - f(x_0, y_0)| < |f(x, y) - f(x_0, y)| + |f(x_0, y) - f(x_0, y_0)| < \epsilon.$$

We apply this theorem to $(\partial f/\partial y)(x, y)$. Let (x_0, y_0) be an interior point of $B \times I$. Suppose first $S_N(x_0) \neq +\infty$. By the continuity of S_N, we can find a neighborhood of x_0, in B, such that $S_N(x) \neq +\infty$. Using the fact that

$$(\partial f/\partial y)(x, y) = S_N(x) e^{\nu S_N(x)} \cos \alpha S_N(x)$$

for $y \in I$ and x in this neighborhood, it is easy to show that the continuity in x is uniform with respect to y. Now, if x_0 is such that $S_N(x_0) = +\infty$, then by the continuity of S_N, given $M > 0$, we can find $\delta > 0$ such that $|x - x_0| < \delta$ implies $S_N(x) > M$. For any $y \in I$:

$$\left|\frac{\partial f}{\partial y}(x, y) - \frac{\partial f}{\partial y}(x_0, y)\right| = |S_N(x) e^{\nu S_N(x)} \cos S_N(x) - 0| < M e^{\nu M}$$

if we choose M so large that $-My > 1$ for all $y \in I$. The right-hand side can be made arbitrarily small for all $y \in I$ by choosing M large enough, hence the continuity at x_0 is uniform in y.

Then

$$\frac{\partial C_r(\beta, \alpha)}{\partial \beta} = \int_\Delta S_N e^{\beta S_N} \cos \alpha S_N \, d\mathbf{r}_1 \cdots d\mathbf{r}_N.$$

Similar arguments show that

$$\frac{\partial C_i(\beta, \alpha)}{\partial \beta} = \int_\Delta S_N e^{\beta S_N} \sin \alpha S_N \, d\mathbf{r}_1 \cdots d\mathbf{r}_N,$$

$$\frac{\partial C_r(\beta, \alpha)}{\partial \alpha} = -\int_\Delta S_N e^{\beta S_N} \sin \alpha S_N \, d\mathbf{r}_1 \cdots d\mathbf{r}_N,$$

$$\frac{\partial C_i(\beta, \alpha)}{\partial \alpha} = +\int_\Delta S_N e^{\beta S_N} \cos \alpha S_N \, d\mathbf{r}_1 \cdots d\mathbf{r}_N.$$

It is clear that

$$\frac{\partial C_r}{\partial \beta} = \frac{\partial C_i}{\partial \alpha}, \qquad \frac{\partial C_r}{\partial \alpha} = -\frac{\partial C_i}{\partial \beta} \quad \text{so that}$$

C is an analytic function of z for $\mathrm{Re}\, z < 0$.

Characterizing Coherent States of the Radiation Field*

THOMAS F. JORDAN†
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania
(Received 10 January 1966)

Simple proofs are given for two properties of a Bose field discovered recently in the quantum-theoretic description of optical coherence. The first is the theorem of Glauber and Titulaer that first-order coherence means that only one mode is excited. The second is the theorem of Aharanov, Falkoff, Lerner, and Pendleton that eigenstates of the annihilation operators are characterized by their ability to factor when the system is divided into two channels. The restriction of the latter to the case of a single excited mode is removed.

TWO simple properties of a Bose field have been discovered recently in the quantum-theoretic description of optical coherence. The two theorems proved in this paper are intended to provide elementary demonstrations of these properties. They are obtained by considering states of the field with respect to different sets of modes. Let a_i and their adjoints $a_i^\dagger$ be annihilation and creation operators satisfying the commutation relations,

$$a_i a_k - a_k a_i = 0, \qquad a_i a_k^\dagger - a_k^\dagger a_i = \delta_{ik}$$

for a set of independent (or "orthogonal") modes j. For each unitary matrix u

$$b_r = \sum_i u_{ri} a_i$$

and

$$b_r^\dagger = \sum_i u_{ri}^* a_i^\dagger$$

are annihilation and creation operators for another set of modes r.

First-order coherence for a state of the radiation field[1] means that there are complex numbers z_i such that

$$\langle a_i^\dagger a_k \rangle = z_i^* z_k.$$

Then

$$\langle b_r^\dagger b_s \rangle = \sum_{ik} u_{ri}^* u_{sk} \langle a_i^\dagger a_k \rangle = (\sum_i u_{ri} z_i)^* (\sum_k u_{sk} z_k)$$

so the definition of first-order coherence is the same with respect to a set of modes defined by any unitary matrix u if all the sums

$$\sum_i u_{ri} z_i$$

are finite, as is the case when

$$\sum_i |z_i|^2$$

is finite, which means that the total number of photons represented by the operator

$$\sum_i a_i^\dagger a_i$$

has a finite expectation value. Glauber and Titulaer[2] have discovered and proved the following.

Theorem: A state for which $\langle \sum_i a_i^\dagger a_i \rangle$ is finite has the property of first-order coherence if and only if there is a set of modes (with annihilation and creation operators b_r and $b_r^\dagger$ defined by a unitary matrix u as above) of which only one mode is excited.

Proof: If only the single mode $r = 1$ is excited, then

$$\langle b_r^\dagger b_s \rangle = \delta_{r1}\, \delta_{s1} \langle b_1^\dagger b_1 \rangle,$$

which satisfies the definition of first-order coherence. To establish the converse, suppose that the state has first-order coherence and let

$$u_{1i} = z_i^* (\sum_k |z_k|^2)^{-\frac{1}{2}}$$

assuming that

$$\sum_k |z_k|^2 = \langle \sum_k a_k^\dagger a_k \rangle$$

is finite. Then

$$\sum_i |u_{1i}|^2 = 1.$$

* Supported by the U. S. Army Research Office, Durham, North Carolina.
† Alfred P. Sloan Research Fellow.
[1] R. J. Glauber, Phys. Rev. 130, 2529 (1963); R. J. Glauber and U. M. Titulaer, *ibid.* 140, B676 (1965); T. F. Jordan, Helv. Phys. Acta 37, 697 (1964); N. Mukunda and T. F. Jordan, J. Math. Phys. 7, 849 (1966).
[2] R. J. Glauber and U. M. Titulaer, Phys. Rev. 145, 1041 (1966). At the Physics of Quantum Electronics Conference in San Juan, Puerto Rico, June 1965, I learned, from Glauber and Titulaer, the statement but not the proof of this theorem. That is why I happen to have made a proof which seems to me to be sufficiently different from the original proof to be of some interest. I want to leave no doubt, however, that credit for the original statement and proof of this theorem is due to Glauber and Titulaer.

We can find a unitary matrix u such that $u_{ri} = u_{1i}$ for $r = 1$. Then

$$\sum_i u_{ri}z_i = (\sum_k |z_k|^2)^{\frac{1}{2}} \sum_i u_{ri}u_{1i}^* = (\sum_k |z_k|^2)^{\frac{1}{2}} \delta_{r1}$$

and thus

$$\langle b_r^\dagger b_s \rangle = \delta_{r1}\,\delta_{s1} \sum_k |z_k|^2.$$

In particular,

$$\langle b_r^\dagger b_r \rangle = 0$$

for $r \neq 1$, which means that only the single mode $r = 1$ is excited. This completes the proof of the theorem.

Let $|z\rangle$ be a normalized eigenvector of the annihilation operators a_i

$$a_i\,|z\rangle = z_i\,|z\rangle,$$
$$\langle z\,|\,z\rangle = 1,$$

with z denoting the set of complex eigenvalues z_i.[3] Then

$$b_r\,|z\rangle = \sum_i u_{ri}a_i\,|z\rangle = (\sum_i u_{ri}z_i)\,|z\rangle,$$

so $|z\rangle$ is an eigenvector also of the annihilation operators b_r for a set of modes defined by any unitary matrix u if all the sums

$$\sum_i u_{ri}z_i$$

are finite, as is the case when

$$\sum_i |z_i|^2 = \langle z|\sum_i a_i^\dagger a_i\,|z\rangle$$

is finite. Consider a division of such a set of modes into two subsets. Any product of the annihilation operators b_r and the creation operators $b_r^\dagger$ factors uniquely into two products of operators for the two subsets. For a pure state represented by the vector $|z\rangle$, the expectation value of such a product factors into the expectation values of the two products for the two subsets of modes. We may think of this division as an analysis of the field with respect to two channels. For any such division, the state represented by the vector $|z\rangle$ factors into two independent states for the two channel subsystems. The converse statement that this property is characteristic of the eigenvectors of the annihilation operators was discovered by Aharanov, Falkoff, Lerner, and Pendleton,[4] was developed by Glauber and Titulaer,[5] and is demonstrated by the following.

Theorem: A state for which $\langle\sum_i a_i^\dagger a_i\rangle$ is finite is a pure state represented by an eigenvector of the annihilation operators if it has the property that

$$\langle b_r^\dagger b_s\rangle = \langle b_r^\dagger\rangle\,\langle b_s\rangle$$

for $r \neq s$ for every set of modes r defined by a unitary matrix u for which u_{ri} is nonzero for at most two values of j for each r.

Proof: Taking u to be the identity matrix and considering any two modes, say modes 1 and 2, we have

$$\langle a_1^\dagger a_2\rangle = \langle a_1^\dagger\rangle\,\langle a_2\rangle,$$
$$\langle a_2^\dagger a_1\rangle = \langle a_2^\dagger\rangle\,\langle a_1\rangle.$$

Then, taking u to be the identity matrix except for $r, j = 1, 2$, we get

$$\begin{aligned}0 &= \langle b_1^\dagger b_2\rangle - \langle b_1^\dagger\rangle\,\langle b_2\rangle \\ &= u_{11}^*u_{21}(\langle a_1^\dagger a_1\rangle - \langle a_1^\dagger\rangle\,\langle a_1\rangle) \\ &\qquad + u_{12}^*u_{22}(\langle a_2^\dagger a_2\rangle - \langle a_2^\dagger\rangle\,\langle a_2\rangle).\end{aligned}$$

Since unitarity requires that

$$u_{11}^*u_{21} + u_{12}^*u_{22} = 0,$$

we conclude that

$$\langle a_1^\dagger a_1\rangle - \langle a_1^\dagger\rangle\,\langle a_1\rangle = \langle a_2^\dagger a_2\rangle - \langle a_2^\dagger\rangle\,\langle a_2\rangle.$$

In general, we can conclude that

$$\langle a_i^\dagger a_i\rangle - \langle a_i^\dagger\rangle\,\langle a_i\rangle$$

is the same for all j. Since this number is nonnegative, it must be zero for $\langle\sum_i a_i^\dagger a_i\rangle$ to be finite. Thus we have

$$\langle(a_i - \langle a_i\rangle)^\dagger(a_i - \langle a_i\rangle)\rangle = \langle a_i^\dagger a_i\rangle - \langle a_i^\dagger\rangle\,\langle a_i\rangle = 0$$

for all j. Let the density matrix representing the state be

$$\rho = \sum_n w_n\,|\phi_n\rangle\,\langle\phi_n|,$$

with ϕ_n being orthonormal vectors and w_n being positive numbers that sum to one. We have

$$\begin{aligned}\sum_n w_n([a_i - \langle a_i\rangle]\phi_n,\,[a_i - \langle a_i\rangle]\phi_n) \\ = \langle(a_i - \langle a_i\rangle)^\dagger(a_i - \langle a_i\rangle)\rangle = 0\end{aligned}$$

for all j. For each n, either $w_n = 0$ or

$$(a_i - \langle a_i\rangle)\phi_n = 0$$

[3] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).

[4] Y. Aharanov, D. Falkoff, E. Lerner, and H. Pendleton (to be published) prove this theorem for a pure state with only one excited mode using the complete factoring of the state.

[5] Ref. 2; Glauber and Titulaer prove this theorem for any state with only one excited mode.

for all j. The latter means that ϕ_n is an eigenvector of each a_j with eigenvalue $\langle a_j \rangle$. Since there are no two linearly independent eigenvectors for the same set of eigenvalues, the only possibility is that w_n is nonzero for only one n, and the state is a pure state represented by an eigenvector of the annihilation operators a_j. This completes the proof of the theorem.

This theorem requires only factoring of the expectation values of operators bilinear in the creation and annihilation operators for various divisions of particular sets of modes into two subsets. Thus, it follows as a corollary to this theorem that the complete factoring of the state for divisions into two subsets of the more general sets of modes discussed earlier is implied by the factoring of the expectation values of the bilinear operators for these particular sets of modes.

ACKNOWLEDGMENT

I am grateful to U. M. Titulaer for a very helpful suggestion regarding the proof of the second theorem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

The Discontinuities of the Triangle Graph as a Function of an Internal Mass. II

C. KACSER*

Department of Physics and Astronomy, University of Maryland, College Park, Maryland

(Received 7 April 1966)

We consider the discontinuities of the triangle-graph amplitude as a function of an internal mass variable. These discontinuities are important, since they form the kernel of the Aitchison–Anisovich integral equation, which is derived from the Khuri–Treiman three-body final-state-interaction dispersion relation. We evaluate the discontinuities by explicitly performing the Feynman α integrations. We also discuss their analytic continuations. Finally we consider the applicability of the Cutkosky rules to such an internal mass variable discontinuity. It is argued that these rules must be modified in two ways. One of these is straightforward, having to do with the appearance of spacelike masses. The other is more involved and is a consequence of the results of homology theory. We apply the *modified* Cutkosky rules to the triangle-graph discontinuities and obtain the same results as found by the direct method, so confirming the modifications which we have made.

1. PREAMBLE

ONE approach to the problem of three-body production and decay processes is to use a dispersion relation of the Khuri–Treiman type.[1] Aitchison[2,3] has shown how this can be transformed into a single-variable integral equation, using a method first applied by Anisovich[4] in the nonrelativistic problem. This equation lends itself to numerical solution on a computer, and work is being actively carried out on this.[5,6] The kernel of this integral equation is identical with the sum of the discontinuities of the triangle graph taken with respect to an *internal* mass. It is therefore of importance to study these discontinuities.

In Ref. 3 these discontinuities were evaluated by using Cutkosky's rules.[7] These discontinuities consist of three terms, Δ_1, Δ_2, and Δ_3, corresponding to three different branch points. The application of Cutkosky's rules for these discontinuities runs into difficulties associated with spacelike "masses", which were not treated fully in Ref. 3. Furthermore, Aitchison's evaluation of Δ_3 actually contains several errors. Aitchison did correctly note that Cutkosky's rules do not provide a complete specification of

* This work was supported in part by the National Science Foundation under Grant NSF GP-3221.

[1] N. N. Khuri and S. B. Treiman, Phys. Rev. 119, 1115 (1960).

[2] I. J. R. Aitchison, Nuovo Cimento 35, 434 (1965).

[3] I. J. R. Aitchison, Phys. Rev. 137, B1070 (1965); also Erratum (to be published). Note that the f and Δ_i of that paper differ in normalization from those of the present paper. Specifically, $f_{\text{Aitch}} = -2\pi^2 f_{\text{Kac}}$, $\Delta_{i,\text{Aitch}} = \pi^{-1}\Delta_{\text{Kac}}$. Note that $\Delta_{\text{Aitch}} = [4(\pi i)^3]^{-1}[f_{\text{Aitch}}(\lambda^2 - i\epsilon) - f_{\text{Aitch}}(\lambda^2 + i\epsilon)]$. We apologize for the confusion in conventions. (See also Ref. 9.)

[4] V. V. Anisovich, Zh. Eksperim. i Teor. Fiz. 44, 1593 (1963) [English transl.: Soviet Phys.—JETP 17, 1072 (1963)].

[5] I. J. R. Aitchison (to be published).

[6] I. Duck and F. C. Khanna, Nucl Phys. 77, 609 (1966).

[7] R. E. Cutkosky, J. Math. Phys. 1, 429 (1960). See also W. B. Rolnick, Phys. Rev. Letters 16, 544 (1966).

for all j. The latter means that ϕ_n is an eigenvector of each a_j with eigenvalue $\langle a_j \rangle$. Since there are no two linearly independent eigenvectors for the same set of eigenvalues, the only possibility is that w_n is nonzero for only one n, and the state is a pure state represented by an eigenvector of the annihilation operators a_j. This completes the proof of the theorem.

This theorem requires only factoring of the expectation values of operators bilinear in the creation and annihilation operators for various divisions of particular sets of modes into two subsets. Thus, it follows as a corollary to this theorem that the complete factoring of the state for divisions into two subsets of the more general sets of modes discussed earlier is implied by the factoring of the expectation values of the bilinear operators for these particular sets of modes.

ACKNOWLEDGMENT

I am grateful to U. M. Titulaer for a very helpful suggestion regarding the proof of the second theorem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

The Discontinuities of the Triangle Graph as a Function of an Internal Mass. II

C. KACSER*

Department of Physics and Astronomy, University of Maryland, College Park, Maryland

(Received 7 April 1966)

We consider the discontinuities of the triangle-graph amplitude as a function of an internal mass variable. These discontinuities are important, since they form the kernel of the Aitchison–Anisovich integral equation, which is derived from the Khuri–Treiman three-body final-state-interaction dispersion relation. We evaluate the discontinuities by explicitly performing the Feynman α integrations. We also discuss their analytic continuations. Finally we consider the applicability of the Cutkosky rules to such an internal mass variable discontinuity. It is argued that these rules must be modified in two ways. One of these is straightforward, having to do with the appearance of spacelike masses. The other is more involved and is a consequence of the results of homology theory. We apply the *modified* Cutkosky rules to the triangle-graph discontinuities and obtain the same results as found by the direct method, so confirming the modifications which we have made.

1. PREAMBLE

ONE approach to the problem of three-body production and decay processes is to use a dispersion relation of the Khuri–Treiman type.[1] Aitchison[2,3] has shown how this can be transformed into a single-variable integral equation, using a method first applied by Anisovich[4] in the nonrelativistic problem. This equation lends itself to numerical solution on a computer, and work is being actively carried out on this.[5,6] The kernel of this integral equation is identical with the sum of the discontinuities of the triangle graph taken with respect to an *internal* mass. It is therefore of importance to study these discontinuities.

In Ref. 3 these discontinuities were evaluated by using Cutkosky's rules.[7] These discontinuities consist of three terms, Δ_1, Δ_2, and Δ_3, corresponding to three different branch points. The application of Cutkosky's rules for these discontinuities runs into difficulties associated with spacelike "masses", which were not treated fully in Ref. 3. Furthermore, Aitchison's evaluation of Δ_3 actually contains several errors. Aitchison did correctly note that Cutkosky's rules do not provide a complete specification of

* This work was supported in part by the National Science Foundation under Grant NSF GP-3221.

[1] N. N. Khuri and S. B. Treiman, Phys. Rev. 119, 1115 (1960).

[2] I. J. R. Aitchison, Nuovo Cimento 35, 434 (1965).

[3] I. J. R. Aitchison, Phys. Rev. 137, B1070 (1965); also Erratum (to be published). Note that the f and Δ_i of that paper differ in normalization from those of the present paper. Specifically, $f_{\text{Aitch}} = -2\pi^2 f_{\text{Kac}}$, $\Delta_{i,\text{Aitch}} = \pi^{-1}\Delta_{\text{Kac}}$. Note that $\Delta_{\text{Aitch}} = [4(\pi i)^3]^{-1}[f_{\text{Aitch}}(\lambda^2 - i\epsilon) - f_{\text{Aitch}}(\lambda^2 + i\epsilon)]$. We apologize for the confusion in conventions. (See also Ref. 9.)

[4] V. V. Anisovich, Zh. Eksperim. i Teor. Fiz. 44, 1593 (1963) [English transl.: Soviet Phys.—JETP 17, 1072 (1963)].

[5] I. J. R. Aitchison (to be published).

[6] I. Duck and F. C. Khanna, Nucl Phys. 77, 609 (1966).

[7] R. E. Cutkosky, J. Math. Phys. 1, 429 (1960). See also W. B. Rolnick, Phys. Rev. Letters 16, 544 (1966).

the discontinuity; and that they had to be supplemented by using the prescriptions of homology theory.[8]

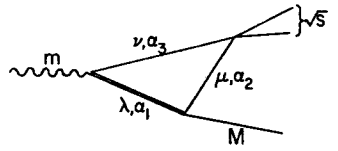
In a further paper by Aitchison and the present author,[9] the physical significance of the discontinuities was discussed, and Δ_1 was in fact evaluated by using well-known properties of the triangle graph. However, Δ_2 and Δ_3 were not discussed there, and in fact the error in Δ_3 was not noted.

The purpose of the present paper is to evaluate all three discontinuities, Δ_1, Δ_2, and Δ_3, by using a direct method that totally avoids the use of Cutkosky's rules. (Furthermore we treat the most general unequal mass case, while both Refs. 3 and 9 considered a somewhat degenerate case.) Once we know Δ_1, Δ_2, and Δ_3 as evaluated by first principles, we return to a reconsideration of their evaluation by means of Cutkosky's rules. We state a new generalization of Cutkosky's rules for the case of spacelike masses; and we reemphasize the necessity of following the *modified* prescription which arises from homology theory. Finally, we apply these modified Cutkosky rules to the evaluation of Δ_1, Δ_2, and Δ_3. We find confirmation of our modified prescription by obtaining the same results in each case as we obtained by the direct method.

The detailed structure of the paper is as follows. In Sec. 2 we define the problem more carefully in terms of the Feynman integral for the triangle graph. In Secs. 3–6 we evaluate Δ_1, Δ_2, and Δ_3 explicitly by performing the Feynman α integrations. This evaluation is initially performed for the case that the external two-particle mass variable s is not yet on its cut, so that the Δ_i are real. In Sec. 7 we then consider the analytic continuation in s, following the method of Ref. 3.

There are also three appendices. In Appendix A we reconsider the analytic continuation in s by using the known analytic properties of the triangle-graph amplitude. In Appendix B we turn to the question of the Cutkosky rules. We first discuss why they need to be modified when applied to internal mass discontinuities, and then explicitly evaluate Δ_1, Δ_2, and, in particular Δ_3, using these rules. The results agree with those of the direct method. Finally, in Appendix C we evaluate Δ_3 using the method of Aitchison,[3] corrected where necessary, and once again we obtain the same final expression for Δ_3.

FIG. 1. The basic triangle graph.



2. STATEMENT OF PROBLEM

We consider the Feynman graph of Fig. 1, and are in fact interested in the fully physical decay region specified by $m > \lambda + \nu$, $\lambda > \mu + M$, $(m - M)^2 > s > (\nu + \mu)^2$. Here, for a real decay $m \to \lambda + \nu \to M + \nu + \mu$, λ corresponds to an intermediate resonance mass, with a negative imaginary part.[10] However, in the integral equation one must consider all real λ^2 in the range $-\infty < \lambda^2 < (M - \mu)^2$.

We define the "triangle graph" f as[11]

$$f = \frac{1}{\pi^2 i}\int \frac{d^4k}{\Pi(m_i^2 - q_i^2 - i\epsilon)} = \int_0^1 d\alpha_1 \int_0^1 d\alpha_2 \int_0^1 d\alpha_3\, \delta(1 - \alpha_1 - \alpha_2 - \alpha_3)/D \quad (1)$$

with

$$D = \alpha_1\lambda^2 + \alpha_2\mu^2 + \alpha_3\nu^2 - \alpha_1\alpha_2 M^2 - \alpha_2\alpha_3 s - \alpha_3\alpha_1 m^2 - i\epsilon. \quad (2)$$

Then, using standard methods,[11–15] but working in s and λ^2 as variables, one readily generalizes the results of Bronzan and Kacser[16] to find the following sets of singularities:

(1) The leading Landau singularity surface,[17]

$$\Gamma \equiv s\lambda^2(s + \lambda^2) - s\lambda^2(m^2 + M^2 + \mu^2 + \nu^2) + \lambda^2(m^2 - M^2)(\nu^2 - \mu^2) + s(m^2 - \nu^2)(M^2 - \mu^2) + m^4\mu^2 + M^4\nu^2 + \mu^4 m^2 + \nu^4 M^2 - M^2 m^2(\nu^2 + \mu^2) - \mu^2\nu^2(m^2 + M^2) = 0, \quad (3)$$

[8] D. Fotiadi, M. Froissart, J. Lascoux, and F. Pham, Ecole Polytechnique preprint (unpublished), and related works; also Topology 4, 159 (1965).

[9] I. J. R. Aitchison and C. Kacser [Nuovo Cimento 40, 576 (1965)] use the same definition of f and the Δ_i as we do here. However (owing to confusion engendered by the remarks of Ref. 3 of the present paper), a factor of π has been dropped in their Eqs. (6), (7), (9), (10), (12), and in the final equation; that is, the right-hand sides of each of those equations should be multiplied by π. Further, the symbol L is used for our R.

[10] The use of this "isobar" approximation has been discussed fully by I. J. R. Aitchison and C. Kacser, Phys. Rev. 133, B1239 (1964).

[11] R. Karplus, C. M. Sommerfield, and E. H. Wichmann, Phys. Rev. 111, 1187 (1958).

[12] L. D. Landau, Nucl. Phys. 13, 181 (1959).

[13] R. J. Eden, "Lectures on the Use of Perturbation Methods in Dispersion Theory," University of Maryland, Physics Department, Technical Report No. 211 (1961).

[14] R. J. Eden and G. C. Polkinghorne, *Brandeis Summer School Lecture Notes, 1961* (W. A. Benjamin, Inc., New York, 1962).

[15] J. Tarski, J. Math. Phys. 1, 149 (1960).

[16] J. B. Bronzan and C. Kacser, Phys. Rev. 132, 2703 (1963); C. Kacser, *ibid.* 132, 2712 (1963); see also Refs. 3 and 9.

[17] This is given by the expression $1 - x^2 - y^2 - z^2 - 2xyz = 0$ with $x = (s - \nu^2 - \mu^2)/(2\mu\nu)$, $y = (m^2 - \nu^2 - \lambda^2)/(2\nu\lambda)$, and $z = (M^2 - \lambda^2 - \mu^2)/(2\lambda\mu)$. It is identical with the Kibble cubic [T. W. B. Kibble, Phys. Rev. 117, 1159 (1960)] for the process $\mu + \nu \to m + M$, with s the direct channel invariant and λ^2 the crossed $m + \nu \to M + \mu$ channel invariant.

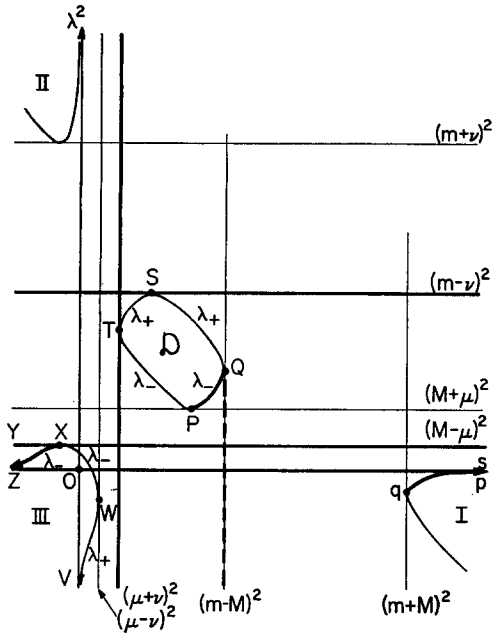


FIG. 2. The various singularity surfaces in the real (s,λ^2) plane, for $m > M + \nu + \mu$; throughout this paper we take the specific case $M > \mu$ and $\nu > \mu$ to determine the asymptotic behavior at $s = 0$ and $\lambda^2 = 0$. The singularities are the heavy lines, and the arcs PQ, ZX and pq, and the complex surfaces joining these arcs. The second-type singularity is shown dashed.

(2) $\alpha_1 = 0$ end point: $s = (\mu + \nu)^2$,
(3) $\alpha_2 = 0$ end point: $\lambda^2 = (m - \nu)^2$,
(4) $\alpha_3 = 0$ end point: $\lambda^2 = (M - \mu)^2$,
(5) double end point: $\alpha_2 = \alpha_3 = 0$: $\lambda^2 = 0$,
(6) second-type[18] or non-Landau singularity $s = (m - M)^2$.

Using the standard asymptote and tangency conditions,[15] one finds that Γ is as shown in Fig. 2, in which the various singularities are also indicated. (We take $M > \mu$ for definiteness.) For $s < (\mu + \nu)^2$, the only singularities in λ^2 are at $(m - \nu)^2$, $(M - \mu)^2$, and 0. Further, *for* $s < (\mu + \nu)^2$ *and* $\lambda^2 > m^2 + \nu^2$, *D can never vanish and f is real*, so that the λ^2 cuts must go towards minus infinity, and hence we can write

$$f(s,\lambda^2) = -\frac{1}{\pi}\int_{-\infty}^{(m-\nu)^2} d\lambda'^2 \frac{\Delta_1(s,\lambda'^2)}{\lambda'^2 - \lambda^2 + i\epsilon} - \frac{1}{\pi}\int_{-\infty}^{(M-\mu)^2} d\lambda'^2 \frac{\Delta_2(s,\lambda^2)}{\lambda'^2 - \lambda^2 + i\epsilon} - \frac{1}{\pi}\int_{-\infty}^{0} d\lambda'^2 \frac{\Delta_3(s,\lambda'^2)}{\lambda'^2 - \lambda^2 + i\epsilon}; \quad -\infty < s \leq (\mu + \nu)^2. \tag{4}$$

[18] See Ref. 16.

Here, the minus signs are inserted so that, even though the physical limit is taken in the sense $\lambda^2 - i\epsilon$, yet

$$\text{Im } f(s,\lambda^2) = \Delta_1\theta[(m-\nu)^2 - \lambda^2] + \Delta_2\theta[(M-\mu)^2 - \lambda^2] + \Delta_3\theta(-\lambda^2); \quad s < (\mu + \nu)^2. \tag{5}$$

Our purpose here is to evaluate Δ_1, Δ_2, and Δ_3, *initially for* $s < (\mu + \nu)^2$, but by analytic continuation for $s < (m - M)^2$. The method is based on Eqs. (5) and (1), and is modeled on the work of Barton and Kacser.[19]

3. GEOMETRICAL DESCRIPTION OF THE CURVE $D = 0$

From Eq. (1), Imf can only arise from those parts of the α-intergation region for which $D = 0$. Hence we investigate the curve $D = 0$ in the $\alpha_1\alpha_3$ space, using $\alpha_2 = 1 - \alpha_1 - \alpha_3$, and looking particularly in the integration region $0 \leq \alpha_1 \leq 1$, $0 \leq \alpha_3 \leq \alpha_1$. We ignore the term $i\epsilon$.

We have

$$D(\alpha_1,\alpha_3) = s\alpha_3^2 + (M^2 + s - m^2)\alpha_3\alpha_1 + M^2\alpha_1^2 + (\nu^2 - \mu^2 - s)\alpha_3 + (\lambda^2 - \mu^2 - M^2)\alpha_1 + \mu^2. \tag{6}$$

Then $D = 0$ is a hyperbola for $s < (m - M)^2$. Also $D(1, 0) = \lambda^2$, $D(0, 1) = \nu^2$, and $D(0, 0) = \mu^2$. Further,

$$D(\alpha_1, 0) = M^2\alpha_1^2 + (\lambda^2 - M^2 - \mu^2)\alpha_1 + \mu^2.$$

Hence $D(\alpha_1, 0) = 0$

$$\text{for} \quad \alpha_1 = \frac{M^2 + \mu^2 - \lambda^2 \pm k(M^2,\lambda^2,\mu^2)}{2M^2}. \tag{7}$$

Here,

$$k^2(a^2,b^2,c^2) \equiv a^4 + b^4 + c^4 - 2a^2b^2 - 2b^2c^2 - 2c^2a^2 = [a^2 - (b-c)^2][a^2 - (b+c)^2] = \text{cyclic perms} = (a^2 - b^2 - c^2)^2 - 4b^2c^2 = \text{cyclic perms}. \tag{8}$$

Unless stated otherwise, k is always to be taken as positive if $k^2 > 0$. (Once we continue in s, in Sec. 7, k will be specified with respect to the various branch cuts.) Equation (7) is plotted in Fig. 3, again for the case $M > \mu$. We see that, for $\lambda^2 < (M - \mu)^2$, D has two real intercepts with the α_1 axis, and that

[19] G. Barton and C. Kacser, Nuovo Cimento 21, 593 (1961), in particular, the Appendix. Beware of errors in factors of 2, and also the omission of the second-type singularity, cf. Ref. 13.

they both lie in the interval $0 \leq \alpha_1 \leq 1$ for $0 \leq \lambda^2 \leq (M-\mu)^2$. Similarly, one has

$$D(\alpha_1, 1-\alpha_1) = m^2\alpha_1^2 + (\lambda^2 - m^2 - \nu^2)\alpha_1 + \nu^2,$$

which can be obtained from (7) by the substitutions $M \to m$, $\mu \to \nu$ (thus preserving $m > \nu$), so that D has two real intercepts with the $\alpha_1 + \alpha_3 = 1$ boundary for $\lambda^2 < (m-\nu)^2$, both lying in the interval $0 \leq \alpha_2 \leq 1$ for $0 \leq \lambda^2 \leq (m-\nu)^2$. Finally similar analysis shows that $D(0, \alpha_3)$ never vanishes for real α_3 if $(\mu+\nu)^2 > s_2 > (\mu-\nu)^2$. For most of the remainder of this paper, we restrict ourselves to this range of s, $(\mu+\nu)^2 > s > (\mu-\nu)^2$. For this range, $D(\alpha_1, \alpha_3) = 0$ is a hyperbola, whose asymptotes both have positive slope.[20]

With this information, we can locate the curve $D(\alpha_1, \alpha_3) = 0$ with respect to the (α_1, α_3) integration region for all real λ^2, as shown in Fig. 4. The transition in Fig. 4 from (b) to (c) can be investigated by setting $\lambda^2 = (M-\mu)^2$, and expanding D about $\alpha_1 = \mu/M$, $\alpha_3 = 0$; however, the need for continuity between Fig. 4 (c) and (d) really uniquely determines (c). In Fig. 4 the "topology" is significant, but whether an individual arc has a tangent parallel to either the α_1 or the α_3 axis is not necessarily correctly described. Recall however that we are looking at one branch of a hyperbola, whose asymptotes both have positive slope.

The details of Fig. 4 correspond exactly to the singularities already found by means of the Landau analysis. Thus singularity (3) α_2 end point, $\lambda^2 = (m-\mu)^2$ is the transition from (a) to (b) in the figure; singularity (4) is the transition from (b) to (c); and singularity (5) is the transition from (c) to (d). From the viewpoint of Fig. 4, the reason that singularities are some combination of end point and pinch is simply that the only way the "topology" of D can change relative to the integration boundary is by an arc of D either becoming tangent to the boundary, or by slipping off at a vertex of the boundary. The non-Landau second-type singularity

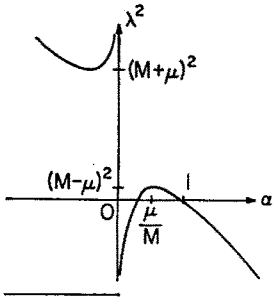


FIG. 3. The curve $D(\alpha_1, 0) = 0$ in the real (α_1, λ^2) plane, for the case $M > \mu$.

[20] They are given by $2s\,\alpha_3/\alpha_1 = m^2 - s - M^2 \pm k(s, m^2, M^2) > 0$, since $m^2 > (\sqrt{s} + M)^2$, $s > 0$.

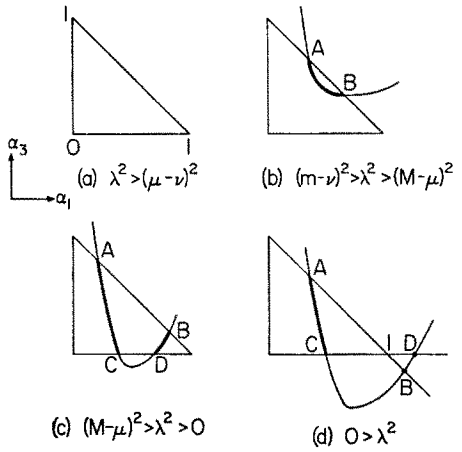


FIG. 4. The curve $D(\alpha_1, \alpha_3) = 0$ and the integration region $0 < \alpha < 1$, $0 < \alpha_3 < \alpha_1$, shown for various real λ^2 with $(\mu+\nu)^2 > s > (\mu-\nu)^2$. The heavy arcs show those regions that contribute to Im f.

$s = (m-M)^2$ corresponds to the discontinuous transition of D in which it changes from being a hyperbola to being an ellipse.

Finally, the "leading" Landau singularity of complete coincident pinches corresponds to the case when the (α_1, α_3) hyperbola degenerates to two crossing straight lines, the point of intersection giving coincident pairs of values for each of α_1, α_2, and α_3 simultaneously. Thus, rewriting Eq. (6) in the form

$$D(\alpha_1, \alpha_3) \equiv a\alpha_1^2 + 2h\alpha_1\alpha_3 + b\alpha_3^2 + 2g\alpha_1 + 2f\alpha_3 + c, \qquad (6')$$

this degeneracy condition is simply given by[21]

$$\Delta \equiv \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} = 0. \qquad (8a)$$

The equation $\Delta \equiv 0$ is identical with the usual determinant equation for Γ.[22]

For the moment, with $(\mu+\nu)^2 > s > (\mu-\nu)^2$, we disregard the leading singularities. We can then immediately identify the discontinuity functions Δ_1, Δ_2, and Δ_3. We rewrite Eq. (1) as

$$f = \int_0^1 d\alpha_1 \int_0^{\alpha_1} d\alpha_3 \frac{1}{D(\alpha_1, \alpha_3)}$$

so that Imf arises from those parts of the curve

[21] Recall the standard form $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$; see for instance G. Salmon, *A Treatise on Conic Sections* (Longsmans Green and Company, Inc., London, 1879), 6th ed.

[22] Because, with $D = \sum c_{ij}\beta_i\beta_j$, with the identification $\beta_1 = \alpha_1$, $\beta_2 = 1$, and $\beta_3 = \alpha_3$; then if one sets $\beta_2 = \alpha_1 + \alpha_2 + \alpha_3$ to obtain $D = \sum a_{ij}\alpha_i\alpha_j$, one readily finds $\det a_{ij} \equiv \det c_{ij} \equiv \Delta$.

$D(\alpha_1, \alpha_3) = 0$ which lie within the integration region. For the interval $(m-\nu)^2 > \lambda^2 > (M-\mu)^2$ this corresponds to the complete arc AB of Fig. 4(b). Similarly, for the interval $(M-\mu)^2 > \lambda^2 > 0$ this corresponds to the arcs AC + DB of Fig. 4(c). But this can be rewritten as AB − CD. Finally, for $\lambda^2 < 0$, we can write the contribution as AB − CD + BD. Thus, in fact, we can write

$(\mu+\nu)^2 > s > (\mu-\nu)^2$:

$$\Delta_1 = \operatorname{Im} \int_{AB} d\alpha_1 \int d\alpha_3 \frac{1}{D(\alpha_1, \alpha_3)}; \quad (m-\nu)^2 \geq \lambda^2, \tag{9}$$

$$\Delta_2 = -\operatorname{Im} \int_{CD} d\alpha_1 \int d\alpha_3 \frac{1}{D(\alpha_1, \alpha_3)}; \quad (M-\mu)^2 \geq \lambda^2, \tag{10}$$

$$\Delta_3 = \operatorname{Im} \int_{BD} d\alpha_1 \int d\alpha_3 \frac{1}{D(\alpha_1, \alpha_3)}; \; 0 \geq \lambda^2. \tag{11}$$

Equations (9)–(11) solve our problem for the specified range of s, leaving only the explicit evaluation to be performed. This is straightforward, but tedious, and we only give the main steps. Before doing so, we remark from Fig. 4 that Δ_1 clearly has $\lambda^2 = (m-\nu)^2$ as a branch point, and that Δ_2 has a branch point at $\lambda^2 = (M-\mu)^2$. However, Δ_3 is regular at $\lambda^2 = 0$.

4. EXPLICIT EVALUATION OF Δ_1

The arc AB of $D(\alpha_1, \alpha_3) = 0$ is part of the "upper" branch of a hyperbola both of whose asymptotes have positive slope. Hence, any line of the form $\alpha_1 + \alpha_3 = \text{const}$ intersects AB either twice or not at all. It is therefore convenient to rotate axes so that we first integrate along such lines. Thus we define

$$\alpha = (1/\sqrt{2})(\alpha_1 + \alpha_3), \qquad \beta = (1/\sqrt{2})(\alpha_1 - \alpha_3)$$

whence

$$\begin{aligned} D &\equiv a\beta^2 + b\beta + c \equiv a(\beta-\beta_1)(\beta-\beta_2) \\ &= (\tfrac{1}{2}m^2)\beta^2 + \beta[\alpha(s-M^2) + (1/\sqrt{2}) \\ &\quad \times (\nu^2 + M^2 - s - \lambda^2)] + [\tfrac{1}{2}\alpha^2(2s + 2M^2 - m^2) \\ &\quad + (\alpha/\sqrt{2})(\nu^2 - 2\mu^2 - M^2 - s + \lambda^2) + \mu^2 - i\epsilon], \end{aligned} \tag{12}$$

where we have restored the $i\epsilon$ of Eq. (2). Then, if we define $\beta_2 > \beta_1$, we see that $\operatorname{Im}\beta_2 > 0$, $\operatorname{Im}\beta_1 < 0$. Hence

$$\begin{aligned} \Delta_1 &= \operatorname{Im} \int_{\alpha_0}^{1/\sqrt{2}} d\alpha \int_{-\infty}^{\infty} d\beta \frac{1}{a(\beta-\beta_1)(\beta-\beta_2)} \\ &= \operatorname{Im} \int_{\alpha_0}^{1/\sqrt{2}} \frac{d\alpha}{a} \pi i \frac{[\operatorname{sgn}(\operatorname{Im}\beta_2) - \operatorname{sgn}(\operatorname{Im}\beta_1)]}{\beta_2 - \beta_1} \\ &= \operatorname{Im} 2\pi i \int_{\alpha_0}^{1/\sqrt{2}} \frac{d\alpha}{|(b^2-4ac)^{\frac{1}{2}}|}, \end{aligned} \tag{13}$$

where α_0 is the *larger* of the two α roots of $\beta_2 = \beta_1$. (Recall that we are on the upper branch of the hyperbola.) Thus we set

$$\begin{aligned} X &\equiv b^2 - 4ac \equiv A\alpha^2 + B\alpha \\ &\quad + C \equiv A(\alpha-\alpha_0)(\alpha-\alpha') \\ &= \alpha^2 k^2(s, m^2, M^2) + \alpha\sqrt{2}\{(s-M^2) \\ &\quad \times (\nu^2 + M^2 - s - \lambda^2) \\ &\quad + m^2(2\mu^2 - \nu^2 + M^2 + s - \lambda^2)\} \\ &\quad + \tfrac{1}{2}\{(\nu + M^2 - \lambda^2 - s)^2 - 4m^2\mu^2\} \end{aligned} \tag{14}$$

with $\alpha_0 > \alpha'$. Then

$$\begin{aligned} \Delta_1 &= 2\pi i \int_{\alpha_0}^{1/\sqrt{2}} (A\alpha^2 + B\alpha + C)^{-\frac{1}{2}} d\alpha \\ &= \frac{2\pi i}{|A^{\frac{1}{2}}|} \ln\{2\,|(AX)^{\frac{1}{2}}| + 2A\alpha + B\}\Big|_{\alpha_0}^{1/\sqrt{2}}. \end{aligned} \tag{15}$$

After some algebra, one finds that

$$X|_{1/\sqrt{2}} = \tfrac{1}{2}k^2(m^2, \nu^2, \lambda^2)$$

and

$$\{2\,|(AX)^{\frac{1}{2}}| + 2A\alpha + B\}|_{1/\sqrt{2}} = \sqrt{2}\,(|U^{\frac{1}{2}}| - R), \tag{16}$$

where

$$\begin{aligned} U &\equiv U(m^2, M^2; \nu^2, \mu^2) \\ &= k^2(s, m^2, M^2)k^2(m^2, \nu^2, \lambda^2), \end{aligned} \tag{17}$$

and

$$\begin{aligned} R &\equiv R(m^2, M^2; \nu^2, \mu^2) \\ &= -m^4 + m^2(\nu^2 - 2\mu^2 + \lambda^2 + M^2 + s) \\ &\quad + (s-M^2)(\lambda^2 - \nu^2) \\ &= 2m^2(\lambda^2 + M^2 - \mu^2) \\ &\quad - (M^2 + m^2 - s)(m^2 + \lambda^2 - \nu^2). \end{aligned} \tag{18}$$

Also, one finds that[23]

$$\begin{aligned} \{2(AX)^{\frac{1}{2}} + 2A\alpha + B\}|_{\alpha_0} &= |(B^2 - 4AC)^{\frac{1}{2}}| \\ &= \sqrt{2}\,2m\,|\Gamma^{\frac{1}{2}}|. \end{aligned} \tag{19}$$

[23] That $B^2 - 4AC = 8m^2\Gamma$ is not unexpected, since this is the α discriminant of X, which in turn is the β discriminant of $D(\alpha, \beta)$. Hence, we can apply Lemma 1B of Theorem 2 in Ref. 15 to show that this repeated discriminant is proportional to the determinant of the coefficients of $D(\alpha, \beta)$ taken in the sense of Eqs. (6′) and (8a). Such a determinant is invariant under a rotation of axes, so that it is identical to Δ, and hence related to Γ (cf. Ref. 22).

But

$$R^2 - U = 4m^2\Gamma, \tag{20}$$

and, in the region $\lambda^2 \ll 0$, $(\mu + \nu)^2 > s > (\mu - \nu)^2$, we find that $R < 0$, $R^2 - U > 0$. Thus

$$\Delta_1 = \operatorname{Im} \frac{2\pi i}{|k(s, m^2, M^2)|} \ln\left(\frac{|U^{\frac{1}{2}}| - R}{|2m\Gamma^{\frac{1}{2}}|}\right)$$
$$= \frac{\pi}{|k(s, m^2, M^2)|} \ln\left(\frac{R - |U^{\frac{1}{2}}|}{R + |U^{\frac{1}{2}}|}\right).$$

Here both of $R \pm U^{\frac{1}{2}} < 0$, so that the logarithm is real. Neither of $R \pm U^{\frac{1}{2}}$ can change sign as λ^2 varies in $-\infty < \lambda^2 \leq (m - \nu)^2$ since $\Gamma \neq 0$, so that finally

$$\Delta_1 = \frac{\pi}{|k(s, m^2, M^2)|} \ln\left(\frac{R - |U^{\frac{1}{2}}|}{R + |U^{\frac{1}{2}}|}\right);$$
$$-\infty < \lambda^2 \leq (m - \nu)^2,$$
$$(\mu + \nu)^2 \geq s \geq (\mu - \nu)^2. \tag{21}$$

We consider the analytic continuation in s of this result in a later section.

5. EVALUATION OF Δ_2

This proceeds very similarly to that of Δ_1, so that we give only the briefest outline. The arc CD is intersected either twice or not at all by a line $\alpha_3 =$ const, so that we work in (α_1, α_3) as variables. We rewrite Eq. (6) as

$$D(\alpha_1, \alpha_3) \equiv M^2(\alpha_1 - a_1)(\alpha_1 - a_2),$$

with

$$a_2 > a_1, \qquad \operatorname{Im} a_2 > 0, \qquad \operatorname{Im} a_1 < 0.$$

(Do not confuse α and a.) Then

$$\Delta_2 = -\operatorname{Im} \int_{\alpha_{3-}}^{0} d\alpha_3 \frac{2\pi i}{M^2(a_2 - a_1)}$$
$$= -\operatorname{Im} \int_{\alpha_{3-}}^{0} \frac{d\alpha_3\, 2\pi i}{(p\alpha_3^2 + q\alpha_3 + r)^{\frac{1}{2}}}$$
$$= -\operatorname{Im} \frac{2\pi i}{|p^{\frac{1}{2}}|} \ln\{2(pY)^{\frac{1}{2}} + p\alpha_3 + q\}\Big|_{\alpha_{3-}}^{0}. \tag{22}$$

Here

$$Y \equiv p\alpha_3^2 + q\alpha_3 + r \equiv p(\alpha_3 - \alpha_{30})(\alpha_3 - \alpha_3')$$
$$= k^2(s, m^2, M^2)\alpha_3^2 + 2R'\alpha_3 + k^2(\lambda^2, \mu^2, M^2) \tag{23}$$

with

$$\alpha_{30} > \alpha_3'$$

and

$$R' \equiv R'(m^2, M^2; \nu^2, \mu^2)$$
$$= -M^4 + M^2(\mu^2 - 2\nu^2 + \lambda^2 + m^2 + s)$$
$$+ (s - m^2)(\lambda^2 - \mu^2)$$
$$= 2M^2(\lambda^2 + m^2 - \nu^2)$$
$$- (M^2 + m^2 - s)(M^2 + \lambda^2 - \mu^2)$$
$$= R(M^2, m^2; \mu^2, \nu^2), \tag{24}$$

where Eq. (24) should be compared with Eq. (18). Then one readily finds that

$$[2(pY)^{\frac{1}{2}} + 2p\alpha_3 + q]|_0 = 2\{|U'^{\frac{1}{2}}| + R'\}, \tag{25}$$

where

$$U' \equiv U'(m^2, M^2; \nu^2, \mu^2) = k^2(s, m^2, M^2)k^2(M^2, \mu^2, \lambda^2)$$
$$= U(M^2, m^2; \mu^2, \nu^2) \tag{26}$$

[cf. Eq. (17)]. Also

$$[2(pY)^{\frac{1}{2}} + 2p\alpha_3 + q]|_{\alpha_{3-}}$$
$$= |(q^2 - 4pr)^{\frac{1}{2}}| = 2\,|(R'^2 - U')^{\frac{1}{2}}|, \tag{27}$$

where[24]

$$R'^2 - U' = 4M^2\Gamma. \tag{28}$$

Hence one readily finds

$$\Delta_2 = \frac{\pi}{|k(s, m^2, M^2)|} \ln\left(\frac{R' - |U'^{\frac{1}{2}}|}{R + |U'^{\frac{1}{2}}|}\right);$$
$$-\infty < \lambda^2 \leq (M - \mu)^2,$$
$$(\mu + \nu)^2 \geq s \geq (\mu - \nu)^2. \tag{29}$$

Equation (29) is identical with Eq. (21) under the substitutions $M^2 \rightleftharpoons m^2$ and $\mu^2 \rightleftharpoons \nu^2$. We note that $R' > |U'^{\frac{1}{2}}| > 0$.

6. EVALUATION OF Δ_3.

We finally come to Δ_3, which is the discontinuity function for which Cutkosky's rules are no longer a complete specification; and thus the present evaluation is the main content of this work. Regrettably the algebra is heavy, but can be conquered if one believes strongly enough in the "simplicity" of nature.

We proceed exactly as in Sec. 4, and work with the rotated variables α and β.

[The use of α and β ensures that any line $\alpha =$ const cuts the arc BD only once [cf. Fig. 4(d)].

[24] Remarks similar to those of Ref. 23 apply here. The "factorizations" (20) and (28) are interesting variants of the more usual factorizations of Γ such as $\Gamma \equiv s[\lambda^2 - \lambda_+(s)][\lambda^2 - \lambda(s)]$, and can in fact be written as $\Gamma \equiv m^2(\mu^2 - \mu_-^2)(\mu^2 - \mu_+^2)$ and $\Gamma \equiv M^2(\nu^2 - \nu_-^2)(\nu^2 - \nu_+^2)$, respectively.

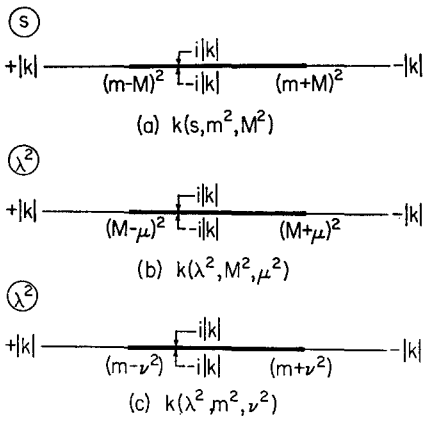


FIG. 5. The cuts determining (a) $k(s, m^2, M^2)$, (b) $k(\lambda^2, \mu^2, M^2)$, and (c) $k(\lambda^2, \nu^2, m^2)$.

If one were to use the variables (α_1, α_3) one would have to distinguish between the two cases

$$\left.\frac{d\alpha_3}{d\alpha_1}\right|_B \gtrless 0.$$

This transition does occur for some s for sufficiently negative λ^2, though it can be handled by analytic continuation in λ^2 of the result for the case

$$\left.\frac{d\alpha_3}{d\alpha_1}\right|_B < 0.$$

We used the (α_1, α_3) variables as an independent check of our result.]

Then [cf. Eq. (11), Fig. 4(d); and Eqs. (13), (15)]

$$\Delta_3 = \operatorname{Im}\int_{\alpha_B}^{\alpha_D} d\alpha \frac{\pi i}{(b^2-4ac)^{\frac{1}{2}}}$$
$$= \operatorname{Im}\frac{\pi i}{|A^{\frac{1}{2}}|}\ln\{|2(AX)^{\frac{1}{2}}| + 2A\alpha + B\}\Big|_{\alpha_B}^{\alpha_D}, \quad (30)$$

since now only β_2 contributes. Here

$$\alpha_B = \frac{1}{\sqrt{2}}, \quad (31)$$
$$\alpha_D = \frac{1}{\sqrt{2}}\frac{1}{2M^2}[M^2+\mu^2-\lambda^2+|k(M^2,\lambda^2,\mu^2)|].$$

Then [cf. Eq. (16)]

$$\{2|(AX)^{\frac{1}{2}}| + 2A\alpha + B\}_{1/\sqrt{2}} = \sqrt{2}(|U^{\frac{1}{2}}| - R) \geq 0 \quad (32)$$

and

$$(8M^4X)^{\frac{1}{2}}|_{\alpha D} = R' + (m^2+M^2-s)|k(M^2,\lambda^2,\mu^2)| \geq 0. \quad (33)$$

(This is an example of the "simplicity" of nature.) One then finds that

$$\sqrt{2}M^2[2(AX)^{\frac{1}{2}} + 2A\alpha + B]|_{\alpha D} = [m^2+M^2-s+|k(s,m^2,M^2)|] \times [R' + |U'^{\frac{1}{2}}|] \geq 0 \quad (34)$$

(another example). Furthermore,

$$(m^2+M^2-s)^2 - k^2(s,m^2,M^2) = 4m^2M^2. \quad (35)$$

Hence one can combine Eqs. (30), (32), (34), and (35), together with Eqs. (20) and (28), to finally obtain

$$\Delta_3 = \frac{\pi}{2|k(s,m^2,M^2)|}\ln\left[\frac{(R'+|U'^{\frac{1}{2}}|)(R+|U^{\frac{1}{2}}|)(m^2+M^2-s+|k(s,m^2,M^2)|)}{(R'-|U'^{\frac{1}{2}}|)(R-|U^{\frac{1}{2}}|)(m^2+M^2-s-|k(s,m^2,M^2)|)}\right];$$
$$\lambda^2 \leq 0, \qquad (\mu+\nu)^2 \geq s > (\mu-\nu)^2. \quad (36)$$

7. ANALYTIC CONTINUATION OF THE Δ_i

Equations (21), (29), and (36) determine the Δ_i for all relevant real λ^2, but only for the restricted s range $(\mu+\nu)^2 \geq s \geq (\mu-\nu)^2$. In this range the arguments of each logarithm are positive,[25] and the Δ_i are each real. We wish to continue the Δ_i to all s in $-\infty \leq s \leq (m-M)^2$. Here we follow the method of Ref. 3. In Appendix A we verify our conclusions by an independent method.

We first define the various k functions in the whole of the appropriate complex planes, as determined by the cuts of Fig. 5. We may then remove all the "absolute value" signs in Eqs. (21), (29), and (36) if we define

$$U^{\frac{1}{2}} = k(s,m^2,M^2)k(m^2,\nu^2,\lambda^2),$$
$$U'^{\frac{1}{2}} = k(s,m^2,M^2)k(M^2,\mu^2,\lambda^2). \quad (37)$$

We then observe that, *provided* the logarithms remain on their principal sheets, each of the Δ_i are regular at $s = (m \pm M)^2$. We further see that Δ_1 has a square root branch point at its threshold $\lambda^2 = (m-\nu)^2$, and similarly, Δ_2 has one at its threshold $\lambda^2 = (M-\mu)^2$. However, Δ_3 is completely regular at its threshold $\lambda^2 = 0$.

Let us now continue the Δ_i in s. Then as long as

[25] Though note that both $R \pm |U|^{1/2} < 0$.

the logarithms remain on their principal sheets, the only new singularities that can arise occur at $R^2 = U$ or at $R'^2 = U'$; i.e., they can only occur on the leading singularity surface $\Gamma = 0$. We must therefore determine, for each point on Γ, which of $R = +U^{\frac{1}{2}}$ and $R = -U^{\frac{1}{2}}$ applies, and similarly which of $R' = \pm U'^{\frac{1}{2}}$. For definiteness, we refer to the labeled arcs of Γ in Fig. 2 as

$$\text{TSQ}:\lambda_+^2(s),\quad \text{TPQ}:\lambda_-^2(s),\quad \text{ZXW}:\lambda_-^2(s),\quad \text{VW}:\lambda_+^2(s), \tag{38}$$

where the continuity of each of λ_+^2 and λ_-^2 as distinguishable roots of Γ is implied, and is then in fact assured by giving s a small positive imaginary part. Further, the small imaginary part to be associated with either $\lambda_+^2(s)$ or $\lambda_-^2(s)$ can then be read off Fig. 2 by using the sign of $\partial\lambda^2/\partial s$ appropriate to the arc under consideration.

We observe that $R = 0$ is a rectangular hyperbola in (s, λ^2) with asymptotes $\lambda^2 = \nu^2 - m^2 < 0$, and $s = M^2 - m^2 < 0$, lying in the $(+, +)$ and $(-, -)$ quadrants. Now, we are only concerned with those parts of Γ which bound the regions $\mathfrak{D}$ and III; (hereafter called $\Gamma_{\mathfrak{D}}$ and Γ_{III}) so that, on these parts of Γ, $U = 0$ only at S and at Q. Thus the hyperbola $R = 0$ intersects $\Gamma_{\mathfrak{D}}$ only at S and at Q, while it lies "under" the complete arc Γ_{III}. Hence, since $R < 0$ between its two branches, we immediately see that $R < 0$ on Γ_{III} and on STPQ, while $R > 0$ on SQ. Since $U^{\frac{1}{2}} \geq 0$ throughout these regions, we see that

$$\begin{aligned} R &= +U^{\frac{1}{2}}:\text{SQ}, \\ R &= -U^{\frac{1}{2}}:\text{STPQ, VWXZ}. \end{aligned} \tag{39}$$

Furthermore $R^2 < U$ inside $\mathfrak{D}$, while outside $\mathfrak{D}$ and Γ, $R^2 > U$.

We are now ready to perform the analytic continuation of Δ_1 in s using the technique of Appendix B of Ref. 3. We start in the region $(\mu - \nu)^2 \leq s \leq (\mu + \nu)^2$, all $\lambda^2 \leq (m - \nu)^2$, in which $R < 0$, with $R^2 > U$.

In this region Δ_1 must be real (since for such s, f is real for λ^2 not on its cut). Thus in this region we write

$$\begin{aligned} &\frac{1}{\pi} k(s, m^2, M^2)\Delta_1 \\ &= \ln\left(\frac{R - U^{\frac{1}{2}}}{R + U^{\frac{1}{2}}}\right) = \ln\left(\frac{U^{\frac{1}{2}} - R}{-U^{\frac{1}{2}} - R}\right) \\ &= \ln\left[\frac{(U^{\frac{1}{2}} - R)^2}{R^2 - U}\right] = \ln\left[\frac{(U^{\frac{1}{2}} - R)^2}{s(\lambda^2 - \lambda_-^2)(\lambda^2 - \lambda_+^2)}\right] \\ &= 2\ln(U^{\frac{1}{2}} - R) - \ln s \\ &\quad - \ln(\lambda^2 - \lambda_-^2) - \ln(\lambda^2 - \lambda_+^2). \end{aligned} \tag{40}$$

In this s region $(\mu - \nu)^2 < s < (\mu + \nu)^2$, $\lambda_-^2(s)$ and $\lambda_+^2(s)$ are complex conjugates, and, by continuity from the definitions for $s > (\mu + \nu)^2$, we see that $\text{Im}\,\lambda_+^2(s) > 0$, $\text{Im}\,\lambda_-^2(s) < 0$. Thus the appropriate choice of branches in Eqs. (40), which ensures that Δ_1 is real, is given by taking each logarithm on its principal branch, with the cuts from $\lambda_+^2(s)$ and $\lambda_-^2(s)$ each extending parallel to the real λ^2 axis to $-\infty$.

We now increase s into the region $s_{\text{T}} < s < s_{\text{S}}$, where s_{T} is the s value at the point T [in this case $s_T = (\mu + \nu)^2$] and similarly for s_{S}. In this continuation of Eq. (40), for fixed λ^2, the branch points $\lambda_\pm^2(s)$ come down onto the real λ^2 axis with

$$\begin{aligned} &\lambda_+^2 \in \text{ST}, \quad \text{Im}\,\lambda_+^2(s + i\epsilon) > 0, \\ &\lambda_-^2 \in \text{TP} \quad \text{Im}\,\lambda_-^2(s + i\epsilon) < 0. \end{aligned} \tag{41}$$

Furthermore, $R + U^{\frac{1}{2}}$ passes through zero on $\mathfrak{D}$; however, the form of Eq. (40) is such that the factor $\ln(U^{\frac{1}{2}} - R)$ remains real for $s < s_{\text{S}}$. Hence we readily find that, for $s_{\text{T}} < s < s_{\text{S}}$, Δ_1 is real outside $\mathfrak{D}$, and acquires an imaginary part of $+ i\pi/k(s, m^2, M^2)$ inside $\mathfrak{D}$.

We next continue to $s_{\text{S}} < s < s_{\text{P}}$. As s moves past S, $\lambda_+^2(s + i\epsilon)$ moves clockwise around the λ^2 contour end point $\lambda^2 = (m - \nu)^2$, and now acquires a small negative imaginary part. The λ^2 contour is undistorted by this, as is shown in Fig. 6. Hence $\ln(\lambda^2 - \lambda_+^2)$ now has an imaginary part $-i\pi$ for $\lambda^2 < \lambda_+^2$, and $-2i\pi$ for $\lambda_+^2 < \lambda^2 < (m - \nu)^2$. However, for this latter region, we are above the arc SQ, so that the factor $2\ln(U^{\frac{1}{2}} - R)$ must also be considered. The argument of this logarithm changes from being positive to being negative as we cross the arc SQ to the right. In order to determine the proper branch to associate with this logarithm, we re-express Eq. (40) in a form which allows us to continue in s across SQ with no difficulty. That is, we transform Eq. (40) to

$$\begin{aligned} &\frac{1}{\pi} k(s, m^2, M^2)\Delta_1 \\ &= \ln\left[\frac{s(\lambda^2 - \lambda_-^2)(\lambda^2 - \lambda_+^2)}{(-U^{\frac{1}{2}} - R)^2}\right] \\ &= \ln s + \ln(\lambda^2 - \lambda_-^2) + \ln(\lambda^2 - \lambda_+^2) \\ &\qquad - 2\ln(-U^{\frac{1}{2}} - R). \end{aligned} \tag{42}$$

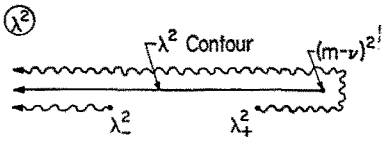


FIG. 6. The branch cuts associated with $\lambda_\pm^2(s)$ for $s_{\text{S}} < s < s_{\text{P}}$.

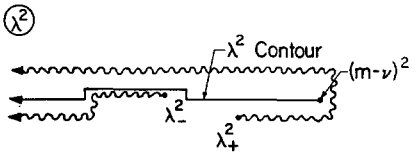


FIG. 7. The branch cuts and λ^2 contour for $s_P < s < s_Q$.

For $s < s_S$ inside $\mathfrak{D}$, we *know* that this expression has an imaginary part of $+i\pi$. But, with the same $\lambda_\pm^2(s)$ cuts as before, this implies that $-2 \ln (-U^{\frac{1}{2}} - R)$ must have an imaginary part of $+2\pi i$ (recall $-U^{-\frac{1}{2}} - R < 0$ here, but changes sign on STPQ). Hence, as s increases past s_S, and λ_+^2 circles around $(m - \nu)^2$, we find the imaginary part of $k\Delta_1/\pi$ remains $+i\pi$ in the region $\lambda_-^2 < \lambda^2 < \lambda_+^2$, and remains zero for $\lambda_+^2 < \lambda^2$. For $\lambda^2 < \lambda_-^2$ Eq. (42) cannot be used, since the factor $-2 \ln (-U^{\frac{1}{2}} - R)$ changes its imaginary part; however, Eq. (40) can be used in this region, and shows that Δ_1 is real for $\lambda^2 < \lambda_-^2$.

We next continue in s past s_P, and consider the arc PQ near which expression (40) is appropriate. At P, λ_-^2 circles around $(M + \mu)^2$ and moves into the upper half λ^2 plane. However, since Δ_1 is part of the spectral function of the dispersion relation (4) for f, the contour of integration must also be distorted upwards, (see Fig. 7) and hence remains *above* λ_-^2. Hence even though Im $\lambda_-^2(s + i\epsilon) > 0$, no change occurs in the effective prescription for Δ_1, i.e., Im $(1/\pi)k(s, m^2, M^2)\Delta_1 = \pi$ inside $\mathfrak{D}$, $= 0$ otherwise.[26]

In a precisely similar manner, and using Eq. (40), we find that inside III, Δ_1 has an imaginary part Im $\Delta_1 = -\pi/k(s, m^2, M^2)$.[27] In sum

$$\lambda^2 < (m - \nu)^2, \quad s < (m - M)^2:$$

$$\text{Im } \Delta_1 = \begin{cases} +\pi^2/k(s, m^2, M^2), & (s, \lambda^2) \in \mathfrak{D}, \\ -\pi^2/k(s, m^2, M^2), & (s, \lambda^2) \in \text{III}, \\ 0, & \text{otherwise.} \end{cases} \quad (43)$$

[26] However, one very important change has occurred. For all parts of $\Gamma_{\mathfrak{D}}$ other than PQ, the singularity of Δ_1 has been located at the appropriate $\lambda^2{}_\pm(s + i\epsilon)$ as determined by Γ. Thus, since the singularities of f are on the opposite side of the λ^2 contour to those of Δ_1, f has no singularities on QSTP in the curve limit, and hence the complex surfaces sprouting from QSTP are nonsingular. However, for the arc PQ, Δ_1 effectively has a singularity at $\lambda^2{}_- - i\epsilon'$, so that f has one at $\lambda^2{}_- + i\epsilon'$, which is the curve limit. Hence, the complex surface sprouting from PQ is singular on the physical sheet of f. This result agrees with that of Ref. 16.

[27] Again, for the arc ZX, $\lambda^2{}_-(s + i\epsilon)$ attempts to move into the upper-half complex λ^2 plane, but instead simply slightly displaces the λ^2 contour upwards. Hence the remarks of Ref. 26 apply. In fact, the singularity surface of f, which leaves PQ, first connects to ZX, and then continues through complex points to pq. (We remark again that we have assumed $M > \mu$ for definiteness. This determines that X is on Γ_{III} rather than on Γ_{I}.)

We can treat Δ_2 in a manner analogous to that used on Δ_1. Now one finds that $R' = 0$ at X, and $R' < 0$ on ZX, $R' > 0$ on XWV. Thus, on XWV, one writes

$$\text{XWV}: \frac{1}{\pi} k(s, m^2, M^2)\Delta_2$$

$$= \ln \left(\frac{R' - U'^{\frac{1}{2}}}{R' + U'^{\frac{1}{2}}} \right)$$

$$= \ln s(\lambda^2 - \lambda_+^2)(\lambda^2 - \lambda_-^2) - 2 \ln (R' + U'^{\frac{1}{2}}), \quad (44)$$

where $\ln (R' + U'^{\frac{1}{2}})$ is regular for $s_X < s < s_W$ with $R' + U'^{\frac{1}{2}} > 0$. Thus Im $(1/\pi)k(s, m^2, M^2)\Delta_2 = +i\pi$ in III. However, in the same region, we have [cf. Eq. (40), now $U^{\frac{1}{2}} - R > 0$]

$$\text{XWV}: \frac{1}{\pi} k(s, m^2, M^2)\Delta_1$$

$$= 2 \ln (U^{\frac{1}{2}} - R) - \ln s(\lambda^2 - \lambda_+^2)(\lambda^2 - \lambda_-^2), \quad (45)$$

so that one sees directly that the imaginary parts of Δ_1 and Δ_2 cancel; and further that neither XW nor WV is a singularity of f in any limit.

Finally, as s moves through s_X, λ_-^2 circles anticlockwise around the λ^2 contour end point at $\lambda^2 = (M - \mu)^2$, ending in the upper half-plane with no distortion on the λ^2 contour. (This is similar to the behavior of λ_+^2 near s_S.) At the same time, as s decreases across XS, $R' + U'^{\frac{1}{2}}$ becomes negative, so that the form of Eq. (43) must be replaced by the equivalent form,

$$\frac{1}{\pi} k(s, m^2, M^2)\Delta_2 = 2 \ln (R' - U'^{\frac{1}{2}})$$

$$- \ln s - \ln (\lambda^2 - \lambda_-^2) - \ln (\lambda^2 - \lambda_+^2) \quad (46)$$

with $2 \ln (R' - U'^{\frac{1}{2}})$ having an imaginary part of $+2\pi i$ inside Γ_{III} for $s_X < s < s_W$ (notice that $R' - U'^{\frac{1}{2}}$ changes from negative to positive as we cross WX out of region III). Then form (46) enables us to continue in s below s_X, and shows that the $-2\pi i$ which arises from $\ln (\lambda^2 - \lambda_-^2)$ for $\lambda^2 > \lambda_-^2$, is canceled by the $+2\pi i$ from $2 \ln (R' - U'^{\frac{1}{2}})$. Thus, to summarize,

$$\lambda^2 < (M - \mu)^2, \quad s < (m - M)^2:$$

$$\text{Im } \Delta_2 = \begin{cases} \pi^2/k(s, m^2, M^2), & (s, \lambda^2) \in \text{III}, \\ 0, & \text{otherwise.} \end{cases} \quad (47)$$

The slight extra complication in the region XYZ arises because X belongs to Γ_{III}. If we had assumed $M < \mu$, we would have found that X belonged to

Γ_I.[28] In the equal mass case $M = \mu$, the point X at which Γ touches $\lambda^2 = (M - \mu)^2$ is at infinity, and no such complication arises.

Finally we come to Δ_3. Since, in fact

$$\Delta_3 = -\tfrac{1}{2}(\Delta_1 + \Delta_2) + \frac{\pi}{2k(s, m^2, M^2)} \times \ln\left[\frac{m^2 + M^2 - s + k(s, m^2, M^2)}{m^2 + M^2 - s - k(s, m^2, M^2)}\right] \quad (48)$$

and the last part of Eq. (46) is regular throughout the region under consideration, the complete specification of Δ_3 is straightforward using (43) and (47).

In Appendix A, we verify these results by an independent method.

8. CONCLUSIONS

Section 7 ends our direct evaluation of the triangle-graph internal mass discontinuities Δ_i. The problem has been stated and solved. However, the Δ_i are of great importance in the Aitchison–Anisovich[2-4] integral equation, which is derived from the Khuri–Treiman[1] equation, and which considers the problem of 3-body final-state interaction effects. Further, it has recently been argued by Pasquier and Aitchison[29] that this equation provides a way of unitarizing any three-body amplitude, since they prove that the Khuri–Treiman equation does satisfy *three*-body unitarity.

It is therefore worthwhile to consider the Δ_i further. Hence, in Appendix A, we show how the analytic continuation of the Δ_i can be obtained by using the known analytic properties of the triangle-graph amplitude. Further, in Appendix B, we generalize Cutkosky's rules for evaluating discontinuities of Feynman amplitudes, so that they can be applied to internal mass variables. One of the two necessary generalizations makes use of homology theory.[8] Yet this generalization is needed. (The moral is clear.) With these generalizations we briefly evaluate the Δ_i using the Cutkosky rules, and find agreement with the direct evaluation. Finally, in Appendix C we reevaluate Δ_3 using the method of Aitchison,[3] corrected where necessary, and again confirm previous results.

[28] By inspection of Eq. (3), we see that near $s \to \infty$, Γ behaves as $\lambda^2 = -(M^2 - \mu^2)(m^2 - \nu^2)/s$. Since $m^2 > \nu^2$, $M^2 \lesseqgtr \mu^2$ determines whether X belongs to Γ_{III} or Γ_I.

[29] R. Pasquier, "Some Analytic Properties in the Total Energy Variable of Khuri–Treiman Type Amplitudes." Preprint, July, 1965 (Laboratoire de Physique Nucleaire, B. P. No. 1, Orsay, S et 0, France); I. J. R. Aitchison and R. Pasquier, "Three Body Unitarity and Khuri–Treiman Amplitudes." Preprint, March, 1966 (High-Energy Physics Group, Cavendish Laboratory, Cambridge, England). See also the works of G. Bonnevay cited in the latter preprint.

Notes added in proof.

1. R. Pasquier (private communication; to be published) has been able to derive the representation

$$\Delta_2 + \Delta_3 = \pi \int_{-\infty}^{s-(m^2,\lambda^2)} \frac{ds'}{(s' - s)k(s', m^2, M^2)},$$

by interchanging the order of the dispersive and the angular projection integrations in the Khuri–Treiman equation. This interchange then leads to an equation which can be immediately compared with Eqs. (19) and (23) given by Aitchison.[3] One can then unambiguously identify Δ_1, and the combination $(\Delta_2 + \Delta_3)$. (Pasquier works in the equal-mass case. Hence, since each of Δ_2 and Δ_3 then appears in a λ^2 integration starting at $\lambda^2 = 0$, only the combination can be identified.) The identification of Δ_1 is identical with that given in Ref. 9 [i.e., Eq. (A1) of the present paper]. Pasquier is able to determine the appropriate s' contour for *all* relevant λ^2, for each of Δ_1 and $(\Delta_2 + \Delta_3)$, since the interchange of orders of integration requires that the contours be chosen in a "compatible" manner (cf. Ref. 29, where this concept is introduced).

By straightforward though tedious algebra, Pasquier is able to evaluate his representation for $(\Delta_2 + \Delta_3)$. He obtains the same result as given in the present paper for this combination. Thus this work confirms the validity of the modified Cutkosky rules, and also the evaluation of Δ_3. Further, since the contours are prescribed (but see the second paper cited in Ref. 29), Pasquier is able to determine the appropriate branch of each logarithm in Δ_1 and $(\Delta_2 + \Delta_3)$. His results agree with those given in the body of the present paper, and in Appendix A.

Presumably Pasquier's work can be generalized to the unequal-mass case.

2. R. Pasquier (private communication) has provided clarification on the analytic continuation of $\Delta_1(s, \lambda^2)$ in s in the region $\lambda_T^2 < \lambda^2 < \lambda_S^2$, as discussed in Appendix A. He first remarks that for $s_S < s < s_P$, we can follow all the physical λ^2 values by using $\lambda^2 + i\epsilon$ [cf. Eq. (A2)]. Specifically for such an s, as λ^2 increases from $-\infty$ to $(m - \nu)^2$, it passes above each of $\lambda_-^2(s)$ and $\lambda_+^2(s)$. As λ^2 increases in this clockwise sense past $\lambda_-^2(s)$, $s_-(\lambda^2)$ decreases and passes *under* s, leading to the configuration shown in Fig. 8(c). Similarly as λ^2 increases further past $\lambda_+^2(s)$, $s_+(\lambda^2)$ also decreases and passes under s, so that the λ^2 contour joining $s_-(\lambda^2)$ to $s_+(\lambda^2)$ becomes totally disengaged from s, in which case it can then be straightened out into the rectilinear contour joining $s_-(\lambda^2)$ to $s_+(\lambda^2)$. Thus for $s_S < s < s_P$,

$\Delta_1(s, \lambda^2)$ is imaginary only for $\lambda_-^2(s) < \lambda^2 < \lambda_+^2(s)$. This agrees with the remarks of Appendix A.

Pasquier now points out that as s decreases past s_B, having a positive imaginary part, $\lambda_-^2(s)$ completely loops λ_B^2 in a clockwise sense. Then in the representation (A2), the λ'^2 contour remains undistorted, but as λ'^2 increases, it effectively now passes *below* $\lambda_-^2(s)$. Thus, for $s < s_B$, as λ^2 increases from $-\infty$ to $(m - \nu)^2$, it first passes over $\lambda_-^2(s)$, and then passes *under* $\lambda_+^2(s)$. As λ^2 passes $\lambda_-^2(s)$, $s_-(\lambda^2)$ decreases and passes under s, leading to a configuration as in Fig. 8(c). Then as λ^2 passes *below* $\lambda_+^2(s)$, $s_-(\lambda^2)$ increases again past s, but again it passes *under* s, so that for $\lambda^2 > \lambda_+^2(s)$, the s' contour joining $s_-(\lambda^2)$ to $s_+(\lambda^2)$ is again *totally disengaged* from the pole at s. Thus again it can be straightened out to the rectilinear contour joining $s_-(\lambda^2)$ to $s_+(\lambda^2)$. This then corresponds to the physical limit shown in Fig. 8(d); so that, in particular, $\Delta_1(s, \lambda^2)$ is real for $s < s_-(\lambda^2)$, even for $\lambda^2 > \lambda_T^2$.

The author is happy to acknowledge very helpful correspondence with Dr. Roger Pasquier.

ACKNOWLEDGMENTS

The author is extremely grateful to I. J. R. Aitchison for much helpful correspondence, and, by proxy, also to R. Pasquier, He is grateful to Dr. G. Moorhouse, and Dr. T. G. Pickavance, Director of the Rutherford High Energy Laboratory, Science Research Council, England, for extending him the hospitality of the Laboratory during September 1965, at which time this work was conceived; and finally, he is grateful to the National Science Foundation for financial assistance under travel grant GP 4953.

APPENDIX A. ALTERNATIVE TREATMENT OF THE ANALYTIC CONTINUATION OF THE Δ_i

Our discussion in Sec. 7 was based on the explicit form of the functions Δ_1, Δ_2, and Δ_3. However, one of these functions Δ_1 also has an integral representation which was obtained in Ref. 9, and it is interesting to study how that representation leads to the same results. In Ref. 9, it was shown that

$$\Delta_1(s, \lambda^2) = \pi \int_{s_-(\lambda^2)}^{s_+(\lambda^2)} \frac{ds'}{(s' - s)k(s', m^2, M^2)}. \qquad \text{(A1)}$$

[Reference 9 only considers the equal mass case $\mu = \nu = M$; also we have inserted the missing factor of π here.] This form is to hold with k taken real and positive (i.e., in agreement with the cut specification of Fig. 5), and, in particular, with an undistorted s' contour of integration, for $\lambda_0^2 < \lambda^2 < (m - \nu)^2$. (This avoids the complications arising from the second-type singularity.) However, this form was only derived for $s < (\mu + \nu)^2$, so that at first sight it is not clear how to extend this both in s and in λ^2.

The clue is provided by the representation of f given by Eq. (4). We already *know* the singularity structure of $f(s, \lambda^2)$ for all s and λ^2, as given by Refs. 16 and 10. It therefore follows from Eq. (4) that the singularity structure of $\Delta_1(s, \lambda^2)$ in λ^2 for fixed s is fully prescribed. It is possible to "invert" this information to yield the singularity structure of $\Delta_1(s, \lambda^2)$ in s for fixed λ^2. With this information, the integral representation (A1) can be analytically continued in s. Since the direct evaluation of the representation (A1) was already performed in Ref. 9, and obtained the same result as given in Eq. (21), the only specific ambiguities concern the analytic continuations, i.e., the "proper" branches to be associated with the logarithm. By considering the prescribed contours, we show that the results of the present method agree with those of Sec. 7.

We first note that the representation of Eq. (4) can be written

$$f(s, \lambda^2) = -\frac{1}{\pi} \int_{-\infty}^{(m-\nu)^2} \frac{d\lambda'^2\, \Delta_1(s, \lambda'^2 + i\epsilon)}{(\lambda'^2 + i\epsilon) - \lambda^2} + (\Delta_2, \Delta_3 \text{ terms}). \qquad \text{(A2)}$$

That is, if the physical limit of $f(s, \lambda^2 - i\epsilon)$ has a singularity, then this is a corresponding singularity of Δ_1 in the λ^2 limit $\Delta_1(s, \lambda^2 + i\epsilon)$.

Now, of course, the representation (A2) originally assumed that s was real, with $-\infty < s \leq (\mu + \nu)^2$. However, our aim is to analytically continue the representation in s. In this continuation, as long as no λ^2 singularity of $\Delta_1(s, \lambda^2)$ crosses the λ'^2 integration contour, no new singularity of $f(s, \lambda^2)$ occurs (at least on its first physical sheet). That is, if $f(s, \lambda^2)$ has any λ^2 physical sheet singularities not on $-\infty < \lambda^2 < (m - \nu)^2$, it follows that a λ^2 singularity of $\Delta(s, \lambda^2)$ must have crossed the λ^2 cut; however, any singularities of $\Delta(s, \lambda^2)$ which are not first-sheet singularities of $f(s, \lambda^2)$ *cannot* have crossed the λ^2 cut.

We now recall[16,10] that the *only* physical sheet singularities of $f(s, \lambda^2)$ are those associated with the complex surface PQpq with corresponding imaginary parts for s and λ^2. Thus, in the physical s limit $s + i\epsilon$, the only λ^2 *physical*-sheet singularities are associated with the arc PQ. On this, $\lambda_-^2(s)$ is singular on the physical sheet at $\lambda_-^2(s + i\epsilon) = \lambda_-^2(s) + i\epsilon'$. Further, if we consider $s \pm i\Delta$, Δ finite, then, *if* it is on this

surface PQpq $\lambda^2(s \pm i\Delta)$ has a finite positive (negative) imaginary part, and is still a physical-sheet singularity of $f(s, \lambda^2)$. But there are *no* other singularities of $f(s_-\lambda^2)$.

However, when we consider $\Delta_1(s, \lambda^2)$, then the representation (A1) shows us that Δ_1 necessarily has singularities at $s = s_\pm(\lambda^2)$, or $\lambda^2 = \lambda_\pm^2(s)$. That is, $\Delta_1(s, \lambda^2)$ necessarily has singularities on the *whole* of the (complex) surface Γ.

We also have another very useful piece of information, namely that, for $-\infty < s < (\mu - \nu)^2$, Δ_1 must be real, since for such s, $\Delta_1 = \text{Im } f(s, \lambda^2)$ for $(M - \mu)^2 < \lambda^2 < (m - \nu)^2$; and $f(s, \lambda^2)$ is real for both s and λ^2 off their cuts.

These two pieces of information enable us to determine Δ_1 nearly completely, as shown in Fig. 8. In that figure, we consider 4 different ranges of λ^2, in the last three λ^2 having an infinitesimal positive imaginary part. Let us first disregard Fig. 8(d). Then the prescriptions follow from the representation (A1), with $\lambda'^2 + i\epsilon$ being the proper λ'^2 limit. In particular, in Fig. 8 (c) we know that, from the original domain of definition of the representation (A1), the integral is to be taken over an undistorted contour on the first sheet of $k(s', m^2, M^2)$, at least for $s < (\mu + \nu)^2$. The $\lambda^2 + i\epsilon$ limit then puts the s' contour slightly below the real axis, so that s can then continue along the real axis for all $s < (m - M)^2 = s_1$. Thus for this λ^2 range, $\text{Im } \Delta_1 = \pi^2/k$ for $s_-(\lambda^2) < s < s_+(\lambda^2)$, 0 otherwise.

The transition in Fig. 8 from (c) to (b) follows by continuity, since $s_+(\lambda^2 + i\epsilon)$ loops clockwise round $s = s_1$ as $\lambda^2 + i\epsilon$ decreases past λ_Q^2. One then again finds that $\text{Im } \Delta_1 = \pi^2/k$ for $s_-(\lambda^2) < s < s_+(\lambda^2)$, 0 otherwise, for one can rewrite the representation in this region as

$$\Delta_1(s, \lambda^2) = \pi\left[\int_{s_-(\lambda^2)}^{s_1} + \int_{s_+(\lambda^2)}^{s_1}\right] \frac{1}{(s' - s)} \times \frac{1}{k(s', m^2, M^2)|_{\text{first sheet}}}, \qquad \text{(A3)}$$

where the two contours $s_- \to s_1$ and $s_+ \to s_1$ are below and above the real axis, respectively.

In Fig. 8, on going from (b) to (a), i.e., $(M - \mu)^2 < \lambda^2 < (M + \mu)^2$, the representation (A3) still holds, but now $s_+ = (s_-)^*$, so that Δ_1 now becomes purely real for all real $s < (\mu + \nu)^2$. This agrees with the reality condition previously stated, and extends it somewhat. It also *confirms* the distorted s' contours of Fig. 8 (a) and (b).

We notice that our prescriptions in Fig. 8 agree with Eq. (A1) over *undistorted* contours, for $\lambda_Q^2 < \lambda^2 < \lambda_T^2$, for which the second-type singularity is irrelevant [see paragraph after Eq. (A1)]. The distorted contours in Fig. 8 (a) and (b) are for $\lambda^2 < \lambda_Q^2$ and show the second-type singularity. The undistorted contour prescription also holds for $\lambda_T^2 < \lambda^2 < \lambda_S^2$, at least for $s < (\mu + \nu)^2$. This now corresponds to the range in Fig. 8 (d). For this range, we now assert as in Fig. 8 (d) that the physical s limit is from *above* the effectively *un*distorted s' contour. (In the figure, the distorted s' contour follows the trajectory of $s_-(\lambda^2)$, but we still take the physical limit from above, so that this is equivalent to the undistorted contour.)

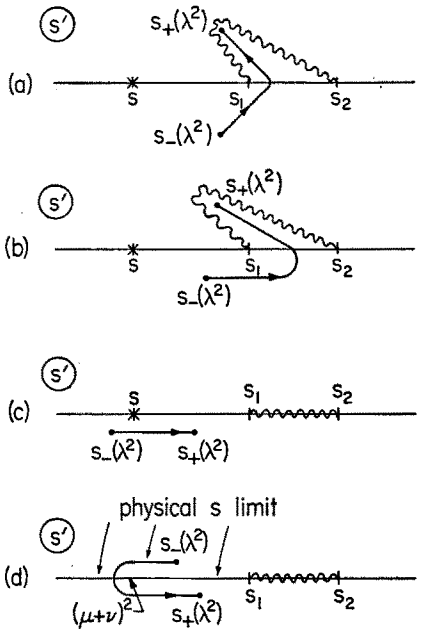


FIG. 8. The s' contour for the integral representation Eq. (A1) of $\Delta_1(s, \lambda^2)$. (a) $(M - \mu)^2 < \lambda^2 < (M + \mu)^2$, λ^2 real; (b) $(M + \mu)^2 < \lambda^2 < \lambda_Q^2$, $\lambda^2 + i\epsilon$ limit; (c) $\lambda_Q^2 < \lambda^2 < \lambda_T^2$, $\lambda^2 + i\epsilon$ limit; (d) $\lambda_T^2 < \lambda^2 < \lambda_S^2 = (m - \nu)^2$, $\lambda^2 + i\epsilon$ limit. We have defined $s_1 = s_Q = (m - M)^2$ and $s_2 = s_q = (m + M)^2$. In (a) $s_\pm$ have finite imaginary parts, while in (b)–(d) their imaginary parts are infinitesimal.

At this point, we really have to return to the analysis of Sec. 7. If we were to take the distorted s' contour literally, we would allow the possibility of the s pole being looped by it for $s_T < s < s_S$, and so leading to $\text{Im } \Delta_1 = 2\pi^2/k$ for $s_T < s < s_-(\lambda^2)$. But we would then contradict the conclusion of Sec. 7 discussed in the paragraphs around Eqs. (40) and (41), in which the s continuity was used to argue that no $2\pi^2/k$ contribution arises. The apparent difficulty here is a manifestation of the need for "compatible" contours, discussed in Ref. 29. (See also Note 2 added in proof.) With our prescrip-

tion, Im $\Delta_1 = \pi^2/k$ for $s_-(\lambda^2) < s < s_+(\lambda^2)$, 0 otherwise. Thus (with this slight retreat to Sec. 7) the conclusions of this appendix agree fully with those of Sec. 7.

Parenthetically, we can readily see how the arc PQpq becomes a first-sheet singularity of $f(s, \lambda^2)$. $\Delta_1(s, \lambda^2)$ of course always has singularities at $s = s_\pm(\lambda^2)$; i.e., at $\lambda^2 = \lambda_\pm^2(s)$. As s increases past s_P with small positive imaginary part (the physical limit) $\lambda_-^2(s)$ moves clockwise around $\lambda^2 = \lambda_P^2 = (M + \mu)^2$, circling from below the λ'^2 contour and pushing it upwards infinitesimally. As s now moves upwards to acquire a finite positive imaginary part, $\lambda_-^2(s)$ also moves upwards and carries the λ'^2 contour with it. This then produces a *first*-sheet λ^2 singularity at $\lambda_-^2(s)$. Thus the $(+, +)$ arc of PQpq is a *first*-sheet singularity of f, similarly, so is the $(-, -)$ arc of PQpq.

APPENDIX B. EVALUATION OF THE Δ_i USING A (NECESSARILY) MODIFIED FORM OF THE CUTKOSKY RULES

Preamble

Cutkosky[7] has given a complete method for evaluating the discontinuity of any Feynman graph. This method was originally proposed for discontinuities across cuts in an external mass variable. However, there is no obvious reason why it cannot be applied as is to an *internal* mass discontinuity, such as the Δ_i. In fact, Cutkosky's method was so used for the first time in Ref. 3. In that paper, Aitchison did evaluate Δ_1, Δ_2, and Δ_3. His Δ_1 and Δ_2 agree with the ones obtained in the present paper; however, the evaluation of Δ_3 given in Appendix A of that paper contains several errors. None the less, Aitchison did correctly remark that the Cutkosky prescription needed to be modified in a way which is implied by the work of Fotiadi, Froissart, Lascoux, and Pham,[8] in their studies of Feynman diagrams using homological methods. This homological prescription differs from that of Cutkosky as regards the proper domain of integration which is to be used. Otherwise the rules are identical.

On close inspection, it turns out that, even as regards Δ_1 and Δ_2, the Cutkosky domain of integration must be somewhat modified. In the present appendix, we therefore first briefly sketch the modified Cutkosky method of evaluation of Δ_1 and Δ_2. We then treat Δ_3 in more detail. Our results for each of Δ_1, Δ_2, and Δ_3 agree with those obtained in the main text of this paper.

The modification of the Cutkosky domains has to do with the Cutkosky prescription $\delta_+(q^2 - m^2) \equiv \delta(q^2 - m^2)\theta(q_0)$, or, in our case $\delta(q^2 - \lambda^2)\theta(q_0)$. We find that the "proper" root restriction $\theta(q_0)$ must be relaxed when considering an internal mass variable such as λ^2. The reason is related to the fact that such variables have cuts which typically include $\lambda^2 < 0$. A straightforward application of the Feynman prescription $\lambda^2 - i\epsilon$ for $\lambda^2 > 0$ does, of course, lead to the inclusion of only one q_0-pole of the Feynman propagator $(q^2 - \lambda^2 + i\epsilon)^{-1}$, and such a restriction is then equivalent to the restriction $\theta(q_0)$. However, for $\lambda^2 < 0$, the propagator q_0-poles can enter the complex plane, in which case the restriction $\theta(q_0)$ becomes ill defined. In fact, in that case the poles will be complex conjugates of each other, and one will automatically only pick up one such pole in performing the Feynman integral, even if one ignores the Cutkosky $\theta(q_0)$. Thus pragmatically *one should ignore those $\theta(q_0)$ which effectively cause a cutoff in the Cutkosky domain of integration only because an internal mass is then negative.* A straightforward example and application of this idea is presented in our consideration of Δ_1.

However, a further problem arises when we consider Δ_3. In that case we again have a Cutkosky prescription $\theta(q_0)\delta(q^2 - \lambda^2)$ with $\lambda^2 < 0$. This time there are no other δ-functions, so that we really do have a q_0-integral to perform. Yet, even disregarding the $\theta(q_0)$, we have no way of integrating $\delta(q_0^2 - \mathbf{q}^2 + |\lambda|^2)$ throughout $0 < |\mathbf{q}| < \infty$ without q_0 becoming imaginary in part of the range of integration. One possible choice would then be to integrate over $(-\lambda^2)^{\frac{1}{2}} < |\mathbf{q}| < \infty$. However, the work of Fotiadi *et al.*[8] indicates that one should always integrate over some closed cycle which *vanishes* as one approaches the branch point in question. (That the cycle be closed is related to the fact that the discontinuity is the difference between two integration contours, and hence necessarily has a closed integration path. That the cycle should vanish at the branch point seems plausible, but not mandatory.) In our case, these conditions suggest setting $q_0 = ik$, and integrating in the k, $|\mathbf{q}|$ plane around the circumference of the circle $k^2 + \mathbf{q}^2 = l^2$, where $l^2 = -\lambda^2$. However, if we let $\hat{\mathbf{q}}$ range over all directions, we must in fact restrict this integration to the *semi*circle $|\mathbf{q}| > 0$. This interpretation of the closed-cycle condition differs from that of Aitchison. As we see below, our interpretation does lead to the same Δ_3 as already found. It is clear that the results of homology theory have practical impact on sordid computational problems, and that this theory deserves further study.

Evaluation of Δ_1 and Δ_2

Recalling the normalization of f given in Eq. (1), an immediate application of Cutkosky's rules leads to

$$\begin{aligned}\Delta_1 &= \frac{1}{2i}\,\mathrm{disc}\, f\Big|_{\lambda^2=(m-\nu)^2} \\ &= \frac{1}{2i}\times\frac{1}{\pi^2 i}\times(2\pi i)^2\int d^4k\;\delta_+(\mu^2\text{ leg}) \\ &\qquad\times\delta_+(\lambda^2\text{ leg})\times(\nu^2\text{ leg propagator}) \\ &= \frac{1}{2i}\times\frac{1}{\pi^2 i}\times(2\pi i)^2\int d^4k\;\theta(\rho_0-k_0) \\ &\qquad\times\delta[(\rho-k)^2-\nu^2]\,\frac{\theta(k_0)\;\delta[k^2-\lambda^2]}{\mu^2-(k-\sigma)^2-i\epsilon}\,, \qquad \text{(B1)}\end{aligned}$$

where the various momenta are defined in Fig. 9. Equation (B1) is Lorentz-invariant, and we therefore evaluate it in the ρ rest frame, with $\rho_0 = m$, and

$$\sigma_0 = \frac{m^2+M^2-s}{2m}\,, \qquad |\mathfrak{d}| = \frac{k(m^2, M^2, s)}{2m}. \qquad \text{(B2)}$$

Then

$$\begin{aligned}\Delta_1 &= \frac{1}{\pi^2 i}\times(2\pi i)^2 \\ &\qquad\times\int_0^{2\pi}d\phi\int_{-1}^{1}d\cos\theta\int_{-\infty}^{\infty}dk_0\int_0^{\infty}|\mathbf{k}|^2\,d\,|\mathbf{k}| \\ &\qquad\times\theta(m-k_0)\theta(k_0)\times\delta[m^2+\lambda^2-\nu^2-2mk_0] \\ &\qquad\times\frac{\delta[k_0^2-|\mathbf{k}|^2-\lambda^2]}{\mu^2-\lambda^2-M^2+2k_0\sigma_0-2\,|\mathbf{k}|\,|\mathfrak{d}|\cos\theta-i\epsilon}\,, \qquad \text{(B3)}\end{aligned}$$

where we have taken $\mathfrak{d}$ along the z axis. Because of the two delta-functions, we see that

$$k_0 = \frac{m^2+\lambda^2-\nu^2}{2m}\,, \qquad |\mathbf{k}|^2 = \frac{k^2(m^2,\lambda^2,\nu^2)}{4m^2}. \qquad \text{(B4)}$$

Thus $|\mathbf{k}|^2 = 0$ at the branch point $\lambda^2 = (m-\nu)^2$, at which $k_0 = m-\nu$. This lies within the domain of each θ-function. However, in general, these restrictions become

$$\begin{aligned}\theta(m-k_0) &\equiv \theta[m-(m^2+\lambda^2-\nu^2)/2m] \\ &= \theta(m^2+\nu^2-\lambda^2) \qquad \text{(B5a)}\end{aligned}$$

and

$$\theta(k_0) \equiv \theta(m^2+\lambda^2-\nu^2). \qquad \text{(B5b)}$$

While the restriction introduced by $\theta(m-k_0)$ has no limiting effect, the second restriction arising from $\theta(k_0)$ implies that Δ_1 *vanishes* for $\lambda^2 < \nu^2 - m^2 < 0$. But we know that this is not a singularity of f. Hence this second restriction (B5b) *must be ignored*. It arises from a $\theta(q_0)$, where q_0 is part of a 4-vector q which is *spacelike*. In fact, for such a spacelike vector, the restriction $\theta(q_0)$ is not even Lorentz-invariant, which provides another good reason for ignoring it. Once this step is taken, the evaluation is trivial, leading to

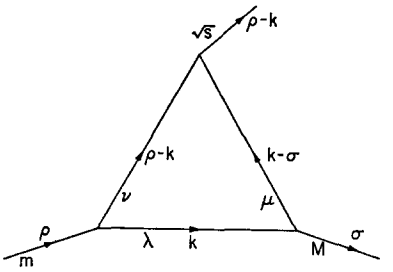


FIG. 9. The 4-momenta appropriate to the evaluation of the Δ_i.

$$\begin{aligned}\Delta_1 &= \frac{1}{2i}\times\frac{1}{\pi^2 i}\times(2\pi i)^2\times 2\pi\int_{-1}^{1}d\cos\theta\times\frac{1}{2m}\times\frac{|\mathbf{k}|}{2}\times\frac{1}{[\mu^2-\lambda^2-M^2+2k_0\sigma_0-2\,|\mathbf{k}|\,|\mathfrak{d}|\cos\theta-i\epsilon]} \\ &= -\frac{\pi}{k(s,m^2,M^2)}\ln\left\{\left[\mu^2-\lambda^2-M^2+2\left(\frac{m^2+\lambda^2-\nu^2}{2m}\right)\left(\frac{m^2+M^2-s}{2m}\right)-\frac{k(m^2,\lambda^2,\nu^2)k(s,m^2,M^2)}{2m^2}\right]\right. \\ &\qquad\left.\times\left[\mu^2-\lambda^2-M^2+2\left(\frac{m^2+\lambda^2-\nu^2}{2m}\right)\left(\frac{m^2+M^2-s}{2m}\right)+\frac{k(m^2,\lambda^2,\nu^2)k(s,m^2,M^2)}{2m^2}\right]^{-1}\right\} \\ &= -\frac{\pi}{k(s,m^2,M^2)}\ln\left\{\frac{-R-U^{\frac{1}{2}}}{-R+U^{\frac{1}{2}}}\right\} = \frac{\pi}{k(s,m^2,M^2)}\ln\left\{\frac{R-U^{\frac{1}{2}}}{R+U^{\frac{1}{2}}}\right\}, \qquad \text{(B6)}\end{aligned}$$

which agrees with Eq. (21).

In a precisely similar manner, one can evaluate Δ_2 and obtain the result of Eq. (29). It is very clear that the Cutkosky $\theta(q_0)$ restriction should be relaxed for such cases. Needless to say Cutkosky himself did not actually consider the case when an internal mass might become "spacelike".

Evaluation of Δ_3

In a precisely similar way

$$\Delta_3 = \frac{1}{2i}\,\mathrm{disc}\, f\Big|_{\lambda^2=0} = \frac{1}{2i} \times \frac{1}{\pi^2 i} \times (2\pi i) \times \frac{\int d^4k\ \theta(k_0)\ \delta(k^2 - \lambda^2)}{[\mu^2 - (k-\sigma)^2 - i\epsilon][\nu^2 - (\rho - k)^2 - i\epsilon]}. \tag{B7}$$

We again work in the ρ rest frame, and take $\mathfrak{d}$ to be along the z axis. Then

$$\Delta_3 = \frac{1}{2i} \times \frac{1}{\pi^2 i} \times (2\pi i) \times \int_0^{2\pi} d\phi \int_{-1}^{1} d\cos\theta \int_{-\infty}^{\infty} dk_0 \int_0^{\infty} |\mathbf{k}|^2\, d\,|\mathbf{k}|$$
$$\times \frac{\theta(k_0)\ \delta(k_0^2 - \mathbf{k}^2 - \lambda^2)}{[\mu^2 - \lambda^2 - M^2 + 2k_0\sigma_0 - 2\,|\mathbf{k}|\,|\mathfrak{d}|\cos\theta - i\epsilon][\nu^2 - m^2 - \lambda^2 + 2mk_0 - i\epsilon]}. \tag{B8}$$

Now however $\lambda^2 < 0$, so that we immediately drop the restriction $\theta(k_0)$. But we are still faced with the question of the domain of integration, and we follow the guide of homology theory.[8] Thus we set $k_0 = i\kappa$, and $\lambda^2 = -l^2$, to obtain $\delta(l^2 - \mathbf{k}^2 - \kappa^2)$. We then integrate around the circumference of this circle. In practice, this is most easily done by changing variables by

$$|\mathbf{k}| = r\sin\chi, \qquad k_0 = i\kappa = ir\cos\chi,$$

with $0 \le r < \infty$, $0 \le \chi < \pi$, which leads to

$$\Delta_3 = \frac{1}{\pi i} \times 2\pi \times \int_{-1}^{1} d\cos\theta \int_0^{\infty} r\,dr \int_0^{\pi} d\chi\, ir^2 \sin^2\chi$$
$$\times \frac{\delta(r^2 - l^2)}{[\mu^2 - \lambda^2 - M^2 + 2ir\cos\chi\,\sigma_0 - 2r\sin\chi\,|\mathfrak{d}|\cos\theta - i\epsilon][\nu^2 - m^2 - \lambda^2 + 2mir\cos\chi - i\epsilon]}. \tag{B9}$$

The r integration is absorbed by the delta-function, and the χ integral can then be performed by going to $t = \tan\tfrac{1}{2}\chi$, whence

$$\Delta_3 = \frac{1}{2}\int_{-1}^{1} d\cos\theta \int_{-\infty}^{\infty} \frac{2\,dt}{1+t^2}\, l^2 4t^2$$
$$\times [(\mu^2 - \lambda^2 - M^2 - i\epsilon)(1+t^2) + 2il\sigma_0(1-t^2) - 2l\,|\mathfrak{d}|\cos\theta 2t]^{-1}$$
$$\times [(\nu^2 - m^2 - \lambda^2 - i\epsilon)(1+t^2) + 2mil(1-t^2)]^{-1}. \tag{B10}$$

Here originally the t integration went from 0 to ∞; but, by using the symmetry of the integrand under $\theta \to \pi - \theta$, $t \to -t$, we have replaced the t limits by $-\infty < t < +\infty$. We thus have a closed contour integral in t, since the integrand $\to t^{-4}$ as $t \to \infty$. The integrand has poles at

$$t_1 = i, \quad t_2 = -i, \quad t_3 = t_\nu = i\left[\frac{(m-\lambda)^2 - \nu^2 + i\epsilon}{(m+\lambda)^2 - \nu^2 + i\epsilon}\right]^{\frac{1}{2}}, \quad t_4 = -t_3$$
$$\left.\begin{matrix} t_5 \\ t_6 \end{matrix}\right\} = \frac{-2i\lambda\,|\mathfrak{d}|\cos\theta \mp i[(\mu^2 - \lambda^2 - M^2 - i\epsilon)^2 - 4\lambda^2(\sigma_0^2 - \mathfrak{d}^2\cos^2\theta)]^{\frac{1}{2}}}{\mu^2 - \lambda^2 - M^2 - i\epsilon - 2\lambda\sigma_0}. \tag{B11}$$

Here, we have set $il = \lambda$ in some of the expressions. (By inspection of (B8), we see that the final result depends only on λ^2, so that the choice $il = \pm\lambda$ is immaterial provided it is made consistently.) We also define

$$a = \mu^2 - \lambda^2 - M^2 - 2\lambda\sigma_0 - i\epsilon, \qquad b = \nu^2 - (m+\lambda)^2 - i\epsilon. \tag{B12}$$

Then we have

$$\Delta_3 = \int_{-1}^{1} d\cos\theta \int_{-\infty}^{\infty} \frac{4l^2t^2}{(1+t^2)[at^2 + 4i\lambda\,|\mathfrak{d}|\cos\theta\, t + (a + 4\lambda\sigma_0)][bt^2 + (b + 4m\lambda)]}$$
$$= \int_{-1}^{1} d\cos\theta \int_{-\infty}^{\infty} dt\, \frac{4l^2}{ab}\, \frac{t^2}{\prod_{i=1}^{6}(t - t_i)}. \tag{B13}$$

The t integration can then be performed by closing the contour in the upper half-plane and enclosing t_1, t_3, and t_5. Here we are working on the basis of the pole locations for $\lambda^2 \approx 0$, which are then independent of θ; and by analytic continuation we preserve the same choice of poles for all $\lambda^2 \ll 0$.[30] Then

$$\begin{aligned}\Delta_3 &= 2\pi i \int_{-1}^{1} d\cos\theta \frac{4l^2}{ab}(R_1 + R_3 + R_5) \\ &= I_1 + I_3 + I_5,\end{aligned} \tag{B14}$$

where

$$R_1 = (-)ab\{2i[4\lambda\sigma_0 - 4\lambda\,|\mathfrak{d}|\cos\theta][4m\lambda]\}^{-1}, \tag{B15}$$

$$R_3 = t_p^2 a\{(1 + t_p^2)(2t_p) \times [at_p^2 + 4\lambda i\,|\mathfrak{d}|\cos\theta\, t_p + (a + 4\lambda\sigma_0)]\}^{-1}, \tag{B16}$$

and

$$R_5 = t_5^2 b\{(1 + t_5^2)(t_5 - t_6) \times [bt_5^2 + (b + 4m\lambda)]\}^{-1}. \tag{B17}$$

The dependence of each of R_1 and R_3 on $\cos\theta$ is explicit, and the $\cos\theta$ integration can immediately be performed. Thus

$$\begin{aligned}I_1 &= \frac{(2\pi i)4l^2(-)}{(2i)4m\lambda(-4\lambda\,|\mathfrak{d}|)} \ln\left(\frac{\sigma_0 - |\mathfrak{d}|}{\sigma_0 + |\mathfrak{d}|}\right) \\ &= \frac{\pi}{2k(s, m^2, M^2)} \times \ln\left[\frac{m^2 + M^2 - s + k(s, m^2, M^2)}{m^2 + M^2 - s - k(s, m^2, M^2)}\right].\end{aligned} \tag{B18}$$

Similarly, one finds

$$I_3 = \frac{\pi}{2k(s, m^2, M^2)} \ln\left(\frac{R + U^{\frac{1}{2}}}{R - U^{\frac{1}{2}}}\right), \tag{B19}$$

where R and U are defined in (17) and (18).

The evaluation of I_5 requires more circumspection, since t_5 is an irrational function of $\cos\theta$, and R_5 is a nontrivial function of t_5. We change variables to $\tau = t_5$, so that

$$\begin{aligned}\tau = t_5, \quad \cos\theta &= i\left[\frac{a(\tau^2 + 1) + 2\lambda\sigma_0}{4\lambda\,|\mathfrak{d}|\,\tau}\right], \\ d\cos\theta &= i\left(\frac{a\tau^2 - a - 2\lambda\sigma_0}{4\lambda\,|\mathfrak{d}|\,\tau^2}\right) d\tau.\end{aligned} \tag{B20}$$

Also

$$\begin{aligned}t_5 - t_6 &= 2[t_5 + (2i\lambda\,|\mathfrak{d}|\cos\theta/a)] \\ &= (a\tau^2 - a - 2\lambda\sigma_0)/(a\tau) \\ &= [(d\cos\theta)/i(d\tau)](4\lambda\,|\mathfrak{d}|\,\tau/a)\end{aligned} \tag{B21}$$

and

$$\begin{aligned}\tau_+ &= \tau(\cos\theta = +1) \\ \tau_- &= \tau(\cos\theta = -1) \\ &= [\mp 2i\lambda\,|\mathfrak{d}| - ik(\mu^2, \lambda^2, M^2)]/a.\end{aligned} \tag{B22}$$

We then obtain

$$\begin{aligned}I_5 &= \frac{\pi\lambda}{|\mathfrak{d}|}\int_{\tau_-}^{\tau_+} \frac{d(\tau^2)}{(1 + \tau^2)[b\tau^2 + b + 4m\lambda]} \\ &= \frac{\pi\lambda}{|\mathfrak{d}|}\int_{\tau_-}^{\tau_+} \frac{d\tau^2}{4m\lambda}\left(\frac{1}{1 + \tau^2} - \frac{b}{b\tau^2 + b + 4m\lambda}\right) \\ &= \frac{\pi}{2k(s, m^2, M^2)}[\ln(1 + \tau^2) - \ln(b\tau^2 + b + 4m\lambda)]_{\tau_-}^{\tau_+} \\ &= \frac{\pi}{2k(s, m^2, M^2)} \ln\left[\frac{b + 4m\lambda(1 + \tau_-^2)^{-1}}{b + 4m\lambda(1 + \tau_+^2)^{-1}}\right].\end{aligned} \tag{B23}$$

For the final step, we use the remarkable facts that

$$\begin{aligned}(1 + \tau_\pm^2)\frac{a^2}{4\lambda} &= 2\lambda M^2 - \sigma_0(\mu^2 - \lambda^2 - M^2) \mp |\mathfrak{d}|\,k(\lambda^2, M^2, \mu^2) \\ &= M^2 a^2[2\lambda M^2 - \sigma_0(\mu^2 - \lambda^2 - M^2) + |\mathfrak{d}|\,k(\lambda^2, M^2, \mu^2)]^{-1}.\end{aligned} \tag{B24}$$

Hence one readily finds that

$$b + 4m\lambda(1 + \tau_\pm^2)^{-1} = (1/2M^2)(-R' \pm U'^{\frac{1}{2}}), \tag{B25}$$

so that finally

$$I_5 = \frac{\pi}{2k(s, m^2, M^2)} \ln\left(\frac{R' + U'^{\frac{1}{2}}}{R' - U'^{\frac{1}{2}}}\right). \tag{B26}$$

On gathering together (B14), (B18), (B19), and (B26) we obtain precisely the same result as already given in Sec. 6.

Summary

The analysis of this appendix hence shows (at least in our specific case) that internal mass discontinuities can be calculated according to Cutkosky's rule, provided two important modifications are made in the interpretation. These are:

1. Ignore the "proper" pole prescription associ-

[30] Here we are assuming that $M > \mu$, so that $a < 0$ for $\lambda^2 \approx 0$. Our final results are independent of this assumption, as can be seen by analytic continuation, since $M = \mu$ is not a singularity in M.

ated with $\delta_+(q^2 - m^2)$ for $m^2 < 0$. Replace this by the simple prescription $\delta(q^2 - m^2)$ for $m^2 < 0$.

2. When integrating over the q^2 associated with such a $\delta(q^2 - m^2)$ for $m^2 < 0$, the q_0, $|\mathbf{q}|$ domain of integration must be re-interpreted into the complex plane, in such a way that one is integrating over a closed cycle which *vanishes* as one approaches the singularity in question. This result shows the importance of homological methods in practical considerations of Feynman graphs. The moral here is clear.

APPENDIX C. EQUIVALENCE OF Δ_3 WITH AITCHISON'S FORM (*Added in proof*)

As already noted, Aitchison's evaluation[3] of Δ_3 contains several errors. However, his basic method of evaluating Δ_3 is valid, and in this appendix we show that his corrected result is in fact equivalent to the one given here.

We commence with Eq. (B10), and perform the $\cos\theta$ integration. Then

$$\Delta_3 = \frac{i\lambda}{|\mathfrak{d}|}\int_{-\infty}^{\infty}\frac{t\,dt}{(t^2+1)b(t^2-t_p^2)} \times \ln\left\{\frac{(t-t_{5+})(t-t_{6+})}{(t-t_{5-})(t-t_{6-})}\right\}, \tag{C1}$$

where

$$\left.\begin{aligned} t_{5\pm} &= t_5|_{\cos\theta=\pm1} = -\frac{1}{a}[ik(\lambda^2,M^2,\mu^2) \pm 2i\lambda\,|\mathfrak{d}|], \\ t_{6\pm} &= t_6|_{\cos\theta=\pm1} \\ &= -\frac{1}{a}[-ik(\lambda^2,M^2,\mu^2) \pm 2i\lambda\,|\mathfrak{d}|], \end{aligned}\right\} \tag{C2}$$

and

$$t_p = i\left\{\frac{(m-\lambda)^2-\nu^2}{(m+\lambda)^2-\nu^2}\right\}^{\frac{1}{2}} = \frac{ik(\lambda^2,m^2,\nu^2)}{(m+\lambda)^2-\nu^2} = \frac{-ik(\lambda^2,m^2,\nu^2)}{b} = \frac{i[(m-\lambda)^2-\nu^2]}{k(\lambda^2,m^2,\nu^2)}. \tag{C3}$$

We again locate the poles and branch points for $\lambda^2 \approx 0$, and preserve these same locations relative to the integration contour for $\lambda^2 \ll 0$ by analytic continuation. Then the integrand has poles at i and at t_p in the upper half-plane, at $-i$ and at $-t_p$ in the lower half-plane; and branch points at t_{5+} and t_{5-} in the upper half-plane, and at t_{6+} and t_{6-} in the lower half-plane (cf., Aitchison, Fig. 15). Then using

$$\frac{t}{(t^2+1)(t^2-t_p^2)} = \frac{1}{2(1+t_p^2)}\left\{\left[\frac{1}{t-t_p}-\frac{1}{t-i}\right] + \left[\frac{1}{t+t_p}-\frac{1}{t+i}\right]\right\},$$

where we have split the poles into the two half-planes, we immediately obtain

$$\Delta_3 = \frac{\pi}{2k(s,m^2,M^2)}\ln\left\{\left(\frac{t_p-t_{6+}}{t_p-t_{6-}}\right)\left(\frac{t_p+t_{5-}}{t_p+t_{5+}}\right) \times \left(\frac{i+t_{5+}}{i+t_{5-}}\right)\left(\frac{i-t_{6-}}{i-t_{6+}}\right)\right\} = \frac{\pi}{2k(s,m^2,M^2)}\ln\left\{\frac{(t_p+t_{5-})(i+t_{5+})}{(t_p+t_{5+})(i+t_{5-})}\right\}^2. \tag{C4}$$

(Here, we have used $t_{5\pm} = -t_{6\mp}$.) [Eq. (C4) is essentially the same as Aitchison's expression for $\tilde{\Delta}_3(s,\lambda)$ given above his Eq. (A8). Aitchison incorrectly integrated χ from 0 to 2π, instead of only from 0 to π. Thus, he incorrectly set $\Delta_3 = \tilde{\Delta}_3(s,\lambda) + \tilde{\Delta}_3(s,-\lambda)$—his Eq. (A4), when in fact $\Delta_3 \equiv \tilde{\Delta}_3(s,\lambda)$. Actually $\tilde{\Delta}_3(s,\lambda)$ is a function of λ^2.]

While (C4) is correct, and is in fact only dependent on λ^2, this property is not explicit. Now, from (C2) and (C3)

$$t_{5\pm}(\lambda) = -1/t_{5\pm}(-\lambda), \quad t_p(\lambda) = -1/t_p(-\lambda),$$

where $t_{5\pm}(\lambda) \equiv t_{5\pm}$, and $t_{5\pm}(-\lambda)$ is obtained from $t_{5\pm}(\lambda)$ by the substitution $\lambda \to -\lambda$; similarly for $t_p(-\lambda)$. Hence, after trivial manipulation, we can rewrite (C4) as

$$\Delta_3 = \frac{\pi}{2k(s,m^2,M^2)}\ln\frac{A_+B_+}{A_-B_-} \tag{C5}$$

with

$$\left.\begin{aligned} A_+ &= [t_p(\lambda)+t_{5-}(\lambda)][t_p(-\lambda)+t_{5-}(-\lambda)], \\ B_+ &= [i+t_{5+}(\lambda)][i+t_{5+}(-\lambda)] \end{aligned}\right\} \tag{C6}$$

and where A_- and B_- are obtained from A_+ and B_+ by the replacement $|\mathfrak{d}| \to -|\mathfrak{d}|$, under which $t_{5\pm} \to t_{5\mp}$.

To reduce A_+ and B_+ further, we introduce the following condensed notation:

$$\left.\begin{aligned} k_\mu &= k(\lambda^2,M^2,\mu^2), & K_\mu &= \lambda^2+M^2-\mu^2, \\ k_\nu &= k(\lambda^2,m^2,\nu^2), & K_\nu &= \lambda^2+m^2-\nu^2, \\ k_s &= k(M^2,m^2,s), & K_s &= M^2+m^2-s. \end{aligned}\right\} \tag{C7}$$

Then

$$2m\sigma_0 = K_s, \qquad 2m\,|\mathfrak{d}| = k_s, \tag{C8}$$

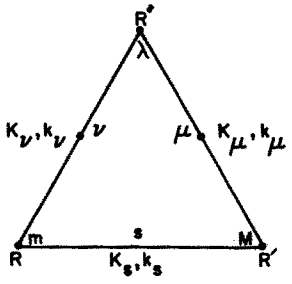
and

$$\left.\begin{aligned} R &= 2m^2K_\mu - K_sK_\nu, & U^{\frac{1}{2}} &= k_sk_\nu \\ R' &= 2M^2K_\nu - K_sK_\mu, & U'^{\frac{1}{2}} &= k_sk_\mu \end{aligned}\right\} \tag{C9}$$

[(cf., Eqs. (18), (17), (24), and (26)]. We also define

$$R'' = 2\lambda^2K_s - K_\mu K_\nu, \quad U''^{\frac{1}{2}} = k_\mu k_\nu. \tag{C10}$$

There is a very useful symmetry here, as indicated in a triangle



(C11)

We then find

$$A_+ = \left[\frac{i(K_\nu - 2m\lambda)}{k_\nu} + \frac{i(mk_\mu - \lambda k_s)}{mK_\mu + \lambda K_s}\right] \times [\text{ditto } \lambda \to -\lambda]$$

$$= \left[\frac{i\{m(R'' - U''^{\frac{1}{2}}) + \lambda(R + U^{\frac{1}{2}})\}}{k_\nu(mK_\mu + \lambda K_s)}\right] \times [\text{ditto } \lambda \to -\lambda]$$

$$= -\frac{[m^2(R'' - U''^{\frac{1}{2}})^2 - \lambda^2(R + U^{\frac{1}{2}})^2]}{k_\nu^2(m^2K_\mu^2 - \lambda^2K_s^2)}. \quad \text{(C12)}$$

But we also have the property

$$R^2 - U = 4m^2\Gamma, \qquad R'^2 - U' = 4M^2\Gamma,$$
$$R''^2 - U'' = 4\lambda^2\Gamma$$

[cf., Eqs. (3), (20), (28), and the symmetry (C11)]. Hence

$$m^2(R'' - U''^{\frac{1}{2}})^2 - \lambda^2(R + U^{\frac{1}{2}})^2$$
$$= \frac{1}{4\Gamma}\{(R^2 - U)(R'' - U''^{\frac{1}{2}})^2 - (R''^2 - U'')(R + U^{\frac{1}{2}})^2\}$$
$$= \frac{-1}{2\Gamma}(R + U^{\frac{1}{2}})(R'' - U''^{\frac{1}{2}})(RU''^{\frac{1}{2}} + R''U^{\frac{1}{2}}). \quad \text{(C13)}$$

Similarly, we find

$$B_+ = \left[i + \frac{i(mk_\mu + \lambda k_s)}{(mK_\mu + \lambda K_s)}\right][\text{ditto } \lambda \to -\lambda]$$

$$= \left[\frac{m(K_\mu + k_\mu) + \lambda(K_s + k_s)}{-i(mK_\mu + \lambda K_s)}\right] \times [\text{ditto } \lambda \to -\lambda]$$

$$= -\frac{[m^2(K_\mu + k_\mu)^2 - \lambda^2(K_s + k_s)^2]}{(m^2K_\mu^2 - \lambda^2K_s^2)}. \quad \text{(C14)}$$

We now use the property

$$K_\mu^2 - k_\mu^2 = 4\lambda^2M^2, \qquad K_\nu^2 - k_\nu^2 = 4\lambda^2m^2, \qquad K_s^2 - k_s^2 = 4m^2M^2 \quad \text{(C15)}$$

to obtain

$$m^2(K_\mu + k_\mu)^2 - \lambda^2(K_s + k_s)^2$$
$$= \frac{1}{4M^2}\{(K_s^2 - k_s^2)(K_\mu + k_\mu)^2 - (K_\mu^2 - k_\mu^2)(K_s + k_s)^2\}$$
$$= \frac{1}{2M^2}(K_s + k_s)(K_\mu + k_\mu)(K_sk_\mu - K_\mu k_s). \quad \text{(C16)}$$

Finally (using brute force multiplication), we observe that

$$(K_sk_\mu - K_\mu k_s)(RU''^{\frac{1}{2}} + R''U^{\frac{1}{2}})$$
$$= (K_sk_\mu - K_\mu k_s)[(2m^2K_\mu - K_sK_\nu)k_\nu k_\mu + (2\lambda^2K_s - K_\mu K_\nu)k_sk_\nu]$$
$$= 2k_\nu(\lambda^2K_s^2 - m^2K_\mu^2)(R' + U'^{\frac{1}{2}}), \quad \text{(C17)}$$

where (C15) was used to eliminate k_μ^2 and k_s^2.

Putting Eqs. (C12), (C13), (C14), (C16), and (C17) together, we then obtain

$$A_+B_+ = \frac{(R+U^{\frac{1}{2}})(R''-U''^{\frac{1}{2}})(K_s+k_s)(K_\mu+k_\mu)(R'+U'^{\frac{1}{2}})}{2M^2\Gamma k_\nu(m^2K_\mu^2-\lambda^2K_s^2)}. \quad \text{(C18)}$$

To obtain A_-B_- from A_+B_+ we change $|\delta|$ to $-|\delta|$, i.e., $k_s \to -k_s$. In this change $U^{\frac{1}{2}} \to -U^{\frac{1}{2}}$, $U'^{\frac{1}{2}} \to -U'^{\frac{1}{2}}$, $k_s \to -k_s$, and *no* other quantities change. Hence

$$\frac{A_+B_+}{A_-B_-} = \frac{(R + U^{\frac{1}{2}})(R' + U'^{\frac{1}{2}})(K_s + k_s)}{(R - U^{\frac{1}{2}})(R' - U'^{\frac{1}{2}})(K_s - k_s)}. \quad \text{(C19)}$$

Equations (C5) and (C19) are exactly equivalent to Eqs. (36) and (48). Hence Aitchison's equation for $\tilde{\Delta}_3(s, \lambda)$ is equivalent to our expression for Δ_3, except for minor algebraic errors (cf, his erratum, to be published).

Continuous Degenerate Representations of Noncompact Rotation Groups. II

N. Limić,* J. Niederle,† and R. Raczka‡
International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy
(Received 22 March 1966)

Three principal continuous series of most degenerate unitary irreducible representations of an arbitrary noncompact rotation group $SO(p, q)$ have been derived and their properties discussed in detail. The corresponding harmonic functions have been constructed.

1. INTRODUCTION

In our previous paper,[1] the discrete series of most degenerate representations of an arbitrary noncompact rotation group $SO_0(p, q)$ were derived and their properties discussed. In the present work, we go into the continuous series of most degenerate representations of these groups and the corresponding harmonic functions.

The harmonic functions for the Lorentz group $SO(3, 1)$ were investigated in detail by Dolginov and his co-workers.[2] An arbitrary Lorentz-type group was considered by Vilenkin,[3] who derived the irreducible unitary representations of the class one of $SO(n, 1)$. The construction of harmonic functions, which carry one continuous series of the most degenerate representations of any $SO_0(p, q)$ group, was given in Ref. 4. In this paper, we present three continuous series of most degenerate representations of an arbitrary noncompact $SO_0(p, q)$ group, and we also construct explicitly a set of corresponding harmonic functions. These harmonic functions are characterized only by discrete numbers connected with the representation of the maximal compact subgroup $SO(p) \times SO(q)$ of the group $SO_0(p, q)$.

In Sec. 2, two series of continuous representations of the $SO_0(p, q)$ $p \geq q > 1$ and corresponding harmonic functions related to the hyperboloids are constructed. The same problem for the Lorentz-type groups $SO_0(p, 1)$ is considered in Sec. 3. The continuous representations and corresponding harmonic functions related to the cone of an arbitrary $SO_0(p, q)$ group are investigated in Sec. 4. Section 5 is devoted to proof of the irreducibility of the derived series of representations. Finally, in Sec. 6, we discuss some features of the derived representations and harmonic functions. For instance, it turns out that, except for one series of representations of the Lorentz-type groups, there exist two irreducible representations of $SO_0(p, q)$ differing by parity for any definite eigenvalue of the Casimir operator.

The completeness relations of our harmonic functions, the corresponding decomposition of quasi-regular representations, and the connection with the Gel'fand–Kostiučenko triplet will be treated in detail in a subsequent article (Part III).

In what follows, we use the conventional terminology, that is, we speak about representations of the group $SO_0(p, q)$ on the Hilbert space $\mathfrak{H}$, although we derived only representations of the Lie algebra $\mathfrak{R}$ of the considered group on definite vector space $\mathfrak{D}$, which is dense in the Hilbert space $\mathfrak{H}$. However, in Part III of our series of articles, it will be shown that our local representations induce the global irreducible unitary representations of the group $SO_0(p, q)$.

2. CONTINUOUS SERIES OF MOST DEGENERATE REPRESENTATIONS OF $SO_0(p, q)$ GROUPS ($p \geq q > 1$) RELATED TO HYPERBOLOIDS

For the most degenerate representations of $SO_0(p, q)$, the ring of invariant operators of the corresponding Lie algebra is generated only by one independent operator. Following the procedure explained in Sec. 2 of Ref. 1, we can represent it as the Laplace–Beltrami operator on the definite vector space $\mathfrak{D}$, which is dense in the Hilbert space of functions, the domain of which are the following homogeneous spaces of rank one,

$$X_+ = SO_0(p, q)/SO_0(p - 1, q), \qquad X_- = SO_0(p, q)/SO_0(p, q - 1). \tag{2.1}$$

* On leave of absence from Institute Rudjer Bošković, Zagreb.
† On leave of absence from Institute of Physics of the Czechoslovak Academy of Sciences, Prague.
‡ On leave of absence from Institute of Nuclear Research, Warsaw.
[1] R. Raczka, N. Limić, and J. Niederle, J. Math. Phys. **7**, 1861 (1966).
[2] A. Z. Dolginov, Zh. Eksperim. i Teor. Fiz. **30**, 746 (1956) [English transl.: Soviet Phys.—JETP **3**, 589 (1956)]; A. Z. Dolginov and J. N. Toptygin, Zh. Eksperim. i Teor. Fiz. **37**, 1441 (1959) [English transl.: Soviet Phys.—JETP **10**, 1022 (1960)]; A. Z. Dolginov and A. N. Moskalev, Zh. Eksperim. i Teor. Fiz. **37**, 1697 (1959) [English transl.: Soviet Phys.—JETP **10**, 1202 (1960)].
[3] N. Ya. Vilenkin, Trudy Moskov. Mat. Obšhćhestva **12**, 185 (1963).
[4] J. Fischer, J. Niederle, and R. Raczka, J. Math. Phys. **7**, 816 (1966). J. Niederle and R. Raczka, International Center for Theoretical Physics Preprint IC/65/89, Trieste (1965).

Then we solve the eigenvalue problem for this invariant operator. It is obvious that the generalized Fourier images of its eigenfunctions carry continuous representations of the considered group. The irreducibility of these representations is proved in Sec. 5.

The homogeneous space X_+ and X_-, defined in (2.1), can be realized by the hyperboloids H_q^p and H_p^q, respectively.[1] The hyperboloid H_b^a is determined by the equation

$$(x^1)^2 + \cdots + (x^a)^2 - (x^{a+1})^2 - \cdots - (x^{a+b})^2 = 1 \qquad (2.2)$$

and is imbedded in $(a+b)$-dimensional Minkowski space $M^{(a,b)}$. By using the biharmonic coordinate system defined in [(3.4)–(3.9) of Ref. 1], and considering the properties of the metric tensor $g_{\alpha\beta}(H_q^p)$ on the hyperboloid H_q^p, we can write the Laplace–Beltrami operator $\Delta(H_q^p)$ in the form [see (3.10) of Ref. 1]

$$\Delta(H_q^p) = \frac{-1}{\cosh^{p-1}\theta \sinh^{q-1}\theta}\frac{\partial}{\partial\theta}\cosh^{p-1}\theta\cdot\sinh^{q-1}\theta\frac{\partial}{\partial\theta} + \frac{\Delta(S^{p-1})}{\cosh^2\theta} - \frac{\Delta(S^{q-1})}{\sinh^2\theta}, \qquad \theta\in[0,\infty), \qquad (2.3)$$

where $\Delta(S^{p-1})[\Delta(S^{q-1})]$ is the Laplace–Beltrami operator on the sphere S^{p-1} $[S^{q-1}]$ of the compact rotation group $SO(p)[SO(q)]$ defined by (A3) and (A4) of Ref. 1. If we represent the eigenfunctions of $\Delta(H_q^p)$ as a product of the eigenfunctions of $\Delta(S^{p-1})$, $\Delta(S^{q-1})$, and a function $V^{\Lambda}_{l_{[p/2]},l_{[q/2]}}$, we obtain the following differential equation[5] for the latter function:

$$\left[\frac{-1}{\cosh^{p-1}\theta\cdot\sinh^{q-1}\theta}\frac{d}{d\theta}\cosh^{p-1}\theta\cdot\sinh^{q-1}\theta\frac{d}{d\theta} - \frac{l_{[\frac12 p]}(l_{[\frac12 p]}+p-2)}{\cosh^2\theta} + \frac{\tilde{l}_{[\frac12 q]}(\tilde{l}_{[\frac12 q]}+q-2)}{\sinh^2\theta} - \Lambda^2 - \left(\frac{p+q-2}{2}\right)^2\right]\cdot V^{\Lambda}_{l_{[p/2]},l_{[q/2]}}(\theta) = 0. \qquad (2.4)$$

Here, $l_{[\frac12 p]}(l_{[\frac12 p]}+p-2)[\tilde{l}_{[\frac12 q]}(\tilde{l}_{[\frac12 q]}+q-2)]$ are eigenvalues of the operator $\Delta(S^{p-1})[\Delta(S^{q-1})]$ on the sphere $S^{p-1}[S^{q-1}]$. Therefore, $l_{[\frac12 p]}[l_{[\frac12 q]}]$ are non-negative integers, except the lowest case $p=2[q=2]$ when by definition $l_1\equiv m_1[\tilde{l}_1\equiv\tilde{m}_1]$, and $m_1[\tilde{m}_1]$ is integer. $\Lambda^2+[\frac12(p+q-2)]^2$, $\Lambda\in[0,\infty)$ is our ansatz for eigenvalues of $\Delta(H_q^p)$ corresponding to the continuous series of representations. It is shown in Part III that in this way the whole continuous spectrum of $\Delta(H_q^p)$ is obtained.

The solution of Eq. (2.4), regular at the origin, is given by the function

$$V^{\Lambda}_{l_{[p/2]},l_{[q/2]}}(\theta) = (N^{-\frac12})\cdot\tanh^{|l_{[q/2]}|}\theta\cdot\cosh^{[-\frac12(p+q-2)+i\Lambda]}\theta\cdot{}_2F_1\{\tfrac12[|\tilde{l}_{[\frac12 q]}|+|l_{[\frac12 p]}|-i\Lambda+\tfrac12(p+q-2)], \cdot\tfrac12[|\tilde{l}_{[\frac12 q]}|-|l_{[\frac12 p]}|-i\Lambda+\tfrac12(q-p+2)];|\tilde{l}_{[\frac12 q]}|+\tfrac12 q;\tanh^2\theta\}, \qquad (2.5)$$

with

$$N = 2\pi\left|\frac{\Gamma(|\tilde{l}_{[\frac12 q]}|+\frac12 q)\cdot\Gamma(i\Lambda)}{\Gamma\{\frac12[i\Lambda+|l_{[\frac12 p]}|+|\tilde{l}_{[\frac12 q]}|+\frac12(p+q-2)]\}\cdot\Gamma\{\frac12[i\Lambda+|\tilde{l}_{[\frac12 q]}|-|l_{[\frac12 p]}|+\frac12(q-p+2)]\}}\right|^2.$$

The eigenfunctions of the Laplace–Beltrami operator $\Delta(H_q^p)$ are then harmonic functions

$$Y^{\Lambda,l_2,\cdots,l_{[p/2]},l_2,\cdots,l_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}(\theta,\omega,\tilde{\omega})$$

of the form

$$Y^{\Lambda,l_2,\cdots,l_{[p/2]},l_2,\cdots,l_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}(\theta,\omega,\tilde{\omega}) = V^{\Lambda}_{l_{[p/2]},l_{[q/2]}}(\theta)\cdot Y^{l_2,\cdots,l_{[p/2]}}_{m_1,\cdots,m_{[p/2]}}(\omega)\cdot Y^{l_2,\cdots,l_{[q/2]}}_{\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}(\tilde{\omega}), \qquad (2.6)$$

where $V^{\Lambda}_{l_{[p/2]},l_{[q/2]}}(\theta)$ is given in (2.5),

$$Y^{l_2,\cdots,l_{[p/2]}}_{m_1,\cdots,m_{[p/2]}}(\omega) \quad\text{and}\quad Y^{l_2,\cdots,l_{[q/2]}}_{\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}(\tilde{\omega})$$

are, respectively, the eigenfunctions of $\Delta(S^{p-1})$ and $\Delta(S^{q-1})$. They can be expressed as a product of the usual d-functions of angular momenta and exponential functions [see (A9) and (3.18) of Ref. 1]. For instance, $Y^{l_2,\cdots,l_{[p/2]}}_{m_1,\cdots,m_{[p/2]}}(\omega)$ looks like

[5] Here and elsewhere we keep the notation from Ref. 1. Let us keep in mind that the brackets are defined as follows:

$$[\tfrac12 a] = \begin{cases}\frac12 a & \text{if } a=2r\\ \frac12(a-1) & \text{if } a=2r+1\end{cases} \qquad r=1,2,\cdots$$

and

$$\{\tfrac12 a\} = \begin{cases}\frac12 a & \text{if } a=2r\\ \frac12(a+1) & \text{if } a=2r+1\end{cases} \qquad r=1,2,\cdots.$$

$$Y^{l_2,\cdots,l_{[p/2]}}_{m_1,\cdots,m_{[p/2]}}(\omega) \equiv \begin{cases} Y^{l_2,\cdots,l_r}_{m_1,\cdots,m_r}(\omega) = (N_r^{-\frac{1}{2}}) \prod_{k=2}^{r} \sin^{2-k}(\vartheta^k)\cdot d^{J_k}_{M_k;M_{k'}}(2\vartheta^k)\cdot \prod_{k=1}^{r} \exp im_k\varphi^k & \text{if } p = 2r, \\ Y^{l_2,\cdots,l_{r+1}}_{m_1,\cdots,m_r}(\omega) = (N_{r+1}^{-\frac{1}{2}}) \sin^{1-r}(\vartheta^{r+1})\cdot d^{J_{r+1}}_{M_{r+1};0}(\vartheta^{r+1}) & \\ \qquad \cdot \prod_{k=2}^{r} \sin^{2-k}(\vartheta^k)\cdot d^{J_k}_{M_k;M_{k'}}(2\vartheta^k)\cdot \prod_{k=1}^{r} \exp im_k\varphi^k, & \text{if } p = 2r+1, \end{cases} \tag{2.7}$$

with the normalization factors

$$N_r = 2\pi^r \prod_{k=2}^{r} (l_k + k - 1)^{-1}, \tag{2.8}$$

$$N_{r+1} = 4\pi^r(2(l_{r+1} + r) - 1)^{-1}\cdot \prod_{k=2}^{r} (l_k + k - 1)^{-1},$$

and

$$J_k = \tfrac{1}{2}(l_k + k - 2), \quad M_k = \tfrac{1}{2}(m_k + l_{k-1} + k - 2),$$

$$M_k' = \tfrac{1}{2}(m_k - l_{k-1} - k + 2) \text{ for } k = 2, 3, \cdots, r,$$

$$J_{r+1} = l_{r+1} + r - 1, \text{ and } M_{r+1} = l_r + r - 1.$$

The harmonic functions

$$Y^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}(\theta, \omega, \tilde{\omega})$$

constitute an orthogonal set of functions with respect to the measure $d\mu(\theta, \omega, \tilde{\omega})$ induced by our coordinate system on the hyperboloid H^p_q

$$d\mu(\theta, \omega, \tilde{\omega}) = [\bar{g}(H^p_q)]^{\frac{1}{2}}\, d\omega\cdot d\tilde{\omega}\cdot d\theta = d\mu(\omega)\cdot d\mu(\tilde{\omega})\cdot d\mu(\theta), \tag{2.9}$$

where

$$d\mu(\omega) = \begin{cases} \prod_{k=2}^{r} \cos(\vartheta^k)\cdot \sin^{(2k-3)}(\vartheta^k)\cdot d\vartheta^k\cdot \prod_{k=1}^{r} d\varphi^k, & \text{for } p = 2r, \\ \sin^{2r-1}(\vartheta^{r+1})\, d\vartheta^{r+1}\cdot \prod_{k=2}^{r} \cos(\vartheta^k)\cdot \sin^{(2k-3)}(\vartheta^k)\cdot d\vartheta^k\cdot \prod_{k=1}^{r} d\varphi^k, & \text{for } p = 2r+1, \end{cases}$$

$$d\mu(\theta) = \cosh^{p-1}\theta\cdot \sinh^{q-1}\theta\cdot d\theta,$$

and the expression for $d\mu(\tilde{\omega})$ is the same as for $d\mu(\omega)$, but in twiddle variables.

Let us construct now the carrier space of representations belonging to the group $SO_0(p, q)$. The generalized Fourier transform with respect to the eigenfunction (2.6) of a function

$$f(\theta, \omega, \tilde{\omega}) = P(x^1, \cdots, x^{p+q}) \exp\left[-\sum_{i=1}^{p+q} (x^i)^2\right],$$

where $P(x^1, \cdots x^{p+q})$ is an arbitrary polynomial in x^i, and x^i are expressed in our biharmonic coordinates on H^p_q [(3.4)–(3.9) of Ref. 1], has the form

$$\chi^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}} = \langle Y^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}, f\rangle = \int_{H^p_q} \overline{Y^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}(\theta, \omega, \tilde{\omega})} \cdot f(\theta, \omega, \tilde{\omega})\cdot d\mu(\theta, \omega, \tilde{\omega}). \tag{2.10}$$

All such Fourier transforms determine the Hilbert spaces $\mathfrak{H}^{\Lambda,+}_{p,q}$ and $\mathfrak{H}^{\Lambda,-}_{p,q}$ of vectors

$$\chi^{\Lambda,\pm} \equiv \left\{\chi^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}};\ l_{[\frac{1}{2}p]} + \tilde{l}_{[\frac{1}{2}q]} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix}\right\}$$

for which

$$||\chi^{\Lambda,\pm}||^2 = \sum_{\substack{l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]} \\ m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}} |\chi^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}|^2 < \infty,$$

where $l_{[\frac{1}{2}p]} + \tilde{l}_{[\frac{1}{2}q]}$ is even or odd, respectively. The scalar product in the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,q}$ is defined by

$$(\chi^{\Lambda,\pm}, \psi^{\Lambda,\pm}) = \sum \overline{\chi^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}}\cdot \psi^{\Lambda,l_2,\cdots,l_{[p/2]},\tilde{l}_2,\cdots,\tilde{l}_{[q/2]}}_{m_1,\cdots,m_{[p/2]},\tilde{m}_1,\cdots,\tilde{m}_{[q/2]}}, \tag{2.11}$$

where the sum is taken over all integers $(l_2, \cdots, l_{[\frac{1}{2}p]}, \tilde{l}_2, \cdots, \tilde{l}_{[\frac{1}{2}q]}, m_1, \cdots, m_{[\frac{1}{2}p]}, \tilde{m}_1, \cdots, \tilde{m}_{[\frac{1}{2}q]})$ with $l_{[\frac{1}{2}p]} + \tilde{l}_{[\frac{1}{2}q]}$ even or odd, respectively. The Hilbert space $\mathfrak{H}^{\Lambda,+}_{p,q}$ ($\mathfrak{H}^{\Lambda,-}_{p,q}$) is an eigenspace of the parity operator with the eigenvalue $+1(-1)$ and can be decomposed as follows:

$$\mathfrak{H}^{\Lambda,\pm}_{p,q} = \sum_{l_{[p/2]},\tilde{l}_{[q/2]}=0}^{\infty} \oplus\, \mathfrak{H}^{\Lambda,\pm}_{p,q;l_{[p/2]},\tilde{l}_{[q/2]}},$$

where $\mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(\frac{1}{2}p)},l_{(\frac{1}{2}q)}}$ are finite-dimensional vector spaces determined by $\chi^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}$ with fixed $l_{(\frac{1}{2}p)}$ and $\tilde{l}_{(\frac{1}{2}q)}$.

The continuous series $C^{\Lambda,\pm}_{p,q}$ of the representations of the group $SO_0(p, q)$, $p > q \geq 1$, on the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,q}$ is induced by the representation of the corresponding Lie algebra on the vectors $\chi^{\Lambda,\pm}$

$$\begin{aligned} L_{ij}\chi^{\Lambda,\pm} &\equiv \left\{\langle Y^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}, L_{ij}f\rangle;\ l_{(\frac{1}{2}p)} + \tilde{l}_{(\frac{1}{2}q)} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix}\right\}, \\ B_{st}\chi^{\Lambda,\pm} &\equiv \left\{\langle Y^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}, B_{st}f\rangle;\ l_{(\frac{1}{2}p)} + \tilde{l}_{(\frac{1}{2}q)} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix}\right\}, \end{aligned} \tag{2.12}$$

where again

$$f(\theta, \omega, \tilde{\omega}) = P(x^1, \cdots, x^{p+q}) \exp\left[-\sum_{i=1}^{p+q}(x^i)^2\right].$$

[$P(x^1, \cdots, x^{p+q})$ is a polynomial and x^i are expressed in the biharmonic coordinates on the hyperboloid H^p_q in (3.4)–(3.9) of Ref. 1] and L_{ij}, B_{st} are elements of the Lie algebra of the compact and noncompact type, respectively [Ref. 1, Eqs. (6.2), (6.3)].

The proof of the irreducibility of the derived series $C^{\Lambda,\pm}_{p,q}$ of the continuous representations is given in Sec. 5.

The harmonic functions on the hyperboloid H^q_p can be obtained by exchanging p, $l_{(\frac{1}{2}p)}$ and q, $\tilde{l}_{(\frac{1}{2}q)}$, respectively, and vice versa only in the function $V^{\Lambda}_{l_{(p/2)},\tilde{l}_{(q/2)}}(\theta)$ contained in the harmonic function $Y^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}$ on the hyperboloid H^p_q. The continuous series $C^{\Lambda,\pm}_{q,p}$ of the most degenerate irreducible unitary representations of $SO_0(p, q)$, $p \geq q > 1$, on $\mathfrak{H}^{\Lambda,\pm}_{q,p}$ are then constructed by the same procedure as described above.

3. CONTINUOUS SERIES OF MOST DEGENERATE REPRESENTATIONS OF $SO(p, 1)$ GROUPS RELATED TO THE HYPERBOLOIDS

The spaces X_+ and X_- (2.1) can be realized now by hyperboloids H^p_1 and H^1_p, respectively.[1] The biharmonic coordinates on H^p_1 and H^1_p are introduced again as in [Ref. 1, Sec. 3]. We consider the Lorentz-type groups separately because the range of θ on the hyperboloid H^p_1 is $(-\infty, \infty)$, and therefore the solution of the eigenvalue problem of $\Delta(H^p_1)$ is different from the corresponding one in the previous case. On the hyperboloid H^1_p, the range of θ is from zero to infinity since we restrict ourselves to the upper sheet of the hyperboloid H^1_p. Of course, the upper sheet of H^1_p is a transitive manifold only under the proper $SO_0(p, 1)$ group, i.e., under the group of transformations $g = (g_{ik})$, for which g_{11} is positive.

The Laplace–Beltrami operator $\Delta(H^p_1)$ has the form [Ref. 1, Eq. (5.2)]

$$\Delta(H^p_1) = \frac{-1}{\cosh^{p-1}\theta}\frac{\partial}{\partial\theta}\cosh^{p-1}\theta\frac{\partial}{\partial\theta} + \frac{\Delta(S^{p-1})}{\cosh^2\theta}, \qquad \theta \in (-\infty, \infty), \tag{3.1}$$

where $\Delta(S^{p-1})$ is the Laplace–Beltrami operator for the $SO(p)$ group. Representing the eigenfunctions of $\Delta(H^p_1)$ as a product of eigenfunctions of $\Delta(S^{p-1})$ and a function $V^{\Lambda}_{l_{(p/2)}}(\theta)$, we obtain the following differential equation for $V^{\Lambda}_{l_{(p/2)}}(\theta)$:

$$\left[\frac{1}{\cosh^{p-1}\theta}\frac{d}{d\theta}\cosh^{p-1}\theta\frac{d}{d\theta} + \frac{l_{(\frac{1}{2}p)}(l_{(\frac{1}{2}p)} + p - 2)}{\cosh^2\theta} + \Lambda^2 + \left(\frac{p-1}{2}\right)^2\right]\cdot V^{\Lambda}_{l_{(p/2)}}(\theta) = 0, \tag{3.2}$$

where $l_{(\frac{1}{2}p)}(l_{(\frac{1}{2}p)} + p - 2)$ and $\Lambda^2 + \frac{1}{2}(p-1)^2$ are eigenvalues of $\Delta(S^{p-1})$ and $\Delta(H^p_1)$, respectively. Both independent solutions of Eq. (3.2) are regular at the origin and can be taken as orthogonal functions ${}_{1,2}V^{\Lambda}_{l_{(p/2)}}(\theta)$ in the form

$$\begin{aligned} {}_1V^{\Lambda}_{l_{(p/2)}}(\theta) &= -2({}_1K^{-\frac{1}{2}})\cdot\tanh\theta\cdot\cosh^{-[\frac{1}{2}(p-1)+i\Lambda]}\theta \\ &\quad\cdot {}_2F_1\{\tfrac{1}{2}[i\Lambda + l_{(\frac{1}{2}p)} + \tfrac{1}{2}(p+1)], \tfrac{1}{2}[i\Lambda - l_{(\frac{1}{2}p)} - \tfrac{1}{2}(p-5)]; \tfrac{3}{2}; \tanh^2\theta\} \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} {}_2V^{\Lambda}_{l_{(p/2)}}(\theta) &= ({}_2K^{-\frac{1}{2}})\cosh^{-[\frac{1}{2}(p-1)+i\Lambda]}\theta \\ &\quad\cdot {}_2F_1\{\tfrac{1}{2}[i\Lambda + l_{(\frac{1}{2}p)} + \tfrac{1}{2}(p-1)], \tfrac{1}{2}[i\Lambda - l_{(\frac{1}{2}p)} - \tfrac{1}{2}(p-3)]; \tfrac{1}{2}; \tanh^2\theta\}, \end{aligned}$$

with

$$ {}_1K = \frac{\pi\{\cosh(\pi\Lambda) - (-1)^{l_{(\frac{1}{2}p)}}\cdot\cos[\frac{1}{2}(p-1)\pi]\}\cdot|\Gamma\{\frac{1}{2}[i\Lambda + l_{(\frac{1}{2}p)} + \frac{1}{2}(p-1)]\}|^2}{\sinh(\pi\Lambda)\cdot|\Gamma\{\frac{1}{2}[i\Lambda + l_{(\frac{1}{2}p)} + \frac{1}{2}(p+1)]\}|^2}, $$

$$ {}_2K = \frac{\pi\{\cosh(\pi\Lambda) + (-1)^{l_{(\frac{1}{2}p)}}\cdot\cos[\frac{1}{2}(p-1)\pi]\}\cdot|\Gamma\{\frac{1}{2}[i\Lambda + l_{(\frac{1}{2}p)} + \frac{1}{2}(p+1)]\}|^2}{\sinh(\pi\Lambda)\cdot|\Gamma\{\frac{1}{2}(i\Lambda + l_{(\frac{1}{2}p)} + \frac{1}{2}(p-1)]\}|^2}, $$

and $\Lambda \in [0, \infty)$, $l_{(\frac{1}{2}p)}$ is non-negative integer except for the case $p = 2$ when $l_1 \equiv m_1$ and m_1 is integer. The eigenfunctions of the Laplace–Beltrami operator $\Delta(H_1^p)$ are harmonic functions orthogonal with respect to the measure $d\mu(\omega, \theta) = [\bar{g}(H_1^p)]^{\frac{1}{2}}d\omega\, d\theta$. Their form is

$$ {}_{1,2}Y^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}(\theta, \omega) = {}_{1,2}V^{\Lambda}_{l_{(p/s)}}(\theta)\cdot Y^{l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}(\omega), \qquad (3.4) $$

where $Y^{l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}(\omega)$ are eigenfunctions of $\Delta(S^{p-1})$ expressed in Eq. (2.7) and ${}_{1,2}V^{\Lambda}_{l_{(p/s)}}(\theta)$ is given in (3.3).

The construction of the carrier space of the representation of the group $SO_0(p, 1)$ is analogous to the previous case. Thus, the generalized Fourier transform with respect to the eigenfunction (3.4) of a function

$$ f(\theta, \omega) = P(x^1, \cdots, x^{p+1}) \exp\left[-\sum_{i=1}^{p+1}(x^i)^2\right], $$

where $P(x^1, \cdots, x^{p+1})$ is an arbitrary polynomial; x^i, expressed in biharmonic coordinates on H_1^p[Ref. 1, Eqs. (3.4)–(3.9)], has the form

$$ {}_{1,2}\chi^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}} = \langle {}_{1,2}Y^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}, f\rangle = \int_{H_1^p} \overline{{}_{1,2}Y^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}}(\theta, \omega)\cdot f(\theta, \omega)\cdot d\mu(\theta, \omega). \qquad (3.5) $$

All such Fourier transforms determine the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,1}$ of vectors

$$ \chi^{\Lambda,\pm} \equiv \left\{ {}_{\alpha}\chi^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}};\ \alpha + l_{(\frac{1}{2}p)} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix} \right\} $$

for which

$$ ||\chi^{\Lambda,\pm}||^2 = \sum_{\substack{\alpha, l_2, \cdots, l_{(p/s)}\\ m_1, \cdots, m_{[p/s]}}} |{}_{\alpha}\chi^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}|^2 < \infty, $$

and for which $\alpha + l_{(\frac{1}{2}p)}$ is, respectively, even or odd. The scalar product in the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,1}$ is defined as

$$ (\chi^{\Lambda,\pm}, \psi^{\Lambda,\pm}) = \sum \overline{{}_{\alpha}\chi^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}}\cdot {}_{\alpha}\psi^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}, \qquad \chi^{\Lambda,\pm}, \psi^{\Lambda,\pm} \in \mathfrak{H}^{\Lambda,\pm}_{p,1}, \qquad (3.6) $$

where the sum is taken over all integers $l_2, \cdots, l_{(\frac{1}{2}p)}$, $m_1, \cdots, m_{(\frac{1}{2}p)}$, and α and $\alpha + l_{(\frac{1}{2}p)}$ is even or odd, respectively. The Hilbert space $\mathfrak{H}^{\Lambda,+}_{p,1}(\mathfrak{H}^{\Lambda,-}_{p,1})$ is again the eigenspace of the parity operator with the eigenvalue $+1$ (-1). The structure of the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,1}$ has the form

$$ \mathfrak{H}^{\Lambda,\pm}_{p,1} = \sum_{l_{(p/s)}} \oplus\, \mathfrak{H}^{\Lambda,\pm}_{p,1,l_{(p/s)}}, \qquad (3.7) $$

where $\mathfrak{H}^{\Lambda,\pm}_{p,1,l_{(p/s)}}$ are the finite-dimensional vector spaces containing ${}_{\alpha}\chi^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}$ with fixed $l_{(\frac{1}{2}p)}$.

The continuous series $C^{\Delta,\pm}_{p,1}$ of the representations of the group $SO_0(p, 1)$ on the Hilbert space $\mathfrak{H}^{\Delta,\pm}_{p,1}$ is induced by the representation of the corresponding Lie algebra on the vectors $\chi^{\Delta,\pm}$:

$$ \begin{aligned} L_{ij}\chi^{\Delta,\pm} &\equiv \left\{ \langle {}_{\alpha}Y^{\Delta, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}, L_{ij}f\rangle;\ \alpha + l_{(\frac{1}{2}p)} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix} \right\}, \\ B_{st}\chi^{\Delta,\pm} &\equiv \left\{ \langle {}_{\alpha}Y^{\Delta, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}, B_{st}f\rangle;\ \alpha + l_{(\frac{1}{2}p)} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix} \right\}, \end{aligned} \qquad (3.8) $$

where $f(\theta, \omega)$ is a function as in (3.5) and L_{ij}, B_{st} are, respectively, the representations of the generators of the compact and noncompact one-parameter subgroups of the group $SO_0(p, 1)$.

The Laplace–Beltrami operator $\Delta(H_p^1)$ on the hyperboloid H_p^1 has the form (See Ref. 1, Sec. 5)

$$ \Delta(H_p^1) = \frac{-1}{\sinh^{p-1}\theta}\frac{\partial}{\partial\theta}\sinh^{p-1}\theta\,\frac{\partial}{\partial\theta} - \frac{\Delta(S^{p-1})}{\sinh^2\theta}, \qquad \theta \in [0, \infty). \qquad (3.9) $$

Since $\Delta(H_p^1)$ has again the continuous spectrum of the form $-\Lambda^2 - [\frac{1}{2}(p-1)]^2$, and eigenvalues of $\Delta(S^{p-1})$ are $-l_{(\frac{1}{2}p)}\cdot(l_{(\frac{1}{2}p)} + p - 2)$, the eigenfunctions of $\Delta(H_p^1)$ can be expressed as

$$ Y^{\Lambda, l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}(\theta, \omega) = V^{\Lambda}_{l_{(p/s)}}(\theta)\cdot Y^{l_2, \cdots, l_{(p/s)}}_{m_1, \cdots, m_{[p/s]}}(\omega), \qquad (3.10) $$

where now

$$V^{\Lambda}_{l_{[p/2]}}(\theta) = (K^{-\frac{1}{2}}) \tanh^{|l_{[p/2]}|}\theta \cdot \cosh^{[i\Lambda-\frac{1}{2}(p-1)]}\theta$$
$$\cdot {}_2F_1\{\tfrac{1}{2}[|l_{[\frac{1}{2}p]}| - i\Lambda + \tfrac{1}{2}(p-1)], \tfrac{1}{2}[|l_{[\frac{1}{2}p]}| - i\Lambda + \tfrac{1}{2}(p+1)]; |l_{[\frac{1}{2}p]}| + \tfrac{1}{2}p; \tanh^2\theta\}, \quad (3.11)$$
$$K = \left| \frac{(2\pi)^{\frac{1}{2}}\Gamma(i\Lambda)\cdot\Gamma(|l_{[\frac{1}{2}p]}| + \frac{1}{2}p)}{\Gamma\{\frac{1}{2}[i\Lambda + \frac{1}{2}(p-1) + |l_{[\frac{1}{2}p]}|]\}\cdot\Gamma\{\frac{1}{2}[i\Lambda + \frac{1}{2}(p+1) + |l_{[\frac{1}{2}p]}|]\}} \right|^2,$$

$Y^{l_2,\cdots,l_{[p/2]}}_{m_1,\cdots,m_{[p/2]}}(\omega)$ is given in (2.6), and $\Lambda \in [0, \infty)$, $l_{[\frac{1}{2}p]}$ being a non-negative integer except for the case $p = 2$ when, by definition, $l_1 \equiv m_1$, and m_1 is integer.

The series $C^{\Lambda}_{1,p}$ of the continuous most degenerate irreducible unitary representations of the group $SO_0(p, 1)$ on the Hilbert space $\mathfrak{H}^{\Lambda}_{1,p}$ are easily obtained from those constructed in Sec. 2 by omitting dependence on all twiddle variables.

As proved in Sec. 5, we construct two series, $C^{\Lambda,\pm}_{p,1}$ and $C^{\Lambda}_{1,p}$, of irreducible unitary representations related to the hyperboloids H^p_1 and H^1_p, respectively.

4. CONTINUOUS SERIES OF MOST DEGENERATE REPRESENTATIONS OF $SO(p, q)$ GROUPS RELATED TO THE CONE

In this section, we derive the continuous series of most degenerate representations of an arbitrary $SO_0(p, q)$ group on the Hilbert space $\mathfrak{H}(X)$ of functions the domain X of which is the following homogeneous space of rank one under the action of $SO_0(p, q)$[6]:

$$X = SO_0(p, q)/T^{p+q-2} \textcircled{S} SO(p-1, q-1). \quad (4.1)$$

Here, T^{p+q-2} is the group of translations in the $(p + q - 2)$-dimensional Minkowski space $M^{p-1,q-1}$. The homogeneous space X can be realized by the cone C^p_q defined as

$$(x^1)^2 + \cdots + (x^p)^2 - (x^{p+1})^2 - \cdots - (x^{p+q})^2 = 0.$$

Following the general procedure described in Ref. 1, Sec. 1, we have to introduce first the biharmonic coordinate system on the cone C^p_q. Then we would try to find the metric tensor $g_{\alpha\beta}(C^p_q)$ on the cone C^p_q and construct the Laplace–Beltrami operator. However, it turns out that the metric tensor is singular on the cone, and hence the Laplace–Beltrami operator does not exist. Therefore, we have to construct the second-order Casimir operator $C^{(2)}$ directly from the algebra. Calculating the Cartan metric tensor from the Lie algebra $\mathfrak{R}$ [see Ref. 1 (6.1)] of the group $SO_0(p, q)$, we easily find that the Casimir operator has the form

$$C^{(2)} = \frac{-Q_2}{2(p+q-2)}, \qquad Q_2 = \sum_{i<j} L^2_{ij} - \sum_{s<t} B^2_{st}. \quad (4.2)$$

[6] The authors are grateful to Dr. O. Nachtman for a valuable discussion on the group of motion on the cone.

The biharmonic coordinate system on the cone is introduced as

$$x^k = r\cdot x'^k, \qquad k = 1, 2, \cdots, p,$$
$$x^l = r\cdot \bar{x}^l, \qquad l = p+1, p+2, \cdots, p+q, \quad (4.3)$$

where x'^k, $\bar{x}^l$ have the same structure as in formulas (3.5)–(3.7) of Ref. 1. We represent now the Lie algebra $\mathfrak{R}$ of the group $SO_0(p, q)$ with respect to the parametrization (4.3) by the operators L_{ij} and B_{st}:

$$L_{ij} = \left(x^i \frac{\partial\varphi^1}{\partial x^j} - x^j \frac{\partial\varphi^1}{\partial x^i}\right)\frac{\partial}{\partial\varphi^1} + \cdots$$
$$+ \left(x^i \frac{\partial\vartheta^{(\frac{1}{2}p)}}{\partial x^j} - x^j \frac{\partial\vartheta^{(\frac{1}{2}p)}}{\partial x^i}\right)\frac{\partial}{\partial\vartheta^{(\frac{1}{2}p)}}, \quad (4.4)$$

where $i, j = 1, 2, \cdots, p$. The analogous expressions hold for L_{ij}, $i, j = p+1, p+2, \cdots, p+q$.

$$B_{st} = x'^s\bar{x}^t r\frac{\partial}{\partial r}$$
$$+ x^s\frac{\partial\tilde{\varphi}^1}{\partial x^t}\frac{\partial}{\partial\varphi^1} + \cdots + x^s\frac{\partial\tilde{\vartheta}^{(\frac{1}{2}q)}}{\partial x^t}\frac{\partial}{\partial\tilde{\vartheta}^{(\frac{1}{2}q)}}$$
$$+ x^t\frac{\partial\varphi^1}{\partial x^s}\frac{\partial}{\partial\varphi^1} + \cdots + x^t\frac{\partial\vartheta^{(\frac{1}{2}p)}}{\partial x^s}\frac{\partial}{\partial\vartheta^{(\frac{1}{2}p)}}, \quad (4.5)$$

where $s = 1, 2, \cdots, p$; $t = p+1, p+2, \cdots, p+q$; and x'^s, $\bar{x}^t$ are defined in (4.3). The operators L_{ij} are r-independent and have the same form as in previous cases. The corresponding representation of the invariant operator Q_2 has the form

$$Q_2 = -\left[r^2\frac{\partial^2}{\partial r^2} + (p+q-1)r\frac{\partial}{\partial r}\right]. \quad (4.6)$$

The left-invariant measure on the cone is given by

$$d\mu(r, \omega, \tilde{\omega}) = r^{p+q-3}\, dr\cdot d\mu(\omega)\cdot d\mu(\tilde{\omega}), \quad (4.7)$$

where $d\mu(\omega)$ is defined by (2.8).

From here we pursue again our general procedure, i.e., we first solve the eigenvalue problem for the invariant operator Q_2 (4.6). The eigenfunctions of the operator Q_2 are the solutions of the differential equation

$$\left\{r^2\frac{d^2}{dr^2} + (p+q-1)\cdot r\cdot\frac{d}{dr} + \Lambda^2 + [\tfrac{1}{2}(p+q-2)]^2\right\}\chi(r) = 0, \quad (4.8)$$

where we put the ansatz $\Lambda^2 + [\frac{1}{2}(p+q-2)]^2$, $\Lambda \in (-\infty, \infty)$ for the spectrum of the operator Q_2. It will be shown in Part III of our work that we do not lose any part of the spectrum of Q_2 in this way. Hence, the general form of the eigenfunction of the operator Q_2 has the form $r^\alpha \varphi(\omega, \tilde{\omega})$, where $\alpha = -\frac{1}{2}(p+q-2) + i\Lambda$, and $\varphi(\omega, \tilde{\omega})$ is a function which can be chosen in an arbitrary way. It is convenient for our purpose to restrict eigenfunctions of the invariant operator Q_2 to be the harmonic functions of the following form:

$$Y^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}(r, \omega, \tilde{\omega}) = (2\pi)^{-\frac{1}{2}} r^\alpha \cdot Y^{l_2, \cdots, l_{[p/2]}}_{m_1, \cdots, m_{[p/2]}}(\omega) \cdot Y^{l_2, \cdots, l_{[q/2]}}_{\tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}(\tilde{\omega}), \quad (4.9)$$

where $\alpha = -\frac{1}{2}(p+q-2) + i\Lambda$, and the functions $Y^{l_2, \cdots, l_{[p/2]}}_{m_1, \cdots, m_{[p/2]}}(\omega)$ are defined in (2.7) for $p > 1$ (for $p = 1$ this function equals one).

The generalized Fourier transform with respect to the eigenfunction (4.9) of a function

$$f(r, \omega, \tilde{\omega}) = P(x^1, \cdots, x^{p+q}) \exp\left[-\sum_{i=1}^{p+q} (x^i)^2\right],$$

where $P(x_1, \cdots, x^{p+q})$, an arbitrary polynomial in x^i [x^i are defined in our biharmonic coordinates (4.3) on the cone C^p_q], is given by

$$\chi^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}} = \langle Y^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}, f\rangle = \int_{C^p_q} \overline{Y^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}}(r, \omega, \tilde{\omega}) \cdot f(r, \omega, \tilde{\omega}) \cdot d\mu(r, \omega, \tilde{\omega}). \quad (4.10)$$

For $p \geq q > 1$, as in previous cases, all such generalized Fourier transforms determine the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,q}$, of the vectors

$$\chi^{\Lambda,\pm} = \{\chi^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[p/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}\},$$

for which

$$||\chi^{\Lambda,\pm}||^2 = \sum |\chi^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}|^2 < \infty,$$

where $l_{[\frac{1}{2}p]} + l_{[\frac{1}{2}q]}$ is even or odd, respectively. The Hilbert space $\mathfrak{H}^{\Lambda;\pm}_{p,q}$ can be decomposed in the form

$$\mathfrak{H}^{\Lambda,\pm}_{p,q} = \sum_{\substack{l_{[p/2]}, l_{[q/2]}=0 \\ l_{[p/2]}+l_{[q/2]}=\begin{cases}\text{even}\\ \text{odd}\end{cases}}}^{\infty} \oplus \mathfrak{H}^{\Lambda,\pm}_{p,q;l_{[p/2]}, l_{[q/2]}}, \quad (4.11)$$

where $\mathfrak{H}^{\Lambda,\pm}_{p,q;l_{[p/2]}, l_{[q/2]}}$ are the finite-dimensional vector spaces determined by $\chi^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}$ with fixed $l_{[\frac{1}{2}p]}$ and $l_{[\frac{1}{2}q]}$.

The scalar product in the Hilbert space $\mathfrak{H}^{\Lambda;\pm}_{p,q}$ is defined by

$$(\chi^{\Lambda,\pm}, \psi^{\Lambda,\pm}) = \sum \overline{\chi^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}} \cdot \psi^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}, \quad (4.12)$$

where the sum is taken over all integers $l_2, \cdots l_{[\frac{1}{2}p]}$, $l_2, \cdots l_{[\frac{1}{2}q]}$, $m_1, \cdots, m_{[\frac{1}{2}p]}$, $\tilde{m}_1, \cdots, \tilde{m}_{[\frac{1}{2}q]}$, with $l_{[\frac{1}{2}p]} + l_{[\frac{1}{2}q]}$ even or odd, respectively.

The representation of the series $C^{\Lambda;\pm}_{p,q}$ on the Hilbert space $\mathfrak{H}^{\Lambda;\pm}_{p,q}$ is defined by the representation of the algebra on the vectors $\chi^{\Lambda,\pm}$

$$L_{ij}\chi^{\Lambda,\pm} \equiv \left\{\langle Y^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}, L_{ij} f\rangle;\ l_{[\frac{1}{2}p]} + l_{[\frac{1}{2}q]} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix}\right\},$$
$$B_{st}\chi^{\Lambda,\pm} \equiv \left\{\langle Y^{\Lambda, l_2, \cdots, l_{[p/2]}, l_2, \cdots, l_{[q/2]}}_{m_1, \cdots, m_{[p/2]}, \tilde{m}_1, \cdots, \tilde{m}_{[q/2]}}, B_{st} f\rangle;\ l_{[\frac{1}{2}p]} + l_{[\frac{1}{2}q]} = \begin{pmatrix}\text{even}\\ \text{odd}\end{pmatrix}\right\}, \quad (4.13)$$

with $f(r, \omega, \tilde{\omega})$ as in (4.10).

For $q = 1$, the corresponding representations are constructed in a way completely analogous to that used in Sec. 3. Let us keep in mind that in this case we have restricted ourselves only to the proper $SO(p,1)$ group (upper sheet of the cone C^p_1).

5. IRREDUCIBILITY

A. The Representations Related to the Hyperboloid for $p \geqq q > 2$

The maximal compact subalgebra (consisting the generators L_{ij}) with any generator B_{st} of a one-parameter noncompact subgroup generates the whole algebra. Therefore, the problem of the irreducibility of the representation of the algebra can be solved by considering only the set of generators L_{ij} together with one of the generators B_{st}.

For the proof, we take the generator $B_{p;p+q}$ and represent it by the definition (2.12) on the vectors (2.10) which determine the Hilbert space $\mathfrak{H}^{\Lambda;+}_{p,q} \oplus \mathfrak{H}^{\Lambda;-}_{p,q}$. Calculating the explicit form of the operator $B_{p,p+q}$ from the expression (6.3) of Ref. 1, we easily find that it can map an arbitrary vector $f_{l_{[p/2]}, l_{[q/2]}} \in \mathfrak{H}^{\Lambda,\pm}_{p,q;l',l_{[p/2]}, l_{[q/2]}}$ only to such vectors $f^B_{l_{[p/2]}', l_{[q/2]}}$, for which

$$l'_{[\frac{1}{2}p]} = l_{[\frac{1}{2}p]} \pm 1 \quad \text{and} \quad l'_{[\frac{1}{2}q]} = l_{[\frac{1}{2}q]} \pm 1.$$

Hence, we have at least two invariant subspaces

$\mathfrak{H}^{\Lambda,\pm}_{p,q}$ with respect to the representation of the algebra for the same eigenvalue of the Casimir operator. In what follows, we show that any of these is irreducible, i.e., they do not contain the invariant subspaces with respect to the representation of the algebra.

To show that there is no invariant subspace in the vector space determined by the vectors (2.9) with respect to the algebra R, it is sufficient to find vectors

$$\chi^{\Lambda,l_{(p/2)},\tilde{l}_{(q/2)}}_{m_{(p/2)},\tilde{m}_{(q/2)}} \in \mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(p/2)},\tilde{l}_{(q/2)}},$$

$$l_{(\frac{1}{2}p)},\ \tilde{l}_{(\frac{1}{2}q)} = 0, 1, 2, \cdots,$$

such that every $B_{p,p+q}\chi^{\Lambda,l_{(p/2)},\tilde{l}_{(q/2)}}_{m_{(p/2)},\tilde{m}_{(q/2)}}$ has nonvanishing components in four neighbouring subspaces $\mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(p/2)}\pm1,\tilde{l}_{(q/2)}\pm1}$. Choosing

$$\chi^{\Lambda,l_{(p/2)},\tilde{l}_{(q/2)}}_{m_{(p/2)},\tilde{m}_{(q/2)}} = \chi^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}},$$

where

$$m_2 = \tilde{m}_2 = \cdots = m_{(\frac{1}{2}p)} = \tilde{m}_{(\frac{1}{2}q)} = 0$$

and

$$|m_1|,\ |\tilde{m}_1|,\ l_2,\ \tilde{l}_2,\ \cdots,\ l_{(\frac{1}{2}p)-1},\ \tilde{l}_{(\frac{1}{2}q)-1}$$

have the minimal possible values, we calculate from (6.3) of Ref. 1 and (2.12):

$$\begin{aligned}
B_{p,p+q}\chi^{\Lambda,l,\tilde{l}}_{0,0} = {} & -\cdot|\tilde{l} + l + \tfrac{1}{2}(p+q-2) + i\Lambda|\cdot A_+(l)\cdot A_+(\tilde{l})\cdot[N(l+1,\tilde{l}+1)/N(l,\tilde{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda,l+1,\tilde{l}+1} \\
& + |\tilde{l} - l + \tfrac{1}{2}(p+q-2) + i\Lambda - p|\cdot A_+(l)\cdot A_-(\tilde{l})\cdot[N(l+1,\tilde{l}-1)/N(l,\tilde{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda,l+1,\tilde{l}-1} \\
& - |\tilde{l} - l + \tfrac{1}{2}(p+q-2) + i\Lambda - p + 2|\cdot A_-(l)\cdot A_+(\tilde{l})\cdot[N(l-1,\tilde{l}+1)/N(l,\tilde{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda,l-1,\tilde{l}+1} \\
& + |\tilde{l} + l + \tfrac{1}{2}(p+q-2) + i\Lambda - 2|\cdot A_-(l)\cdot A_-(\tilde{l})\cdot[N(l-1,\tilde{l}-1)/N(l,\tilde{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda,l-1,\tilde{l}-1},
\end{aligned} \tag{5.1}$$

where $l \equiv l_{(\frac{1}{2}p)}$, $\tilde{l} \equiv \tilde{l}_{(\frac{1}{2}q)}$, $A_\pm(l)$ are defined by (6.7) or (6.8) of Ref. 1, $N(l,\tilde{l}) \equiv N(l_{(\frac{1}{2}p)}\tilde{l}_{(\frac{1}{2}q)}) = N_{(\frac{1}{2}p)}\cdot N_{(\frac{1}{2}q)}$ [$N_{(\frac{1}{2}p)}$ and $N_{(\frac{1}{2}q)}$ are defined by the expressions (3.21) of Ref. 1], and

$$\phi^{\Lambda,l_{(p/2)},\tilde{l}_{(q/2)}} = \begin{cases} \chi^{\Lambda,l_{(p+1)/2},\tilde{l}_{(q+1)/2}}, & \text{for odd } p \text{ and } q, \\ \frac{1}{2}(i^{-1})[\chi^{\Lambda,l_{p/2},\tilde{l}_{(q+1)/2}}_{1,0} - \chi^{\Lambda,l_{p/2},\tilde{l}_{(q+1)/2}}_{-1,0}], & \text{for even } p \text{ and odd } q, \\ -\frac{1}{4}\chi^{\Lambda,l_{p/2},\tilde{l}_{q/2}}_{1,1} + \frac{1}{4}\chi^{\Lambda,l_{p/2},\tilde{l}_{q/2}}_{1,-1} + \frac{1}{4}\chi^{\Lambda,l_{p/2},\tilde{l}_{q/2}}_{-1,1} - \frac{1}{4}\chi^{\Lambda,l_{p/2},\tilde{l}_{q/2}}_{-1,-1}, & \text{for even } p \text{ and } q. \end{cases} \tag{5.2}$$

The only coefficients in the expression (5.1) which can vanish for non-negative integers $l_{(\frac{1}{2}p)}$, $\tilde{l}_{(\frac{1}{2}q)}$ are $A_-(l_{(\frac{1}{2}p)})$ and $A_-(\tilde{l}_{(\frac{1}{2}q)})$. If $p(q)$ is odd, then the coefficient $A_-(l_{\frac{1}{2}(p+1)})[A_-(\tilde{l}_{(\frac{1}{2}q+1)})]$ is zero for $l_{\frac{1}{2}(p+1)} = 0[\tilde{l}_{\frac{1}{2}(q+1)} = 0]$ in accordance with the fact that we have the representation. If $p(q)$ is even, $A_-(l_{\frac{1}{2}p}) = 0$ $[A_-(\tilde{l}_{\frac{1}{2}q}) = 0]$ for $l_{\frac{1}{2}p} = 1$ $[\tilde{l}_{\frac{1}{2}q} = 1]$. This does not mean that the representation is reducible, because the mapping

$$\mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(p/2)},\tilde{l}_{(q/2)}} \xrightarrow{B} \mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(p/2)}+1,\tilde{l}_{(q/2)}+1}$$

always exists and the operator $B_{p,p+q}$ is skew-symmetric on the vectorspace determined by the vectors (2.10).

Thus, we proved that the second-order Casimir operator is not sufficient to specify the irreducible representations. The complete specification of the irreducible representations is achieved by the commutative invariant algebra, generated by the second-order Casimir operator and an operator P, the eigenvalues of which are ±1. Let us show that the operator P is the representation of the parity operator $px^k = -x^k$, $k = 1, 2, \cdots, p+q$. From the explicit form of the harmonic functions, we easily calculate the representation of the parity operator on any harmonic function

$$\begin{aligned}
& PY^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}(\Omega) \\
& = Y^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}(p\Omega) \\
& = (-1)^{l_{(p/2)}+\tilde{l}_{(q/2)}}\cdot Y^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}.
\end{aligned} \tag{5.3}$$

Then the representation of the parity operator on the space $\mathfrak{H}^{\Lambda,\pm}_{p,q}$ is defined by the expression

$$\begin{aligned}
& P\chi^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}} \\
& = \langle Y^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}, Pf\rangle \\
& = (-1)^{l_{(p/2)}+\tilde{l}_{(q/2)}}\cdot\chi^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}}_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}.
\end{aligned} \tag{5.4}$$

It follows that the vectors $\chi^{\Lambda,\pm}$ of the space $\mathfrak{H}^{\Lambda,\pm}_{p,q}$ are common eigenvectors of the operator $\Delta(H^p_q)$ with the eigenvalue $\Lambda^2 + \frac{1}{2}(p+q-2)]^2$ and the operator P with the eigenvalue ±1, respectively.

Completely analogous proof holds for the series $C^{\Lambda,\pm}_{q,p}$ of the representations related to the hyperboloid H^q_p.

B. The Representations Related to the Hyperboloid for $q = 2$

The proof of the irreducibility of the representations of the continuous series $C^{\Lambda,\pm}_{p,2}$ and $C^{\Lambda,\pm}_{2,p}$ does not differ from the previous one as the vector

$$B_{p,p+2}\chi_{0,\tilde{m}_1}^{\Lambda,l_{(p/2)}},$$

where

$$\chi_{0,\tilde{m}_1}^{\Lambda,l_{(p/2)}} = \chi_{m_1,\cdots,m_{(p/2)},\tilde{m}_1}^{\Lambda,l_2,\cdots,l_{(p/2)}}$$

with $|m_1|, l_2, \cdots, l_{(\frac{1}{2}p)-1}$ having minimal possible values, has essentially the same structure as in the previous case.

C. The Representations Related to the Hyperboloid for $q = 1$

The proof of the irreducibility of the representations of the continuous series $C^{\Lambda}_{1,p}$ can be obtained by specifying the one which is derived in A above. Omitting completely the indices $\tilde{l}_2, \cdots, \tilde{l}_{(\frac{1}{2}q)}$, $\tilde{m}_1, \cdots \tilde{m}_{(\frac{1}{2}q)}$, in A, we obtain

$$\begin{aligned} B_{1,1+p}\chi_0^{\Lambda,l_{(p/2)}} = &- |l_{(\frac{1}{2}p)} + \tfrac{1}{2}(p-1) + i\Lambda|\cdot A_+(l_{(\frac{1}{2}p)})[N(l_{(\frac{1}{2}p)}+1)/N(l_{(\frac{1}{2}p)})]^{\frac{1}{2}}\phi^{\Lambda,l_{(p/2)}+1} \\ &- |l_{(\frac{1}{2}p)} + \tfrac{1}{2}(p+1) + i\Lambda|\cdot A_-(l_{(\frac{1}{2}p)})[N(l_{(\frac{1}{2}p)}-1)/N(l_{(\frac{1}{2}p)})]^{\frac{1}{2}}\phi^{\Lambda,l_{(p/2)}-1}, \end{aligned} \tag{5.5}$$

where $A_\pm(l_{(p/2)})$ are defined as before, $N(l_{(p/2)}) = N_{(p/2)}$ is determined in (3.21) of Ref. 1, and

$$\phi^{\Lambda,l_{(p/2)}} = \begin{cases} \chi_0^{\Lambda,l_{(p+1)/2}}, & \text{for odd } p, \\ \tfrac{1}{2}(i^{-1})(\chi_1^{\Lambda,l_{p/2}} - \chi_{-1}^{\Lambda,l_{p/2}}), & \text{for even } p. \end{cases} \tag{5.6}$$

For the representations of the series continuous $C^{\Lambda,\pm}_{p,1}$ we choose, as in the case of discrete representations,[1] the vector

$${}_{1,2}\chi_{m_{(p/2)}}^{\Lambda,l_{(p/2)}} = {}_{1,2}\chi_{m_1,\cdots,m_{(p/2)}}^{\Lambda,l_2,\cdots,l_{(p/2)}},$$

where $m_2 = \cdots = m_{(\frac{1}{2}p)} = 0$ and $|m_1|, l_2, \cdots, l_{(\frac{1}{2}p)-1}$ have the minimal possible values.

$$\begin{aligned} B_{p,p+1}\cdot{}_1\chi_0^{\Lambda,l} = &- |\tfrac{1}{2}(p-1) + l + i\Lambda|\cdot A_+(l)\cdot[N(l+1)/N(l)]^{\frac{1}{2}}{}_2\phi^{\Lambda,l+1} \\ &- |\tfrac{1}{2}(p-3) + l + i\Lambda|\cdot A_-(l)\cdot[N(l-1)/N(l)]^{\frac{1}{2}}{}_2\phi^{\Lambda,l-1}, \\ B_{p,p+1}\cdot{}_2\chi_0^{\Lambda,l} = &|\tfrac{1}{2}(p-1) + l + i\Lambda|\cdot A_+(l)\cdot[N(l+1)/N(l)]^{\frac{1}{2}}{}_1\phi^{\Lambda,l+1} \\ &+ |\tfrac{1}{2}(p-3) + l + i\Lambda|\cdot A_-(l)\cdot[N(l-1)/N(l)]^{\frac{1}{2}}{}_1\phi^{\Lambda,l-1}, \end{aligned} \tag{5.7}$$

where $A_\pm(l)$ are defined as before, $N(l_{(\frac{1}{2}p)}) = N_{(\frac{1}{2}p)}$ is defined in (3.21) of Ref. 1, and

$${}_{1,2}\phi^{\Lambda,l_{(p/2)}} = \begin{cases} {}_{1,2}\chi_0^{\Lambda,l_{(p+1)/2}}, & \text{for odd } p \\ \tfrac{1}{2}(i^{-1})({}_{1,2}\chi_1^{\Lambda,l_{p/2}} - {}_{1,2}\chi_{-1}^{\Lambda,l_{p/2}}), & \text{for even } p. \end{cases} \tag{5.8}$$

Analyzing the coefficients in the expressions (5.5) and (5.7), one can verify again that there is no invariant subspace, i.e., the representations of the series $C^{\Lambda,\pm}_{p,1}$ and $C^{\Lambda}_{1,p}$ are irreducible. Hence, the irreducible representations of the series $C^{\Lambda,\pm}_{p,1}$ on the Hilbert space $\mathfrak{H}^{\Lambda,\pm}_{p,1}$ are again characterized by the eigenvalue of both the Casimir operator $\Delta(H_1^p)$ and the parity operator P.

D. The Representations Related to the Cone

By the same argument as in the previous section, we first establish the existence of at least two invariant subspaces $\mathfrak{H}^{\Lambda,+}_{p,q}$ and $\mathfrak{H}^{\Lambda,-}_{p,q}$ with respect to the representation of the group for the fixed value of the Casimir operator Q_2. Then we prove their irreducibility as before (for instance, for $q > 2$). As we have mentioned, the representations L_{ij} (4.4) of the generators of the compact one-parameter subgroups have the same form on $\mathfrak{H}^{\Lambda,\pm}_{p,q}$ as in Sec. 2. Moreover, since the operators L_{ij} are reduced by the subspaces $\mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(p/2)},l_{(q/2)}}$ exactly to the same operators as in Sec. 2, the subspaces $\mathfrak{H}^{\Lambda,\pm}_{p,q;l_{(p/2)},l_{(q/2)}}$ are irreducible with respect to the maximal compact subgroup. Now, we proceed as before, i.e., we consider the operator $B_{p,p+q}$ and the vector

$$\chi_{m_{(p/2)},\tilde{m}_{(q/2)}}^{\Lambda,l_{(p/2)},\tilde{l}_{(q/2)}} = \chi_{m_1,\cdots,m_{(p/2)},\tilde{m}_1,\cdots,\tilde{m}_{(q/2)}}^{\Lambda,l_2,\cdots,l_{(p/2)},\tilde{l}_2,\cdots,\tilde{l}_{(q/2)}},$$

where $m_2 = \cdots = m_{(\frac{1}{2}p)} = \tilde{m}_2 = \cdots = \tilde{m}_{(\frac{1}{2}q)} = 0$ and $|m_1|, |\tilde{m}_1|, l_2, \tilde{l}_2, \cdots, l_{(\frac{1}{2}p)-1}, \tilde{l}_{(\frac{1}{2}q)-1}$ have the minimal possible values. Using the definitions (4.5) and (4.13), we easily compute the following expression:

$$
\begin{aligned}
B_{p,p+q}\cdot\chi_{0,0}^{\Lambda;l,\bar{l}} = {} & +[i\Lambda - l - \bar{l} - \tfrac{1}{2}(p+q-2)]\cdot A_+(l)\cdot A_+(\bar{l})\cdot[N(l+1,\bar{l}+1)/N(l,\bar{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda;l+1,\bar{l}+1} \\
& +[i\Lambda - l + \bar{l} - \tfrac{1}{2}(p-q+2)]\cdot A_+(l)\cdot A_-(\bar{l})\cdot[N(l+1,\bar{l}-1)/N(l,\bar{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda;l+1,\bar{l}-1} \\
& +[i\Lambda + l - \bar{l} + \tfrac{1}{2}(p-q-2)]\cdot A_-(l)\cdot A_+(\bar{l})\cdot[N(l-1,\bar{l}+1)/N(l,\bar{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda;l-1,\bar{l}+1} \\
& +[i\Lambda + l + \bar{l} + \tfrac{1}{2}(p+q-6)]\cdot A_-(l)\cdot A_-(\bar{l})\cdot[N(l-1,\bar{l}-1)/N(l,\bar{l})]^{\frac{1}{2}}\cdot\phi^{\Lambda;l-1,\bar{l}-1}, \qquad (5.9)
\end{aligned}
$$

where $A_\pm(l)$, $N(l, \bar{l})$ are defined as in the expression (5.1) and $\phi^{\Lambda;l,\bar{l}}$ is defined as in expression (5.2).

Using the same analysis as before, we can check that there are no invariant subspaces of the vector spaces determined by the vectors (4.10) with respect to the representation of the algebra.

The representations relate to the cone for $q = 1$ and $q = 2$. The corresponding proofs of irreducibility obtain as before.

The unitarity of the representations of the group $SO_0(p, q)$ on Hilbert spaces $\mathfrak{H}_{p,q}^{\Lambda;\pm}$ related to all three homogeneous spaces will be proved in Part III of our work.

6. SUMMARY

Three most degenerate principal continuous series of the irreducible unitary single-valued representations of an arbitrary noncompact rotation group $SO_0(p, q)$ have been constructed. These series are related to three homogeneous spaces of rank one under the action of $SO_0(p, q)$, i.e., to the hyperboloids H_q^p and H_p^q, and to the cone C_q^p.

Generally, the most degenerate continuous series of irreducible unitary representations of $SO_0(p, q)$ are characterized by two numbers, Λ and P. The former determines the eigenvalue of the second-order Casimir operator and the latter is the eigenvalue of the parity operator.

In particular cases the situation is as follows:

(i) $SO_0(p, q)$, $p \geq q > 1$—The constructed representations are determined by both Λ and P. Λ is real from the ranges $(0, \infty)$ and $(-\infty, \infty)$ for representations related respectively to the hyperboloids and to the cone; P has the value ± 1.

(ii) $SO_0(p, 1)$—Two series of representations of the Lorentz-type group related to the hyperboloid H_1^p and to the upper sheet of the cone C_1^p are also characterized by both Λ and P, whereas the representations of the series related to the upper sheet of the hyperboloid H_p^1 are characterized only by the number Λ. The ranges of Λ is again $(0, \infty)$ and $(-\infty, \infty)$ for representations related respectively to the hyperboloids and to the cone; P is equal to ± 1.

The harmonic functions of the derived three continuous series of representations have been explicitly constructed. They are labeled by numbers Λ, P, from the corresponding ranges mentioned above, and by a set of integers $l_2, \cdots, l_{[\frac{1}{2}p]}, m_1, \cdots, m_{[\frac{1}{2}p]}, \bar{l}_2, \cdots, \bar{l}_{[\frac{1}{2}q]}, \bar{m}_1, \cdots, \bar{m}_{[\frac{1}{2}q]}$, which determine the eigenvalues of the maximal set of compact commuting operators defined in (7.8) of Ref. 1.

ACKNOWLEDGMENTS

The authors are grateful to Professor Abdus Salam, Professor P. Budini, and the IAEA for the hospitality extended to them at the International Centre for Theoretical Physics, Trieste, Italy.

It is also a pleasure to thank Professor K. Maurin and Dr. J. Fischer for interesting discussions.

Bi-Orthogonality Relations for Solving Half-Space Transport Problems

N. J. McCormick* and I. Kuščer
Institute of Physics, University of Ljubljana, Ljubljana, Yugoslavia
(Received 29 November 1965)

The method of singular eigenfunction expansions is applied to time-independent, one-speed, half-space transport problems with anisotropic scattering. Adjoint eigenfunctions are constructed such that a set of half-range bi-orthogonality relations is valid. These relations lead to the expansion coefficients in a direct manner. The adjoint eigenfunctions are also used to express the half-space albedo operator which relates the emerging angular density to the ingoing one.

I. INTRODUCTION

In the mathematically equivalent theories of one-speed neutron transport and of monochromatic radiative transfer, several analytical methods have been developed for handling problems in plane geometry.[1,2] Most work in this field has so far been limited to the simplest scattering laws, in particular, to isotropic scattering. The primary reason for this limitation is that, for more complicated scattering laws, all such methods soon become lengthy and beset with complications.[3,4] Problems with highly anisotropic scattering are, however, of practical interest, especially in the optics of turbid media. The measured scattering function of natural fog[5] shows a forward to backward ratio of 40 : 1 and can be satisfactorily fitted only by a polynomial of at least eighth order.[6] The need for handling anisotropic scattering again arises with a simple model in slowing down theory, where a Mellin transform in the energy variable leads to an equation of the one-speed type.[7]

Any kind of anisotropic scattering can, of course, be handled numerically by the Wick–Chandrasekhar or the S_N methods of discrete ordinates. However, it might be useful to gain more insight into the properties of solutions by having some analytical method fully worked out for an arbitrary scattering law. This will be done with Case's method,[8] which utilizes an expansion of the angular density (intensity) in terms of the eigenfunctions of the homogeneous transport equation. Completeness properties of the set of eigenfunctions were first proved by Case for isotropic scattering, and by Mika[9] for anisotropic scattering.

If scattering is isotropic, application of Case's method to half-space problems can be simplified by the use of half-range orthogonality relations for the eigenfunctions.[10] So far this auxiliary technique has been generalized to linearly anisotropic scattering (where the scattering function is linear in the cosine of the scattering angle).[11] There, bi-orthogonality relations involving certain "adjoint" eigenfunctions had to be employed. In the present work,[12] this technique is generalized to the case where the scattering law is described by an arbitrary polynomial. Azimuthal dependence of the solutions will also be permitted.

Section II deals with the homogeneous transport equation and its eigenfunctions. Also included is a set of equations for Chandrasekhar's H-functions, which are basic for the application of singular eigenfunctions to half-space problems. Sections III and IV (the latter for a nonabsorbing medium) contain the derivation of the bi-orthogonality relations and of associated equations involving the eigenfunctions and their adjoints. This derivation provides some reduction of effort over what was previously done for isotropic[10] and linearly anisotropic scattering.[11] It turns out that the construction of the adjoints in the general case is somewhat tedious, but, after this is done, the solutions of standard half-space problems can be expressed in an exact and compact way by the adjoints and the

* National Science Foundation Postdoctoral Fellow.
[1] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).
[2] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).
[3] I. Kuščer, J. Math. & Phys. **34**, 256 (1956); **37**, 52 (1958).
[4] I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge University Press, London, 1960).
[5] E. D. Spencer, J. Opt. Soc. Am. **50**, 584 (1960).
[6] S. Pahor (unpublished).
[7] R. E. Marshak, Rev. Mod. Phys. **19**, 185 (1947); see also J. J. McInerney, Nucl. Sci. Eng. **22**, 215 (1965).
[8] K. M. Case, Ann. Phys. (N. Y.) **9**, 1 (1960); *Recent Developments in Neutron Transport Theory*, Michigan Memorial Phoenix Project Report, The University of Michigan (1961).
[9] J. R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).
[10] I. Kuščer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. (N. Y.) **30**, 411 (1964).
[11] N. J. McCormick and I. Kuščer, J. Math. Phys. **6**, 1939 (1965).
[12] Preliminary results reported at the Ankara International Summer School of Transport Theory (1965).

H-functions. The scheme to be followed in the computation of these two sets of functions is summarized in Sec. V.

II. EIGENFUNCTIONS

The transport equation (equation of transfer) has the form

$$\left(\mu \frac{\partial}{\partial x} + 1\right) I(x, \mu, \varphi) = \frac{1}{4\pi} \int_{-1}^{1} d\mu' \int_{0}^{2\pi} d\varphi' \, p(\cos\theta) I(x, \mu', \varphi'), \quad (1)$$

where μ is the cosine of the polar angle with respect to the positive x axis, and φ is the azimuthal angle. The scattering (phase) function $p(\cos\theta)$ is assumed to be a polynomial of order N,

$$p(\cos\theta) = \sum_{l=0}^{N} \varpi_l P_l(\cos\theta), \quad (2)$$

$$0 < \varpi_0 < 1, \quad |\varpi_l| < 2l + 1, \quad l = 1, 2, \cdots N. \quad (3)$$

We defer the case $\varpi_0 = 1$ until Sec. IV.

Application of the spherical harmonics addition theorem to Eq. (2) gives[1,13]

$$p(\cos\theta) = p^0(\mu, \mu') + 2 \sum_{m=1}^{N} p^m(\mu, \mu')(1 - \mu^2)^{\frac{1}{2}m} \times (1 - \mu'^2)^{\frac{1}{2}m} \cos m(\varphi - \varphi'), \quad (4)$$

where

$$p^m(\mu, \mu') = \sum_{l=m}^{N} c_l^m p_l^m(\mu) p_l^m(\mu'), \quad (5)$$

$$p_l^m(\mu) = (d^m/d\mu^m) P_l(\mu), \quad (6)$$

$$p_m^m(\mu) = \prod_{n=0}^{m-1} (2n + 1), \quad \text{(See Ref. 14)} \quad (7)$$

$$c_l^m = \varpi_l \frac{(l - m)!}{(l + m)!}. \quad (8)$$

Here and henceforth, $l = m, m + 1, \cdots N$.

We seek particular solutions of Eq. (1) of the form

$$I(x, \mu, \varphi) = e^{-x/\nu} \phi^m(\nu, \mu)(1 - \mu^2)^{\frac{1}{2}m} \begin{Bmatrix} \cos m\varphi \\ \sin m\varphi \end{Bmatrix}, \quad (9)$$

and obtain the following equations for the eigenfunctions $\phi^m(\nu, \mu)$:

$$(\nu - \mu)\phi^m(\nu, \mu) = \tfrac{1}{2}\nu \int_{-1}^{1} p^m(\mu, \mu')\phi^m(\nu, \mu') \, dm(\mu'), \quad (10)$$

$$m = 0, 1, \cdots N,$$

where, for the sake of brevity, the symbol

$$dm(\mu) \equiv (1 - \mu^2)^m \, d\mu \quad (11)$$

has been introduced. Let us normalize the eigenfunctions by

$$\int_{-1}^{1} \phi^m(\nu, \mu) \, dm(\mu) = 1. \quad (12)$$

For convenience in future notation, we suppress the superscript m with $\phi^m(\nu, \mu)$ and other functions. All relevant quantities in any of the equations must be understood as referring to the particular azimuthal component under consideration.

To the interval $-1 < \nu < 1$, there corresponds a continuous set of singular eigenfunctions

$$\phi(\nu, \mu) = \tfrac{1}{2}\nu g(\nu, \mu) \mathcal{P} \frac{1}{\nu - \mu} + \lambda(\nu)(1 - \nu^2)^{-m} \delta(\nu - \mu). \quad (13)$$

The $\mathcal{P}$ refers to Cauchy principal values of integrals of $\phi(\nu, \mu)$. Here we have used the definitions

$$\lambda(\nu) = 1 - \tfrac{1}{2}\nu \mathcal{P} \int_{-1}^{1} \frac{g(\nu, \mu) \, dm(\mu)}{\nu - \mu}, \quad (14)$$

$$g(\nu, \mu) = \sum_{l=m}^{N} c_l g_l(\nu) p_l(\mu). \quad (15)$$

The functions $g_l(\nu)$ are defined by the equation

$$g_l(\nu) = \int_{-1}^{1} \phi(\nu, \mu) p_l(\mu) \, dm(\mu), \quad (16)$$

and obey the recurrence relation[1,13]

$$(l + 1 - m)g_{l+1}(\nu) - h_l \nu g_l(\nu) + (l + m)g_{l-1}(\nu) = 0, \quad (17)$$

$$h_l = 2l + 1 - \varpi_l. \quad (18)$$

By virtue of (7) and (12),

$$g_m^m(\nu) = \prod_{n=0}^{m-1} (2n + 1), \quad (19)$$

so it follows from (17) that the $g_l(\nu)$ are polynomials of order $(l - m)$, alternatively even and odd.

In addition to the set of continuum eigenfunctions, there may also exist discrete modes of the form

$$\phi(\pm\nu_i, \mu) = \tfrac{1}{2}\nu_i \frac{g(\pm\nu_i, \mu)}{\nu_i \mp \mu}. \quad (20)$$

[13] T. W. Mullikin, Astrophys. J. 139, 379 (1964).

[14] In this and subsequent equations, substitute 1 and 0 for the symbols $\Pi^k{}_{l=n}$ and $\Sigma^k{}_{l=n}$, respectively, whenever $k < n$.

The discrete eigenvalues are determined as roots of

$$\Lambda(\pm\nu_j) = 0, \tag{21}$$

$$\Lambda(z) = 1 - \tfrac{1}{2}z\int_{-1}^{1}\frac{g(z,\mu)\,dm(\mu)}{z-\mu} = 1 - \tfrac{1}{2}z\int_{-1}^{1}\frac{g(\mu,\mu)\,dm(\mu)}{z-\mu}. \tag{22}$$

It can be proved that these eigenvalues are real, and we assume that they are outside the interval $(-1, 1)$. Let ν_j, $j = 1, 2, \cdots M$, denote the positive eigenvalues, ordered in decreasing magnitude. Their number is limited, as can be shown in the manner of Mika,[9] by

$$M \leq N - m + 1. \tag{23}$$

We exclude degeneracies in the ν_j's, since any such exceptional case requires special treatment.[15]

The boundary values of $\Lambda(z)$ on the cut $(-1, 1)$ can be expressed as

$$\Lambda^{\pm}(\nu) \equiv \lim_{\epsilon\to 0^+}\Lambda(\nu \pm i\epsilon) = \lambda(\nu) \pm \tfrac{1}{2}i\pi\nu g(\nu,\nu)(1-\nu^2)^m. \tag{24}$$

Also of importance is the expansion

$$\Lambda(z) = 1 - \sum_{n=0}^{\infty}\frac{\eta_{2n}}{z^{2n}}, \qquad |z| > 1, \tag{25}$$

where

$$\eta_{2n} = \frac{1}{2}\int_{-1}^{1}\mu^{2n}g(\mu,\mu)\,dm(\mu). \tag{26}$$

By induction it has been shown[13] that

$$\Lambda(\infty) = 1 - \eta_0 = \prod_{l=m}^{N}\frac{h_l}{2l+1}. \tag{27}$$

In applying Case's method of solution, we need the H-function of Chandrasekhar[1] for the "characteristic function" $\frac{1}{2}g(\mu,\mu)(1-\mu^2)^m$. A closed-form expression can be quoted,

$$H^{-1}(-z) = [\Lambda(\infty)]^{\frac{1}{2}}(1-z)^{-M}\prod_{j=1}^{M}(\nu_j - z) \times \exp\left[\frac{1}{2\pi i}\int_0^1 \ln\frac{\Lambda^+(\nu)}{\Lambda^-(\nu)}\frac{d\nu}{\nu - z}\right], \tag{28}$$

which is equivalent to an expression derived by Chandrasekhar.[1] A translation from this notation to that used by Case[8] and others[9-11] is effected by

$$H^{-1}(-z) = [\Lambda(\infty)]^{\frac{1}{2}}\prod_{j=1}^{M}(\nu_j - z)X(z), \tag{29}$$

$$\mu H(\mu) = 2\varpi_0^{-1}[\Lambda(\infty)]^{\frac{1}{2}}\prod_{j=1}^{M}(\nu_j - \mu)\gamma(\mu). \tag{30}$$

The notation with the H-function has some advantage over that involving the X or γ.

We also need the following equations[1,8]:

$$H^{-1}(z)H^{-1}(-z) = \Lambda(z), \tag{31}$$

$$H^{-1}(\nu_j)\left[\frac{dH^{-1}(-z)}{dz}\right]_{z=\nu_j} = \left[\frac{d\Lambda(z)}{dz}\right]_{z=\nu_j}, \tag{32}$$

$$\frac{1}{H(z)} = 1 - \tfrac{1}{2}z\int_0^1\frac{g(\mu,\mu)H(\mu)\,dm(\mu)}{z+\mu}, \tag{33}$$

$$-\frac{1}{2\pi i}\left[\frac{1}{H^+(-\mu)} - \frac{1}{H^-(-\mu)}\right] = \tfrac{1}{2}\mu H(\mu)g(\mu,\mu)(1-\mu^2)^m, \tag{+34}$$

$$\frac{1}{2}\left[\frac{1}{H^+(-\mu)} + \frac{1}{H^-(-\mu)}\right] = H(\mu)\lambda(\mu). \tag{+35}$$

The restriction of the variable μ to the interval $(0, 1)$ in (34), (35), and in some of the subsequent equations is indicated by $+$ with the equation numbers.

From (33), we obtain the expansion

$$\frac{1}{H(-z)} = 1 - \sum_{n=0}^{\infty}\frac{\beta_n}{z^n}, \qquad |z| > 1, \tag{36}$$

where

$$\beta_n = \frac{1}{2}\int_0^1\mu^n g(\mu,\mu)H(\mu)\,dm(\mu). \tag{37}$$

Using Eqs. (25), (31), and (36), we find a set of useful relations,[4]

$$\eta_{2n} - 2\beta_{2n} + \sum_{k=0}^{2n}(-1)^k\beta_k\beta_{2n-k} = 0. \tag{38}$$

The β_n can be, in an obvious way, expressed in terms of the moments of the H-function,

$$\alpha_n = \int_0^1\mu^n H(\mu)\,dm(\mu). \tag{39}$$

III. BI-ORTHOGONALITY RELATIONS

In half-space problems, we are led to expansions of the following form[8,9]:

$$I(x,\mu) = \sum_{j=1}^{M}a_j\phi(\nu_j,\mu)e^{-x/\nu_j} + \int_0^1 A(\nu)\phi(\nu,\mu)e^{-x/\nu}\,d\nu + f(x,\mu). \tag{40}$$

The term $f(x, \mu)$ embodies (1) any eigenmodes increasing with x, if they are permitted by the boundary condition at $x \to \infty$, and (2) a particular solu-

[15] R. Goldstein, Nucl. Sci. Eng. 18, 412 (1964).

tion of the inhomogeneous equation in the event that sources are present. The expansion coefficients a_i and $A(\nu)$ have to be determined from the boundary condition at $x = 0$, that is, from

$$\psi(\mu) = \sum_{i=1}^{M} a_i\phi(\nu_i, \mu) + \int_0^1 A(\nu)\phi(\nu, \mu)\, d\nu, \qquad (+41)$$

where

$$\psi(\mu) \equiv I(0, \mu) - f(0, \mu). \qquad (42)$$

The existence of such expansions is guaranteed by the half-range completeness theorem. A proof of this theorem has been given by Mika[9] for $m = 0$, and can be generalized to any m.

We wish to facilitate the determination of the coefficients a_i and $A(\nu)$ by constructing a set of half-range bi-orthogonality relations. To this end, an appropriate weight function must be found and an "adjoint" $\tilde{\phi}(\nu, \mu)$ specified for each $\phi(\nu, \mu)$.

Inspection of the derivations in previous work[10,11] leads to the conjecture that the weight function again will have an analogous form (except for the change in notation). Thus, the bi-orthogonality relations should read as

$$\int_0^1 \phi(\nu, \mu)\tilde{\phi}(\nu', \mu)\mu H(\mu)\, dm(\mu) = 0, \quad \nu \neq \nu'. \qquad (+43)$$

The plus sign with the equation number now refers to ν and ν' and indicates here, as well as in some further equations, that positive eigenvalues must be taken, $\nu, \nu' \in (0, 1)$ or $= +\nu_j$, $j = 1, 2, \cdots M$. Otherwise, the letter ν will be used to denote any eigenvalue, positive or negative.

It will later be necessary to carry out some transformations of the integral in (43), where we have to rely upon the identities (34) and (35). Our choice of the weight function essentially hinges upon these identities, which are a consequence of $H(z)$ being a solution of the appropriate Hilbert problem.

We are left with the task of constructing the function $\tilde{\phi}(\nu, \mu)$ such that Eq. (43) will hold. Let us first see what happens to the integral in that equation if ϕ is substituted for $\tilde{\phi}$. The derivation will be amenable to some useful generalizations if complex variables $z, z' \notin (0, 1)$, $z \neq z'$, are substituted instead of the real ν, ν'. That is, we wish to evaluate the integral

$$\int_0^1 \frac{g(z, \mu)}{z - \mu}\frac{g(z', \mu)}{z' - \mu}\mu H(\mu)\, dm(\mu) \equiv J(z, z'). \qquad (44)$$

Partial-fraction analysis immediately leads to

$$J(z, z') = \frac{1}{z - z'}\int_0^1 g(z, \mu)g(z', \mu) \times \left[\frac{1}{z' - \mu} - \frac{1}{z - \mu}\right]\mu H(\mu)\, dm(\mu). \qquad (45)$$

Now, in view of Eq. (15), J can be expressed in terms of the integrals

$$\bar{K}_l(z) \equiv \int_0^1 p_l(\mu)\frac{g(z, \mu)}{z - \mu}\mu H(\mu)\, dm(\mu). \qquad (46)$$

These functions can be reduced to simpler quantities by the following manipulation:

$$\bar{K}_l(z) = \int_0^1 p_l(\mu)\frac{g(z, \mu) - g(\mu, \mu)}{z - \mu}\mu H(\mu)\, dm(\mu) + \int_0^1 p_l(\mu)\frac{g(\mu, \mu)}{z - \mu}\mu H(\mu)\, dm(\mu). \qquad (47)$$

In the last term, we substitute the left-hand side of Eq. (34) and then change the integration path into a contour encircling the interval (0, 1). After the contour is blown up, we collect the residue at $\mu = z$. The integral over the large loop gives the principal part (the polynomial part), $2[p_l(z)/H(-z)]_{as}$, of the function $2p_l(z)/H(-z)$ appearing in the integrand. The final result is

$$\bar{K}_l(z) = K_l(z) - 2p_l(z)/H(-z), \qquad (48)$$

where $K_l(z)$ is the following polynomial of maximum order $N - m - 1$ if $l < N$, and $N - m$ if $l = N$:

$$K_l(z) = \int_0^1 p_l(\mu)\frac{g(z, \mu) - g(\mu, \mu)}{z - \mu}\mu H(\mu)\, dm(\mu) + 2\left[\frac{p_l(z)}{H(-z)}\right]_{as}. \qquad (49)$$

It is useful to know for later purpose that, if z is substituted by any of the positive eigenvalues, we have the simpler expression

$$K_l(\nu) = \frac{2}{\nu}\int_0^1 p_l(\mu)\phi(\nu, \mu)\mu H(\mu)\, dm(\mu). \qquad (+50)$$

For $\nu = \nu_j$, the proof of this equation is immediate from Eq. (46) because $H^{-1}(-\nu_j) = 0$, and the last term in (48) therefore vanishes. The continuum eigenvalues are included by allowing z in (46) to approach the interval (0, 1) from above and from below. The arithmetic mean gives a principal-value integral. Then we add on both sides the contribution due to the δ-term in the eigenfunction $\phi(\nu, \mu)$. It is this contribution which now cancels the last term in (48), in view of the identity (35).

With Eqs. (15), (46), and (48) used in (45), we express $J(z, z')$ as

$$J(z, z') = L(z, z') - \frac{2}{z - z'}\left[\frac{g(z, z')}{H(-z')} - \frac{g(z', z)}{H(-z)}\right], \quad (51)$$

where L is a symmetric polynomial of maximum order $N - m - 1$ in each variable, namely:

$$L(z, z') = \frac{1}{z - z'} \sum_{l=m}^{N} c_l[g_l(z)K_l(z') - g_l(z')K_l(z)]. \quad (52)$$

From Eqs. (45) and (51) we deduce that, for any two different positive eigenvalues, the relation

$$\frac{4}{\nu\nu'} \int_0^1 \phi(\nu, \mu)\phi(\nu', \mu)\mu H(\mu)\, dm(\mu) = L(\nu, \nu') \quad (+53)$$

holds. For two discrete eigenvalues, this is immediately clear because the last term in (51) vanishes. The extension to the continuum eigenvalues is carried out in the same way as in the proof of Eq. (50).

With Eq. (53), it becomes clear how the adjoint $\tilde{\phi}(\nu, \mu)$ has to be constructed. It should differ from $\phi(\nu, \mu)$ by an amount

$$\tfrac{1}{2}\nu B(\nu, \mu) \equiv \tilde{\phi}(\nu, \mu) - \phi(\nu, \mu) \quad (54)$$

such that the corresponding contribution to the integral (43) cancels the right-hand side of Eq. (53). That is, we want $B(\nu, \mu)$ to fit the condition

$$\frac{2}{\nu} \int_0^1 \phi(\nu, \mu)B(\nu', \mu)\mu H(\mu)\, dm(\mu) = -L(\nu, \nu'). \quad (+55)$$

If this is satisfied, the bi-orthogonality relation (43) holds, with $\tilde{\phi}$ defined by Eq. (54).

In view of Eq. (52), and thinking of a Legendre expansion of $B(\nu, \mu)$, we see that this function must be a polynomial of maximum order $N - m - 1$ in each variable. Therefore, we write

$$B(\nu', \mu) = \sum_{l=m}^{N-1} B_l(\nu')p_l(\mu), \quad (56)$$

and we try to determine the coefficients from the condition (55). In view of Eq. (50), we have

$$\sum_{l=m}^{N-1} B_l(\nu')K_l(\nu) = -L(\nu, \nu'). \quad (57)$$

In the simple case $N = 1$, $m = 0$, the L, K_0, and B are constants; from Eq. (57), we obtain

$$B = \varpi_1 h_0 \varpi_0 \alpha_1/(2 - \varpi_0\alpha_0), \quad (58)$$

which has been shown[16] to equal the cB of Ref. 11.

[16] N. J. McCormick, Ph.D. thesis, The University of Michigan (unpublished).

The sole condition for the existence and uniqueness of $B(\nu, \mu)$ is that the K_l's in Eq. (57) are linearly independent. Unfortunately, it has not been possible so far to guarantee this linear independence for all physically possible scattering functions, although simple examples indicate that linear dependence could occur at most for special values of the ϖ_l's. For what follows, we assume that linear independence of the K_l polynomials does hold; for any particular case, one should verify the validity of this assumption numerically.

It may be added that the troubles here are similar to those encountered with other methods for handling transport problems with anisotropic scattering. Either explicitly or tacitly, in each of these methods[3,4,9,17] the nonvanishing of some determinant has been assumed without proof.

Two further useful equations are obtained by generalizing the integral in (55) to complex variables. Using (56), (46), (48), and (57), we find

$$\int_0^1 \frac{g(z, \mu)}{z - \mu} B(z', \mu)\mu H(\mu)\, dm(\mu) = -L(z, z') - \frac{2B(z', z)}{H(-z)}. \quad (59)$$

Combination of this result with (44) and (51) enables us to write

$$\int_0^1 \frac{g(z, \mu)}{z - \mu}\left[\frac{g(z', \mu)}{z' - \mu} + B(z', \mu)\right]\mu H(\mu)\, dm(\mu) = -\frac{2g(z, z')}{(z - z')H(-z')} - \frac{2}{H(-z)}\left[\frac{g(z', z)}{z' - z} + B(z', z)\right]. \quad (60)$$

If two different positive eigenvalues are substituted for z, z' in this equation, the bi-orthogonality relations follow, as already mentioned. On the other hand, we may also substitute negative eigenvalues. All such relations can be summarized in a single equation,

$$\int_0^1 \phi(\nu, \mu)\tilde{\phi}(\nu', \mu)\mu H(\mu)\, dm(\mu) = -\frac{\nu'\phi(\nu, \nu')}{H(-\nu')}\Theta(\nu') - \frac{\nu\tilde{\phi}(\nu', \nu)}{H(-\nu)}\Theta(\nu), \quad \nu \neq \nu'. \quad (61)$$

Here we have introduced the following step function:

$$\Theta(\nu) = \begin{cases} 0 & \text{for } 0 < \nu < 1, \\ 1 & \text{otherwise.} \end{cases} \quad (62)$$

[17] S. Pahor, Nucl. Sci. Eng. 26, 192 (1966).

It may be mentioned that Eq. (61) encompasses a set of eight formulas worked out previously[11] for the special case $N = 1$, $m = 0$.

We are left with the determination of the normalization constants. Those for the discrete eigenfunctions follow from (60) if we put $z = \nu_i$, and then let z' approach the same value. After application of the identity (32), the result is

$$\int_0^1 \phi(\nu_i, \mu)\tilde{\phi}(\nu_i, \mu)\mu H(\mu)\, dm(\mu) = \tfrac{1}{2}\nu_i^2 g(\nu_i, \nu_i)H(\nu_i)\left[\frac{d\Lambda(z)}{dz}\right]_{z=\nu_i}. \quad (63)$$

Before we can deduce the normalization constant for the continuum modes, we must decide what meaning to attach to the product of two singular eigenfunctions in the event that ν and ν' are allowed to merge. Following an established practice,[8-10] we use

$$\mathcal{P}\frac{1}{\nu - \mu}\,\mathcal{P}\frac{1}{\nu' - \mu} = \frac{1}{\nu - \nu'}\left[\mathcal{P}\frac{1}{\nu' - \mu} - \mathcal{P}\frac{1}{\nu - \mu}\right] + \pi^2\,\delta(\nu - \mu)\,\delta(\nu' - \mu), \quad (64)$$

which for $\nu' \to \nu$ is a definition. With this definition, the Poincaré–Bertrand formula permits the formal inversion of orders of integration in certain double integrals.[18]

All the deductions which led to Eq. (60), and specifically to Eq. (61), remain valid also for $\nu' \to \nu$, $0 < \nu < 1$, except for two additional contributions. One contribution, $\lambda^2(\nu)\nu H(\nu)(1 - \nu^2)^{-m}\delta(\nu - \nu')$, originates from the product of the δ-functions involved in ϕ and $\tilde{\phi}$, and another,

$$[\tfrac{1}{2}\nu g(\nu, \nu)]^2 \nu H(\nu)(1 - \nu^2)^m \pi^2\,\delta(\nu - \nu'),$$

is due to the last term in (64). The identity (24) enables us to write the final result as

$$\int_0^1 \phi(\nu, \mu)\tilde{\phi}(\nu', \mu)\mu H(\mu)\, dm(\mu) = \Lambda^+(\nu)\Lambda^-(\nu)\nu H(\nu)(1 - \nu^2)^{-m}\,\delta(\nu - \nu'), \quad \nu, \nu' \in (0, 1). \quad (65)$$

The three equations (61), (63), and (65) are all we need in several typical half-space problems for calculating the expansion coefficients in (40). It might also be mentioned that these equations simplify the determination of the expansion coefficients for slab problems.[19,20] In some applications, it becomes necessary to derive a few more auxiliary formulas, as was done for $N = 1$.[11]

If we are interested solely in the angular distribution emerging at $x = 0$, we need only Eq. (61) for $0 < \nu < 1$ or $\nu = \nu_i$, and $-1 < \nu' < 0$. Instead of handling each single $\phi(\nu, \mu)$ separately, we may use the whole expansion (41) and get, after renaming the variables,

$$\psi(-\mu) = \mu^{-1}H(\mu) \times \int_0^1 \psi(\mu')\tilde{\phi}(-\mu, \mu')\mu' H(\mu')\, dm(\mu'). \quad (+66)$$

Here the left-hand side denotes the value of the expansion (41) (the coefficients of which we now need not know) with $-\mu$ substituted for μ. Thus the relation serves to determine the emerging angular distribution from the ingoing one.[11] Equation (66) may therefore be called the "switching relation", or we may say that the integral operator in (66) represents the albedo operator for a half-space. A slightly generalized technique can be used to determine both the reflected and the transmitted angular densities for a slab, after the expansion coefficients are calculated.

As an example of the use of Eq. (66), we consider the albedo problem, defined by the boundary conditions

$$I(0, \mu, \varphi) = \delta(\mu - \mu_0)\,\delta(\varphi - \varphi_0), \quad (+67)$$

$$I(x, \mu, \varphi) \to 0, \qquad x \to \infty. \quad (68)$$

The azimuthal Fourier expansion[13] may be written as

$$I(x, \mu, \varphi) = \sum_{m=0}^{N}(2 - \delta_{0m})I^m(x, \mu) \times (1 - \mu^2)^{m/2}(1 - \mu_0^2)^{m/2}\cos m(\varphi - \varphi_0) + \delta(\mu - \mu_0)e^{-x/\mu_0}\left[\delta(\varphi - \varphi_0) - \frac{1}{2\pi} \times \sum_{m=0}^{N}(2 - \delta_{0m})\cos m(\varphi - \varphi_0)\right]. \quad (69)$$

The boundary condition for I^m is

$$I^m(0, \mu) = \frac{\delta(\mu - \mu_0)}{2\pi(1 - \mu_0^2)^m}. \quad (+70)$$

According to Eq. (41), where now $f = 0$, we

[18] I. Kuščer and N. J. McCormick, Nucl. Sci. Eng. 23, 404 (1965).

[19] G. J. Mitsis, Nucl. Sci. Eng. 17, 55 (1963).

[20] N. J. McCormick and M. R. Mendelson, Nucl. Sci. Eng. 20, 462 (1964).

substitute $I^m(0, \mu')$ for $\psi(\mu')$ in the switching relation (66) and obtain

$$I^m(0, -\mu) = (2\pi\mu)^{-1}\mu_0 H^m(\mu_0)H^m(\mu)\tilde{\phi}^m(-\mu, \mu_0). \quad (+71)$$

The entire emerging angular distribution is then given by

$$I(0, -\mu, \varphi) = \frac{\mu_0}{2\pi\mu}\sum_{m=0}^{N}(2 - \delta_{0m})H^m(\mu)H^m(\mu_0) \times \tilde{\phi}^m(-\mu, \mu_0)(1 - \mu^2)^{m/2} \times (1 - \mu_0^2)^{m/2}\cos m(\varphi - \varphi_0). \quad (+72)$$

The functions

$$4\pi\mu(2 - \delta_{0m})I^m(0, -\mu)(1 - \mu^2)^{m/2}(1 - \mu_0^2)^{m/2}$$

and $4\pi\mu I(0, -\mu, \varphi)$ for $\mu > 0$ equal Chandrasekhar's $S^m(\mu, \mu_0)$ and $S(\mu, \varphi; \mu_0, \varphi_0)$, respectively. In terms of these functions, relation (66) and its azimuth-dependent generalization can be derived by physical intuition, as shown by Chandrasekhar.[1]

The reciprocity theorem of Chandrasekhar says that $\mu I^m(0, -\mu)$ is a symmetric function of μ and μ_0, so that

$$\mu_0\tilde{\phi}^m(-\mu, \mu_0) = \mu\tilde{\phi}^m(-\mu_0, \mu). \quad (73)$$

An expression similar to (71) has already been derived by Busbridge,[4] namely

$$I^m(0, -\mu) = \frac{\mu_0 H^m(\mu)H^m(\mu_0)}{4\pi(\mu + \mu_0)} \times \sum_{l=m}^{N}(-1)^{l+m}c_l^m q_l^m(\mu)q_l^m(\mu_0), \quad (+74)$$

where $q_l^m(\mu)$ are polynomials. We conclude that

$$2\mu^{-1}(\mu + \mu_0)\tilde{\phi}^m(-\mu, \mu_0) = g_l^m(-\mu, \mu_0) - (\mu + \mu_0)B(-\mu, \mu_0) = \sum_{l=m}^{N}(-1)^{l+m}c_l^m q_l^m(\mu)q_l^m(\mu_0). \quad (75)$$

Pahor[17] has recently found a practical method for obtaining Busbridge's q_l^m, so that B could then be constructed from (75).

IV. THE NONABSORBING MEDIUM

In the case with no absorption ($\varpi_0 = 1$ and therefore $h_0 = 0$), the azimuth-independent problem ($m = 0$) requires special consideration because ν_1 and $-\nu_1$ merge at infinity. The approach when $\varpi_0 \to 1$ is like $\nu_1 \approx [h_0 h_1]^{-\frac{1}{2}}$. Consequently, two of the eigenmodes become identical,

$$\lim_{\varpi_0 \to 1}\phi(\pm\nu_1, \mu)e^{\mp x/\nu_1} = \tfrac{1}{2}. \quad (76)$$

A new linearly independent eigenmode arises in the following way:

$$\lim_{\varpi_0 \to 1}\tfrac{1}{2}h_1\nu_1[\phi(-\nu_1, \mu)e^{x/\nu_1} - \phi(\nu_1, \mu)e^{-x/\nu_1}] = \tfrac{1}{2}(h_1 x - 3\mu). \quad (77)$$

Once again we are omitting the superscript ($m = 0$).

The function $\Lambda(z)$ now has a double zero at infinity, and we see from Eqs. (27) and (38) that

$$\eta_0 = 1, \qquad \beta_0 = 1, \qquad \beta_1^2 = \eta_2. \quad (78)$$

The value of η_2 now becomes important and is given by the equation[3]

$$\eta_2 = \frac{1}{3}\prod_{l=2}^{N}\frac{h_l}{2l + 1}. \quad (79)$$

The polynomials $g_l(\nu)$ are of order $l - 2$ instead of l. Therefore, the polynomials $K_l(z)$ as defined in Eq. (49) are of maximum order $N - 3$ or $l - 1$, whichever is greater; if $N \leq 2$, K_0 vanishes. Obviously the lowering of the order destroys the previously assumed linear independence of the K_l. However, because the maximum order of $L(\nu, \nu')$ now is only $N - 2$, it is sufficient to assume that linear independence holds after any one of the K_l's, $l = 0, 1, \cdots N - 2$, is omitted. Under this assumption, the expansion (57) exists, but it is not uniquely determined.

Except for the ensuing ambiguity in $B(\nu, \mu)$, we can now take over from the previous section the formulas for the eigenfunctions $\phi(\nu, \mu)$, with the exclusion of the limits of $\phi(\nu_1, \mu)$ and $\tilde{\phi}(\nu_1, \mu)$. The corresponding relations for the latter functions will be derived separately.

By using $\nu = \nu_1$ in Eq. (61), we obtain in the limit

$$\int_0^1 \tilde{\phi}(\nu, \mu)\mu H(\mu)\, d\mu = -\frac{\nu}{H(-\nu)}\Theta(\nu). \quad (80)$$

This gives us a supplementary condition upon $B(\nu, \mu)$. After the expansion (56) is inserted into (80) with $\nu > 0$, we have

$$K_0(\nu) + \sum_{l=0}^{N-1}B_l(\nu)k_l = 0, \quad (81)$$

where

$$k_l = \lim_{\varpi_0 \to 1}\nu_1 K_l(\nu_1) = \int_0^1 P_l(\mu)\mu H(\mu)\, d\mu. \quad (82)$$

We assume that condition (81) and Eq. (57) determine $B(\nu, \mu)$ uniquely. For example, for $N = 2$,

we find in this way

$$B(\nu, \mu) = \frac{3\varpi_2}{4}\left(\frac{\alpha_2}{\alpha_1} - \mu\right). \tag{83}$$

From Eq. (61), supplemented by (80), we see that the switching relation (66) remains valid, with $\psi(\mu)$ defined as before.

We also have to introduce

$$\tilde{\phi}_1(\mu) \equiv \lim_{\varpi_0 \to 1} \tilde{\phi}(\nu_1, \mu) = \tfrac{1}{2}[1 + b(\mu)], \tag{84}$$

where

$$b(\mu) = \lim_{\varpi_0 \to 1} \nu_1 B(\nu_1, \mu). \tag{85}$$

Equation (61) shows that

$$\int_0^1 \phi(\nu, \mu)\tilde{\phi}_1(\mu)\mu H(\mu)\, d\mu = -\frac{\nu}{H(-\nu)}\tilde{\phi}_1(\nu)\Theta(\nu). \tag{86}$$

Then, using the expansion

$$b(\mu) = \sum_{l=0}^{N-2} b_l P_l(\mu), \tag{87}$$

we derive from (86) with $\nu > 0$,

$$K_0(\nu) + \sum_{l=0}^{N-2} b_l K_l(\nu) = 0, \tag{88}$$

which determines $\tilde{\phi}_1(\mu)$ up to a constant factor. This constant will now be evaluated from the normalization constant corresponding to (63) for $j = 1$.

First, we derive from the recursion formula for $g_l(\nu)$ that

$$\lim_{\varpi_0 \to 1} g_l(\nu_1)P_l(\nu_1) = \prod_{n=1}^{l} \frac{2n-1}{h_n}. \tag{89}$$

It follows by inductive reasoning that

$$\lim_{\varpi_0 \to 1} g(\nu_1, \nu_1) = \prod_{n=1}^{N} \frac{2n+1}{h_n} = \frac{1}{h_1\eta_2}, \tag{90}$$

in view of Eq. (79). Finally, from (63), with (25) and (36),

$$\int_0^1 \tilde{\phi}_1(\mu)\mu H(\mu)\, d\mu = \frac{1}{h_1\beta_1} = \frac{1}{2}\left[\alpha_1 + \sum_{l=0}^{N-2} b_l k_l\right]. \tag{91}$$

Equations (88) and (91) determine the b_l's, and hence the function $\tilde{\phi}_1(\mu)$. As an example, we quote the result for $N = 2$:

$$\tilde{\phi}_1(\mu) = 1/h_1\beta_1\alpha_1. \tag{92}$$

With respect to Milne's problem where the eigenmode (77) appears in the expansion, we also need to evaluate integrals involving $\mu\tilde{\phi}$. From Eqs. (77) (with $x = 0$) and (61), we find

$$\int_0^1 \mu\tilde{\phi}(\nu, \mu)\mu H(\mu)\, d\mu = -\frac{\nu^2}{H(-\nu)}\Theta(\nu) - \tfrac{2}{3}h_1\beta_1\nu\tilde{\phi}_1(-\nu), \tag{93}$$

where we have used Eq. (36) and the symmetry relation (73), with an obvious generalization in the variables. Equation (93), together with Eqs. (80), (86), and (91), represent a set of additional formulas needed for $\varpi_0 = 1$.

For the integral

$$C \equiv \int_0^1 \mu\tilde{\phi}_1(\mu)\mu H(\mu)\, d\mu \tag{94}$$

no expression like the $1/h_1\beta_1$ in Eq. (91) has been found, so we have to evaluate it in terms of the moments of the H-function, which must be computed anyway.

As an example, we consider some results for the Milne problem, for which

$$I(0, \mu) = 0, \tag{+95}$$

$$I(x, \mu) \sim x, \qquad x \to \infty. \tag{96}$$

We now have the expansion

$$I(x, \mu) = \tfrac{1}{2}(h_1 x - 3\mu) + \tfrac{1}{2}a_1 + \sum_{j=2}^{M} a_j\phi(\nu_j, \mu)e^{-x/\nu_j} + \int_0^1 A(\nu)\phi(\nu, \mu)e^{-x/\nu}\, d\nu. \tag{97}$$

The two leading terms in (97) give the asymptotic density

$$\rho_{as}(x) = 2\pi(h_1 x + a_1), \tag{98}$$

which vanishes at the extrapolated endpoint, $x = -z_0$. After a_1 is evaluated by applying (86), (91), and (94) to the expansion (97) for $x = 0$, we find

$$z_0 = 3C\beta_1. \tag{99}$$

To determine the angular density emerging at $x = 0$, we use the switching relation (66) and Eq. (93) to show that

$$I(0, -\mu) = h_1\beta_1 H(\mu)\tilde{\phi}_1(\mu). \tag{+100}$$

With expression (92) inserted, this equation reproduces a result of Chandrasekhar[1] for $N = 2$.

V. CALCULATIONAL SCHEME

As shown by the two examples in Secs. III and IV, the final solutions of standard half-space prob-

lems can be expressed in closed form by the H-functions and the adjoints $\bar{\phi}^m(\nu, \mu)$. The latter involve the polynomials $B^m(\nu, \mu)$, the coefficients of which have to be computed numerically. Only for the simplest cases (i.e., for $N - m = 0$ or 1, and for $N = 2$ if $\varpi_0 = 1$, $m = 0$) is it practical to express the $B^m(\nu, \mu)$ directly in terms of the coefficients ϖ_l of the scattering function. Even then the H-functions and a few of their moments have to be calculated numerically for each particular scattering law of interest.

For any scattering law, numerical computations could proceed according to the following scheme. First, the coefficients of $g_l^m(\nu)$ are found from Eqs. (17) and (19). [The coefficients for $g_l^0(\nu)$ are explicitly known.[3]] This enables us to express $\Lambda^m(z)$ from Eq. (22) either numerically or in terms of the function $\tanh^{-1}(z^{-1})$. The discrete eigenvalues ν_j^m are then determined from (21). (For ν_1^0, Holte's power series[3] could be used if absorption is weak.) The values $[d\Lambda^m(z)/dz]_{z=\nu_j^m}$ are also to be computed. Similarly, $\lambda^m(\nu)$ and the boundary values are found in terms of $\tanh^{-1}(\nu)$ through Eqs. (14) and (24), respectively.

The functions $H^m(\mu)$, $0 \leq \mu \leq 1$, and the values of $H^m(\nu_j^m)$ can be computed either from the nonlinear integral equation,[1,21,22] Eq. (33), or from the closed-form expression (28). A modified form of the latter, probably better suited for computation, is[17]

$$\frac{1}{H^m(\mu)} = (1 + \mu)^{-M} \prod_{j=1}^{M} (1 + \mu/\nu_j^m)$$

$$\times \exp\left[\frac{1}{2\pi i}\int_0^1 \ln\frac{\Lambda^{m+}(\nu)}{\Lambda^{m-}(\nu)}\left\{\frac{1}{\nu + \mu} - \frac{1}{\nu}\right\} d\nu\right], \quad (101)$$

All further computation makes use of the numerically computed moments α_n^m of the H-functions, defined in Eq. (39), for $n = 0, 1, \cdots 3(N - m)$. In an obvious way, the quantities β_n^m of Eq. (37) may then be calculated for $n = 0, 1, \cdots (N - m)$. [For $\varpi_0 = 1$ and $m = 0$, only the moments α_n^0, $n = 1, 2, \cdots (3N - 2)$, and the β_n^0, $n = 1, 2, \cdots N$, are required.] Numerical accuracy may be checked by use of the identities (38).

With the values of α_n^m and β_n^m, the coefficients of the polynomials $K_l^m(\nu)$, $l = 0, 1, \cdots N$, easily follow from (49). Linear independence should then be verified. Finally, after calculating the coefficients of $L^m(\nu, \nu')$ from Eq. (52), the determination of the coefficients of $B^m(\nu, \mu)$ is carried out using Eq. (57). For $\varpi_0 = 1$, $m = 0$, Eq. (81) must also be used to find $B^0(\nu, \mu)$. In this event, some additional work is required for the determination of the coefficients of $b(\mu)$ from Eqs. (88) and (91).

The determination of the K- and L-polynomials could be avoided if the Busbridge polynomials $q_l^m(\mu)$ are determined by the method of Pahor.[17] The coefficients of $B^m(\nu, \mu)$ would then be found by use of Eq. (75).

ACKNOWLEDGMENTS

We wish to thank Dr. S. Pahor for stimulating discussions. One of us (N.J.M.) wishes to express his gratitude to the members of the Institute of Physics, University of Ljubljana, for the hospitality shown him.

APPENDIX. REMARK ABOUT FULL-RANGE ORTHOGONALITY RELATIONS

As shown by Mika,[9] Eq. (10) immediately leads to full-range orthogonality of the full set of eigenfunctions, with the weight function $\mu(1 - \mu^2)^m$. In particular, for $\nu, \nu' \in (-1, 1)$,

$$\int_{-1}^{1} \phi(\nu, \mu)\phi(\nu', \mu)\mu\, dm(\mu)$$

$$= \Lambda^+(\nu)\Lambda^-(\nu)\nu(1 - \nu^2)^{-m}\, \delta(\nu - \nu'), \quad (A1)$$

where the superscripts m are again omitted. To this we can add a convenient expression for the normalization constants for the discrete eigenfunctions:

$$\int_{-1}^{1} [\phi(\pm\nu_i, \mu)]^2 \mu\, dm(\mu)$$

$$= \pm\tfrac{1}{2}\nu_i^2 g(\nu_i, \nu_i)[d\Lambda(z)/dz]_{z=\nu_i}. \quad (A2)$$

These relations enable us to solve the problem of a plane source in an infinite medium.

The proof of the last relation is obtained by applying the reasoning of Sec. III to full-range integrals. Denoting the various functions modified in this sense by the same symbols as in that section, we introduce first

$$\int_{-1}^{1} \frac{g(z, \mu)}{z - \mu}\frac{g(z', \mu)}{z' - \mu}\mu\, dm(\mu) \equiv J(z, z'). \quad (A3)$$

After a partial fraction analysis, J can be expressed in terms of the integrals

$$\bar{K}_l(z) = \int_{-1}^{1} p_l(\mu)\frac{g(z, \mu)}{z - \mu}\mu\, dm(\mu). \quad (A4)$$

[21] T. W. Mullikin, Astrophys. J. 139, 1267 (1964).
[22] S. Pahor and I. Kuščer, Astrophys. J. 143, 888 (1966).

Analysis similar to that leading to Eq. (48) then shows that

$$\bar{K}_l(z) = K_l(z) - 2p_l(z)\Lambda(z), \tag{A5}$$

where $K_l(z)$ is again a polynomial. We have

$$K_l(\nu) = \frac{2}{\nu}\int_{-1}^{1} p_l(\mu)\phi(\nu, \mu)\mu\, dm(\mu) \tag{A6}$$

whenever ν is an eigenvalue, as shown by the aid of Eqs. (21) and (24). The recursion formula for $p_l(\mu)$ and Eqs. (16) and (17) show that

$$K_l(\nu) = \frac{2h_l}{2l+1}\, g_l(\nu). \tag{A7}$$

Therefore, proceeding as in Sec. III, we finally find that

$$J(z, z') = -\frac{2}{z - z'}\,[g(z, z')\Lambda(z') - g(z', z)\Lambda(z)]. \tag{A8}$$

After we substitute $z = \pm\nu_j$ and carry out the limit $z' \to \pm\nu_j$, Eq. (A2) immediately follows.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

Theory of Paramagnetic Impurities in Semiconductors*

DANIEL C. MATTIS AND ELLIOTT H. LIEB†
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 21 February 1966)

In this paper, a model of a paramagnetic impurity in a semiconductor (or of an F' center in an alkali halide) is proposed. It is an exactly soluble form of the quantum-mechanical 3-body problem. Specifically, we deal with 2 interacting particles in any number of dimensions in an attractive external potential, and present the qualitative features of the resulting eigenvalues and eigenfunctions. We find algebraically the conditions for a magnetic moment to appear (e.g., for an F' center to become unstable with respect to an F center) and discover that even a large 2-body electronic repulsion U does not cause a moment to appear when the one-electron bound state orbits about the impurity are sufficiently great. Conversely, in the case of small, tightly bound orbits, beyond a certain value of U, the impurity does in fact become magnetic in the ground state. Using the exact ground-state solution, we show that a perturbation-theoretic expansion in powers of U has a finite radius of convergence.

1. INTRODUCTION

THE problems associated with magnetic impurities in metals have received a great deal of attention,[1] but are still far from reaching a rigorous solution. By contrast, we have readily found an exactly soluble model of paramagnetic (donor or acceptor) impurities in semiconductors, which can be rigorously analyzed over an entire range of parameters with rather interesting results. The present paper is the first report on the theory of this model, giving features of the eigenstates and of the magnetic properties. Quantitative numerical results as well as transport properties (i.e., scattering cross section), statistical mechanics, and other features of this model will be reported later.

It has long been known that the Coulomb repulsion among electrons in impurity states of a semiconductor cannot be safely neglected. More than eleven years ago Brooks[2] wrote, "... band (i.e., Bloch) states have the property that the corresponding wavefunctions are spread throughout the crystal. Thus there is practically no price, in terms of extra electrostatic interaction, for putting two electrons in the same (Bloch) state. This is the condition for the applicability of Fermi statistics in its simple form. In the case of localized states, however, a very different situation obtains. Even though an

* This work was supported by the United States Air Force Office of Scientific Research under Grant AFOSR-107566 (D. M.) and AFOSR-71364, 50866 (E. L.).

† Present address: Physics Department, Northeastern University, Boston, Massachusetts.

[1] The quantum- and statistical-mechanical aspects have been discussed by P. W. Anderson, Phys. Rev. 124, 41 (1961), and recently extended by, among others, A. C. Hewson, Phys. Rev. 144, 420 (1966), and by J. R. Schrieffer and D. C. Mattis, *ibid.* 140, A1412 (1965). Transport properties have been analyzed by J. Friedel, *Metallic Solid Solutions* (W. A. Benjamin, Inc., New York, 1963), and most recently by D. J. Kim, Phys. Rev. (to be published).

[2] H. Brooks, *Advances in Electronics and Electron Physics*, L. Marton, Ed. (Academic Press Inc., New York, 1955), Vol. VII. See also C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York, 1956), 2nd ed., p. 359.

Analysis similar to that leading to Eq. (48) then shows that

$$\bar{K}_l(z) = K_l(z) - 2p_l(z)\Lambda(z), \tag{A5}$$

where $K_l(z)$ is again a polynomial. We have

$$K_l(\nu) = \frac{2}{\nu}\int_{-1}^{1} p_l(\mu)\phi(\nu, \mu)\mu\, dm(\mu) \tag{A6}$$

whenever ν is an eigenvalue, as shown by the aid of Eqs. (21) and (24). The recursion formula for $p_l(\mu)$ and Eqs. (16) and (17) show that

$$K_l(\nu) = \frac{2h_l}{2l+1}\, g_l(\nu). \tag{A7}$$

Therefore, proceeding as in Sec. III, we finally find that

$$J(z, z') = -\frac{2}{z - z'}\,[g(z, z')\Lambda(z') - g(z', z)\Lambda(z)]. \tag{A8}$$

After we substitute $z = \pm\nu_j$ and carry out the limit $z' \to \pm\nu_j$, Eq. (A2) immediately follows.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

Theory of Paramagnetic Impurities in Semiconductors*

DANIEL C. MATTIS AND ELLIOTT H. LIEB†
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 21 February 1966)

In this paper, a model of a paramagnetic impurity in a semiconductor (or of an F' center in an alkali halide) is proposed. It is an exactly soluble form of the quantum-mechanical 3-body problem. Specifically, we deal with 2 interacting particles in any number of dimensions in an attractive external potential, and present the qualitative features of the resulting eigenvalues and eigenfunctions. We find algebraically the conditions for a magnetic moment to appear (e.g., for an F' center to become unstable with respect to an F center) and discover that even a large 2-body electronic repulsion U does not cause a moment to appear when the one-electron bound state orbits about the impurity are sufficiently great. Conversely, in the case of small, tightly bound orbits, beyond a certain value of U, the impurity does in fact become magnetic in the ground state. Using the exact ground-state solution, we show that a perturbation-theoretic expansion in powers of U has a finite radius of convergence.

1. INTRODUCTION

THE problems associated with magnetic impurities in metals have received a great deal of attention,[1] but are still far from reaching a rigorous solution. By contrast, we have readily found an exactly soluble model of paramagnetic (donor or acceptor) impurities in semiconductors, which can be rigorously analyzed over an entire range of parameters with rather interesting results. The present paper is the first report on the theory of this model, giving features of the eigenstates and of the magnetic properties. Quantitative numerical results as well as transport properties (i.e., scattering cross section), statistical mechanics, and other features of this model will be reported later.

It has long been known that the Coulomb repulsion among electrons in impurity states of a semiconductor cannot be safely neglected. More than eleven years ago Brooks[2] wrote, "... band (i.e., Bloch) states have the property that the corresponding wavefunctions are spread throughout the crystal. Thus there is practically no price, in terms of extra electrostatic interaction, for putting two electrons in the same (Bloch) state. This is the condition for the applicability of Fermi statistics in its simple form. In the case of localized states, however, a very different situation obtains. Even though an

* This work was supported by the United States Air Force Office of Scientific Research under Grant AFOSR-107566 (D. M.) and AFOSR-71364, 50866 (E. L.).

† Present address: Physics Department, Northeastern University, Boston, Massachusetts.

[1] The quantum- and statistical-mechanical aspects have been discussed by P. W. Anderson, Phys. Rev. 124, 41 (1961), and recently extended by, among others, A. C. Hewson, Phys. Rev. 144, 420 (1966), and by J. R. Schrieffer and D. C. Mattis, *ibid.* 140, A1412 (1965). Transport properties have been analyzed by J. Friedel, *Metallic Solid Solutions* (W. A. Benjamin, Inc., New York, 1963), and most recently by D. J. Kim, Phys. Rev. (to be published).

[2] H. Brooks, *Advances in Electronics and Electron Physics*, L. Marton, Ed. (Academic Press Inc., New York, 1955), Vol. VII. See also C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York, 1956), 2nd ed., p. 359.

electron may be allowed two directions of spin in a localized state, once the state is occupied by an electron of either spin, it cannot then be occupied by an electron of opposite spin, because the electrostatic repulsion of the two localized charge distributions would raise the energy of the second electron. . .". If only a single electron can be bound to the impurity, the latter is perforce a paramagnetic impurity of one Bohr magneton. The present model is designed to explain precisely how this single binding may or may not occur.

Several superficial difficulties have to be overcome in the process, for, in general, the problem of two interacting particles in an attractive potential well (such as the two electrons in helium atom or in the hydrogen molecule) has no solution in closed form. Fortunately for us, the present calculation for the solid is more tractable than the usual two-electron problems, and it *is* possible to obtain an explicit solution in closed form, as we show below.

In all such problems involving two electrons, one readily proves[3] that the ground-state wavefunction is nodeless; it then must be symmetric under interchange of the spatial coordinates of the two particles, and it must belong to spin $S = 0$. Does this preclude magnetic behavior? The answer, surprisingly is no, provided the 2-particle repulsion U exceeds a critical magnitude U_c. In that case, one of the electrons is ionized and the energy splitting between the ground state and the lowest $S = 1$ state ceases to be finite and becomes $O(N^{-2})$, where N = number of atoms in the crystal. The ground state can then be taken as an arbitrary combination of triplet and singlet, so that, in effect, there is one Bohr magneton localized on the impurity, and one uncorrelated Bohr magneton on the second, wandering, particle. The net localized spin of the impurity is then $1\mu_B$, the largest value attainable in the present model. One of the results obtained below is an expression for U_c in terms of the one-electron parameters (viz., band structure and binding energy of the impurity potential well). When U_c is infinite, then the impurity is always entirely nonmagnetic.

[3] E. Lieb and D. Mattis, Phys. Rev. 125, 164 (1962). This theorem and its consequences are discussed also in D. Mattis, *The Theory of Magnetism* (Harper and Row, Publishers, Inc., New York, 1965), Chap. 4. The 2-electron problem considered earlier by J. C. Slater, H. Statz, and G. F. Koster, Phys. Rev. 91, 1323 (1953), also indicated that, without Hund's rule coupling, the ground state belongs to $S = 0$, on the basis of a model similar to the present model with $v \equiv 0$. A different but also exactly soluble model of two electrons in an harmonic oscillator potential, interacting by Coulomb repulsion, was solved by N. Kestner and O. Sinanoglu, Phys. Rev. 128, 2687 (1962). Their interesting result cannot be directly compared with ours, however, because it has no scattering state, no unbound solution, and no magnetism.

After the present manuscript was essentially completed, strong experimental evidence for the existence of such paramagnetic impurities in n-type InSb was reported by Katayama and Tanaka[4] (viz., the existence of a resistance-minimum anomaly such as have been commonly observed in metals containing paramagnetic impurities, but much larger than in these metals). An earlier indication of this was provided by work on the thermoelectric power by Khosla and Sladek.[4] Both series of experiments were preceded by a theory due to Toyozawa,[5] based on the Hartree–Fock approximation. As we see below, this approximation can only be valid for very deep donor levels.

As another application, we recall that the F center consists of an electron bound to a vacancy in an alkali–halide crystal, whereas the F' center consists of *two* electrons bound to the same vacancy. The energetic stability and capture cross section of the F' center for the second electron may also in principle be calculated by the present methods, extending recent approximation schemes.[6]

2. DESCRIPTION OF THE MODEL

In this section, we describe the general model and its general solutions, explaining the steps whereby the latter are obtained. In the following section a one-dimensional example is explicitly worked out using this method.

For definiteness, we discuss a donor-type impurity and 2 electrons in the conduction band of a semiconductor. (The analysis for an acceptor-type impurity and 2 holes in the valence band is, *mutatis mutandis*, formally identical.) The electrons move from one Wannier site to the next, with overlap matrix elements $K(\mathbf{R}_i - \mathbf{R}_j)$. The band structure, given by the Bloch energies $\epsilon(\mathbf{k})$ (the Fourier transforms of the K's) is therefore

$$\epsilon(\mathbf{k}) = \frac{1}{N} \sum_{i,j} K(\mathbf{R}_i - \mathbf{R}_j) \cos \mathbf{k}\cdot\mathbf{R}_{ij}. \qquad (2.1)$$

In addition, there is the potential of an impurity at the origin: $-v(\mathbf{R}_i)$ which is assumed to be deep enough to have *one and only one* bound state.

[If the potential has *no* bound state there can be no localized spin, as the probability that either of the two electrons is in the vicinity of

[4] Y. Katayama and S. Tanaka, Phys. Rev. Letters 16, 129 (1966); R. Khosla and R. Sladek, Phys. Rev. Letters 15, 1521 (1965).

[5] Y. Toyozawa, J. Phys. Soc. Japan 17, 986 (1962). [See parenthetical statement after Eq. (2.22).]

[6] S. Y. La and R. H. Bartram, Phys. Rev. 144, 670 (1966), and references therein.

the impurity is just $O(1/N) \sim$ zero. (This is quite different from the cases of interest in a metal, where electrons are always available in the vicinity of any site.) When the potential has two or more bound states, the analysis becomes more complicated than envisaged in the present calculation but it can be done. A summary of our findings in the more complicated situation is this: if the bound states are a *degenerate set*, then the 2-electron ground state of the impurity is likely to be a triplet or doublet magnetic state, in agreement with Hund's rule. Otherwise, the impurity is generally a nonmagnetic singlet *for all values of the two-electron repulsion*. This situation, descriptive of multilevel traps such as gold in germanium, can in fact be handled by the present methods although we do not further consider it in the present paper.[7]]

Before introducing the two-particle interaction, let us solve for the one-electron eigenstates for the band structure (2.1) in the presence of the impurity potential $-v$. This can always be done by straightforward methods (an explicit solution in the one-dimensional case is displayed in the following section) and results in a set of orthonormal eigenfunctions labeled by an ascending quantum number $r = 0, 1, 2, \cdots$ and spin quantum number $m = \pm\frac{1}{2}$:

$$\phi_{r,m} = \sum_i f_r(R_i) c^*_{i,m} \mid 0) \tag{2.2}$$

with energy eigenvalues: e_r;

$r = 0$ for the bound state, $r = 1, 2, \cdots$ for continuum states, arranged in the sequence $e_r \leq e_{r+1}$, and interlacing the Bloch energies ϵ_r. (2.3)

Two-particle states are merely antisymmetrized product states,

$$\Phi^0_{r,m;r',m'} = 2^{-\frac{1}{2}}\{[\sum_i f_r(R_i) \sum_j f_{r'}(R_j) \pm \sum_j f_r(R_i) \sum_i f_{r'}(R_i)]\} c^*_{i,m} c^*_{i,m'} \mid 0) \tag{2.4}$$

with energy eigenvalues:

$E(r, r') = e_r + e'_r$; $r = r' = 0$ for the two-particle bound state, $r = 0$ and $r' \neq 0$ (or vice-versa) for the one-particle bound state, and $r \neq 0$ and $r' \neq 0$ for the completely ionized states of the impurity. (2.5)

[7] For more background into the one-electron problem, the nature of the bound states, etc., we refer to G. F. Koster and J. C. Slater, Phys. Rev. 95, 1167 (1954); *ibid.* 96, 1208 (1954); and M. Lax, *ibid.* 94, 1391 (1954). Interaction with field bands is discussed by L. J. Sham, Phys. Rev. 150, 720 (1966).

Of all these two-particle states, *only the one-particle bound states possess a localized magnetic moment* of one Bohr magneton, the others having a localized magnetic moment which is either identically zero (two-particle bound states) or virtually zero, $O(1/N)$, (for the completely ionized states).

The two-particle eigenstates in presence of the two-particle interaction can be written as

$$\Phi = \sum_{i,j} F(R_i, R_j) c^*_{i,m} c^*_{j,m'} \mid 0), \tag{2.6}$$

and are antisymmetric or symmetric under the interchange of R_i and R_j according to whether they belong to eigenvalue $S = 1$ or $S = 0$ of total spin.

We obtain the correct F's by a Green's function technique. First we write the complete two-body Green's function,

$$G_E(ij \mid i'j') = \sum_{r,r'} \frac{f^*_r(i) f^*_{r'}(j) f_r(i') f_{r'}(j')}{E - e_r - e_{r'}} \tag{2.7}$$

and then, in terms of the repulsive two-body interaction $U(R_i, R_j) \geq 0$, we obtain

$$F(ij) = \sum_{i',j'} U(i', j') G_E(i'j' \mid ij) F(i'j'), \tag{2.8}$$

a system of linear coupled equations which are to be solved for the wavefunction F and the energy eigenvalue E. If U is nonvanishing over the entire crystal, this poses an intractable problem which must be solved by any of the variety of approximate techniques used in scattering theory. Although it is justified to neglect this long-range interaction, there are good reasons, outlined by Anderson,[1] not to neglect the Coulomb repulsion *near the impurity*. There, the atomic orbitals, hence the Wannier orbitals, tend to be more compact, and the electrostatic repulsion cannot be ignored, particularly in the case of magnetic atoms. If U is nonvanishing over Z distinct sites in the neighborhood of the impurity (located at the origin), the solution of (2.8) reduces to a $Z^2 \times Z^2$ determinantal equation. In what follows, we assume the simplest model,[8] i.e., an interaction only at the origin [i.e., $Z = 1$, with $U(0, 0) \equiv U$, and $U(i, j) = 0$ for R_i and R_j not both at the origin].

With the assumption of such a local repulsion, the wavefunctions are explicitly given as

$$F(ij) = UG(00 \mid ij) F(00). \tag{2.9}$$

We immediately note that all the antisymmetric solutions (i.e., the triplet states, in addition to those

[8] The simplest model displays the qualitative features of the general model; we have also determined that Anderson's model[1] yields similar results in the present context, when a one-particle bound state exists.

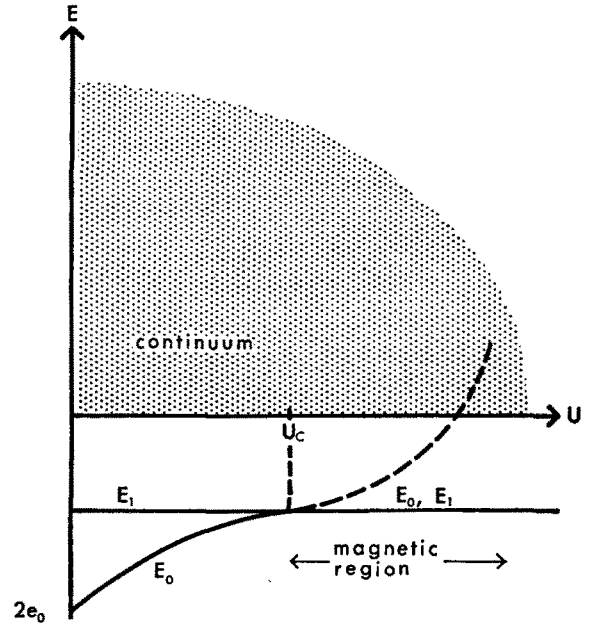


FIG. 1. Energy eigenvalues as functions of repulsive 2-body interaction U. Lowest singlet solution E_0 and triplet solution E_1 merge for $U \geq U_c$, resulting in a localized moment of 1 μ_B. The dotted line (spurious) indicates solution of Eq. (2.17) in range $U > U_c$. Note that, although curve shown does not indicate it, $dE_0/dU = 0$ when $U = U_c$. Range of energies above E_1 is 1-particle scattering states continuum. Shaded range labeled "continuum" refers to the 2-particle scattering states.

singlet states which have a node at the origin) have the unperturbed energy eigenvalues

$$E = e_r + e_{r'} \tag{2.10}$$

thus compensating the vanishing numerator in (2.9) by a vanishing denominator. These wavefunctions are given in (2.4). Excluding such trivial solutions from present considerations, let us now consider those solutions for which $F(00) \neq 0$. First, we calculate the magnitude of this parameter required to normalize the wavefunctions:

$$1 = \sum_{i,j} |F(ij)|^2 = |F(00)|^2 U^2 \sum_{r,r'} \frac{|f_r(0) f_{r'}(0)|^2}{(E - e_r - e_{r'})^2},$$

i.e.,

$$F(00) = \left\{ U^2 \sum_{r,r'} \frac{|f_r(0) f_{r'}(0)|^2}{(E - e_r - e_{r'})^2} \right\}^{-\frac{1}{2}}, \tag{2.11}$$

where we have made use of the orthogonality relation,

$$\sum_i f_r^*(i) f_{r'}(i) = \delta_{r,r'}. \tag{2.12}$$

Next, we solve for the eigenvalues E by setting $R_i = R_j = 0$ in the eigenvalue equation (2.9) and obtain:

$$\frac{1}{U} = \sum_{r,r'} \frac{|f_r(0) f_{r'}(0)|^2}{E - e_r - e_{r'}}. \tag{2.13}$$

This is a standard eigenvalue equation with the usual *interlacing* properties: there is an eigenvalue E between each pair of neighboring unperturbed energy levels $E(r, r')$ given in (2.5). Thus all *but one* of the energy levels may be displaced by at most $O(1/N)$, no matter how large U is allowed to become. The exception is the 2-particle bound state, which corresponded to $r = r' = 0$ for the unperturbed electrons. If we denote its energy eigenvalue by E_0, we must have

$$2e_0 < E_0 < e_0 + e_1 \equiv E_1. \tag{2.14}$$

The first inequality reflects the fact that a repulsive interaction can only raise the energy,[9] and the second is proved by setting $E_0 = e_0 + e_1 - x$. As x is varied from the value $e_1 - e_0$ to zero, the right-hand side of (2.13) varies from $+\infty$ to $-\infty$, which ensures that at some intermediate value a solution to the equation *must* exist.

What we are to discover is that, in the limit $N \to \infty$ *the range of* x required to change the right-hand side of (2.13) from some finite positive value to $-\infty$ *is only* $O(1/N)$. Consequently, the solution E_0 of (2.13) must have the general features shown in Fig. 1: it is an increasing function of U below a certain value denoted U_c. For values of U larger than this, $E_0 = e_0 + e_1 = \text{const}$, in the limit $N \to \infty$. As a result, the ground-state energy is a nonanalytic, albeit continuous, function of U at $U = U_c$, *which defines* U_c *as*: precisely the magnitude of the repulsive potential required to singly ionize the impurity. *Thus, for* $U \geq U_c$, *the impurity possesses a magnetic moment of one Bohr magneton.*

The above is not a general property of the above eigenvalue equation, but follows, rather, from detailed consideration of both numerator and denominator in this equation. We find it important to consider the normalization of the unperturbed states $f_r(i)$, so as to find their amplitudes at the origin, and we find that the continuum states behave differently from the bound state in one very significant way.

Thus, we note that, while the bound-state amplitude at the impurity site is $f_0(0) \sim 0(1)$, the scattering-state amplitudes are $O(N^{-\frac{1}{2}})$, and we therefore write

$$|f_r(0)|^2 \equiv n^2(r)/N, \qquad r \geq 1, \tag{2.15}$$

[9] Differentiating (2.13) with respect to U yields $dE/dU \geq 0$ for *all* eigenvalues E.

which defines $n(r)$ a quantity $O(1)$. The eigenvalue equation (2.13) now reads

$$\frac{1}{U} = \frac{|f_0(0)|^4}{E - 2e_0} + |f_0(0)|^2 \frac{2}{N} \sum_{r\geq 1} \frac{n^2(r)}{E - e_0 - e_r} + \left(\frac{1}{N}\right)^2 \sum_{\substack{r\geq 1 \\ r'\geq 1}} \frac{n^2(r)n^2(r')}{E - e_r - e_{r'}}. \quad (2.16)$$

In solving this equation for E_0, it is permissible to proceed to the limit $N \to \infty$ and replace sums by integrals, provided E_0 does not exceed $E_1 \equiv e_0 + e_1$. Let us define U_c as that value of U for which $E_0 = E_1$. Then, $E_0 = E_1$ for the entire range $U_c \leq U \leq \infty$, whereas for $U < U_c$ the correct solution is obtained from the equation

$$\frac{1}{U} = \frac{|f_0(0)|^4}{E_0 - 2e_0} + 2|f_0(0)|^2 \int dr \frac{g(r)n^2(r)}{E_0 - e_0 - e(r)} + \iint dr\, dr' \frac{g(r)g(r')n^2(r)n^2(r')}{E_0 - e(r) - e(r')}, \quad U \leq U_c, \quad (2.17)$$

where $g(r) \equiv$ density of states factor required to change a sum to an integral.

If we interpret the integrals as principal parts integrations, this equation also has a (spurious) solution for E_0 in the range $U > U_c$, indicated by the dotted line in Fig. 1, which merely shows that *the limiting process $N \to \infty$ must be taken carefully*, for we have already seen that the correct solution in this range is $E_0 = E_1$.

We may also examine the results of perturbation theory, by expanding the exact solution in powers of U. Letting

$$E_0 = 2e_0 + y$$

define the energy shift y, we readily solve for this quantity in (2.17):

$$y = |f_0(0)|^4 U(1 - UQ)^{-1} = |f_0(0)|^4 U(1 + UQ + \cdots),$$

where Q, the correlation terms in (2.17), are given by

$$Q = 2|f_0(0)|^2 \int dr \frac{g(r)n^2(r)}{E_0 - e_0 - e(r)} + \iint dr\, dr' \frac{g(r)g(r')n^2(r)n^2(r')}{E_0 - e(r) - e(r')}$$

and, in the Rayleigh–Schrödinger perturbation theory, may be evaluated using approximate values of y, computed to the desired order in powers of U. Clearly, the expansion ceases to exist, and perturbation theory becomes meaningless, once

$$U|Q| \geq 1.$$

The Brillouin–Wigner perturbation theory, in which Q is not approximated, appears to have a somewhat larger radius of convergence. Further examination of this point would be of interest, but is outside the scope of the present investigation.

We now proceed to a very useful simplification, which enables the double integral in (2.17) to be formally eliminated from the theory. This is especially valuable for numerical computations.

Simplification of some Integrals

We define $I_N(W_0)$ as

$$I_N(W_0) \equiv -\frac{1}{N} \sum_{r\geq 1} \frac{n^2(r)}{W_0 - e_r}, \qquad W_0 \leq e_1, \quad (2.18)$$

and the limiting function $I(W_0) \equiv I_\infty(W_0)$ as

$$I(W_0) = -\int dr \frac{g(r)n^2(r)}{W_0 - e(r)}. \quad (2.19)$$

The negative signs are introduced to make $I > 0$. Next, we consider the *ground-state eigenvalue* W_0 of the *one-electron* Hamiltonian with the *original* impurity potential $-v(R_i)$ *plus* a perturbing *one*-body potential $\lambda\delta_{i,0}$ *added thereto*, resulting in $-v(R_i) + \lambda\delta_{i,0}$. It may be assumed[7] that we know W_0 as a function of λ, or conversely, that we know $\lambda(W_0)$ [this is of course easiest if $v(R_i)$ itself was nonvanishing only at the origin, as in the example of the following section]. At any rate, we know $W_0 \leq e_1$. We may use the *one*-body Green's function to obtain the integral equation,

$$\frac{1}{\lambda(W_0)} = \frac{|f_0(0)|^2}{W_0 - e_0} + \frac{1}{N} \sum_{r\geq 1} \frac{n^2(r)}{W_0 - e_r} \quad (2.20)$$

by complete analogy to the preceding calculation. But now we have the advantage of knowing W_0 and $\lambda(W_0)$, and thus can solve for

$$I_N(W_0) = \frac{|f_0(0)|^2}{W_0 - e_0} - \frac{1}{\lambda(W_0)} \quad (2.21)$$

and we may also easily take the limit $N \to \infty$ to obtain $I(W_0)$ therefrom. Substitution into (2.17) yields the following, simplified, equation[10]

$$\frac{1}{U} = \frac{|f_0(0)|^2}{\lambda(E_0 - e_0)} + \int dr\, g(r) \frac{n^2(r)}{\lambda(E_0 - e_r)}, \quad (2.22)$$

[10] For comparison with other results and recent theories, the following references may be useful: L. D. Fadeev, *Mathematical Aspects of the Three-Body Problem in the Quantum Scattering Theory* (Danial Davey & Company, Inc., New York, 1965); J. Callaway, J. Math. Phys. 5, 783 (1964), and Phys. Rev. 140, A618 (1965); G. V. Skorniakov and K. A. Ter-Martirosin, Zh. Eksperim. i Teor. Fiz. 31, 775 (1956) [English transl.: Soviet Phys.—JETP 4, 648 (1957)]; L. Eyges, J. Math. Phys. 6, 1320 (1965).

which is the desired alternate, and fully equivalent version of (2.17).

[Equation (2.22) may be (crudely) interpreted as an effective, Hartree–Fock type, one-body repulsion, given by a constant $\lambda(W)$,

$$\lambda = U\,|f_0(0)|^2$$

with the integral in (2.22) representing the effects of *correlations*. We see that the Hartree–Fock approximation would be exact if U were replaced by a smaller interaction (the integral can be shown to be negative)

$$\tilde{U} = \left\{U^{-1} - \int dr\; g(r)n^2(r)/\lambda(E_0 - e(r))\right\}^{-1} < U,$$

but since this involves an integral over the "effective coupling constant" λ, $\tilde{U}$ is in general just as difficult to obtain as an exact solution to the problem. Note that, when v is much larger than the band width, the correlation energy becomes negligible and the Hartree–Fock theory is correct for all values of U.]

3. LINEAR CHAIN AS EXPLICITLY SOLUBLE EXAMPLE

The appearance of a localized magnetic moment in the one-dimensional case is all the more striking because of the theorems[3] that the ground-state magnetic moment vanishes. For this reason, it would be of great interest to generalize the present analysis to examine the case of more than two electrons, but we have not completely succeeded in this as yet.

We consider the nearest neighbor coupling, i.e., $K(0) = 1$ and $K(\pm a) = -\frac{1}{2}$ corresponding to the band structure

$$\epsilon(k) = 1 - \cos ka, \tag{3.1}$$

and an attractive potential localized at the origin

$$-v(R_i) = -v\delta_{i,0}, \qquad v \geq 0. \tag{3.2}$$

The one-particle eigenfunctions with a node at the impurity are simply

$$f_r(n) = \left(\frac{2}{N+1}\right)^{\frac{1}{2}} \sin k_r n \tag{3.3}$$

with energies independent of the interaction,

$$e_r = 1 - \cos k_r, \tag{3.4}$$

where periodic boundary conditions imposes

$$\sin k_r(\tfrac{1}{2}N + 1) = \sin k_r(-\tfrac{1}{2}N),$$

which results in

$$k_r = 2\pi r/(N+1), \qquad r = 1, 2, \cdots, \tfrac{1}{2}N. \tag{3.5}$$

Precisely because of the vanishing amplitude at the impurity, however, these functions do not appear in the equations for the ground state of the interacting system. The even eigenfunctions are of course derived from the cosine functions and, introducing the phase shifts φ_r, may be written as

$$f_r(n) = C_r^{-\frac{1}{2}} \cos(k_r\,|n| + \varphi_r). \tag{3.6}$$

The energy is also given by (3.4), but the wave vectors k_r must be recalculated. This is done by studying the $n = 0$ amplitude equation

$$e_r \cos\varphi_r = \cos\varphi_r - \cos(k_r + \varphi_r) - v\cos\varphi_r. \tag{3.7}$$

We use (3.4) to eliminate e_r and obtain

$$\tan\varphi_r = v/\sin k_r, \qquad 0 \leq \varphi_r \leq \pi. \tag{3.8}$$

Periodic boundary conditions give a second relation between φ and k

$$k_r = (2\pi r - 2\varphi_r)/(N+1), \quad r = 1, 2, \cdots, \tfrac{1}{2}N. \tag{3.9}$$

In the limit $N \to \infty$ the normalization constants C_r are independent of r, and are given by

$$C_r = \tfrac{1}{2}N. \tag{3.10}$$

So far we have N out of the total $N+1$ eigenfunctions; the missing one is the bound-state solution, which decays exponentially from the origin and is therefore independent of boundary conditions for large N. For an infinite chain, it has the form

$$f_0(n) = C_0^{-\frac{1}{2}} \exp -\alpha\,|n|, \qquad \alpha \geq 0, \tag{3.11}$$

hence an energy,

$$e_0 = 1 - \cosh\alpha \tag{3.12}$$

a finite amount below the continuum. The eigenvalue is once more determined by the $n = 0$ amplitude equation

$$e_0 = 1 - \exp(-\alpha) - v, \tag{3.13}$$

which, combined with the preceding, yields

$$\sinh\alpha = v \qquad \text{(meaningful only for } v \geq 0) \tag{3.14}$$

or

$$e_0 = 1 - (1+v^2)^{\frac{1}{2}} \quad \text{(note that } e_0 < 0). \tag{3.15}$$

Finally one determines the normalization constant

$$C_0 = 1 + 2\sum_{n>0} \exp -2\alpha n = \coth\alpha. \tag{3.16}$$

All these formulas are valid to within an error $O(\exp -\alpha N)$, and so should hold, with a reasonable choice of parameters, for all but the shortest chains.

Equations (3.6) and (3.8)–(3.10) yield $f_r^2(0)$ [i.e.,

$n^2(r)]$ for the continuum states $r \geq 1$. The bound-state amplitude at $n = 0$ is given by Eqs. (3.11), (3.14), and (3.16). The results are

$$|f_0(0)|^2 = v/(1+v^2)^{\frac{1}{2}}$$

$$\text{and} \quad n^2(r) = 2\left[1 + \frac{v^2}{\sin^2 k_r}\right]^{-1}. \qquad (3.17)$$

Finally, we can readily obtain $\lambda(W_0)$ required for the simplified form (2.22) of the eigenvalue equation, and find

$$\lambda(W_0) = v - [(1 - W_0)^2 - 1]^{\frac{1}{2}}. \qquad (3.18)$$

The eigenvalue equation in question now reads

$$\frac{1}{U} = \frac{v}{(1+v^2)^{\frac{1}{2}}(v - \{[2 - (1+v^2)^{\frac{1}{2}} - E_0]^2 - 1\}^{\frac{1}{2}})} + \frac{1}{\pi}\int_0^\pi \frac{dk\,(1 + v^2/\sin^2 k)^{-1}}{v - [(2 - \cos k - E_0)^2 - 1]^{\frac{1}{2}}} \qquad (3.19)$$

and must be solved for E_0 only in the range

$$2 - 2(1+v^2)^{\frac{1}{2}} \leq E_0 \leq 1 - (1+v^2)^{\frac{1}{2}}. \qquad (3.20)$$

As U is increased from zero to a value U_c, E_0 increases from the left-hand value of the inequality to the right-hand value. For U greater than U_c, Eq. (3.19) is no longer valid and we have,

$$E_0 \equiv 1 - (1+v^2)^{\frac{1}{2}}(\text{ind}^t \text{ of } U \text{ for } U \geq U_c). \qquad (3.21)$$

[This is but repeating the observations made following Eq. (2.14).]

To obtain the magnitude of U_c, we replace E_0 in (3.19) by the upper bound in (3.20):

$$\frac{1}{U_c} = \frac{1}{(1+v^2)^{\frac{1}{2}}} + \frac{1}{\pi} \times \int_0^\pi \frac{dk\,(1 + v^2/\sin^2 k)^{-1}}{v - \{[1 - \cos k + (1+v^2)^{\frac{1}{2}}]^2 - 1\}^{\frac{1}{2}}}. \qquad (3.22)$$

The solution is shown in Fig. 2. From this figure, it is clear that, if v is smaller than a certain value (denoted v_0), then the right-hand side of (3.22) is negative, signifying that no value of the interparticle repulsion will ionize the impurity. This is further illustrated in Fig. 3.

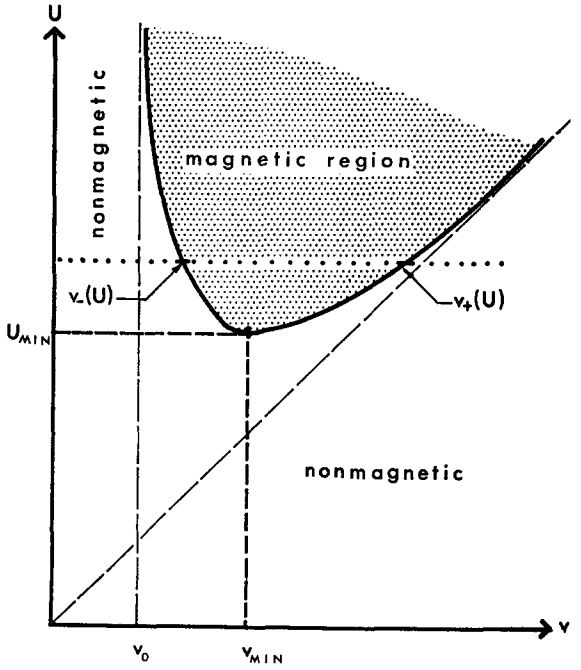


FIG. 2. Schematic solution of Eq. (3.22) for one-dimensional model. For $v \leq v_0$ no finite 2-body repulsion can magnetize the impurity ground state, which will always have a 2-body $S = 0$ bound state. For $v > v_0$, the impurity can be ionized (acquiring 1 Bohr magneton in the ground state) by $U > U_c$, where U_c is the solid line shown. (Dashed lines indicate the asymptotes.) Region of magnetic behavior is indicated by shaded region above the curve U_c, which has its minimum value $U_{\min}$ at $v_{\min}$. The points $v_\pm(U)$ are defined in Fig. 3.

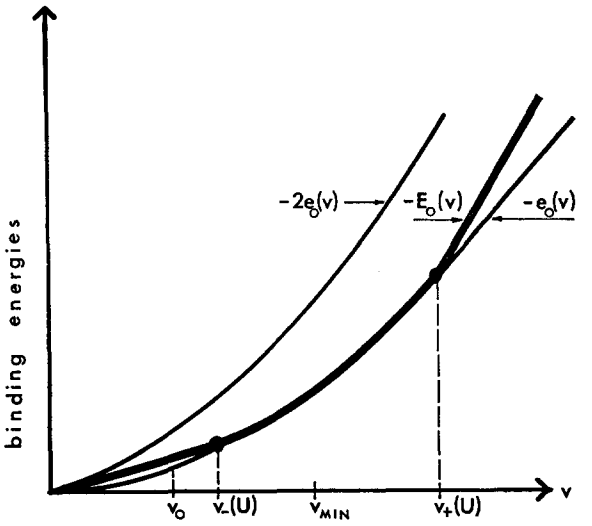


FIG. 3. Schematic solution of Eq. (3.22) for fixed interparticle repulsive potential, U, and variable attractive potential, v. Compare with Fig. 2, especially the points $v_-(U)$ and $v_+(U)$. Note that $v_-(U)$ is always greater than the fixed point v_0, regardless of the magnitude of U, and that v_+ and v_- straddle $v_{\min}$. Finally, note that one can easily prove that dE_0/dv, as well as E_0, is a *continuous* function of v, and therefore $(dE_0/dv) = (de_0/dv)$ at $v = v_\pm$.

4. SUMMARY

In examining the ground state of two interacting electrons about an attractive impurity, we found a variety of possible behavior (as summarized in the first two figures). Generally, there exists a finite critical repulsive interaction U_c such that, if U exceeds U_c, the two electrons cannot be simultaneously bound in the neighborhood of the impurity, and one of them spontaneously ionizes in the ground

state. When this is the case, only one electron, with its single Bohr magneton, is found in this neighborhood. For U not exceeding U_c, however, the electrons are *both* bound to the impurity which therefore has net spin $S = 0$ in the ground state.

However, for the linear chain, we have also found that, when the attractive impurity potential v is *sufficiently small* (less than an amount v_0, as shown in Fig. 2), no finite electronic repulsion can ionize the impurity, i.e., $U_c = \infty$. The reason for this is intuitively obvious; when the impurity potential is weak, the bound-state orbits are very large so that the charge clouds of the two electrons have very little interaction. It then requires little correlation energy to keep the two particles out of each other's way in this limit, and the result is a nonmagnetic ground state for *all values* of the interaction U. In the opposite limit of very large attractive impurity potential v, the electrons become very tightly bound and require $U \gtrsim v$ to be ionized. (Here the approximate Hartree–Fock theory[5,6] should be most reliable.) There exists an *optimum* magnitude of v (denoted by $v_{\min}$, Fig. 2) requiring a minimal interaction energy U (denoted by min) to become magnetic. The detailed nature of these results must be modified somewhat if they are to apply to three dimensions, as bound states do not exist for arbitrarily small v, but otherwise qualitatively similar comments can be made in three dimensions on the basis of our exact results, Eqs. (2.13) or (2.16) and (2.17), or (2.22).

Considering the simplicity of the present model, such a variety of behavior is truly surprising, and illustrative of the virtue of exactly soluble models in the analysis of the complex phenomena of electronic magnetism. Extensions of the present model which have some appeal might include the extension to more than two particles and/or more than one impurity, and, closer to the present analysis, an *explicit* quantitative calculation of the eigenvalues and eigenfunctions in three dimensions, and a calculation of the anomalous scattering cross section[4] in the magnetic case when $U > U_c$.

The capture cross section of traps may also be related to the two-body interaction, which provides a mechanism whereby one electron may become bounded by releasing its energy to a second carrier. The scattering properties in our model are, in any event, expected to have considerably more structure than in the one-electron theory of solids because of the two distinct continua, Fig. 1. We hope to return to these topics subsequently.

Acoustic Scattering from an Interface between Media of Greatly Different Density

K. M. MITZNER
Northrop Norair, Hawthorne, California
(Received 31 March 1966)

The problem of acoustic scattering from a curved interface between two homogeneous media is formulated as two integral equations relating the normal velocity and the velocity potential at the interface. The equations are so chosen as to minimize coupling. When γ, the ratio of the densities, goes to zero, the equations decouple, one becoming the equation of a soft-boundary problem, the other of a hard-boundary problem. For small γ, an approximate solution is constructed by perturbation methods from solutions to the related soft- and hard-boundary problems.

1. INTRODUCTION

MANY practical acoustics problems can be satisfactorily idealized to scattering from an interface along which two homogeneous fluids are in direct contact. Frequently one medium is much more tenuous than the other but not so tenuous that γ, the ratio of the densities,[1] can be set to zero in reflection calculations. In other cases, as when the media are a gas and a liquid, γ is negligibly small in reflection calculations but important in calculating the transmission from one medium into the other.

Thus motivated, we develop here a general integral equation formulation of time-harmonic acoustic scattering from an interface of arbitrary shape between homogeneous media. We then specialize to the case of small γ and, by perturbation methods, construct an approximate solution in terms of the solutions to related soft- and hard-boundary problems.

More specifically, we first formulate the general two-medium problem as two coupled integral equations, the unknowns being the normal velocity and the velocity potential at the interface. The formulation is optimized to give minimal coupling of the two equations. When γ goes to zero, the equations decouple, one becoming the equation for the field in the dense medium with a soft boundary, the other becoming the equation for the field in the tenuous medium with a hard boundary. For small γ, the decoupling is not complete; however, by expanding the fields in formal power series in γ and collecting terms of the same order, we obtain a sequence of equations which can be solved one by one. The terms of the series for the normal velocity are solutions of soft-boundary problems, and the terms of the series for the velocity potential are solutions of hard-boundary problems. The formal series will converge in some cases, but will, in general, be asymptotic.

For complicated interface geometries, especially for problems which must be solved by numerical methods, it is much easier to calculate the first few terms of the perturbation series than to solve the coupled equations directly. Furthermore, the perturbation solution gives us a physical insight into the effect of varying γ.

In Sec. 2 we introduce the integral equations for scattering in a single medium and the boundary conditions at an interface. From these we construct in Sec. 3 the minimal coupling integral equation formulation of the two-medium problem. The perturbation technique for small γ is developed in Sec. 4. Then, in Sec. 5, we work out a simple example, scattering of a plane wave incident on a plane interface from the tenuous side; in this case the perturbation series converges provided the angle of incidence is not too close to grazing.

To avoid unnecessary mathematical detail, we limit our discussion *a priori* to well-defined physically meaningful problems with unique solutions. Furthermore, we consider only interfaces sufficiently smooth and source functions sufficiently well-behaved so that we can use without modification the theorems on limits of potentials given by Kellogg in his classical treatise[2]; the most important constraint thus imposed is that the interface must have continuous curvature.

Harmonic time dependence $e^{-i\omega t}$ is to be understood everywhere.

2. BASIC EQUATIONS AND BOUNDARY CONDITIONS

The Field in a Homogeneous Region

Consider a volume V bounded by surface S with outward normal $\mathbf{n}_0$. Let the volume be filled with a homogeneous medium of density ρ in which the

[1] Without loss of generality, we will always choose $\gamma \leq 1$.

[2] O. D. Kellogg, *Foundations of Potential Theory* (Dover Publications, Inc., New York, 1953), especially Chap. VI.

speed of sound is c. We define in the usual manner a velocity potential U related to the particle velocity $\mathbf{V}$ and the pressure disturbance P by

$$\mathbf{V} = \nabla U, \qquad P = i\omega\rho U. \tag{2.1}$$

This potential satisfies the Helmholtz equation

$$(\nabla^2 + k^2)U = (1/\rho)Q, \tag{2.2}$$

where

$$k = \omega/c \tag{2.3}$$

is the wave number and $Q(\mathbf{r})$ the source term.

The field in V can be expressed in terms of its values on S with the aid of the Green's function for an unbounded volume,

$$G_0(\mathbf{r}, \mathbf{r}') = -(\rho/4\pi R)e^{ikR}, \qquad R = |\mathbf{r} - \mathbf{r}'|, \tag{2.4}$$

which satisfies

$$(\nabla^2 + k^2)G_0 = \rho\,\delta(\mathbf{r} - \mathbf{r}') \tag{2.5}$$

and obeys the radiation condition at infinity. We first define the incident field U_0 as the field which would be produced in an unbounded volume by the sources in V; using the Green's function, this can be written[3]

$$U_0 = (1/\rho^2)\int_V dV'\, G_0 Q'. \tag{2.6}$$

Then, applying Green's second identity[4] to U and G_0, we obtain the well-known equation

$$\rho(U - U_0) = \int_S dS'\,[(\partial G_0/\partial n_0')U' - G_0(\partial U'/\partial n_0')], \tag{2.7}$$

valid at points $\mathbf{r}$ in the interior of V.

Applying Green's first identity[4] to $\mathbf{e}\cdot\nabla U$ and G_0, with $\mathbf{e}$ an arbitrary constant vector, we find

$$\rho(\nabla U - \nabla U_0) = \int_S dS'\,[-\nabla' G_0 \times (\mathbf{n}_0' \times \nabla' U') + k^2 G_0 U' \mathbf{n}_0' + (\nabla' G_0)\,\partial U'/\partial n_0'] \tag{2.8}$$

at an interior point of V; the details of the derivation are given in the Appendix.

Once U and $\partial U/\partial n_0$ are known on S, we can find U and ∇U everywhere from Eqs. (2.7) and (2.8). Thus the scattering problem reduces to determination of the two surface fields. It is well-known that for $\mathbf{r}$ on S these fields satisfy the integral equation

$$\tfrac{1}{2}\rho U - \rho U_0 = \int_{\bar{S}} dS'\,[(\partial G_0/\partial n_0')U' - G_0(\partial U'/\partial n_0')] \tag{2.9}$$

obtained by taking the limit as $\mathbf{r}$ approaches S in Eq. (2.7). Similarly, taking the limit of Eq. (2.8) in the manner described in the Appendix, we find

$$\tfrac{1}{2}\rho\,\partial U/\partial n_0 - \rho\,\partial U_0/\partial n_0 = \int_{\bar{S}} dS'\,[(\mathbf{n}_0 \times \nabla G_0)\cdot(\mathbf{n}_0' \times \nabla' U') + k^2(\mathbf{n}_0\cdot\mathbf{n}_0')G_0 U'] - \int_{\bar{S}} dS'\,(\partial G_0/\partial n_0)(\partial U'/\partial n_0'), \tag{2.10}$$

where $\int_{\bar{S}}$ is the integral over the punctured surface, that is, the limit as $\Delta \to 0$ of the integral over that part of S outside a sphere of radius Δ centered at $\mathbf{r}$. This last equation was first derived by Maue,[5] using a rather more complicated approach.

We can obtain an alternative expression to (2.10) by multiplying Eq. (2.7) by $(\mathbf{n}_0\cdot\nabla)$ and letting $\mathbf{r}$ approach S. We find

$$\tfrac{1}{2}\rho\,\partial U/\partial n_0 - \rho\,\partial U_0/\partial n_0 = \int_{\bar{S}} dS'\left\{\left[\frac{\partial^2}{\partial n_0\,\partial n_0'}\left(G_0 + \frac{\rho}{4\pi R}\right)\right]U' - \frac{\partial G_0}{\partial n_0}\frac{\partial U'}{\partial n_0'}\right\} - \frac{\rho}{4\pi}\lim\left\{\frac{\partial}{\partial n_0}\int_S dS'\left[\frac{\partial}{\partial n_0'}\left(\frac{1}{R}\right)\right]U'\right\}, \tag{2.11}$$

where it can easily be verified that the first integral exists. Both forms (2.10) and (2.11) will prove useful below.

It is convenient at this point to introduce the operators L_i defined by

$$L_a X = -\int_{\bar{S}} dS'\,(1/\rho)(\partial G_0/\partial n_0')X', \tag{2.12a}$$

$$L_b X = \int_S dS'\,(1/\rho)G_0 X', \tag{2.12b}$$

$$L_c X = -\int_{\bar{S}} dS'\,(1/\rho)[(\mathbf{n}_0 \times \nabla G_0)\cdot(\mathbf{n}_0' \times \nabla' X') + k^2(\mathbf{n}_0\cdot\mathbf{n}_0')G_0 X'] = -\int_S dS'\left[\frac{\partial^2}{\partial n_0\,\partial n_0'}(G_0/\rho + 1/4\pi R)\right]X' + \frac{1}{4\pi}\lim\left\{\frac{\partial}{\partial n_0}\int_S dS'\left[\frac{\partial}{\partial n_0'}\left(\frac{1}{R}\right)\right]X'\right\}, \tag{2.12c}$$

[3] For simplicity we indicate functions of the integration point $\mathbf{r}'$ with a prime. Joint functions of $\mathbf{r}$ and $\mathbf{r}'$ are explicitly defined as such. In all other cases the observation point argument $\mathbf{r}$ is to be assumed.

[4] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Sec. 3.3.

[5] A. W. Maue, Z. Physik **126**, 601 (1949).

$$L_d X = \int_S dS' \, (1/\rho)(\partial G_0/\partial n_0) X'. \tag{2.12d}$$

Then Eq. (2.9) can be written in the compact form

$$M = (\tfrac{1}{2} + L_a)U + L_b \, \partial U/\partial n_0 - U_0 = 0, \tag{2.13}$$

and Eqs. (2.10) and (2.11) can both be represented by

$$N = L_c U + (\tfrac{1}{2} + L_d) \, \partial U/\partial n_0 - \partial U_0/\partial n_0 = 0. \tag{2.14}$$

It is clear that Eqs. (2.13) and (2.14) are not independent, for, if they were, we could solve them with no reference to boundary conditions; in fact, both are equivalent to Eq. (2.2) and thus to each other.

Boundary Conditions

Now let us consider two homogeneous media, m_1 occupying volume V_1 and m_2 occupying volume V_2, separated by a boundary S with unit normal

$$\mathbf{n} = \mathbf{n}_{01} = -\mathbf{n}_{02} \tag{2.15}$$

directed from V_1 to V_2. We assume that the media are in direct contact at S, that is, that the effect of the membrane or surface tension layer separating the fluids is negligible. Then pressure and normal velocity are continuous across S provided there are no boundary-layer sources. More generally, the boundary conditions are

$$\partial U_2/\partial n - \partial U_1/\partial n = V_\Delta, \tag{2.16a}$$

$$\rho_2 U_2 - \rho_1 U_1 = -(1/i\omega) P_\Delta, \tag{2.16b}$$

where the discontinuity in normal velocity $V_\Delta(\mathbf{r})$ and the pressure discontinuity $P_\Delta(\mathbf{r})$ represent the effect of boundary-layer sources.

3. MINIMAL COUPLING FORMULATION OF SCATTERING AT AN INTERFACE

We are now ready to formulate the problem of scattering at an interface between two media. In the nondegenerate case, that is, when

$$\gamma = \rho_2/\rho_1 \neq 0, \tag{3.1}$$

the problem is effectively solved once the four functions U_1, U_2, $\partial U_1/\partial n$, and $\partial U_2/\partial n$ are determined on S. The boundary conditions (2.16) supply two of the four equations necessary for this determination.

The other two equations are not unique. We have available two pairs of equivalent equations $M_1 = 0$, $N_1 = 0$, and $M_2 = 0$, $N_2 = 0$, obtained by writing Eqs. (2.13) and (2.14) for both media, and we may combine these four equations in an infinite variety of ways so as to produce two independent equations. Specifically, the formulation can be completed using any two equations of form

$$M_2 + \alpha M_1 = 0, \tag{3.2}$$

$$N_1 - \beta N_2 = 0, \tag{3.3}$$

with α and β finite.

Although all resulting formulations are mathematically equivalent, they are by no means equally desirable for either theoretical or computational purposes. In fact, there is one specific formulation which in most situations simplifies both the expression and the solution of the two medium problem. This we call the minimal coupling formulation because (as we show below) in a very significant sense, it minimizes the coupling between the potential and the normal velocity.

The minimal coupling formulation is obtained by setting

$$\alpha = 1, \qquad \beta = \gamma \tag{3.4}$$

in Eqs. (3.2) and (3.3). Using Eq. (2.16) to eliminate U_1 and $\partial U_2/\partial n$, we then obtain the two basic equations for U_2 and $\partial U_1/\partial n$:

$$\tfrac{1}{2}(1+\gamma)U_2 + (L_{a2} + \gamma L_{a1})U_2 + L_e(\partial U_1/\partial n) = U^*_{02} + U^*_{01} \tag{3.5}$$

and

$$\tfrac{1}{2}(1+\gamma)(\partial U_1/\partial n) + (L_{d1} + \gamma L_{d2})(\partial U_1/\partial n) + \gamma L_f U_2 = U^*_{0n1} + \gamma U^*_{0n2}. \tag{3.6}$$

Here the L_a and L_d are defined by Eqs. (2.12a), (2.12d), and (2.15); the other operators are given by

$$L_e X = (L_{b1} - L_{b2})X = \frac{i}{2\pi} \int_S dS' \, e^{ik_+R}[(\sin k_-R)/R]X' \tag{3.7}$$

and

$$\begin{aligned} L_f X &= (L_{c1} - L_{c2})X \\ &= -\frac{i}{2\pi} \int_S dS' \, \frac{\partial^2}{\partial n \, \partial n'} e^{ik_+R}[(\sin k_-R)/R]X' \\ &= -\frac{i}{2\pi} \int_S dS' \, (e^{ik_+R}/R^3) \\ &\quad \times \{[(1 - ik_+R)\sin k_-R - k_-R \cos k_-R] \\ &\quad \times [\mathbf{n}\cdot\mathbf{n}' - (3 - ik_+R)(\mathbf{n}\cdot\mathbf{s})(\mathbf{n}'\cdot\mathbf{s})] \\ &\quad + [(k_-R)^2 \sin k_-R - ik_+R \\ &\quad \times (\sin k_-R + k_-R \cos k_-R)](\mathbf{n}\cdot\mathbf{s})(\mathbf{n}'\cdot\mathbf{s})\}X'; \end{aligned} \tag{3.8}$$

and the source terms are

$$U_{02}^* = U_{02} + \int_S dS'\,(G_{02}/\rho_2)V_\Delta', \tag{3.9}$$

$$U_{01}^* = U_{01} - \tfrac{1}{2}P_\Delta/i\omega\rho_1 + (1/i\omega\rho_1^2)\int_S dS'\,(\partial G_{01}/\partial n')P_\Delta', \tag{3.10}$$

$$U_{0n1}^* = \partial U_{01}/\partial n + (1/i\omega\rho_1^2)\int_S dS'\,[(\mathbf{n}\times\nabla G_{01})\cdot(\mathbf{n}'\times\nabla' P_\Delta') + k_1^2(\mathbf{n}\cdot\mathbf{n}')G_{01}P_\Delta'], \tag{3.11}$$

and

$$U_{0n2}^* = \partial U_{02}/\partial n - \tfrac{1}{2}V_\Delta + \int_S dS'\,(1/\rho_2)(\partial G_{02}/\partial n)V_\Delta'. \tag{3.12}$$

In the expressions for L_e and L_f, we have used

$$k_+ = \tfrac{1}{2}(k_2 + k_1), \qquad k_- = \tfrac{1}{2}(k_2 - k_1) \tag{3.13}$$

and

$$\mathbf{s} = (\mathbf{r} - \mathbf{r}')/R = \nabla R = -\nabla' R. \tag{3.14}$$

The first form of Eq. (2.12c) is used in obtaining U_{0n1}^* and the second form in obtaining L_f; Chap VI, Theorem 10 of Ref. 2 is used to cancel the limit integrals in the latter calculation.

In order to see the sense in which the above formulation minimizes coupling and why this is important, we must introduce some elementary ideas from the theory of integral equations. An equation is of the first kind in a variable if the variable appears only implicitly, that is, only under the integral sign. An equation is of the second kind in a variable if the variable appears both implicitly and explicitly. Either type of integral equation can be considered as equivalent to an infinite set of simultaneous linear algebraic equations. Equations of the second kind are thus preferable, in both analytical and numerical work, for the same reasons that a matrix with relatively large diagonal elements is preferable to one with all elements of about the same magnitude.

Now Eq. (3.5) is of second kind in U_2 and first kind in $\partial U_1/\partial n$, whereas Eq. (3.6) is of second kind in $\partial U_1/\partial n$ and of first kind in U_2. If L_e and L_f were zero—which is indeed the case when $k_1 = k_2$—we would have two uncoupled equations of second kind. The terms $L_e(\partial U_1/\partial n)$ and $\gamma L_f U_2$ can thus be thought of as coupling terms which complicate the solution of Eqs. (3.5) and (3.6).

If now we compare the expressions for L_f and L_c, we see that L_f is the only operator of form $[L_{c1} - (\beta/\gamma)L_{c2}]$ for which the higher-order singularities in the kernels of the L_c cancel. Only L_f has a kernel which can be integrated without introducing tangential derivatives or special limiting processes as in Eq. (2.12c). Similarly L_e is the only operator of form $(\alpha L_{b1} - L_{b2})$ for which the singularities in the kernels of the L_b cancel. Thus we can characterize the minimal coupling formulation as follows:

> The minimal coupling formulation involves lower-order singularities in the kernels of the coupling terms than does any other formulation generated from Eqs. (3.2) and (3.3).[6]

The cancelation of singularities tends to minimize the overall effect of the coupling terms and to simplify both analytical manipulation and numerical solution of the equations.

Let us turn now to the two degenerate cases $\rho_2 = 0$ and $\rho_1 = \infty$, both corresponding to $\gamma = 0$. When $\rho_2 = 0$, the problem is characterized by the soft-boundary condition

$$\rho_2 U_2 = \rho_1 U_1 - (1/i\omega)P_\Delta = 0. \tag{3.15}$$

The formulation reduces to the single equation

$$(\tfrac{1}{2} + L_{d1})\,\partial U_1/\partial n = U_{0n1}^* \tag{3.16}$$

obtained by substituting Eq. (3.15) in Eq. (2.10). When $\rho_1 = \infty$, the problem is characterized by the hard-boundary condition

$$\partial U_1/\partial n = \partial U_2/\partial n - V_\Delta = 0, \tag{3.17}$$

and the formulation reduces to the single equation

$$(\tfrac{1}{2} + L_{a2})U_2 = U_{02}^* \tag{3.18}$$

obtained by substituting Eq. (3.17) into Eq. (2.9).

We now readily see that, as γ approaches zero, S will look more and more like a soft boundary as seen from V_1 and more and more like a hard boundary as seen from V_2. Starting from this observation we obtain a perturbation technique for small γ in Sec. 4.

4. THE PERTURBATION TECHNIQUE FOR SMALL γ

In the last section we noted the relationsip between the problem of scattering from an interface

[6] The analogous formulation for electromagnetic theory is given in C. Müller, *Grundprobleme der Mathematischen Theorie Elektromagnetischer Schwingungen* (Springer-Verlag, Berlin, 1957), Sec. 23. Coupling is not minimized in the acoustics formulation derived in J. Korringa, J. Math. Phys. 6, 1107 (1965), Sec. 7; here one of the two basic equations is equivalent to Eq. (3.6), but the other corresponds to Eq. (3.2) with $\alpha = 1/\gamma^2$, a choice which does not lead to any simplifications.

when γ is small and the problems of scattering from soft and hard boundaries. We develop now a perturbation technique in which we construct the solution to interface problems with small γ from the solutions to related soft- and hard-boundary problems. First we present the formal derivation and then we discuss its validity and significance.

The formal derivation is straightforward. First we write Eq. (3.6) as

$$(\tfrac{1}{2} + L_{d1})(\partial U_1/\partial n) = U^*_{0n1} + \gamma[U^*_{0n2} - (\tfrac{1}{2} + L_{d2})(\partial U_1/\partial n) - L_f U_2] \quad (4.1)$$

and Eq. (3.5) as

$$(\tfrac{1}{2} + L_{a2})U_2 = U^*_{02} + U^*_{01} - L_e(\partial U_1/\partial n) - \gamma(\tfrac{1}{2} + L_{a1})U_2. \quad (4.2)$$

Since none of the operators L_i or source terms U^*_0 and U^*_{0n} depend on γ, we have thus grouped together on the right all terms of order γ. Next we introduce into the equations formal power series expansions of the unknowns,

$$\partial U_1/\partial n = \sum_{j=0} \gamma^j U^{(j)}_{n1}, \qquad U_2 = \sum_{j=0} \gamma^j U^{(j)}_2, \quad (4.3)$$

and collect terms of the same order in γ.

We thus obtain the equations of the perturbation technique,

$$\begin{aligned} (\tfrac{1}{2} + L_{d1})U^{(j)}_{n1} &= U^{*(j)}_{n1}, \\ (\tfrac{1}{2} + L_{a2})U^{(j)}_2 &= U^{*(j)}_2, \qquad j \geq 0. \end{aligned} \quad (4.4)$$

Here the effective source terms are given by

$$\begin{aligned} U^{*(0)}_{n1} &= U^*_{0n1}, \\ U^{*(1)}_{n1} &= U^*_{0n2} - [(\tfrac{1}{2} + L_{d2})U^{(0)}_{n1} + L_f U^{(0)}_2], \\ U^{*(j)}_{n1} &= -[(\tfrac{1}{2} + L_{d2})U^{(j-1)}_{n1} + L_f U^{(j-1)}_2], \qquad j \geq 2, \end{aligned} \quad (4.5)$$

and

$$\begin{aligned} U^{*(0)}_2 &= U^*_{02} + U^*_{01} - L_e U^{(0)}_{n1}, \\ U^{*(j)}_2 &= -[(\tfrac{1}{2} + L_{a1})U^{(j-1)}_2 + L_e U^{(j)}_{n1}], \quad j \geq 1. \end{aligned} \quad (4.6)$$

The unknowns in Eqs. (4.4) must be solved for in the order $U^{(0)}_{n1}$, $U^{(0)}_2$, $U^{(1)}_{n1}$, $U^{(1)}_2$, etc. Then at each step we have a single equation with a known right-hand side. Comparison with Eqs. (3.16) and (3.18) shows that each $U^{(j)}_{n1}$ is the solution to a soft-boundary problem and each $U^{(j)}_2$ is the solution to a hard-boundary problem. Thus, when the formal series (4.3) converge or give a sufficiently accurate asymptotic approximation, then the problem of scattering from an interface can be reduced to the sequential solution of problems involving soft and hard boundaries.

As $\gamma \to 0$, the limiting solutions for soft and hard boundaries are approached continuously; thus the perturbation theory must be valid, at least asymptotically, for γ sufficiently small.[7] The example of the next section shows that the perturbation series in some cases converges to the exact solution and in others gives rise to asymptotic approximations.

In practice we frequently encounter problems in which U_{01} is the only source. Since $\partial U_1/\partial n$ is given to zeroth order by an equation of form (3.16) and $U_1 = \gamma U_2$ is of higher order, it is meaningful to characterize the situation as scattering from an "almost soft" boundary. Similarly, when U_{02} is the only source, then U_2 is given to zeroth order by an equation of form (3.18) and $\partial U_2/\partial n$ is of higher order; we can characterize this case as an "almost hard" boundary problem.

For a gas–liquid interface, γ is very small and can usually be considered zero in reflection calculations. The field transmitted across an almost soft boundary into the gas m_2 can be calculated accurately from $U^{(0)}_2$ and $U^{(0)}_{n2} = U^{(0)}_{n1}$ without reference to the value of γ. The field transmitted across an almost hard boundary into the liquid m_1 is of order γ and can be calculated accurately from $U^{(1)}_{n1}$ and $U^{(1)}_1 = U^{(0)}_2$.

5. EXAMPLE. SCATTERING FROM AN ALMOST HARD PLANE INTERFACE

Let us now investigate the nature of the perturbation theory by applying it to the well-studied problem of a plane wave incident from the tenuous side on a plane interface. By proper interpretation of the results, we see how, in general, the perturbation series can be either convergent or asymptotic. We also show that an impedance boundary condition does not satisfactorily describe scattering from an almost hard boundary.

Let S be the plane $z = 0$, with $z > 0$ in V_2. Consider a plane wave incident from V_2 at angle θ_i:

$$U_{02} = U_0 \exp\{i\mathbf{k}_2 \cdot \mathbf{r}\} = U_0 \exp\{-ik_2(x \sin\theta_1 + z \cos\theta_1)\}. \quad (5.1)$$

[7] This is not the same as requiring γ to be small compared to unity. The $U_{n1}(j)$ and $U_2(j)$ may be decreasing rapidly with j so that γ is sufficiently small while still greater than unity, or they may be increasing rapidly so that γ can be small compared to unity but still not small enough. The example of Sec. 5 illustrates this point.

Then it is well-known that at the interface

$$\rho U = \frac{2}{1+K\gamma}\,\rho_2 U_0 \exp\{-ik_2 x \sin\theta_i\},$$

$$\partial U/\partial n = -\frac{2iK\gamma}{1+K\gamma}\,k_2\cos\theta_i \times U_0 \exp\{-ik_2 x\sin\theta_i\}, \quad (5.2)$$

with

$$K = [(k_1/k_2)^2\sec^2\theta_i - \tan^2\theta_i]^{\frac{1}{2}}. \quad (5.3)$$

We now show that for $|K\gamma| < 1$ this result can be obtained exactly from the theory of Sec. 4.

This problem is especially simple because $L_a = L_d = 0$. Thus Eqs. (4.4) reduce to

$$U_{n1}^{(j)} = 2U_{n1}^{*(j)}, \qquad U_2^{(j)} = 2U_2^{*(j)}, \quad (5.4)$$

and the effective source terms are also greatly simplified.

The zero-order effective sources are by inspection

$$U_{n1}^{*(0)} = 0, \qquad U_2^{*(0)} = U_{02} = U_0 \exp\{-ik_2 x\sin\theta_i\}. \quad (5.5)$$

To evaluate the higher-order terms we first define

$$E = \exp\{-ik_2(x'-x)\sin\theta_i\} \quad (5.6)$$

and then evaluate the two integrals[8]

$$L_e E = (1/4\pi)\int_0^\infty d\rho'\,(e^{ik_2\rho'} - e^{ik_1\rho'}) \times \int_0^{2\pi} d\phi' \exp\{-i(k_2\sin\theta_i)\rho'\cos\phi'\}$$

$$= \tfrac{1}{2}(k_2\sin\theta_i)^{-\frac{1}{2}}\int_0^\infty d\rho'\,(\rho')^{-\frac{1}{2}}(e^{ik_2\rho'} - e^{ik_1\rho'}),$$

$$J_{01}(k_2\sin\theta_i)\rho'][(k_2\sin\theta_i)\rho']^{\frac{1}{2}} = -\tfrac{1}{2}i\left(\frac{1-K}{K}\right)(k_2\cos\theta_i)^{-1}, \quad (5.7)$$

and

$$L_f E = \frac{i}{2\pi}\int_0^{2\pi} d\phi' \int_0^\infty d\rho' \left[\frac{ik_+\sin k_-\rho'}{\rho'} + \frac{\partial}{\partial\rho'}\left(\frac{\sin k_-\rho'}{\rho'}\right)\right]\exp\{i[k_+\rho' - (k_2\sin\theta_i\cos\phi')\rho']\}$$

$$= -ik_- + \left(\frac{1}{2\pi}\right)k_2\sin\theta_i\int_0^\infty d\rho'\,(1/\rho')\sin(k_-\rho')e^{ik_+\rho'}\int_{-\pi}^{\pi} d\phi''\cos\phi''\exp\{i(k_2\sin\theta_i)\rho'\cos\phi''\}$$

$$= -ik_- + i(k_2\sin\theta_i)^{\frac{1}{2}}\int_0^\infty d\rho'\,(\rho')^{-\frac{1}{2}}\sin(k_-\rho')e^{ik_+\rho'}J_1[(k_2\sin\theta_i)\rho'][(k_2\sin\theta_i)\rho']^{\frac{1}{2}}$$

$$= -\tfrac{1}{2}i(1-K)k_2\cos\theta_i. \quad (5.8)$$

Using these results we obtain

$$U_{n1}^{*(1)} = \partial U_{02}/\partial z - 2L_f U_{02} = -ik_2\cos\theta_i U_{02} - 2U_{02}L_f E = -iKk_2\cos\theta_i U_{02}, \quad (5.9)$$

and the difference equations for the succeeding terms,

$$U_2^{*(j)} = -U_2^{*(j-1)} + i\frac{1-K}{K}\frac{U_{n1}^{*(j)}}{k_2\cos\theta_i}, \quad (5.10a)$$

$$U_{n1}^{*(j)} = -U_{n1}^{*(j-1)} + i(1-K)k_2\cos\theta_i U_2^{*(j-1)}. \quad (5.10b)$$

The pair of equations (5.10) is readily solved to give

$$U_2^{*(j)} = (-K)^j U_{02}, \qquad j \geq 1;$$
$$U_{n1}^{*(j)} = i(-K)^j k_2\cos\theta_i U_{02}, \qquad j \geq 2. \quad (5.11)$$

We thus find

$$\rho U = 2\rho_2 U_{02}\sum_{j=0}^{\infty}(-K\gamma)^j,$$
$$\partial U/\partial n = 2ik_2\cos\theta_i U_{02}\sum_{j=1}^{\infty}(-K\gamma)^j, \quad (5.12)$$

which indeed converge to Eqs. (5.2) for $|K\gamma| < 1$.

We can acquire considerable insight into the nature and limitations of the perturbation technique by a study of the above results. First we note the ratio

$$(\partial U/\partial n)/\rho U = -i\gamma(k_2/\rho_2)[(k_1/k_2)^2 - \sin^2\theta_i]^{\frac{1}{2}}. \quad (5.13)$$

In general k is smaller in the dense medium, so that, no matter how small we make γ, the ratio depends strongly on θ_i. Thus, even in this simple case where there are no internal reflections, the almost hard boundary condition cannot satisfactorily be replaced by a standard impedance boundary condition $\partial U/\partial n = \lambda U$. Such a condition may, however, be valid for the associated almost soft problem.

We next note that for fixed γ we can choose θ_i close enough to grazing so that $|K\gamma| > 1$. If we

[8] In this evaluation we use Ref. 4, Eq. 6.8 (37), and A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Table of Integral Transforms* (McGraw-Hill Book Company, Inc., New York, 1954), Vol. 2, Eq. 8.2 (18) [for Eq. (5.7)] and Eq. 8.4 (15) [for Eq. (5.8)].

admit complex angles of incidence ($\frac{1}{2}\pi + i\theta_I$), then we see that $|K\gamma| > 1$ also holds for complex angles near grazing. Thus the plane wave spectrum of a general source distribution in V_2 will contain a band around grazing for which the perturbation series does not converge. If a given source distribution contains sufficient spectral energy in this band, then the perturbation series for this source also will not converge. Therefore, in general, the perturbation technique gives rise to an asymptotic approximation.

Furthermore, for a concentrated source distribution we would expect best accuracy on that part of the plane below the sources, where grazing waves are least important. By extension, for scattering from a finite body we would expect best accuracy in the "illuminated" region.

6. CONCLUDING REMARKS

The most significant material presented here is the minimal coupling formulation of scattering at a smooth curved interface, given in Sec. 3, and the perturbation technique for the special case of small γ, given in Sec. 4. The minimal coupling formulation is the more fundamental result, for it is the best way to set up most two-medium problems for any range of γ. It can be used as the basis for a numerical approach to the exact problem and as the starting point in developing other approximations, such as for the treatment of small interface irregularities.

The perturbation technique, on the other hand, is of greater immediate importance. It gives a physical insight into the interface scattering problem, a way of directly observing the effect on scattering of changes in γ. Thus we have found with very little effort that transmission from a liquid into a gas is independent of γ and transmission in the opposite direction is proportional to γ; we have also shown that scattering from an almost hard surface cannot be described by a simple impedance boundary condition.

The perturbation technique furthermore reduces the scattering problem to the point where it can be solved numerically for fairly complicated interface geometries using existing methods for soft- and hard-boundary problems.[9] In these methods the integral equation is replaced by a finite set of algebraic equations, and the amount of computation necessary to solve the problem varies roughly as the cube of the number of algebraic equations. For a given accuracy, the exact interface problem requires twice as many equations as a hard- or soft-boundary problem. Thus, for example, the transmission across a gas–liquid interface can be calculated by the perturbation technique with about one fourth the work necessary for an exact solution. When more than one value of γ is to be considered, the computational savings are of course much greater.

The perturbation series converges to the exact solution in some cases but is in general asymptotic. Perturbation solutions with larger domains of convergence and improved rates of convergence can be constructed from this series by various mathematical strategems, such as those discussed by Morse and Feshback.[10] Indeed, we have effectively done this in Sec. 5, where we analytically summed a series, valid for $|K\gamma| > 1$, to get a result valid for all $K\gamma$.

We have assumed throughout that the media are in direct contact. Although this is often a good approximation, there actually must always be a surface-tension layer, membrane, plate, or other separation along the interface to prevent mixing of the fluids. Frequently the dynamics of this separating layer affect the acoustic scattering significantly; in some cases, such as structural vibration problems, the response of this layer to an acoustic wave is of direct interest. It appears that the theory of this paper can be generalized to take into account such layers, and work along these lines is now under way.

ACKNOWLEDGMENT

The author thanks Dr. Kelvin S. H. Lee for his many helpful suggestions on the presentation of this material.

APPENDIX. DERIVATION OF EQUATIONS FOR ∇U AND $\partial U/\partial n_0$

If we apply Green's first identity[4] to $\mathbf{e}\cdot\nabla U$ and G_0, use Eq. (2.5) to eliminate $\nabla^2 G_0$, and then evaluate the integral containing the δ-function, we obtain

$$\mathbf{e}\cdot(\rho\nabla U) = \mathbf{e}\cdot\int_S dS'\,(\partial G_0/\partial n_0')\nabla' U' + \mathbf{e}\cdot\int_V dV'\,k^2 G_0\nabla' U' - \int_V dV'\,\nabla' G_0\cdot\nabla'(\mathbf{e}\cdot\nabla' U'). \quad \text{(A1)}$$

[9] R. P. Banaugh and W. Goldsmith, J. Acoust. Soc. Am. 35, 1590 (1963) (two-dimensional case); G. W. Soules and R. P. Banaugh, Northrop Ventura Report 3515 (1964) (three-dimensional case).

[10] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Pt. II, Sec. 9.1.

To simplify this we require the easily established relationships

$$\nabla'(\mathbf{e}\cdot\nabla'U') = (\mathbf{e}\cdot\nabla')\nabla'U' = \mathbf{e}(Q'/\rho - k^2U') - \nabla' \times (\mathbf{e} \times \nabla'U'), \quad \text{(A2)}$$

$$\int_V dV'\, \nabla'G_0\cdot[\nabla' \times (\mathbf{e} \times \nabla'U')] = -\int_V dV'\, \nabla'\cdot[\nabla'G_0 \times (\mathbf{e} \times \nabla'U')] = \mathbf{e}\cdot\int_S dS'\, (\mathbf{n}_0' \times \nabla'G_0) \times \nabla'U', \quad \text{(A3)}$$

and

$$\nabla G_0 = -\nabla'G_0. \quad \text{(A4)}$$

Using these in Eq. (A1) gives

$$\mathbf{e}\cdot(\rho\nabla U) = \mathbf{e}\cdot\Big\{\int_S dS'\, [(\mathbf{n}_0'\cdot\nabla'G_0)\nabla'U' + (\mathbf{n}_0' \times \nabla'G_0) \times \nabla'U'] + \int_V dV\, k^2\nabla'(G_0U') + \nabla\int_V dV'\, (1/\rho)G_0Q'\Big\}. \quad \text{(A5)}$$

Since this must hold for all values of $\mathbf{e}$, we can suppress the factor $(\mathbf{e}\cdot)$. From Eq. (2.6) we see that the term involving Q' is just $\nabla(\rho U_0)$. The other two integrals can be reduced to the form given in Eq. (2.8) by straightforward application of standard vector identities, and the proof of that equation is then complete.

Now let us multiply Eq. (2.8) by $(\mathbf{n}_0\cdot)$, make the substitution (A.4), and take the limit as $\mathbf{r}$ approaches S. Since the singularity in G_0 is of form $1/R$, the limit can be evaluated using Kellogg's results,[2] specifically Chap. VI, Theorems IV–VI, and the discussion leading up to Theorem V. The result is Eq. (2.10).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

Integral Forms for Quantum-Mechanical Momentum Operators

PETER D. ROBINSON
Mathematics Department, University of York, England
(Received 1 June 1965)

The usual differential form P_0 for the quantum-mechanical momentum operator P which is conjugate to a generalized coordinate q $(\alpha \leq q \leq \beta)$ is, in atomic units, $P_0 = -i(g^{-\frac{1}{4}})\,\partial/\partial q\,(g^{\frac{1}{4}})$, where g is the Jacobian of the transformation from Cartesian to generalized coordinates. However, P_0 is not always self-adjoint on the domain $\mathfrak{D}$ of physically acceptable bound-state wavefunctions, as a proper quantum-mechanical operator should be. An integral form is proposed for P, defined by

$$Pf(q) = (2\pi)^{-\frac{1}{2}}g^{-\frac{1}{4}}(q)\int_{-\infty}^{\infty} \exp(ikq)kF(k)\,dk, \qquad \alpha \leq q \leq \beta,$$

where

$$F(k) = (2\pi)^{-\frac{1}{2}}\int_{\alpha}^{\beta} \exp(-ik\xi)f(\xi)g^{\frac{1}{4}}(\xi)\,d\xi, \qquad f \in \mathfrak{D}.$$

The effect of this integral operator (which is suggested by the ideas of Fourier transforms) differs from that of P_0 only at the end-points of the range of q. In a sense, it is formally equivalent to an operator (suggested by Robinson and Hirschfelder) which is obtained by adding certain delta-function terms to P_0, but it suffers from none of the defects, since delta-functions do not appear explicitly. Various properties of the integral operator are derived. Some discussion of the domain $\mathfrak{D}$ is presented as an appendix.

1. INTRODUCTION

THE customary differential form P_0 for the quantum-mechanical momentum operator P which is conjugate to a real generalized coordinate q is, in atomic units,

$$P_0 = -i(g^{-\frac{1}{4}})\,\partial/\partial q(g^{\frac{1}{4}}), \quad (1)$$

g being the Jacobian of the transformation from Cartesian coordinates to generalized coordinates. Let us suppose that the physically relevant range of the coordinate q is $\alpha \leq q \leq \beta$. Then the inner product (u, v) of any two functions of q is defined by

$$(u, v) \equiv \int_{\alpha}^{\beta} u^*vg\,dq. \quad (2)$$

To simplify this we require the easily established relationships

$$\nabla'(\mathbf{e}\cdot\nabla'U') = (\mathbf{e}\cdot\nabla')\nabla'U' = \mathbf{e}(Q'/\rho - k^2U') - \nabla' \times (\mathbf{e} \times \nabla'U'), \quad \text{(A2)}$$

$$\int_V dV'\, \nabla'G_0\cdot[\nabla' \times (\mathbf{e} \times \nabla'U')] = -\int_V dV'\, \nabla'\cdot[\nabla'G_0 \times (\mathbf{e} \times \nabla'U')] = \mathbf{e}\cdot\int_S dS'\, (\mathbf{n}_0' \times \nabla'G_0) \times \nabla'U', \quad \text{(A3)}$$

and

$$\nabla G_0 = -\nabla'G_0. \quad \text{(A4)}$$

Using these in Eq. (A1) gives

$$\mathbf{e}\cdot(\rho\nabla U) = \mathbf{e}\cdot\Big\{\int_S dS'\, [(\mathbf{n}_0'\cdot\nabla'G_0)\nabla'U' + (\mathbf{n}_0' \times \nabla'G_0) \times \nabla'U'] + \int_V dV\, k^2\nabla'(G_0U') + \nabla\int_V dV'\, (1/\rho)G_0Q'\Big\}. \quad \text{(A5)}$$

Since this must hold for all values of $\mathbf{e}$, we can suppress the factor $(\mathbf{e}\cdot)$. From Eq. (2.6) we see that the term involving Q' is just $\nabla(\rho U_0)$. The other two integrals can be reduced to the form given in Eq. (2.8) by straightforward application of standard vector identities, and the proof of that equation is then complete.

Now let us multiply Eq. (2.8) by $(\mathbf{n}_0\cdot)$, make the substitution (A.4), and take the limit as $\mathbf{r}$ approaches S. Since the singularity in G_0 is of form $1/R$, the limit can be evaluated using Kellogg's results,[2] specifically Chap. VI, Theorems IV–VI, and the discussion leading up to Theorem V. The result is Eq. (2.10).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 11 NOVEMBER 1966

Integral Forms for Quantum-Mechanical Momentum Operators

PETER D. ROBINSON
Mathematics Department, University of York, England
(Received 1 June 1965)

The usual differential form P_0 for the quantum-mechanical momentum operator P which is conjugate to a generalized coordinate q $(\alpha \leq q \leq \beta)$ is, in atomic units, $P_0 = -i(g^{-\frac{1}{4}})\,\partial/\partial q\,(g^{\frac{1}{4}})$, where g is the Jacobian of the transformation from Cartesian to generalized coordinates. However, P_0 is not always self-adjoint on the domain $\mathfrak{D}$ of physically acceptable bound-state wavefunctions, as a proper quantum-mechanical operator should be. An integral form is proposed for P, defined by

$$Pf(q) = (2\pi)^{-\frac{1}{2}}g^{-\frac{1}{4}}(q)\int_{-\infty}^{\infty} \exp(ikq)kF(k)\,dk, \qquad \alpha \leq q \leq \beta,$$

where

$$F(k) = (2\pi)^{-\frac{1}{2}}\int_{\alpha}^{\beta} \exp(-ik\xi)f(\xi)g^{\frac{1}{4}}(\xi)\,d\xi, \qquad f \in \mathfrak{D}.$$

The effect of this integral operator (which is suggested by the ideas of Fourier transforms) differs from that of P_0 only at the end-points of the range of q. In a sense, it is formally equivalent to an operator (suggested by Robinson and Hirschfelder) which is obtained by adding certain delta-function terms to P_0, but it suffers from none of the defects, since delta-functions do not appear explicitly. Various properties of the integral operator are derived. Some discussion of the domain $\mathfrak{D}$ is presented as an appendix.

1. INTRODUCTION

THE customary differential form P_0 for the quantum-mechanical momentum operator P which is conjugate to a real generalized coordinate q is, in atomic units,

$$P_0 = -i(g^{-\frac{1}{4}})\,\partial/\partial q(g^{\frac{1}{4}}), \quad (1)$$

g being the Jacobian of the transformation from Cartesian coordinates to generalized coordinates. Let us suppose that the physically relevant range of the coordinate q is $\alpha \leq q \leq \beta$. Then the inner product (u, v) of any two functions of q is defined by

$$(u, v) \equiv \int_{\alpha}^{\beta} u^*vg\,dq. \quad (2)$$

It follows from (1) and (2), after integrating by parts, that

$$(u, P_0 v) - (P_0 u, v) = -i[guw]_{\alpha-}^{\beta}. \qquad (3)$$

In order that P_0 should be a proper quantum-mechanical operator, it is required to be self-adjoint[1] on the domain $\mathfrak{D}$ of physically acceptable bound-state wavefunctions (some discussion of this domain is given in the Appendix). A necessary condition for the self-adjointness is the vanishing of the boundary term on the right-hand side of Eq. (3) when u and v are functions belonging to $\mathfrak{D}$. However, such boundary terms do not always vanish.

There are thus three possibilities. Firstly, the domain $\mathfrak{D}$ could be restricted so as to exclude functions for which the boundary term in Eq. (3) does not vanish. This is not realistic, as a simple example of nonvanishing boundary terms arises when u and v are σ-type hydrogen-atom wavefunctions expressed in parabolic coordinates.[2,3] Secondly, the generalized momentum operator P may not be a proper quantum-mechanical operator representing an observable. This might well be the case, but it does not seem too unreasonable to suppose that generalized momentum components can, in certain circumstances, be measured, even if the required apparatus has to be rather complicated. Thirdly, the differential operator P_0 may not be a true representation of the quantum-mechanical operator P. It is this third possibility which is pursued in this paper.

An attempt to find a self-adjoint form for P has been made by Robinson and Hirschfelder.[3] They suggest an expression P_1 for P, where

$$P_1 = P_0 - \tfrac{1}{2}i\{\delta_+(q-\alpha) - \delta_-(\beta - q)\} \qquad (4)$$

the delta-functions being similar to Dirac's except that their effects inside an integral are, respectively,

$$\int_{\alpha}^{t} f(q)\ \delta_+(q-\alpha)\, dq = f(\alpha), \qquad \alpha < t \leq \beta,$$

and

$$\int_{t}^{\beta} f(q)\ \delta_-(\beta - q)\, dq = f(\beta), \qquad \alpha \leq t < \beta. \qquad (5)$$

[The inclusion of the delta-function terms is based on the fact that a more fundamental form of (1) is $-i\{\partial/\partial q + \tfrac{1}{2}\ \mathrm{div}\ (h\hat{\mathbf{q}})\}$, where h is the metric scale factor corresponding to q, and $\hat{\mathbf{q}}$ is the unit vector in the direction of q increasing. The delta-functions arise naturally in the interpretation of div $(h\hat{\mathbf{q}})$ at the end-points where $\hat{\mathbf{q}}$ is ill-defined.] It is easy to see that P_1 is formally self-adjoint, insofar as

$$(u, P_1 v) = (P_1 u, v),$$

irrespective of the boundary values of u and v. However, in spite of this, and notwithstanding the fact that P_1 (like P_0) satisfies the necessary commutation requirements, the form (4) is not really satisfactory. Indeed, P_1 is not strictly an operator at all. The delta-function terms are only strictly meaningful when appearing inside an integral over q, and it is not possible to interpret expressions such as P_1^2, or give meanings to inner products like $(P_1 u, P_1 v)$ without an artificial convention defining the square of a delta-function.

These considerations lead us to propose an integral form for the operator P which has the same effect as P_0 when P_0 is self-adjoint, and otherwise is, in a sense, equivalent to P_1. This integral operator does not suffer from the same shortcomings as does P_1, because delta-functions do not appear explicitly.

We develop the integral form in Sec. 2, and derive some of its important properties in Sec. 3. Finally, in Sec. 4, the relationship between corresponding integral and differential operators is discussed. We concentrate on the situation which arises when the end-points α and β are both finite; this is potentially the most unpleasant case. Modifications are obvious if q extends to infinity at either or both ends of its range.

2. THE INTEGRAL FORM FOR P

We are led to the integral form for P by considering the situation in what is effectively generalized momentum space. The starting point is Fourier's integral theorem. Let $f(q) \in \mathfrak{D}$ be a wavefunction, and let $F(k)$ be the 'finite' Fourier transform of $f(q)g^{\frac{1}{2}}(q)$, defined by

$$F(k) = (2\pi)^{-\frac{1}{2}} \int_{\alpha}^{\beta} \exp(-ik\xi) f(\xi) g^{\frac{1}{2}}(\xi)\, d\xi; \qquad (6)$$

$F(k)$ can be regarded as the ordinary Fourier transform of a function $\phi(q)$ defined by

$$\phi(q) = \begin{cases} f(q)g^{\frac{1}{2}}(q), & \alpha \leq q \leq \beta, \\ 0, & -\infty < q < \alpha \quad \text{or} \quad \beta < q < \infty. \end{cases} \qquad (7)$$

[Should α be $-\infty$ or β be $+\infty$, then the appropriate part of Eq. (7) can be omitted.] In Eq. (6) we use

[1] M. H. Stone, *Linear Transformations in Hilbert Space* (American Mathematics Society Publications, New York, 1932), Vol. 15, Chaps. IV and V.
[2] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., Chap. IV.
[3] P. D. Robinson and J. O. Hirschfelder, J. Math. Phys. 4, 348 (1963).

ξ as a dummy variable instead of q; this practice is adopted throughout the paper whenever an integration over the whole range of q is involved, thus avoiding confusion with any q representing a current value.

We make the following assumptions concerning a function $f \in \mathfrak{D}$:

(a) $$\int_\alpha^\beta |f(\xi)|\, g^{\frac{1}{2}}(\xi)\, d\xi < \infty; \tag{8a}$$

(b) $$\int_\alpha^\beta |\xi f(\xi)|\, g^{\frac{1}{2}}(\xi)\, d\xi < \infty; \tag{8b}$$

(c) $$(f, f) \equiv \int_\alpha^\beta |f(\xi)|^2\, g(\xi)\, d\xi < \infty; \tag{8c}$$

(d) $f(\xi)g^{\frac{1}{2}}(\xi)$ is of bounded variation in the neighbourhood of $\xi = q$.

Condition (a) ensures the existence of $F(k)$ for any finite k. Conditions (a) and (d) ensure that Fourier's integral theorem holds (see Titchmarsh[4]), so that

$$f(q) = (2\pi)^{-\frac{1}{2}} g^{-\frac{1}{2}}(q) \int_{-\infty}^{\infty} \exp(ikq) F(k)\, dk, \qquad \alpha < q < \beta. \tag{9}$$

[Even if (a) and (d) do not hold, but (c) does, modified forms of Eqs. (6) and (9) are still available with convergence in mean replacing actual equality.[4]] By Parseval's formula,[4] (c) is equivalent to

$$\int_{-\infty}^{\infty} |F(k)|^2\, dk < \infty, \tag{10}$$

which is one of the conditions which f must satisfy in order that it should belong to $\mathfrak{D}$ (see the Appendix). Condition (c) is inserted to justify a change in the order of integration in Sec. 3A below.

Because of the possible discontinuities at α or β in the function $\phi(q)$ defined by Eq. (7), a factor $\frac{1}{2}$ must be inserted[4] on the left-hand side of Eq. (9) if it is to be true when $q = \alpha$ or $q = \beta$. Thus Eq. (9) does not hold at the end points unless $f(\alpha)g^{\frac{1}{2}}(\alpha)$ or $f(\beta)g^{\frac{1}{2}}(\beta)$ is zero. Although this frequently happens, we do not wish to restrict the function $f(q)$ in this way. In fact, if $f(\alpha)g^{\frac{1}{2}}(\alpha) = f(\beta)g^{\frac{1}{2}}(\beta) = 0$, then the usual differential form P_0 as given in Eq. (1) is self-adjoint (see Sec. 4), and no difficulties arise. Equation (9) can be regarded as an "expansion" of $f(q)$ in terms of the formal eigenfunctions $g^{-\frac{1}{2}}(q) \exp(ikq)$ of the operator P_0, with $F(k)$ playing the role of a wavefunction in k-space.

[4] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, Oxford, 1948), 2nd ed.

Operating on Eq. (9) with P_0, we have

$$P_0 f(q) = -ig^{-\frac{1}{2}} \frac{\partial}{\partial q}(g^{\frac{1}{2}} f) = (2\pi)^{-\frac{1}{2}} g^{-\frac{1}{2}}(q) \times \int_{-\infty}^{\infty} \exp(ikq) kF(k)\, dk, \qquad \alpha < q < \beta. \tag{11}$$

We assume that the integral in Eq. (11) is convergent (see the Appendix); it is a uniformly continuous function of q, and so the differentiation under the integral sign is permissible. Equation (11) is not true in general at the end points α and β because of the discontinuities in the function $\phi(q)$; it does however hold there if $f(\alpha)g^{\frac{1}{2}}(\alpha) = f(\beta)g^{\frac{1}{2}}(\beta) = 0$, and possibly under other circumstances mentioned in Sec. 4. It is readily shown (see Sec. 3B) that the integral operator on the right-hand side of Eq. (11) is self-adjoint, and so we take it as our definition of P; it agrees with the usual form P_0 except at the end points. Thus we stipulate that, for every $f(q) \in \mathfrak{D}$,

$$Pf(q) = (2\pi)^{-\frac{1}{2}} g^{-\frac{1}{2}}(q) \times \int_{-\infty}^{\infty} \exp(ikq) kF(k)\, dk, \qquad \alpha \leq q \leq \beta. \tag{12}$$

It must be emphasized that Eq. (12), which defines P, holds good for the closed interval $\alpha \leq q \leq \beta$, whereas Eqs. (9) and (11) are in general merely true in the open interval $\alpha < q < \beta$.

3. PROPERTIES OF THE INTEGRAL OPERATOR P

We now demonstrate some important properties of the integral operator P, defined by Eq. (12).

A. The Commutator $[P, q] \equiv Pq - qP$ is $-i$

It follows from Eqs. (6) and (12) that

$$P(qf) = (2\pi)^{-1} g^{-\frac{1}{2}}(q) \int_{-\infty}^{\infty} \exp(ikq) k\, dk \times \int_\alpha^\beta \exp(-ik\xi) \xi f(\xi) g^{\frac{1}{2}}(\xi)\, d\xi. \tag{13}$$

Conditions (8a) and (8b) ensure that the ξ-integrals in Eqs. (12) and (13) converge uniformly with respect to k over any finite interval; thus the order of integration can be changed. Doing this, we find that

$$P(qf) - qPf = (2\pi)^{-1} g^{-\frac{1}{2}}(q) \int_\alpha^\beta f(\xi) g^{\frac{1}{2}}(\xi)\, d\xi \times \int_{-\infty}^{\infty} \exp\{ik(q - \xi)\} k(\xi - q)\, dk. \tag{14}$$

The k-integral gives

$$-i\int_{-\infty}^{\infty}\exp\{ik(q-\xi)\}\,dk + i[k\exp\{ik(q-\xi)\}]_{-\infty}^{\infty},$$

and thus, using Eqs. (6) and (9), Eq. (14) becomes

$$P(qf) - qPf = -if + ig^{-\frac{1}{2}}(q)[k\exp(ikq)F(k)]_{-\infty}^{\infty}. \tag{15}$$

Assuming that the boundary term in Eq. (15) vanishes (this matter is discussed in the Appendix), it follows that the operators $[P, q]$ and $-i$ are equivalent when operating on a function $f(q) \in \mathfrak{D}$. Thus the usual commutation rule for a momentum operator is satisfied.

B. P is Self-Adjoint

To show that P is self-adjoint, we must prove that $(u, Pv) = (Pu, v)$ whenever $u, v \in \mathfrak{D}$. Denoting the finite Fourier transforms of $u(q)g^{\frac{1}{2}}(q)$ and $v(q)g^{\frac{1}{2}}(q)$ by $U(k)$ and $V(k)$, we have

$$(u, Pv) = (2\pi)^{-\frac{1}{2}}\int_{\alpha}^{\beta}u^*(q)g^{\frac{1}{2}}(q)\,dq \times\int_{-\infty}^{\infty}\exp(ikq)kV(k)\,dk,$$

$$(Pu, v) = (2\pi)^{-\frac{1}{2}}\int_{\alpha}^{\beta}v(q)g^{\frac{1}{2}}(q)\,dq \times\int_{-\infty}^{\infty}\exp(-ikq)kU^*(k)\,dk.$$

Changing the order of integration in each case (this is justified as in Subsection A above), we see that

$$(u, Pv) = (Pu, v) = \int_{-\infty}^{\infty}kU^*(k)V(k)\,dk. \tag{16}$$

C. The Expectation Value of P for a Real State is Zero

We must show that the expectation value of P for any real state $f(q)$ is zero. This should be so, because the probability current density, being proportional to $(f\nabla f^* - f^*\nabla f)$, is identically zero, and thus the mean value of all momenta should be zero. It is enough to show that (f, Pf) vanishes whenever f is real and $f \in \mathfrak{D}$.

If f is real, then $F^*(k) = F(-k)$. Thus from Eq. (16) with $u = v = f$, we have

$$(f, Pf) = \int_{-\infty}^{\infty}kF(-k)F(k)\,dk = 0, \tag{17}$$

by symmetry.

D. The Meaning of a Function of P

Because of the unitary nature of the transforms (6) and (9), Eq. (11) implies that an operation by P_0 in q-space is unitarily equivalent to a multiplication by k in k-space, provided that $\alpha < q < \beta$. Thus, formally, an operation by a function $\chi(P_0)$ in q-space is equivalent to a multiplication by $\chi(k)$ in k-space, again provided that $\alpha < q < \beta$, i.e.,

$$\chi(P_0)f = (2\pi g)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\exp(ikq)\chi(k)F(k)\,dk, \qquad \alpha < q < \beta. \tag{18}$$

Analogy with the pair of equations (11) and (12) therefore leads to the conjecture that $\chi(P)$ may be defined by

$$\chi(P)f = (2\pi g)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\exp(ikq)\chi(k)F(k)\,dk, \qquad \alpha \le q \le \beta, \tag{19}$$

in those cases where $\chi(P_0)$ is not self-adjoint. Just as in Subsection B above, it can be shown that $\chi(P)$ is formally self-adjoint if χ is a real function. Corresponding to Eq. (16), we have

$$(u, \chi(P)v) = (\chi(P)u, v) = \int_{-\infty}^{\infty}\chi(k)U^*(k)V(k)\,dk. \tag{20}$$

Equation (20) can also be used to give a meaning to inner products like (Pu, Pv), since methods similar to those in Subsection B indicate that $(Pu, Pv) = (u, P^2v)$.

4. THE RELATIONSHIP BETWEEN THE INTEGRAL AND DIFFERENTIAL OPERATORS

It is possible to derive a formal relationship between integral and differential forms for the operator P. We already know that the integral form is equivalent to the usual differential form P_0 in the open interval $\alpha < q < \beta$.

Integrating (6) by parts, we have

$$F(k) = -(2\pi)^{-\frac{1}{2}}\frac{i}{k}\int_{\alpha}^{\beta}\exp(-ik\xi)\frac{\partial}{\partial\xi}(fg^{\frac{1}{2}})\,d\xi + (2\pi)^{-\frac{1}{2}}\frac{i}{k}[\exp(-ik\xi)fg^{\frac{1}{2}}]_{\alpha}^{\beta}. \tag{21}$$

Thus, from (12) it follows that

$$Pf(q) = -i(2\pi)^{-1}g^{-\frac{1}{2}}(q)\int_{-\infty}^{\infty}\exp(ikq)\,dk \times\int_{\alpha}^{\beta}\exp(-ik\xi)\frac{\partial}{\partial\xi}(fg^{\frac{1}{2}})\,d\xi + i(2\pi)^{-1}g^{-\frac{1}{2}}(q)\int_{-\infty}^{\infty}\exp(ikq) \times\{\exp(-ik\beta)f(\beta)g^{\frac{1}{2}}(\beta) - \exp(-ik\alpha)f(\alpha)g^{\frac{1}{2}}(\alpha)\}\,dk. \tag{22}$$

The first term on the right-hand side of Eq. (22) simplifies to $P_0 f(q)$; the second term, which vanishes if $f(\alpha)g^{\frac{1}{2}}(\alpha) = f(\beta)g^{\frac{1}{2}}(\beta) = 0$, can otherwise be expressed formally in terms of delta-functions, viz:

$$\begin{aligned} Pf(q) &= P_0 f(q) + ig^{-\frac{1}{2}}(q)\{\delta(q-\beta)f(\beta)g^{\frac{1}{2}}(\beta) \\ &\quad - \delta(q-\alpha)f(\alpha)g^{\frac{1}{2}}(\alpha)\} \\ &= P_0 f(q) - i\{\delta(q-\alpha) - \delta(q-\beta)\}f(q) \\ &= P_0 f(q) - \tfrac{1}{2}i\{\delta_+(q-\alpha) - \delta_-(\beta-q)\}f(q). \end{aligned} \tag{23}$$

Expression (23) involves the delta functions mentioned in Sec. 1. So we see that the integral form for P is in a sense equivalent to the operator P_1 proposed by Robinson and Hirschfelder; if either $f(\beta)g^{\frac{1}{2}}(\beta)$ or $f(\alpha)g^{\frac{1}{2}}(\alpha)$ is zero, then the corresponding delta-function term is omitted.

We may also omit the delta-function terms in (23), or strictly the last term in Eq. (22), if P is only to be associated with functions belonging to the sub-domain $\mathfrak{D}_0$ of $\mathfrak{D}$ on which the usual differential form P_0 is self-adjoint. In order that P_0 be self-adjoint, it is necessary that $g(\beta)u^*(\beta)v(\beta) = g(\alpha)u^*(\alpha)v(\alpha)$ when $u, v \in \mathfrak{D}_0$ [cf. Eq. (3)]. Thus, if $f(q) \in \mathfrak{D}_0$, we must have $g^{\frac{1}{2}}(\beta)f(\beta) = g^{\frac{1}{2}}(\alpha)f(\alpha)\exp(i\gamma)$, where γ is a real constant. Equation (22) shows that the inner product (u, Pv), for instance, is then the same as

$$(u, P_0 v) + \tfrac{1}{2}i\{g(\beta)u^*(\beta)v(\beta) - g(\alpha)u^*(\alpha)v(\alpha)\},$$

and so $(u, Pv) = (u, P_0 v)$. Thus if the usual form P_0 is self-adjoint, we can take it as the definition of P; it is only when P_0 is not self-adjoint (i.e., when we are concerned with functions belonging to $\mathfrak{D}$ but not to $\mathfrak{D}_0$) that the integral form (12) is necessary for the definition. The integral operator is a kind of *extended* definition of the usual differential operator; this concept of extended definition has been discussed by Friedman,[5] employing the delta-function approach.

Similar formal expressions involving delta functions and their derivatives can be developed to represent powers of P. For example, P^2 is defined by [c.f. Eq. (19)]

$$P^2 f(q) = (2\pi)^{-\frac{1}{2}} g^{-\frac{1}{2}}(q) \int_{-\infty}^{\infty} \exp(ikq)k^2 F(k)\, dk, \qquad \alpha \le q \le \beta. \tag{24}$$

After integrating expression (6) for $F(k)$ twice by parts and substituting in (24), it is found that

$$\begin{aligned} P^2 f(q) &= P_0^2 f(q) + (2\pi)^{-1} g^{-\frac{1}{2}}(q) \\ &\quad \times \int_{-\infty}^{\infty} \exp(ikq)[ik\exp(-ik\xi)fg^{\frac{1}{2}} \\ &\quad + \exp(-ik\xi)\frac{\partial}{\partial\xi}(fg^{\frac{1}{2}})]_\alpha^\beta\, dk. \end{aligned} \tag{25}$$

The last term in (25) can formally be written as

$$g^{-\frac{1}{2}}(q)\left[\left\{\frac{\partial}{\partial q}\delta(q-\xi)\right\}f(\xi)g^{\frac{1}{2}}(\xi) + \delta(q-\xi)\frac{\partial}{\partial\xi}\{f(\xi)g^{\frac{1}{2}}(\xi)\}\right]_{\xi=\alpha}^{\beta}$$

and it can be omitted althogether if we are dealing with functions with respect to which P_0^2 is self-adjoint. The special case when P_0^2 happens to be a kinetic energy operator is an example, for kinetic energy operators are all self-adjoint on the domain $\mathfrak{D}$. This happens when $P_0^2 = -g^{-\frac{1}{2}}\partial^2/\partial q^2(g^{\frac{1}{2}}\ )$ is the same as

$$L \equiv -g^{-1}\frac{\partial}{\partial q}\left\{\frac{g}{h^2}\frac{\partial}{\partial q}\right\} \tag{26}$$

(h being the metric scale factor corresponding to q), as for instance when q is a spherical polar radial coordinate. But, generally P_0^2 is not self-adjoint on the domain $\mathfrak{D}$, and the integral form (24) must be used for P^2.

APPENDIX. THE DOMAIN $\mathfrak{D}$

In theory, there is no ambiguity about the domain $\mathfrak{D}$ containing all physically acceptable bound-state wavefunctions, but in practice, the conditions that a function should belong to $\mathfrak{D}$ are often hard to apply. Kato[6] shows that $\mathfrak{D}$ is the domain of the closure of all Laplace operators, i.e., of all kinetic energy operators, all such operators being self-adjoint on this domain. Kato also gives precise conditions in terms of Fourier transforms under which a wavefunction belongs to $\mathfrak{D}$. If, for example, $\psi(\mathbf{r})$ is a wavefunction describing the behavior of a particle at the vector position $\mathbf{r}$, then the conditions are

$$\int |\Psi(\mathbf{k})|^2\, d\mathbf{k} < \infty, \tag{A1}$$

and

$$\int |\mathbf{k}|^4\, |\Psi(\mathbf{k})|^2\, d\mathbf{k} < \infty, \tag{A2}$$

[5] B. Friedman, *The Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1956), Chap. III.

[6] T. Kato, Transl. Am. Math. Soc. **70**, 195 (1951).

where $\Psi(\mathbf{k})$ is the Fourier transform of $\psi(\mathbf{r})$, defined by

$$\Psi(\mathbf{k}) = (2\pi)^{-\frac{3}{2}} \int \exp(-i\mathbf{k}\cdot\mathbf{r})\psi(\mathbf{r})\, d\mathbf{r}. \quad \text{(A3)}$$

The integrations in (A1)–(A3) are respectively taken over all $\mathbf{k}$-space and all $\mathbf{r}$-space. Generalizations of these conditions hold for wavefunctions of many-particle systems.

The conditions that a wavefunction $f(x)$ of a single Cartesian coordinate x should belong to $\mathfrak{D}$ take the simple forms

$$\int_{-\infty}^{\infty} |\mathfrak{F}(k)|^2\, dk < \infty, \quad \text{(A4)}$$

$$\int_{-\infty}^{\infty} k^4\, |\mathfrak{F}(k)|^2\, dk < \infty, \quad \text{(A5)}$$

where

$$\mathfrak{F}(k) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp(-ikx) f(x)\, dx. \quad \text{(A6)}$$

Unfortunately, the situation is not so straightforward for a function $f(q)$ of the generalized coordinate q. The condition (A4) has its analog in Eq. (10) of the text, which, in the equivalent form (8c), is just the condition that $f(q)$, being a bound-state wavefunction, should be normalizable. However, (A5) has no immediate counterpart. This latter condition is equivalent to saying that

$$(L_x f, L_x f) = (f, L_x^2 f), \quad \text{(A7)}$$

where $L_x \equiv -\partial^2/\partial x^2$ is effectively the kinetic energy operator. The appropriate kinetic energy operator which corresponds to q is L, given by (26), and, instead of (A5), we have

$$(2\pi)^{-\frac{1}{2}} \int_{\alpha}^{\beta} f^*(q) g\, dq \times \int_{-\infty}^{\infty} F(k) L^2 \{g^{-\frac{1}{2}} \exp(ikq)\}\, dk < \infty. \quad \text{(A8)}$$

Equation (A8) simplifies to

$$(2\pi)^{-\frac{1}{2}} \int_{\alpha}^{\beta} f^*(q) g^{\frac{1}{2}}\, dq \times \int_{-\infty}^{\infty} \left\{\sum_{r=0}^{4} a_r k^r\right\} F(k) \exp(ikq)\, dk < \infty, \quad \text{(A9)}$$

where the coefficients a_r are functions of q depending on $g(q)$, $h(q)$, and their derivatives. Condition (A9) most probably implies that

$$\int_{-\infty}^{\infty} k^r F(k) \exp(ikq)\, dk < \infty, \quad r = 0, 1, 2, 3, 4, \quad \text{(A10)}$$

but we cannot be absolutely certain of this without knowing details of $g(q)$ and $h(q)$.

Equation (A10) must hold when $r = 1$ in order to make equations (11) and (12) meaningful and P well-defined. It is also necessary that

$$[kF(k) \exp(ikq)]_{k=-\infty}^{\infty} = 0 \quad \text{(A11)}$$

if the commutation rule $[P, q] = -i$ is to be satisfied [vide Eq. (15)]. (A11) surely holds if (A10) is true for $r = 0, 1, 2, 3, 4$. However, it seems probable that we cannot definitely assert that conditions (A10) with $r = 1$ and (A11) are satisfied, given that $f(q) \in \mathfrak{D}$, and so we must be content with assuming that these conditions are fulfilled in order that the operator P be a proper quantum-mechanical operator.

ACKNOWLEDGMENTS

I should like to thank Dr. J. T. Lewis for suggesting the possibility of integral forms for momentum operators. I am grateful to him, and also to Dr. P. J. Bushell, Professor C. A. Coulson, and Professor J. O. Hirschfelder for useful discussions and for their helpful comments on the manuscript.

Existence of Particlelike Solutions to Nonlinear Field Theories

GERALD ROSEN
Drexel Institute of Technology, Philadelphia, Pennsylvania
(Received 1 February 1966; revised manuscript received 6 March 1966)

A *pseudovirial theorem* is derived for time-independent particlelike solutions of finite energy (singularity-free and spatially localized time-independent solutions) to field theories associated with an action principle. It is shown that a useful necessary condition for the existence of such particlelike solutions is generally obtainable as a corollary to the pseudovirial theorem. This necessary condition is in fact sufficient to preclude existence of any well-localized particlelike solution for all but special field theories with the more common forms of algebraic interaction. On the other hand, strong satisfaction of the necessary condition can lead to model field theories with rigorous closed-form particlelike solutions, as shown by example for a class of Lorentz-covariant theories which feature a real scalar field in interaction with a two-component complex Weyl spinor field. Some of the latter particlelike solutions to the scalar–spinor theory are energetically stable with respect to spatial dilatations, hence likely to be stable in the dynamical sense. A counter example to the more general sufficiency of the strong satisfaction condition is presented, showing that strong satisfaction of the pseudovirial theorem's corollary does not always guarantee the existence of singularity-free particlelike solutions.

I. INTRODUCTION

CONSIDERABLE recent interest has been attached to particlelike solutions to model Lorentz-covariant nonlinear field theories.[1-3] The idea for such a research program is fostered by the well-known and serious practical deficiencies of structureless point-particle representations in classical and quantum field theory. Further motivation for the study of particlelike solutions to model nonlinear theories is provided by the still enigmatic content of unified field theories, where only the manifold of properly bounded and singularity-free solutions is postulated to be of physical significance,[4] an essentially recondite manifold of solutions.

It is the purpose of the present paper to consider the more general question of existence for singularity-free and spatially localized time-independent solutions. Our main result appears as a necessary condition for the existence of any particlelike solution of finite energy, applicable to all field theories associated with an action principle and presented here as a corollary to a *pseudovirial theorem*. This pseudovirial theorem and its corollary are global relationships, like the laws of conservation of total field energy or momentum, but unlike the other more venerable global relationships, the pseudovirial theorem and corollary express nontrivial conditions which must be satisfied by a *time-independent* solution. The corollary is in fact sufficient to preclude existence of any well-localized particlelike solution for all but certain special field theories with the more common forms of algebraic interaction.

In the subsequent section of this paper, we consider a restricted class of Lorentz-covariant model theories obtained by satisfying the pseudovirial theorem's corollary in the strongest possible (local) way for a real scalar field in interaction with a two-component complex Weyl spinor field. Rigorous closed-form singularity-free particlelike solutions are obtainable for this restricted class of model theories. Also noteworthy is the fact that some of the particlelike solutions to the Lorentz-covariant scalar–spinor theories are energetically stable with respect to spatial dilatations, and so it is possible to conjecture their complete dynamical stability.

In the final section of the paper, we consider a related class of theories which feature a real scalar field in interaction with a four-component complex Dirac spinor field. For a restricted class of the latter theories, it is demonstrated that strong satisfaction of the pseudovirial theorem's corollary is insufficient to admit any singularity-free particlelike solution. Thus, we see by example that the strong satisfaction condition is not always sufficient to guarantee such localized solutions.

II. THE PSEUDOVIRIAL THEOREM AND ITS COROLLARY

Time-independent solutions to field theories associated with an action principle satisfy equations of the form

$$\delta E/\delta f(\mathbf{x}) = 0, \tag{1}$$

where the energy functional $E = E[f(\mathbf{x})]$ depends

[1] U. Enz, Phys. Rev. 131, 1392 (1963), and works cited therein.
[2] G. H. Derrick, J. Math. Phys. 5, 1252 (1964).
[3] G. Rosen, J. Math. Phys. 6, 1269 (1965).
[4] A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, New Jersey, 1955), pp. 164–165.

on (real and/or complex) fields represented in (1) by the generic quantity $f(\mathbf{x})$. Let us make a linear decomposition of E,

$$E = \sum_{\{w\}} E^{(w)}, \tag{2}$$

in which $E^{(w)}$ is a functional of "weight" w with respect to changes in the form of $f(\mathbf{x})$ due to spatial dilatations,

$$E^{(w)} = E^{(w)}[f(\mathbf{x})] \equiv \lambda^{-w} E^{(w)}[f(\lambda\mathbf{x})] \tag{3}$$

for all real $\lambda > 0$. Then the *pseudovirial theorem*

$$\sum_{\{w\}} wE^{(w)}[f(\mathbf{x})] = 0 \tag{4}$$

follows immediately from (2) and (3), because the equation

$$\left\{\frac{d}{d\lambda} E[f(\lambda\mathbf{x})]\right\}_{\lambda=1} = 0 \tag{5}$$

is a direct consequence of the Eqs. (1).[5]

Each of the quantities $E^{(w)}[f(\mathbf{x})]$ must be finite in order for (4) to be meaningful, and hence the pseudovirial theorem applies to spatially localized time-independent solutions with no serious singularities. For such "particlelike" solutions, it is generally possible to eliminate *all* spatial derivatives of the fields which appear in the $E^{(w)}[f(\mathbf{x})]$ by evoking the equations (1). A useful corollary to the pseudovirial theorem is thereby obtained in the form

$$\int \mathfrak{F}(f(\mathbf{x}))\, d^3\mathbf{x} = 0, \tag{6}$$

where $\mathfrak{F}(f(\mathbf{x}))$ is a *purely algebraic* function of $f(\mathbf{x})$. Hence, we use the local Eqs. (1) first to establish the global pseudovirial theorem and then again to eliminate spatial derivatives in the global relation provided by the theorem, and in this way an important property of the particlelike solutions to Eqs. (1) is obtained.

It is remarkable that Eq. (6), a necessary condition for the existence of a particlelike solution, is in fact sufficient to preclude existence of such solutions for all but special field theories which feature the common form of algebraic interaction, as illustrated by the following examples.

Example 1: Static particlelike solutions for a self-interacting real scalar field.

$$E = E^{(-1)} + E^{(-3)},$$

where

$$E^{(-1)} \equiv \int (\nabla\theta)^2\, d^3\mathbf{x}, \qquad E^{(-3)} \equiv -g \int |\theta|^p\, d^3\mathbf{x}$$

(g and p positive constant parameters). Equation (1) takes the form

$$\nabla^2\theta + \tfrac{1}{2}gp\,|\theta|^{p-2}\,\theta = 0,$$

and we can bring the statement of the pseudovirial theorem (4),

$$\int [-(\nabla\theta)^2 + 3g\,|\theta|^p]\, d^3\mathbf{x} = 0,$$

into the form of Eq. (6),

$$(p-6)\int |\theta|^p\, d^3\mathbf{x} = 0,$$

which implies that a nontrivial particlelike solution can exist only if $p = 6$, a special case studied previously and shown to admit rigorous singularity-free particlelike solutions.[3] In order to make the manipulations meaningful, we require a $\theta(\mathbf{x})$ of function class C^2 for all $\mathbf{x}$ and such that

$$\lim_{|\mathbf{x}|\to\infty} [|\mathbf{x}|^{3/s}\,\theta(\mathbf{x})] = 0, \qquad s \equiv \min\{p, 6\}.$$

Example 2: Stationary particlelike solutions for a self-interacting complex scalar field.

$$E = E^{(2)} + E^{(0)},$$

where

$$E^{(2)} \equiv \frac{\int \nabla\psi^*\cdot\nabla\psi\, d^3\mathbf{x}}{\int \psi^*\psi\, d^3\mathbf{x}}, \qquad E^{(0)} \equiv \frac{-g\int (\psi^*\psi)^{p/2}\, d^3\mathbf{x}}{\int \psi^*\psi\, d^3\mathbf{x}}$$

(g and p positive constant parameters). Since the pseudovirial theorem gives $E^{(2)} = 0$, we conclude immediately that the nonlinear eigenvalue equation

$$\nabla^2\psi + [\tfrac{1}{2}gp(\psi^*\psi)^{\frac{1}{2}p-1} + E]\psi = 0$$

has no nontrivial solution $\psi(\mathbf{x})$ of function class C^2 for all $\mathbf{x}$ and such that

$$\lim_{|\mathbf{x}|\to\infty} [|\mathbf{x}|^{\frac{1}{2}}\,|\nabla\psi|] = 0,$$

$$\lim_{|\mathbf{x}|\to\infty} [|\mathbf{x}|^{3/t}\,|\psi|] = 0, \qquad t \equiv \min\{p, 2\}.$$

Example 3: Static particlelike solutions in nonrelativistic quantum electrodynamics.

$$E = E^{(-1)} + E^{(-3)},$$

where

$$E^{(-1)} \equiv \int \left[\frac{1}{8\pi}(\nabla\phi)^2 + \frac{\hbar^2}{2m}\nabla\psi^*\cdot\nabla\psi\right] d^3\mathbf{x},$$

$$E^{(-3)} \equiv -e\int \phi\psi^*\psi\, d^3\mathbf{x}$$

[5] The well-known virial theorems for periodic particle motion in classical mechanics and stationary particle states in quantum mechanics can also be established by the spatial dilatation invariance argument used in our proof.

($\hbar$, m, e positive physical constants). Equations (1) take the forms

$$\nabla^2\phi + 4\pi e\psi^*\psi = 0, \qquad \frac{\hbar^2}{2m}\nabla^2\psi + e\phi\psi = 0,$$

and so we find that $E^{(-1)} = -\frac{3}{2}E^{(-3)}$ while the pseudovirial theorem produces $E^{(-1)} = -3E^{(-3)}$. Hence, there is no nontrivial solution with ϕ and ψ of function class C^2 for all $\mathbf{x}$, and such that

$$\lim_{|\mathbf{x}|\to\infty}[|\mathbf{x}|^r\,\phi(\mathbf{x})] = 0, \qquad \lim_{|\mathbf{x}|\to\infty}[|\mathbf{x}|^s\,|\psi(\mathbf{x})|] = 0,$$

$$\lim_{|\mathbf{x}|\to\infty}[|\mathbf{x}|^{\frac{1}{2}}\,|\nabla\psi(\mathbf{x})|] = 0,$$

$$r \geq \tfrac{1}{2}, \qquad s \geq \tfrac{1}{2}, \qquad (r + 2s) \geq 3.$$

Example 4: Static particlelike solutions in relativistic quantum electrodynamics.

$$E = E^{(-1)} + E^{(-2)} + E^{(-3)},$$

where

$$E^{(-1)} \equiv \int \frac{1}{8\pi}(\nabla\phi)^2\, d^3\mathbf{x},$$

$$E^{(-2)} \equiv -i\hbar\int \psi^\dagger\boldsymbol{\alpha}\cdot\nabla\psi\, d^3\mathbf{x},$$

$$E^{(-3)} \equiv \int (m\psi^\dagger\beta\psi - e\phi\psi^\dagger\psi)\, d^3\mathbf{x}.$$

Here, ψ is a four-component complex Dirac spinor field, $\psi^\dagger$ is its Hermitian adjoint, and $\boldsymbol{\alpha}$, β are Hermitian Dirac matrices. In this case, Eqs. (1) are

$$\nabla^2\phi + 4\pi e\psi^\dagger\psi = 0, \quad -i\hbar\boldsymbol{\alpha}\cdot\nabla\psi + (m\beta - e\phi)\psi = 0,$$

and allow the statement of the pseudovirial theorem

$$\int\left[-\frac{1}{8\pi}(\nabla\phi)^2 + 2i\hbar\psi^\dagger\boldsymbol{\alpha}\cdot\nabla\psi + 3(e\phi\psi^\dagger\psi - m\psi^\dagger\beta\psi)\right] d^3\mathbf{x} = 0$$

to be brought into the general form prescribed by Eq. (6),

$$\int (\tfrac{1}{2}e\phi\psi^\dagger\psi - m\psi^\dagger\beta\psi)\, d^3\mathbf{x} = 0.$$

This physically interesting condition does not preclude the existence of a rigorous particlelike solution, and whether or not such a solution can be obtained remains an open question.[6]

[6] A numerical integration calculation has recently been applied to the classical Dirac–Maxwell static field equations by M. Wakano, Progr. Theoret. Phys. (Kyoto) **35**, 1117 (1966). This work supports the existence of a rigorous particlelike solution, but with the total classical field energy E negative.

III. STRONG SATISFACTION OF THE NECESSARY CONDITION: A CLASS OF SOLVABLE MODEL FIELD THEORIES

Let us turn to a more general class of Lorentz-covariant field theories associated with a Lagrangian density of the form[7]

$$\mathfrak{L} = \theta^2 - (\nabla\theta)^2 + i\psi^\dagger\psi + i\psi^\dagger\boldsymbol{\vartheta}\cdot\nabla\psi + \hat{\mathfrak{L}}, \qquad (7)$$

where θ is a real scalar field, ψ is a two-component complex Weyl spinor field ($\psi^\dagger$ its Hermitian adjoint, $\boldsymbol{\vartheta}$ the Pauli matrices), and the interaction Lagrangian density is a generic function

$$\hat{\mathfrak{L}} = \hat{\mathfrak{L}}(\theta, \rho), \qquad \text{where} \qquad \rho \equiv \psi^\dagger\psi. \qquad (8)$$

For time-independent solutions, the energy functional is

$$E = E^{(-1)} + E^{(-2)} + E^{(-3)}, \qquad (9)$$

where

$$E^{(-1)} \equiv \int (\nabla\theta)^2\, d^3\mathbf{x}, \quad E^{(-2)} \equiv -i\int \psi^\dagger\boldsymbol{\vartheta}\cdot\nabla\psi\, d^3\mathbf{x},$$

$$E^{(-3)} \equiv -\int \hat{\mathfrak{L}}\, d^3\mathbf{x},$$

so the field equations (1) are given by

$$\nabla^2\theta + \tfrac{1}{2}(\partial\hat{\mathfrak{L}}/\partial\theta) = 0, \quad i\boldsymbol{\vartheta}\cdot\nabla\psi + (\partial\hat{\mathfrak{L}}/\partial\rho)\psi = 0. \qquad (10)$$

The latter equations can be used to reduce the statement of the pseudovirial theorem

$$\int[-(\nabla\theta)^2 + 2i\psi^\dagger\boldsymbol{\vartheta}\cdot\nabla\psi + 3\hat{\mathfrak{L}}]\, d^3\mathbf{x} = 0 \qquad (11)$$

to the form of Eq. (6) with

$$\mathfrak{F} = -\tfrac{1}{2}\theta(\partial\hat{\mathfrak{L}}/\partial\theta) - 2\rho(\partial\hat{\mathfrak{L}}/\partial\rho) + 3\hat{\mathfrak{L}}. \qquad (12)$$

Now consider the restricted class of scalar-spinor theories for which Eq. (6) with (12)—a necessary condition for the existence of particlelike solutions—is satisfied in the strongest possible way, namely by having the quantity $\mathfrak{F}$ vanish identically, irrespective of the fields for all $\mathbf{x}$. Then according to (12), the "strong satisfaction condition" $\mathfrak{F} \equiv 0$ implies that the interaction Lagrangian density takes the form

$$\hat{\mathfrak{L}} = \theta^2\rho G, \qquad G = G(\kappa), \qquad \kappa \equiv \rho/\theta^4, \qquad (13)$$

[7] Although Lorentz-covariant classical model theories featuring a Weyl spinor field have not been studied heretofore, the literature contains several seemingly related (but not exactly solvable) nonlinear model theories. Interesting work along these lines was initiated by: N. Rosen, Phys. Rev. **55**, 94 (1939); A. C. Menius, Jr., and N. Rosen, Phys. Rev. **62**, 436 (1942), and extended by: R. J. Finkelstein, Phys. Rev. **75**, 1079 (1949). The pseudovirial theorem actually precludes the rigorous existence of singularity-free and spatially localized static or stationary solutions to some of the simpler nonlinear model theories which have been considered in the literature.

and Eqs. (10) become

$$\nabla^2\theta + (3G - 2H)\psi^\dagger\psi\theta = 0, \qquad (14)$$
$$i\vec{\sigma}\cdot\nabla\psi + H\theta^2\psi = 0, \quad H = H(\kappa) \equiv G + \kappa(dG/d\kappa).$$

Note that the arbitrary differentiable real function G, the related function H, and their common argument $\kappa = \psi^\dagger\psi/\theta^4$ all have the dimensions of (energy)$^{-1}$.

Concentrating attention on the restricted class of scalar-spinor theories with the time-independent field equations (14), we find that the rigorous singularity-free spherically symmetric solution is obtainable and expressed in closed form by

$$\theta = \pm(3aH^{-1})^{\frac{1}{2}}(|\mathbf{x}|^2 + a^2)^{-\frac{1}{2}}, \qquad (15)$$
$$\psi = 3a\kappa^{\frac{1}{2}}H^{-1}(|\mathbf{x}|^2 + a^2)^{-\frac{3}{2}}(i\vec{\sigma}\cdot\mathbf{x} + a)u,$$

provided that the equation

$$H^2 + 3\kappa(2H - 3G) = 0 \qquad (16)$$

admits a real positive root $\kappa = \text{const}$ for which $H \neq 0$. In (15), the constant Weyl spinor u is normalized to unity, $u^\dagger u = 1$, and the "size parameter" a has the same sign as H but is otherwise a free (nonzero) constant of homology, stemming from the scale invariance of Eqs. (14):

$$\theta(\mathbf{x}) \to \lambda^{\frac{1}{2}}\theta(\lambda\mathbf{x}), \qquad \psi(\mathbf{x}) \to \lambda\psi(\lambda\mathbf{x})$$

for all real $\lambda > 0$. It is a simple matter to verify that (15) satisfies (14) with κ a constant real positive root of (16), although there is no systematic procedure for obtaining this rigorous solution by straightforward integration of the nonlinear coupled equations.

All of the energy terms in (9) are finite and evaluated easily with (15),

$$E^{(-1)} = \tfrac{1}{4}9\pi^2\,|H^{-1}|, \quad E^{(-2)} = \tfrac{1}{4}27\pi^2\kappa H^{-1}\,|H^{-1}|, \qquad (17)$$
$$E^{(-3)} = -\tfrac{1}{4}27\pi^2\kappa G\,|H^{-3}|,$$

the pseudovirial theorem reproducing the condition (16). The total energy or "particle rest mass" is thus

$$E = \tfrac{3}{2}\pi^2(1 + \tfrac{3}{2}\kappa H^{-1})\,|H^{-1}|, \qquad (18)$$

while the second variation of the energy functional with respect to spatial dilatations is evaluated as

$$\begin{aligned}\delta^2E &\equiv \{(d^2/d\lambda^2)E[\theta(\lambda\mathbf{x}),\ \psi(\lambda\mathbf{x})]\}_{\lambda=1} \\ &= \sum_{w=-1}^{-3}(w^2 - w)E^{(w)}[\theta(\mathbf{x}),\ \psi(\mathbf{x})] \\ &= 2E^{(-1)} + 6E^{(-2)} + 12E^{(-3)} \\ &= -\frac{9\pi^2}{2}(1 + 3\kappa H^{-1})\,|H^{-1}|,\end{aligned} \qquad (19)$$

the sign of the latter quantity relating the energetic stability or instability of the "particle" with respect to changes in the functional form of the solution (15) induced by infinitesimal spatial dilatations. It should be noted that in order for both (18) and (19) to be positive, H evaluated at the root of (16) must satisfy the bounding relation $-3\kappa < H < -\frac{3}{2}\kappa$, and thus, as a consequence of (16), G evaluated at the root of (16) must satisfy the bounding relation $-\kappa < G < -\frac{3}{4}\kappa$.

Qualitatively speaking, stability or instability in the dynamical sense is likely to be determined solely by a spherically symmetric perturbative mode.[3] Therefore, we may conjecture that "energetic stability" of a particlelike solution, expressed by (19) as positive, also indicates the *dynamical stability* of the solution. Support for this tentative mathematical conjecture is provided by the general significance of energy in the dynamics of physical systems, although a formal proof of this conjecture (which must certainly involve complicated qualifying conditions for admissible perturbative modes) is difficult to establish, mainly because admissible dynamical perturbations of the coupled fields cannot be wholly independent, but only independent modulo all conservation laws of the theory implied by the action principle.

We conclude this section with some examples for $G = G(\kappa)$ and brief discussions of the salient properties of the associated particlelike solutions (15).

Example 1: $G \equiv m^{-1}$, where m is a positive constant (dimensions of mass). Obviously this is the simplest admissible form of G. It gives $H = m^{-1}$ and we find that the root of (16) is $\kappa = (3m)^{-1}$. The particle rest mass (18) is $E = (\frac{9}{4}\pi^2)m$, a positive quantity, while the second variation of the energy functional (19) is negative, $\delta^2E = -9\pi^2m$, showing that the particlelike solution is energetically unstable.

Example 2: $G = \gamma m^{-2}\kappa^{-1} - m^{-1}$, γ and $m(\neq 0)$ constants. This is the most general form of G for which the interaction Lagrangian density is composed entirely of terms in positive integer powers of θ and $\psi^\dagger\psi$: $\hat{\mathfrak{L}} = \gamma m^{-2}\theta^6 - m^{-1}\psi^\dagger\psi\theta^2$. In this case, we have $H = -m^{-1}$ and find that the root of (16) is $\kappa = (3\gamma - \frac{1}{3})m^{-1}$, requiring that either $\gamma > \frac{1}{3}$, $m > 0$ or $\gamma < \frac{1}{3}$, $m < 0$ for a solution. The particle rest mass (18) is $E = (\frac{9}{4}\pi^2)(1 - 3\gamma)m$, a positive quantity if either $\gamma < \frac{1}{3}$, $m > 0$ or $\gamma > \frac{1}{3}$, $m < 0$, the latter possibility excluded by the requirement that κ be positive. For (19) we have $\delta^2E =$

$(\frac{9}{2}\pi^2)(9\gamma - 2)m$, a positive quantity if $\gamma > \frac{2}{9}$ with $m > 0$. Hence, κ, E, and $\delta^2 E$ are all positive if $\frac{2}{9} < \gamma < \frac{1}{3}$, $m > 0$.

Example 3: $G = g\kappa^{-1}$, g a positive constant. This is the form of G for a self-interacting scalar field and a free spinor field, a singular case that requires the special treatment provided elsewhere.[3] The particlelike solution, however, can be obtained formally from Eqs. (15) by putting $3\,|H^{-1}| = g^{-\frac{1}{2}}$ and $\kappa = 0$, the correct expressions for (18) and (19) also being given by this substitution. E is positive but $\delta^2 E$ is negative.

Example 4: $G = g\kappa^{\frac{1}{2}}$, g a positive constant. This is the form of G for a self-interacting spinor field and a free scalar field, also a singular case that requires special treatment. Once again the particlelike solution can be obtained formally from Eqs. (15), in this case by the limiting procedure $H^{-1} \to 0$, $3\kappa^{\frac{1}{2}}H^{-1} \to 2g^{-1}$, the correct expressions for (18) and (19) also being secured by this limiting procedure. E is positive but $\delta^2 E$ is negative.

Example 5: $G = -g\kappa^{p+1}$, g and p positive constants. Here we have $H = -g(p + 2)\kappa^{p+1}$ and the unique positive root of (16) is

$$\kappa = [3(2p + 1)/g(p + 2)^2]^{1/p}.$$

The particle rest energy

$$E = [9\pi^2 p/4(2p + 1)]\,|H^{-1}|$$

is a positive quantity, while

$$\delta^2 E = [9\pi^2(1 - p)/2(2p + 1)]\,|H^{-1}|$$

is a positive quantity provided that $p < 1$.

Example 6: $G = \gamma\kappa$. By virtue of (16), the particlelike solutions are obtainable only if the dimensionless coupling constant $\gamma = -\frac{3}{4}$. In this case κ is a free positive parameter, the total energy (18) vanishes, and the quantity (19) is positive, $\delta^2 E = 3\pi^2\kappa^{-1}$.

Example 7: $G = 2m^{-2}\kappa^{-1}[1 - (1 - m\kappa)^{\frac{1}{2}}] - 3m^{-1}$, m a positive constant. This seemingly complicated example is of interest because it gives $H = -3m^{-1}[1 - (1 - m\kappa)^{\frac{1}{2}}]$, and therefore the condition (16) is satisfied identically (that is, as a differential equation). Hence, all positive values of $\kappa \leq m^{-1}$ are admissible. The particle rest energy $E = (\pi^2/4)m$, a quantity independent of κ, and the second variation of the energy functional $\delta^2 E = (\frac{3}{2}\pi^2)\,|1 - (1 - m\kappa)^{-\frac{1}{2}}|^{-1}\,m$ are both positive for $\kappa < m^{-1}$.

IV. INSUFFICIENCY OF THE STRONG SATISFACTION CONDITION FOR A RELATED CLASS OF THEORIES

Strong satisfaction of the pseudovirial theorem's corollary leads to rigorous particlelike solutions for the class of Lorentz-covariant field theories discussed in the preceding section, but the strong satisfaction condition is not always sufficient to guarantee such localized singularity-free solutions. To show by example that the strong satisfaction condition can be insufficient, we consider the Lorentz-covariant theories which feature a real scalar field in interaction with a four-component complex Dirac spinor field, a class of theories closely related to those discussed previously with a Weyl spinor and the generic Lagrangian density (7) with (13).

Time-independent solutions to the class of theories with a Dirac spinor and the strong satisfaction condition are derived from an energy functional of the form

$$E = E^{(-1)} + E^{(-2)} + E^{(-3)},$$

where

$$E^{(-1)} \equiv \int (\nabla\theta)^2\, d^3\mathbf{x}, \quad E^{(-2)} \equiv -i \int \psi^\dagger \boldsymbol{\alpha}\cdot\nabla\psi\, d^3\mathbf{x},$$

$$E^{(-3)} \equiv -\int \theta^2 \psi^\dagger\beta\psi G\, d^3\mathbf{x}. \tag{20}$$

[$G = G(\kappa)$ is a real differentiable function of $\kappa \equiv \psi^\dagger\beta\psi/\theta^4$.] Here, ψ is a four-component complex spinor with $\boldsymbol{\alpha}$ and β the Hermitian Dirac matrices. Equations (1) take a form similar to Eqs. (14) in the case of (20),

$$\nabla^2\theta + (3G - 2H)\psi^\dagger\beta\psi\theta = 0,$$
$$i\boldsymbol{\alpha}\cdot\nabla\psi + H\theta^2\beta\psi = 0, \qquad H \equiv G + \kappa(dG/d\kappa). \tag{21}$$

Now let us require the otherwise arbitrary G to be a function such that H is of definite sign for all (positive or negative) values of κ (e.g., $G = m^{-1}$, a nonzero constant). Then it follows from the relation

$$\nabla\cdot(i\psi^\dagger\beta\boldsymbol{\alpha}\psi) + 2H\theta^2\psi^\dagger\psi = 0 \tag{22}$$

obtained from the field equations (together with $\beta^2 = 1$ and $\boldsymbol{\alpha}\beta = -\beta\boldsymbol{\alpha}$) that no singularity-free and well-localized nontrivial solution exists for Eqs. (21) (with κ not necessarily constant), because we have

$$\int H\theta^2\psi^\dagger\psi\, d^3\mathbf{x} = 0 \tag{23}$$

if

$$\lim_{|\mathbf{x}|\to\infty} [|\mathbf{x}|^2\, \psi^\dagger\psi] = 0. \tag{24}$$

Hence, the strong satisfaction condition is insufficient to guarantee particlelike solutions for this class of field theories with H of definite sign.

Nonexistence of Localized Periodic Solutions to Nonlinear Field Theories

GERALD ROSEN
Drexel Institute of Technology, Philadelphia, Pennsylvania
(Received 28 March 1966)

A simple proof is given which precludes the existence of any spatially localized and temporally periodic singularity-free solution to the c-number field equation for a physically interesting Lorentz-covariant self-interacting scalar theory.

IT has been proved by Jörgens[1] that the nonlinear c-number field equation

$$\ddot{\theta} - \nabla^2\theta + m^2\theta + \tfrac{1}{2}gp\,|\theta|^{p-2}\,\theta = 0 \qquad (1)$$

for a scalar function $\theta = \theta(\mathbf{x}, t)$ (either complex-valued or real-valued and C^2 for all $\mathbf{x}$ and all $t \geq 0$) has a *global solution* for suitably smooth prescribed initial data, if $2 < p < 6$ with m^2 and g nonnegative constant parameters. Does this global existence theorem lend some rigorous mathematical support to the quantum theory models which feature a self-interacting complex or real scalar field and a Lagrangian density

$$\mathfrak{L} = |\dot{\theta}|^2 - |\nabla\theta|^2 - m^2\,|\theta|^2 - g\,|\theta|^p \qquad (2)$$

of the form associated with Eq. (1)? It is shown here that the classical field equation (1) has no global solutions which are localized in space and periodic in time if $p \geq 2$. Hence, there can exist no stationary state in a quantum field theory based on a Lagrangian density (2) with $p \geq 2$ that would correspond in the classical limit $\hbar \to 0$ to a spatially localized and temporally periodic solution to Eq. (1).

To prove the nonexistence of any localized periodic solution to Eq. (1) with $g \geq 0$ and $p \geq 2$, we first note that the action principle

$$\delta I/\delta\theta\,(\mathbf{x}, t) = 0, \qquad I = I[\theta(\mathbf{x}, t)] \equiv \int_0^T \int \mathfrak{L}\, d^3\mathbf{x}\, dt \qquad (3)$$

for a solution periodic in time, $\theta(\mathbf{x}, t + T) = \theta(\mathbf{x}, t)$, implies the global condition

$$\left(\frac{dI[\theta(\lambda\mathbf{x}, t)]}{d\lambda}\right)_{\lambda=1} = 0. \qquad (4)$$

Eq. (4) works out to give

$$\int_0^T \int [-3\,|\dot{\theta}|^2 + |\nabla\theta|^2 + 3m^2\,|\theta|^2 + 3g\,|\theta|^p]\, d^3\mathbf{x}\, dt = 0, \qquad (5)$$

provided that the $\mathbf{x}$ integrations converge over all space[2]; an alternative way to derive Eq. (5) is to multiply Eq. (1) by $\mathbf{x}\cdot\nabla\theta^*$ and integrate the real part of the resulting equation over all $\mathbf{x}$ and over the range 0 to T for t. On the other hand, if we work out

$$\left(\frac{dI[\xi\theta(\mathbf{x}, t)]}{d\xi}\right)_{\xi=1} = 0, \qquad (6)$$

or if Eq. (1) is multiplied by θ^* and integrated, we obtain

$$\int_0^T \int [-|\dot{\theta}|^2 + |\nabla\theta|^2 + m^2\,|\theta|^2 + \tfrac{1}{2}gp\,|\theta|^p]\, d^3\mathbf{x}\, dt = 0, \qquad (7)$$

again provided that the $\mathbf{x}$ integrations converge and T is the period of the solution. Finally, by subtracting $\frac{1}{3}$ of Eq. (5) from Eq. (7), we get

$$\int_0^T \int [\tfrac{2}{3}\,|\nabla\theta|^2 + (\tfrac{1}{2}p - 1)g\,|\theta|^p]\, d^3\mathbf{x}\, dt = 0, \qquad (8)$$

a global relation which implies the necessary condition $p < 2$ for existence of a localized periodic solution with $g > 0$. Therefore, all such global solutions are precluded in a classical field theory[3] with $g \geq 0$ and $p \geq 2$.

[1] K. Jörgens, Z. Math. 77, 295 (1961).

[2] For time-independent solutions, Eq. (5) holds with T arbitrary and exemplifies the general *pseudovirial theorem* discussed by the present author in J. Math. Phys. 7, 2066 (1966).

[3] If the words "such global solutions" are read "rigorous stationary states," the conclusion stated here is believed to produce a valid correspondent in the associated quantum field theory. For a pertinent result in this regard, see, G. Rosen, Phys. Rev. Letters 16, 704 (1966).

Some Observations on Enveloping Algebras of Noncompact Groups*

JOE ROSEN† AND PAUL ROMAN
Department of Physics, Boston University, Boston, Massachusetts
(Received 31 March 1966; revised manuscript received 4 April 1966)

In order to provide the particle physicist with useful mathematical tools (and a large number of interesting identities), the enveloping algebras of the noncompact homogeneous $O(p, q)$, $U(p, q)$, and inhomogeneous $IO(p, q)$, $IU(p, q)$ groups are discussed in some detail. The construction of the generators of $O(p, q + 1)$ and $O(p + 1, q)$ from the enveloping algebra of $IO(p, q)$ and the analog result for the corresponding pseudo-unitary groups are also presented.

I. INTRODUCTION

ONE of the most intriguing and promising recent developments in elementary-particle physics is the recognition of the importance of noncompact groups and Lie algebras. Noncompact groups appear to play various roles, both in the classification of hadron-supermultiplets and, more generally, in diverse attempts to reformulate quantumdynamics. For a review of these emerging novel ideas we refer the reader to Refs. 1–3 . Some of the more specific developments are illustrated in Refs. 4–6.

In addition to the study of the representations of the noncompact groups and their Lie algebras, it appears that a thorough knowledge of the enveloping algebras will constitute a most important tool. For one thing, the Casimir invariants are particular elements of the enveloping algebra, and, apart from their direct relation to physical observables, their knowledge is a necessary requirement when one seeks the systematic classification of representations. Furthermore, as pointed out vividly by Sudarshan,[1] every dynamical variable of a quantized system can be identified with an element of the enveloping algebra of the noninvariance group.[5,7] Moreover, the primitive dynamical variables may be *defined* as certain suitable elements of this enveloping algebra.[1,8]

In view of these remarks, we believe that it would be useful to work out commutation relations involving some lower-degree elements of the enveloping algebras of the most promising noncompact groups, i.e., of the homogeneous and inhomogeneous pseudo-orthogonal and pseudo-unitary groups. A collection of such expressions will be of value for the physicist applying the new notions related to the role of noncompact groups in particle physics, and to provide these formulas is the purpose of the present paper. When constructing these elements of the enveloping algebras, we found a number of interesting identities, and the awareness of such simplifying relations is bound to shortcut many computations.

After general remarks in Sec. II, Sec. III treats the enveloping algebras of the pseudo-orthogonal groups $O(p, q)$. Section IV discusses the enveloping algebras of the inhomogeneous pseudo-orthogonal groups $IO(p, q)$ and shows the construction of the generators of $O(p, q + 1)$ and $O(p + 1, q)$ from elements of the enveloping algebra of $IO(p, q)$. Sections IV and V treat the enveloping algebras of the pseudo-unitary groups $U(p, q)$ and the inhomogeneous pseudo-unitary groups $IU(p, q)$ respectively, the latter section showing the construction of the generators of $U(p, q + 1)$ and $U(p + 1, q)$ from elements of the enveloping algebra of $IU(p, q)$.

II. DEFINITIONS AND GENERALITIES

Let **A** be a Lie algebra of order m characterized by the relation[9,10]

* This research was sponsored by the United States Air Force under Grant No. AF-AFOSR-385-65.

† On leave of absence from Tel-Aviv University, from which a travel grant is gratefully acknowledged.

[1] E. C. G. Sudarshan, in the Proceedings of the Fourth Eastern U. S. Theoretical Physics Conference (Stony Brook, 1965), and Syracuse University Preprint 1206-SU-45.

[2] Y. Ne'eman, in the Proceedings of the Pacific Summer School in Physics (Honolulu, 1965), and Tel-Aviv University Preprint TAUP-2-65.

[3] Y. Ne'eman and Y. Dothan, in the Proceedings of the Athens (Ohio) Conference on Resonant Particles (1965).

[4] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters **17**, 148 (1965).

[5] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Letters **15**, 1041 (1965).

[6] A. O. Barut, Phys. Rev. **135**, B839 (1964); **139**, B1433 (1965); A. O. Barut and A. Böhm, Phys. Rev. **139**, B1107 (1965).

[7] E. C. G. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, Phys. Letters **19**, 322 (1965).

[8] E. C. G. Sudarshan, in the Proceedings of the Toronto Conference on Symmetries (1965).

[9] Summation is to be performed over pairs of equal upper and lower indices.

[10] For real forms of Lie algebras, our structure constants are always real. We fully concur with the view of R. Hermann, *Lie Groups for Physicists* (W. A. Benjamin, Inc., New York, 1966), that the introduction of factors of i masks the algebraic structure without offering any real advantage.

$$[X_i, X_j] = c_{ij}^k X_k \tag{1}$$

among the elements of a basis $X_1, \cdots, X_m$. The enveloping algebra of **A**, to be denoted by $\mathfrak{A}$, consists of all products of all degrees of the elements of **A** with themselves. For a basis of $\mathfrak{A}$ one can take the linearly independent set[11] X_i, X_iX_j, $X_iX_jX_k$, $\cdots$, $(i, j, k, \cdots = 1, \cdots, m$ and $i \leq j \leq k \leq \cdots)$. $\mathfrak{A}$ is thus of infinite order. It can be shown by induction and Eq. (1) that the commutator of two basis elements of $\mathfrak{A}$ from the above sequence is always of the form

$\mathrm{ad}X_i$(qth-degree basis elemnts)

= linear combination of $(p + q - 1)$th- and lower-degree basis elements. (2)

If **A** is represented by complex or real $n \times n$ matrices, the corresponding representation of $\mathfrak{A}$ is not faithful, since it can be at most a representation of the Lie algebra of the complex or real general linear group in n dimensions [$GL(n, C)$ or $GL(n, R)$]. Thus, care must be exercised to distinguish between those algebraic relations within $\mathfrak{A}$ which are true generally and those which are representation-dependent.

The algebra **A** as a vector space serves as the representation space of the adjoint representation of **A**, where the basis element X_i is represented by the linear transformation $\mathrm{ad}X_i$ whose action on the basis vectors of **A** is

$$\mathrm{ad}\, X_i(X_j) \equiv [X_i, X_j] = c_{ij}^k X_k. \tag{3}$$

Similarly, $\mathfrak{A}$ serves as a representation space for **A**. This representation is reducible, but in general not completely reducible, since, according to Eq. (2), the subspace spanned by all basis elements of a certain degree and less is invariant under $\mathrm{ad}X_i$:

$\mathrm{ad}X_i$ (qth-degree basis element)

= linear combination of qth- and lower-degree basis elements. (4)

III. PSEUDO-ORTHOGONAL GROUPS

The n-dimensional pseudo-orthogonal group $O(p, q)$, $p + q = n$, is the real linear group in n dimensions preserving the pseudo-Euclidean metric $[+1, \cdots, +1(p \text{ times}), -1, \cdots, -1(q \text{ times})]$. A basis of the corresponding Lie algebra [of order $\frac{1}{2}n(n-1)$] is provided by the elements $M_{\mu\nu} = -M_{\nu\mu}$ obeying

$$[M_{\mu\nu}, M_{\rho\sigma}] = g_{\mu\rho}M_{\nu\sigma} - g_{\nu\rho}M_{\mu\sigma} + g_{\mu\sigma}M_{\rho\nu} - g_{\nu\sigma}M_{\rho\mu}, \tag{5}$$

where all Greek indices run from 1 to n, $g_{11} = \cdots = g_{pp} = -g_{p+1\,p+1} = \cdots = -g_{nn} = 1$, and $g_{\mu\nu} = 0$ for $\mu \neq \nu$. Raising and lowering of indices is performed, as usual, with $g_{\mu\nu}$ and $g^{\mu\nu}$, where $g^{\mu\nu}g_{\nu\rho} = \delta^\mu_\rho$.

The second-degree basis elements of the enveloping algebra $M_{\mu\nu}M_{\rho\sigma}$ obey

$$[M_{\alpha\beta}, M_{\mu\nu}M_{\rho\sigma}] = g_{\alpha\mu}M_{\beta\nu}M_{\rho\sigma} - g_{\beta\mu}M_{\alpha\nu}M_{\rho\sigma} + g_{\alpha\nu}M_{\mu\beta}M_{\rho\sigma} - g_{\beta\nu}M_{\mu\alpha}M_{\rho\sigma} + g_{\alpha\rho}M_{\mu\nu}M_{\beta\sigma} - g_{\beta\rho}M_{\mu\nu}M_{\alpha\sigma} + g_{\alpha\sigma}M_{\mu\nu}M_{\rho\beta} - g_{\beta\sigma}M_{\mu\nu}M_{\rho\alpha}. \tag{6}$$

Certain linear combinations of these are of interest:

(a) The antisymmetric $(M_{\rho\alpha}M^\alpha_\sigma - M_{\sigma\alpha}M^\alpha_\rho)/(n-2)$ equals $M_{\sigma\rho}$.

(b) The symmetric

$$S_{\rho\sigma} = \tfrac{1}{2}(M_{\rho\alpha}M^\alpha{}_\sigma + M_{\sigma\alpha}M^\alpha_\rho) \tag{7}$$

obeys

$$[M_{\mu\nu}, S_{\rho\sigma}] = g_{\mu\rho}S_{\nu\sigma} - g_{\nu\rho}S_{\mu\sigma} + g_{\mu\sigma}S_{\rho\nu} - g_{\nu\sigma}S_{\rho\mu}. \tag{8}$$

(c) The scalar

$$M^2 = \tfrac{1}{2}S_\alpha{}^\alpha = \tfrac{1}{2}M_{\alpha\beta}M^{\beta\alpha} \tag{9}$$

obeys

$$[M_{\mu\nu}, M^2] = 0, \tag{10}$$

and is therefore a second-degree Casimir operator[12] of $O(p, q)$.

In the defining $(n \times n)$ representation of the Lie algebra of $O(p, q)$, the matrix[13] $\overline{M^2}$ is $(n-1)$ times the unit matrix, and the $\bar{S}'_{\mu\nu} = 2\bar{S}_{\mu\nu}/(n-2)$ obey Eq. (8), with $\bar{S}'_{\mu\nu}$ substituted for $S_{\mu\nu}$, and[14]

$$[\bar{S}'_{\mu\nu}, \bar{S}'_{\rho\sigma}] = -(g_{\mu\rho}\bar{M}_{\nu\sigma} + g_{\nu\rho}\bar{M}_{\mu\sigma} - g_{\mu\sigma}\bar{M}_{\rho\nu} - g_{\nu\sigma}\bar{M}_{\rho\mu}). \tag{11}$$

From Eqs. (5), (8), and (11) it then follows that the matrices $\bar{M}_{\mu\nu}$ and $\bar{S}'_{\mu\nu}$ represent a basis of the Lie algebra of $GL(n, R)$ (of order n^2). This can

[11] A pth-degree product of X's whose factors are not ordered as indicated can be shown to equal a pth-degree basis element plus a linear combination of lesser-degree basis elements by use of Eq. (1).

[12] See E. G. Beltrametti and A. Blasi, Phys. Letters 20, 62 (1966), concerning the number of Casimir operators associated with any Lie algebra.

[13] We use a bar to indicate a matrix representation of a quantity.

[14] This relation does not hold for the algebra elements in general, but only for the matrices representing them here. See comment in Sec. II.

perhaps more easily be seen from the commutation relations of

$$\bar{C}_{\mu\nu} = -\tfrac{1}{2}g_{\mu\mu}(\bar{S}'_{\mu\nu} + \bar{M}_{\mu\nu})\text{(no summation over } \mu\text{)}:$$

$$[\bar{C}_{\mu\nu}, \bar{C}_{\rho\sigma}] = \delta_{\rho\nu}\bar{C}_{\mu\sigma} - \delta_{\mu\sigma}\bar{C}_{\rho\nu}. \tag{12}$$

In addition, the matrices $\bar{M}_{\mu\nu}$ and $i\bar{S}'_{\mu\nu}$ represent a basis of the Lie algebra of the pseudo-unitary group $U(p, q)$ (to be discussed in Sec. V).

IV. INHOMOGENEOUS PSEUDO-ORTHOGONAL GROUPS

The n-dimensional inhomogeneous pseudo-orthogonal group $IO(p, q)$, $p + q = n$, consists of all elements of $O(p, q)$ (see the preceding section) together with all translations in the pseudo-Euclidean space.[15] A basis of the corresponding Lie algebra [of order $\frac{1}{2}n(n + 1)$] is obtained with the elements $M_{\mu\nu} = -M_{\nu\mu}$ and P_μ obeying Eq. (5) and

$$[M_{\mu\nu}, P_\rho] = g_{\mu\rho}P_\nu - g_{\nu\rho}P_\mu, \tag{13}$$

$$[P_\mu, P_\nu] = 0. \tag{14}$$

The second-degree basis elements of the enveloping algebra, $M_{\mu\nu}M_{\rho\sigma}$, $M_{\mu\nu}P_\rho$, $P_\rho M_{\mu\nu}$, and $P_\mu P_\nu$, obey Eq. (6) and

$$\begin{aligned}[M_{\alpha\beta}, M_{\mu\nu}P_\rho] = {} & g_{\alpha\mu}M_{\beta\nu}P_\rho - g_{\beta\mu}M_{\alpha\nu}P_\rho \\ & + g_{\alpha\nu}M_{\mu\beta}P_\rho - g_{\beta\nu}M_{\mu\alpha}P_\rho \\ & + g_{\alpha\rho}M_{\mu\nu}P_\beta - g_{\beta\rho}M_{\mu\nu}P_\alpha,\end{aligned} \tag{15}$$

$$\begin{aligned}[M_{\alpha\beta}, P_\rho M_{\mu\nu}] = {} & g_{\alpha\rho}P_\beta M_{\mu\nu} - g_{\beta\rho}P_\alpha M_{\mu\nu} \\ & + g_{\alpha\mu}P_\rho M_{\beta\nu} - g_{\beta\mu}P_\rho M_{\alpha\nu} \\ & + g_{\alpha\nu}P_\rho M_{\mu\beta} - g_{\beta\nu}P_\rho M_{\mu\alpha},\end{aligned} \tag{16}$$

$$\begin{aligned}[M_{\alpha\beta}, P_\mu P_\nu] = {} & g_{\alpha\mu}P_\beta P_\nu - g_{\beta\mu}P_\alpha P_\nu \\ & + g_{\alpha\nu}P_\mu P_\beta - g_{\beta\nu}P_\mu P_\alpha,\end{aligned} \tag{17}$$

$$\begin{aligned}[P_\alpha, M_{\mu\nu}M_{\rho\sigma}] = {} & -g_{\mu\alpha}P_\nu M_{\rho\sigma} + g_{\nu\alpha}P_\mu M_{\rho\sigma} \\ & - g_{\rho\alpha}M_{\mu\nu}P_\sigma + g_{\sigma\alpha}M_{\mu\nu}P_\rho,\end{aligned} \tag{18}$$

$$\begin{aligned}[P_\alpha, M_{\mu\nu}P_\rho] &= [P_\alpha, P_\rho M_{\mu\nu}] \\ &= -g_{\mu\alpha}P_\nu P_\rho + g_{\nu\alpha}P_\mu P_\rho,\end{aligned} \tag{19}$$

$$[P_\alpha, P_\mu P_\nu] = 0. \tag{20}$$

Interesting linear combinations of these, in addition to those presented in the preceding section, are:

(a) The scalar

$$P^2 = P_\alpha P^\alpha \tag{21}$$

obeys

$$[M_{\mu\nu}, P^2] = [P_\mu, P^2] = 0, \tag{22}$$

so that it is a second-degree Casimir operator of $IO(p, q)$.

(b) The n-vector $(P^\alpha M_{\mu\alpha} - M_{\mu\alpha}P^\alpha)/(n - 1)$ which, however, equals P_μ.

(c) The n-vector

$$J_\mu = \tfrac{1}{2}(P^\alpha M_{\mu\alpha} + M_{\mu\alpha}P^\alpha) \tag{23}$$

obeys

$$[M_{\alpha\beta}, J_\mu] = g_{\alpha\mu}J_\beta - g_{\beta\mu}J_\alpha, \tag{24}$$

$$[P_\alpha, J_\mu] = P_\alpha P_\mu - g_{\alpha\mu}P^2, \tag{25}$$

$$[J_\mu, J_\nu] = -P^2 M_{\mu\nu}, \tag{26}$$

$$[J_\mu, M^2] = J^\alpha M_{\mu\alpha} + M_{\mu\alpha}J^\alpha, \tag{27}$$

$$[P_\mu, M^2] = 2J_\mu. \tag{28}$$

Combining second-degree elements to obtain higher-degree ones, one has:

(d) The scalar $P^\mu J_\mu + J_\mu P^\mu$ is identically zero.

(e) The scalar

$$J^2 = J_\alpha J^\alpha = -[P^\alpha P^\beta S_{\alpha\beta} + \tfrac{1}{4}(n - 1)^2 P^2] \tag{29}$$

obeys

$$[M_{\mu\nu}, J^2] = 0, \tag{30}$$

$$[P_\mu, J^2] = -2P^2 J_\mu, \tag{31}$$

$$[J_\mu, J^2] = -P^2(J^\alpha M_{\mu\alpha} + M_{\mu\alpha}J^\alpha). \tag{32}$$

(f) The scalar

$$Z = P^2 M^2 + J^2. \tag{33}$$

By Eqs. (10), (22), (28), (30), and (31) this obeys

$$[M_{\mu\nu}, Z] = [P_\mu, Z] = 0, \tag{34}$$

making Z a fourth-degree Casimir operator of $IO(p, q)$. In addition, by Eqs. (27) and (32) or just from Eq. (34), it obeys

$$[J_\mu, Z] = 0. \tag{35}$$

We define now

$$N_{\mu\nu} = M_{\mu\nu}, \tag{36}$$

$$N_{\mu\,n+1} = -N_{n+1\,\mu} = \begin{cases} J_\mu/(-P^2)^{\frac{1}{2}} & \text{for } g_{n+1\,n+1} = +1, \\ J_\mu/(+P^2)^{\frac{1}{2}} & \text{for } g_{n+1\,n+1} = -1. \end{cases} \tag{37}$$

The inverse square root in Eq. (37) is assumed to be suitably defined through a power series. In any case, it is clearly meaningful in any irreducible representation of the Lie algebra, since by Schur's

[15] For a discussion of $IO(1, 3)$, the inhomogeneous Lorentz (or Poincaré) group, see, e.g., Yu. M. Shirokov, Zh. Eksperim. i Teor. Fiz. **33**, 861 (1957) [English transl.: Soviet Phys.—JETP **6**, 664 (1958)].

lemma $\overline{P^2}$ is then just a multiple of the unit matrix. (This is, of course, meaningless in a representation for which $\overline{P^2} = 0$.) No ordering problem arises, as J_μ and P^2 commute. One then finds that

$$[N_{ab}, N_{cd}] = g_{ac}N_{bd} - g_{bc}N_{ad} + g_{ad}N_{cb} - g_{bd}N_{ca}, \quad (38)$$

where $a, b, c, d = 1, \cdots, n+1$, i.e., then N_{ab} form a basis of the Lie algebra[16] of $O(p, q, +1)$ or $O(p+1, q)$. The scalar

$$N^2 = \tfrac{1}{2}N_{ab}N^{ba} \quad (39)$$

therefore obeys

$$[N_{ab}, N^2] = 0, \quad (40)$$

since it is the second-degree Casimir operator of $O(p, q+1)$ or $O(p+1, q)$. On the other hand, one finds

$$N^2 = M^2 + J^2/P^2 = Z/P^2, \quad (41)$$

where a remark similar to the one following Eq. (37) holds. This demonstrates the consistency of Eq. (40) with Eqs. (34) and (35). Equation (41) also shows the relation between the second-degree Casimir operator N^2 of $O(p, q+1)$ or $O(p+1, q)$ and the fourth-degree Casimir operator Z of $IO(p, q)$.

V. PSEUDO-UNITARY GROUPS

The n-dimensional pseudo-unitary group $U(p, q)$, $p + q = n$, is the complex linear group in n dimensions preserving the pseudo-unitary metric

$$[+1, \cdots, +1(p \text{ times}), -1, \cdots, -1(q \text{ times})].$$

A basis of the corresponding Lie algebra (of order n^2) is obtained with the elements $E_{\mu\nu} = -E_{\nu\mu}$ and $F_{\mu\nu} = F_{\nu\mu}$ obeying

$$[E_{\mu\nu}, E_{\rho\sigma}] = g_{\mu\rho}E_{\nu\sigma} - g_{\nu\rho}E_{\mu\sigma} + g_{\mu\sigma}E_{\rho\nu} - g_{\nu\sigma}E_{\rho\mu}, \quad (42)$$

$$[F_{\mu\nu}, F_{\rho\sigma}] = g_{\mu\rho}E_{\nu\sigma} + g_{\nu\rho}E_{\mu\sigma} - g_{\mu\sigma}E_{\rho\nu} - g_{\nu\sigma}E_{\rho\mu}, \quad (43)$$

$$[E_{\mu\nu}, F_{\rho\sigma}] = g_{\mu\rho}F_{\nu\sigma} - g_{\nu\rho}F_{\mu\sigma} + g_{\mu\sigma}F_{\rho\nu} - g_{\nu\sigma}F_{\rho\mu}, \quad (44)$$

where all Greek indices run from 1 to n, and again $g_{11} = \cdots = g_{pp} = -g_{p+1\,p+1} = \cdots = -g_{nn} = 1$, $g_{\mu\nu} = 0$ for $\mu \neq \nu$. Comparing with Eqs. (5), (8), and (11), one justifies the remark following Eq. (12) at the end of Sec. III. From Eqs. (43) and (44), it follows that the contraction $F_\alpha{}^\alpha$ obeys

$$[E_{\mu\nu}, F_\alpha{}^\alpha] = [F_{\mu\nu}, F_\alpha{}^\alpha] = 0 \quad (45)$$

and is therefore a first-degree Casimir operator of $U(p, q)$.

The second-degree basis elements of the enveloping algebra are $E_{\mu\nu}E_{\rho\sigma}$, $E_{\mu\nu}F_{\rho\sigma}$, $F_{\mu\nu}E_{\rho\sigma}$, and $F_{\mu\nu}F_{\rho\sigma}$. Interesting linear combinations of these are:

(a) The antisymmetric $(E_{\rho\alpha}E^\alpha_\sigma - E_{\sigma\alpha}E^\alpha_\rho)/(n-2)$, which, however, equals $E_{\sigma\rho}$.

(b) The symmetric

$$T_{\rho\sigma} = \tfrac{1}{2}(E_{\rho\alpha}E^\alpha_\sigma + E_{\sigma\alpha}E^\alpha_\rho), \quad (46)$$

which obeys [cf. Eq. (44)]

$$[E_{\mu\nu}, T_{\rho\sigma}] = g_{\mu\rho}T_{\nu\sigma} - g_{\nu\rho}T_{\mu\sigma} + g_{\mu\sigma}T_{\rho\nu} - g_{\nu\sigma}T_{\rho\mu}. \quad (47)$$

(c) The antisymmetric $(F_{\rho\alpha}F^\alpha_\sigma - F_{\sigma\alpha}F^\alpha_\rho)/(n+1)$ which equals $E_{\rho\sigma}$.

(d) The symmetric

$$U_{\rho\sigma} = \tfrac{1}{2}(F_{\rho\alpha}F^\alpha_\sigma + F_{\sigma\alpha}F^\alpha_\rho), \quad (48)$$

which obeys [cf. Eq. (44)]

$$[E_{\mu\nu}, U_{\rho\sigma}] = g_{\mu\rho}U_{\nu\sigma} - g_{\nu\rho}U_{\mu\sigma} + g_{\mu\sigma}U_{\rho\nu} - g_{\nu\sigma}U_{\rho\mu}. \quad (49)$$

(e) The symmetric

$$V_{\rho\sigma} = \tfrac{1}{4}(E_{\rho\alpha}F^\alpha_\sigma - F_{\rho\alpha}E^\alpha_\sigma + E_{\sigma\alpha}F^\alpha_\rho - F_{\sigma\alpha}E^\alpha_\rho), \quad (50)$$

which obeys [cf. Eq. (44)]

$$[E_{\mu\nu}, V_{\rho\sigma}] = g_{\mu\rho}V_{\nu\sigma} - g_{\nu\rho}V_{\mu\sigma} + g_{\mu\sigma}V_{\rho\nu} - g_{\nu\sigma}V_{\rho\mu}. \quad (51)$$

(f) The antisymmetric

$$W_{\rho\sigma} = \tfrac{1}{4}(E_{\rho\alpha}F^\alpha_\sigma + F_{\rho\alpha}E^\alpha_\sigma - E_{\sigma\alpha}F^\alpha_\rho - F_{\sigma\alpha}E^\alpha_\rho), \quad (52)$$

(g) and the symmetric

$$D_{\rho\sigma} = \tfrac{1}{2}(U_{\rho\sigma} - T_{\rho\sigma}) = \tfrac{1}{4}(F_{\rho\alpha}F^\alpha_\sigma + F_{\sigma\alpha}F^\alpha_\rho - E_{\rho\alpha}E^\alpha_\sigma - E_{\sigma\alpha}E^\alpha_\rho), \quad (53)$$

which obeys

$$[E_{\mu\nu}, W_{\rho\sigma}] = g_{\mu\rho}W_{\nu\sigma} - g_{\nu\rho}W_{\mu\sigma} + g_{\mu\sigma}W_{\rho\nu} - g_{\nu\sigma}W_{\rho\mu} \quad (54)$$

[cf. Eq. (42)],

$$[E_{\mu\nu}, D_{\rho\sigma}] = g_{\mu\rho}D_{\nu\sigma} - g_{\nu\rho}D_{\mu\sigma} + g_{\mu\sigma}D_{\rho\nu} - g_{\nu\sigma}D_{\rho\mu} \quad (55)$$

[cf. Eq. (44)],

$$[F_{\mu\nu}, D_{\rho\sigma}] = g_{\mu\rho}W_{\nu\sigma} + g_{\nu\rho}W_{\mu\sigma} - g_{\mu\sigma}W_{\rho\nu} - g_{\nu\sigma}W_{\rho\mu} \quad (56)$$

[cf. Eq. (43)], and

$$[W_{\mu\nu}, F_{\rho\sigma}] = g_{\mu\rho}D_{\nu\sigma} - g_{\nu\rho}D_{\mu\sigma} + g_{\mu\sigma}D_{\rho\nu} - g_{\nu\sigma}D_{\rho\mu} \quad (57)$$

[cf. Eq. (42)].

[16] This construction was performed in general by A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965), and for $O(4.1)$ by M. Y. Han (preprint, Syracuse University and University of Pittsburgh, 1965). See also A. Böhm, Phys. Rev. **145**, 1212 (1966), and C. Fronsdal, Rev. Mod. Phys. **37**, 211 (1965).

(h) The contraction

$$E^2 = \tfrac{1}{2}T_\alpha{}^\alpha = \tfrac{1}{2}E_{\alpha\beta}E^{\beta\alpha} \tag{58}$$

obeys

$$[E_{\mu\nu}, E^2] = 0, \tag{59}$$

$$[F_{\mu\nu}, E^2] = 4V_{\mu\nu}. \tag{60}$$

(i) The contraction

$$F^2 = \tfrac{1}{2}U_\alpha{}^\alpha = \tfrac{1}{2}F_{\alpha\beta}F^{\beta\alpha} \tag{61}$$

obeys

$$[E_{\mu\nu}, F^2] = 0, \tag{62}$$

$$[F_{\mu\nu}, F^2] = 4V_{\mu\nu}. \tag{63}$$

(j) The quantity

$$D_\alpha{}^\alpha = E^2 - F^2 = \tfrac{1}{2}(E_{\alpha\beta}E^{\beta\alpha} - F_{\alpha\beta}F^{\beta\alpha}) \tag{64}$$

therefore obeys

$$[E_{\mu\nu}, D_\alpha{}^\alpha] = [F_{\mu\nu}, D_\alpha{}^\alpha] = 0 \tag{65}$$

and is a second-degree Casimir operator of $U(p, q)$.

(k) The contraction $V_\alpha{}^\alpha$ is identically zero.

VI. INHOMOGENEOUS PSEUDO-UNITARY GROUPS

The n-dimensional inhomogeneous pseudo-unitary group $IU(p, q)$, $p + q = n$, consists of all elements of $U(p, q)$ (see preceding section) together with all translations in the pseudo-unitary space. A basis of the corresponding Lie algebra [of order $n(n + 2)$] is obtained with the elements $E_{\mu\nu} = -E_{\nu\mu}$, $F_{\mu\nu} = F_{\nu\mu}$, Q_μ, and R_μ obeying Eqs. (42)–(44), and

$$[E_{\mu\nu}, Q_\rho] = g_{\mu\rho}Q_\nu - g_{\nu\rho}Q_\mu, \tag{66}$$

$$[E_{\mu\nu}, R_\rho] = g_{\mu\rho}R_\nu - g_{\nu\rho}R_\mu, \tag{67}$$

$$[F_{\mu\nu}, Q_\rho] = -g_{\mu\rho}R_\nu - g_{\nu\rho}R_\mu, \tag{68}$$

$$[F_{\mu\nu}, R_\rho] = g_{\mu\rho}Q_\nu + g_{\nu\rho}Q_\mu, \tag{69}$$

$$[Q_\mu, Q_\nu] = [Q_\mu, R_\nu] = [R_\mu, R_\nu] = 0. \tag{70}$$

Also

$$[Q_\mu, F_\alpha{}^\alpha] = 2R_\mu, \tag{71}$$

$$[R_\mu, F_\alpha{}^\alpha] = -2Q_\mu. \tag{72}$$

The second-degree basis elements of the enveloping algebra are $E_{\mu\nu}E_{\rho\sigma}$, $E_{\mu\nu}F_{\rho\sigma}$, $F_{\mu\nu}E_{\rho\sigma}$, $F_{\mu\nu}F_{\rho\sigma}$, $E_{\mu\nu}Q_\rho$, $Q_\rho E_{\mu\nu}$, $E_{\mu\nu}R_\rho$, $R_\rho E_{\mu\nu}$, $F_{\mu\nu}Q_\rho$, $Q_\rho F_{\mu\nu}$, $F_{\mu\nu}R_\rho$, $R_\rho F_{\mu\nu}$, $Q_\mu Q_\nu$, $Q_\mu R_\nu$, $R_\mu R_\nu$. Interesting linear combinations of these, in addition to those presented in the preceding section, are:

(a) The quantities

$$Q^2 = Q_\alpha Q^\alpha, \qquad QR = Q_\alpha R^\alpha, \qquad R^2 = R_\alpha R^\alpha, \tag{73}$$

which obey

$$[E_{\mu\nu}, Q^2] = [E_{\mu\nu}, QR] = [E_{\mu\nu}, R^2] = 0, \tag{74}$$

$$[F_{\mu\nu}, Q^2] = -2(Q_\mu R_\nu + R_\mu Q_\nu), \tag{75}$$

$$[F_{\mu\nu}, QR] = 2(Q_\mu Q_\nu - R_\mu R_\nu), \tag{76}$$

$$[F_{\mu\nu}, R^2] = 2(Q_\mu R_\nu + R_\mu Q_\nu). \tag{77}$$

(b) The quantity

$$Q^2 + R^2 = Q_\alpha Q^\alpha + R_\alpha R^\alpha \tag{78}$$

obeying

$$[E_{\mu\nu}, Q^2 + R^2] = [F_{\mu\nu}, Q^2 + R^2] = 0. \tag{79}$$

It also commutes with Q_μ and R_μ [Eq. (70)] and is therefore a second-degree Casimir operator of $IU(p, q)$.

(c) The quantity $\frac{1}{2}(E_{\mu\alpha}Q^\alpha - Q^\alpha E_{\mu\alpha} + F_{\mu\alpha}R^\alpha - R^\alpha F_{\mu\alpha})$ equals Q_μ.

(d) The quantity $\frac{1}{2}(E_{\mu\alpha}R^\alpha - R^\alpha E_{\mu\alpha} - F_{\mu\alpha}Q^\alpha + Q^\alpha F_{\mu\alpha})$ equals R_μ.

(e) The quantities

$$A_\mu = \tfrac{1}{2}(E_{\mu\alpha}Q^\alpha + Q^\alpha E_{\mu\alpha} - F_{\mu\alpha}R^\alpha - R^\alpha F_{\mu\alpha}), \tag{80}$$

$$B_\mu = \tfrac{1}{2}(E_{\mu\alpha}R^\alpha + R^\alpha E_{\mu\alpha} + F_{\mu\alpha}Q^\alpha + Q^\alpha F_{\mu\alpha}) \tag{81}$$

obey

$$[E_{\mu\nu}, A_\rho] = g_{\mu\rho}A_\nu - g_{\nu\rho}A_\mu, \tag{82}$$

$$[E_{\mu\nu}, B_\rho] = g_{\mu\rho}B_\nu - g_{\nu\rho}B_\mu, \tag{83}$$

$$[F_{\mu\nu}, A_\rho] = -g_{\mu\rho}B_\nu - g_{\nu\rho}B_\mu, \tag{84}$$

$$[F_{\mu\nu}, B_\rho] = g_{\mu\rho}A_\nu + g_{\nu\rho}A_\mu \tag{85}$$

[cf. Eqs. (66)–(69)],

$$[A_\mu, F_\alpha{}^\alpha] = 2B_\mu, \tag{86}$$

$$[B_\mu, F_\alpha{}^\alpha] = -2A_\mu \tag{87}$$

[cf. Eqs. (71), (72)],

$$[Q_\mu, A_\alpha] = Q_\mu Q_\alpha - R_\mu R_\alpha - g_{\mu\alpha}(Q^2 + R^2), \tag{88}$$

$$[Q_\mu, B_\alpha] = [R_\mu, A_\alpha] = Q_\mu R_\alpha + R_\mu Q_\alpha, \tag{89}$$

$$[R_\mu, B_\alpha] = R_\mu R_\alpha - Q_\mu Q_\alpha - g_{\mu\alpha}(Q^2 + R^2), \tag{90}$$

$$[Q_\mu, D_\alpha{}^\alpha] = 2A_\mu, \tag{91}$$

$$[R_\mu, D_\alpha{}^\alpha] = 2B_\mu, \tag{92}$$

$$[A_\mu, D_\alpha{}^\alpha] = E_{\mu\alpha}A^\alpha + A^\alpha E_{\mu\alpha} - F_{\mu\alpha}B^\alpha - B^\alpha F_{\mu\alpha}, \tag{93}$$

$$[B_\mu, D_\alpha{}^\alpha] = E_{\mu\alpha}B^\alpha + B^\alpha E_{\mu\alpha} + F_{\mu\alpha}A^\alpha + A^\alpha F_{\mu\alpha}, \tag{94}$$

$$[A_\mu, A_\nu] = [B_\mu, B_\nu] = -(Q^2 + R^2)E_{\mu\nu}, \tag{95}$$

$$[A_\mu, B_\nu] = (Q^2 + R^2)F_{\mu\nu} + g_{\mu\nu}\Lambda, \tag{96}$$

where

$$\Lambda = \tfrac{1}{2}(Q^\alpha B_\alpha + B_\alpha Q^\alpha - R^\alpha A_\alpha - A_\alpha R^\alpha) = 2Q^\alpha R^\beta E_{\alpha\beta} + (Q^\alpha Q^\beta + R^\alpha R^\beta)F_{\alpha\beta}. \quad (97)$$

Taking higher-degree elements, one has:

(f) The quantities

$$A^2 = A_\alpha A^\alpha, \quad AB = A_\alpha B^\alpha, \quad B^2 = B_\alpha B^\alpha \quad (98)$$

obeying

$$[E_{\mu\nu}, A^2] = [E_{\mu\nu}, AB] = [E_{\mu\nu}, B^2] = 0, \quad (99)$$

$$[F_{\mu\nu}, A^2] = -2(A_\mu B_\nu + B_\mu A_\nu), \quad (100)$$

$$[F_{\mu\nu}, AB] = 2(A_\mu A_\nu - B_\mu B_\nu), \quad (101)$$

$$[F_{\mu\nu}, B^2] = 2(A_\mu B_\nu + B_\mu A_\nu) \quad (102)$$

[cf. Eqs. (74)–(77)].

(g) The quantity

$$A^2 + B^2 = A_\alpha A^\alpha + B_\alpha B^\alpha, \quad (103)$$

which obeys

$$[E_{\mu\nu}, A^2 + B^2] = [F_{\mu\nu}, A^2 + B^2] = 0, \quad (104)$$

$$[Q_\mu, A^2 + B^2] = -2(Q^2 + R^2)A_\mu + \Lambda R_\mu + R_\mu \Lambda, \quad (105)$$

$$[R_\mu, A^2 + B^2] = -2(Q^2 + R^2)B_\mu - \Lambda Q_\mu - Q_\mu \Lambda, \quad (106)$$

$$[A_\mu, A^2 + B^2] = (Q^2 + R^2)(F_{\mu\alpha}B^\alpha + B^\alpha F_{\mu\alpha} - E_{\mu\alpha}A^\alpha - A^\alpha E_{\mu\alpha}) + \Lambda B_\mu + B_\mu \Lambda, \quad (107)$$

$$[B_\mu, A^2 + B^2] = -(Q^2 + R^2)(F_{\mu\alpha}A^\alpha + A^\alpha F_{\mu\alpha} + E_{\mu\alpha}B^\alpha + B^\alpha E_{\mu\alpha}) - \Lambda A_\mu - A_\mu \Lambda. \quad (108)$$

(h) The quantity Λ defined in Eq. (97) obeys

$$[E_{\mu\nu}, \Lambda] = [F_{\mu\nu}, \Lambda] = 0, \quad (109)$$

$$[Q_\mu, \Lambda] = 2(Q^2 + R^2)R_\mu, \quad (110)$$

$$[R_\mu, \Lambda] = -2(Q^2 + R^2)Q_\mu, \quad (111)$$

$$[A_\mu, \Lambda] = 2(Q^2 + R^2)B_\mu, \quad (112)$$

$$[B_\mu, \Lambda] = -2(Q^2 + R^2)A^\mu. \quad (113)$$

(i) The quantity

$$\Omega = (Q^2 + R^2)F_\alpha{}^\alpha - \Lambda \quad (114)$$

obeys

$$[E_{\mu\nu}, \Omega] = [F_{\mu\nu}, \Omega] = [Q_\mu, \Omega] = [R_\mu, \Omega] = 0, \quad (115)$$

according to Eqs. (45), (70)–(72), (79), (107)–(109), and is therefore a third-degree Casimir operator of $IU(p, q)$. It also obeys

$$[A_\mu, \Omega] = [B_\mu, \Omega] = 0 \quad (116)$$

by Eqs. (86), (87), (112), and (113) or by Eq. (115).

(j) The quantity

$$\Delta = (Q^2 + R^2)^2 D_\alpha{}^\alpha + (Q^2 + R^2)(A^2 + B^2) - \tfrac{1}{2}\Lambda^2 \quad (117)$$

obeys

$$[E_{\mu\nu}, \Delta] = [F_{\mu\nu}, \Delta] = [Q_\mu, \Delta] = [R_\mu, \Delta] = 0 \quad (118)$$

by Eqs. (65), (91), (92), (104)–(106), (109)–(111), showing that it is a sixth-degree Casimir operator of $IU(p, q)$. In addition, by Eqs. (93), (94), (107), (108), (112), (113) or just by Eq. (118) it obeys

$$[A_\mu, \Delta] = [B_\mu, \Delta] = 0. \quad (119)$$

Define now

$$G_{\mu\nu} = E_{\mu\nu}, \qquad H_{\mu\nu} = F_{\mu\nu}, \quad (120)$$

$$G_{\mu\, n+1} = -G_{n+1\, \mu} = A_\mu/[-\varepsilon(Q^2 + R^2)]^{\frac{1}{2}}, \quad (121)$$

$$H_{\mu\, n+1} = H_{n+1\, \mu} = B_\mu/[-\varepsilon(Q^2 + R^2)]^{\frac{1}{2}}, \quad (122)$$

$$H_{n+1\, n+1} = -\Lambda/\varepsilon(Q^2 + R^2), \quad (123)$$

$$g_{n+1\, n+1} = \varepsilon = \pm 1, \quad (124)$$

where a remark similar to the one following Eq. (37) holds. One then finds

$$[G_{ab}, G_{cd}] = g_{ac}G_{bd} - g_{bc}G_{ad} + g_{ad}G_{cb} - g_{bd}G_{ca}, \quad (125)$$

$$[H_{ab}, H_{cd}] = g_{ac}G_{bd} + g_{bc}G_{ad} - g_{ad}G_{cb} - g_{bd}G_{ca}, \quad (126)$$

$$[G_{ab}, H_{cd}] = g_{ac}H_{bd} - g_{bc}H_{ad} + g_{ad}H_{cb} - g_{bd}H_{ca}, \quad (127)$$

where $a, b, c, d = 1, \cdots, n + 1$. Thus, the G_{ab} and H_{ab} form a basis of the Lie algebra of $U(p, q + 1)$ or $U(p + 1, q)$. The quantity $H_c{}^c$ therefore obeys

$$[G_{ab}, H_c{}^c] = [H_{ab}, H_c{}^c] = 0, \quad (128)$$

since it is a first-degree Casimir operator of $U(p, q + 1)$ or $U(p + 1, q)$. On the other hand,

$$H_c{}^c = F_\alpha{}^\alpha - \Lambda/(Q^2 + R^2) = \Omega/(Q^2 + R^2), \quad (129)$$

where a remark similar to the one following Eq. (37) holds, showing the consistency of Eq. (128) with Eqs. (115) and (116). Eq. (129) also exhibits the relation between the first-degree Casimir operator $H_c{}^c$ of $U(p, q + 1)$ or $U(p + 1, q)$ and the third-degree Casimir operator Ω of $IU(p, q)$. In addition, the quantity

$$G^2 - H^2 = \tfrac{1}{2}(G_{ab}G^{ba} - H_{ab}H^{ba}) \quad (130)$$

obeys

$$[G_{ab}, G^2 - H^2] = [H_{ab}, G^2 - H^2] = 0, \quad (131)$$

since it is a second-degree Casimir operator of $U(p, q + 1)$ or $U(p + 1, q)$. On the other hand, one finds

$$G^2 - H^2 = D_\alpha{}^\alpha + (A^2 + B^2)/(Q^2 + R^2) - \Lambda^2/2(Q^2 + R^2)^2 = \Delta/(Q^2 + R^2)^2, \quad (132)$$

where a remark similar to the one following Eq. (37) holds, demonstrating the consistency of Eq. (131) with Eqs. (118) and (119). Equation (132) shows the relation between the second-degree Casimir operator $G^2 - H^2$ of $U(p, q + 1)$ or $U(p + 1, q)$ and the sixth-degree Casimir operator Δ of $IU(p, q)$.

VI. OTHER GROUPS

Other groups, such as the real and complex general linear groups, [$GL(n, R)$ and $GL(n, C)$] and their inhomogeneous counterparts [$IGL(n, R)$ and $IGL(n, C)$], can be treated similarly to the above procedures, but will not be treated here in detail. We only note that $GL(n, R)$ is especially analogous to $U(p, q)$, $p + q = n$, since both are real forms of $GL(n, C)$.

Analytic Properties of a Class of Potentials and the Corresponding Jost Functions

PIERRE C. SABATIER
Laboratoire de Physique Théorique et Hautes Energies, Orsay, France
(Received 7 April 1966)

Analytical properties of the potential obtained by Newton's method in the inverse scattering problem at fixed energy are thoroughly investigated. It is found that $r\,V(r)$ is analytic in the neighborhood of $r = 0$ and can be continued in the r complex plane as a meromorphic function, the poles of which are of order at least equal to 2. An explicit formula is given to yield the Jost functions for any complex value of the angular momentum ν. They appear to be meromorphic functions with poles located at the negative integers on the real axis. The zeros of $f_2(\nu)$, for large $|\nu|$ and for Re ν and Im $\nu > 0$, are located on a curve which is the boundary of a domain previously shown to contain no Regge pole. The interpolated scattering amplitude is unitary. It behaves for $\nu \to -i\,\infty$ as $e^{\pi|\nu|}$. An important result of this paper is that the class of potentials obtained by Newton's method is much more restricted than one might think. This led the author to look for a more general approach to the inverse scattering problem, which is the subject of a forthcoming publication. In order to illustrate the method of this paper, a detailed study of an example previously introduced by R. G. Newton is given. In the last Appendix, a very remarkable property of the potentials involved in this example is given: the scattering amplitude corresponding to these potentials can be given an exact closed form. Since the corresponding potentials are strongly energy-dependent, it is very likely that this is only a mathematical curiosity.

IN a previous paper,[1] we studied the asymptotic behavior of the potentials derived from a given set of phase shifts in the framework of Newton's method.[2] We showed, in particular, that, if the phase shifts fulfill a very weak condition,[3] namely $|\delta_l| < Cl^{-3-\epsilon}$, as $l \to \infty$, there is one potential, and only one, which goes to zero faster than $r^{-2+\epsilon}$ as $r \to \infty$. All the potentials equivalent to this one go to zero like $r^{-\frac{1}{2}}$ as $r \to \infty$.

The purpose of this paper is to study the class of potentials which can be found by Newton's method. Therefore, we investigate successively the analytic continuation of these potentials $V(r)$ in the complex r plane, and the corresponding Jost functions in the complex l plane. The results for the potentials are the following:

(1) The function $rV(r)$ can be continued in the complex r plane as a meromorphic function, with poles of order at least equal to 2.
(2) There is no pole in a nonvanishing circle centered at the origin.
(3) Potentials with poles on the real positive axis are only special cases.

In a forthcoming paper, we see another particularity of the potentials obtained by Newton's method: they cannot be even analytic functions.

The Jost functions have the following properties:

(1) They can be continued as meromorphic functions in the complex ν plane ($\nu = l$), with simple poles fixed at the negative integers.
(2) They are bounded by $C \exp |\frac{1}{2}\pi\nu|$ on any ray.
(3) In the upper half-plane $f_2(\nu)$ has infinite zeros, which, in general, are distributed for large $|\nu|$ on the curve:

$$\text{Im}\,\nu = \pi^{-1}q\,\text{Log}\,|\text{Re}\,\nu|,$$

where q is a positive integer. This curve is the boundary of a domain (described by Martin[4]), where Regge poles are excluded.

The analytic continuation of the scattering amplitude, while satisfies the unitarity condition, also exhibits Regge poles in the upper half-plane. Furthermore, it behaves on the ray

$$\text{Arg}\,\nu = \tfrac{3}{2}i\pi \quad \text{as} \quad \exp|\pi\nu|.$$

Therefore the conditions of Carlson's theorem are not fulfilled. This is not surprising, since there is an infinity of equivalent potentials.

Those properties led us to look for a method of approach to the inverse scattering problem at fixed energy more general than that of Newton. We have obtained[5] such a method by taking nonvanishing

[1] P. C. Sabatier, J. Math. Phys. 7, 1515 (1966). This paper is hereafter referred to as I, followed, as the case may be, by the number of the formula.
[2] R. G. Newton, J. Math. Phys. 3, 75 (1962).
[3] Throughout this paper, by ϵ we mean a positive number, which can be made arbitrarily small, but not equal to zero. We use C as a general constant. Both are not meant to have the same value every time they are used.
[4] A. Martin, *Progress in Elementary Particle and Cosmic Ray Physics* (North-Holland Publishing Company, Amsterdam, 1965), Chap. I, p. 54.
[5] This method has been given in the author's Doctoral dissertation for Faculté des Sciences, Orsay, and is to be published in J. Math. Phys.

c_l for arbitrary real values of l (not necessarily integers). This, however, is the subject of a forthcoming paper.

1. ANALYTICAL CONTINUATION OF THE POTENTIALS DERIVED FROM THE PHASE SHIFTS IN NEWTON'S METHOD

In Newton's method, the potential is given by the formula

$$V(r) = -2r^{-1}\{d[r^{-1}K(r, r)]/dr\}, \tag{1.1}$$

where $K(r, r)$ is a particular value of $K(r, r')$, which in turn is defined by

$$K(r, r') = f(r, r') - \int_0^r K(r, \rho)f(\rho, r')\rho^{-2}\, d\rho, \tag{1.2}$$

where the function $f(r, r')$ can be obtained from the knowledge of all the phase shifts through its coefficients c_l:

$$f(r, r') = \sum_{l=0}^{\infty} c_l u_l(r)u_l(r'), \tag{1.3}$$

where

$$u_\nu(r) = (\tfrac{1}{2}\pi r)^{\frac{1}{2}}J_{\nu+\frac{1}{2}}(r). \tag{1.4}$$

The derivation of the coefficients c_l from the values of $\tan \delta_l$ is studied thoroughly in I. We showed in particular (Sec. I, 3.2) that, if the phase shifts go to zero as l tends to infinity faster than $l^{-4/3-\epsilon}$, the coefficients c_l are bounded as $l \to \infty$. We see in Sec. 2 that a slightly weaker condition is necessary to legitimate the method used to relate the phase shifts to the c_l. However, the analyses of Sec. 1 are submitted only to a much weaker assumption:

Assumption I:

$$\exists p \Rightarrow |l^{-p}c_l| < C.$$

We now introduce the two functions:

$$\varphi_r(r') \equiv \varphi(r, r') = (rr')^{-1}f(r, r'), \tag{1.5}$$

$$K_r(r') = (rr')^{-1}K(r, r'), \tag{1.6}$$

which enable us to write (1.2) more conveniently in the form

$$K_r(r') = \varphi_r(r') - \int_0^r dr_1\, K_r(r_1)\varphi(r_1, r'). \tag{1.7}$$

1.1. Analytic Continuation of $rV(r)$ in the Neighborhood of $r = 0$

We use the notation z (resp. z') in place of r (resp. r') when we consider these variables in their respective complex plane. We also speak of *the complex plane*, where z and z' are considered only as points in a complex plane. For a given (real) r, Eq. (1.2) is a Fredholm equation. Its kernel, $\varphi(r, r')$, is continued straightforwardly through (1.3) into $\varphi(z, z')$. $\varphi(z, z')$ is an entire function of z (resp. z') for any finite value of z' (resp. z), and therefore is an entire function of z and z'. This results readily from the bounds of the functions $|u_\nu(z)|$ $(< C\ |\frac{1}{2}z|^{\nu+1}/\nu!)$ given for instance in I (Appendix 1), and from the Assumption I on the coefficients c_l. For any couple of values of z_1 and z_2 lying inside a circle $(0, r_0)$ of the complex plane, $|\varphi(z_1, z_2)|$ can be bounded by a finite positive function $C(r_0)$

$$|\varphi(z_1, z_2)| < C(r_0), \qquad \forall\ z_1, z_2 \leq r_0. \tag{1.8}$$

Since $C(r_0)$ goes to a finite limit as $r_0 \to 0$, it is possible to find r_M for which the following formula holds:

$$C(r)r_M < 1, \qquad \forall\ r \leq r_M. \tag{1.9}$$

Now, the Neumann series for (1.7) is

$$K_r(r') = \varphi(r, r') - \int_0^r dr_1\, \varphi(r, r_1)\varphi(r_1, r') + \cdots (-1)^n \times \int_0^r \cdots \int_0^r dr_1 \cdots dr_n\, \varphi(r, r_1) \cdots \varphi(r_n, r'). \tag{1.10}$$

A lower bound for the radius of convergence of the series (1.10) can be obtained by replacing each term by its modulus, and φ by a majorant. The general term is bounded by

$$r^n[C(r)]^n C(r_0), \tag{1.11}$$

where r_0 is a number larger than r and r'; hence the series (1.10) certainly converges for $rC(r) < 1$, or

$$r \leq r_M - \epsilon. \tag{1.12}$$

Suppose now that r and r' are complex and therefore denoted by z, z'. Any term of the series (1.10) is an analytic function of z and z', and its value is independent of the contour $(0, z)$. Using a well-known inequality for complex integrals,[6] and choosing rays issued from the origin as contours of integration, it is easy to see that the bound (1.11) holds, with r replaced by $|z|$, if z and z' lie inside the circle $(0, r_0)$. Hence the series (1.10) is uniformly convergent, and $K_z(z')$ is an analytic function of z (and an entire function of z'), if z lies inside the circle $(0, r_M - \epsilon)$. $zV(z)$ is therefore analytic inside this circle.

1.2. Analytic Continuation of $zV(z)$ for Any Finite z

For any finite value of r [except when -1 happens to be an eigenvalue of (1.7)], it is possible to write

[6] E. C. Titchmarsh, *Theory of Functions* (Oxford University Press, London, 1932).

the solution of this equation in terms of Fredholm determinants[7]

$$K_r(r') = \varphi(r, r') + [\mathfrak{D}(r)]^{-1} \int_0^r \varphi(r, \rho)\, \bar{\mathfrak{D}}_r(\rho, r')\, d\rho, \tag{1.13}$$

where

$$\mathfrak{D}(r) = 1 + \sum_{m=1}^{\infty} (m!)^{-1} \times \int_0^r \cdots \int_0^r K\begin{pmatrix} r_1, r_2 \cdots r_m \\ r_1, r_2 \cdots r_m \end{pmatrix} dr_1 \cdots dr_m, \tag{1.14}$$

$$\bar{\mathfrak{D}}_r(\rho, r') = -\sum_{m=0}^{\infty} (m!)^{-1} \times \int_0^r \cdots \int_0^r K\begin{pmatrix} \rho, r_1 \cdots r_m \\ r', r_1 \cdots r_m \end{pmatrix} dr_1 \cdots dr_m, \tag{1.15}$$

and we recall that the symbol K stands for the following determinant:

$$K\begin{pmatrix} x_1 \cdots x_m \\ y_1 \cdots y_m \end{pmatrix} = \begin{bmatrix} \varphi(x_1, y_1), \varphi(x_1, y_2) \cdots \varphi(x_1, y_m) \\ \varphi(x_2, y_1) \cdots \\ \vdots \\ \varphi(x_m, y_1), \varphi(x_m, y_2) \cdots \varphi(x_m, y_m) \end{bmatrix}. \tag{1.16}$$

From Hadamard's theorem and (1.8), it follows that

$$\int_0^r \cdots \int_0^r \left| K\begin{pmatrix} r_1 \cdots r_m \\ r_1 \cdots r_m \end{pmatrix} \right| dr_1 \cdots dr_m \leq [C(r)]^m m^{m/2} r^m, \tag{1.17}$$

$$\int_0^r \cdots \int_0^r \left| K\begin{pmatrix} \rho, r_1 \cdots r_m \\ r', r_1 \cdots r_m \end{pmatrix} \right| dr_1 \cdots dr_m \leq [C(r)]^m C(r_0)(m+1)^{\frac{1}{2}(m+1)} r^m. \tag{1.18}$$

Consequently (and this is a well-known fact), the series (1.14) and (1.15) converge for any finite value of r. Suppose now that r and r' take complex values z and z'. Any term in the series (1.14) or (1.15) is an analytic function of z and z'. Choosing rays issued from the origin as contours of integration, and since Hadamard's theorem is valid also for determinants with complex elements,[8] the bounds (1.17) and (1.18) hold if only r is replaced by $|z|$, when z and z' lie inside the circle $(0, r)$.

The series (1.14) in the complex form is therefore uniformly convergent for any finite value of z, so that $\mathfrak{D}(z)$ is an entire function of z. Let us now define $\mathfrak{D}(z, z')$ to be

$$\mathfrak{D}(z, z') = \int_0^z \varphi(z, \rho)\, \bar{\mathfrak{D}}_z(\rho, z'). \tag{1.19}$$

$\mathfrak{D}(z, z')$ is an entire function of z and z'. $K_z(z')$ may therefore be continued in the z complex plane, and, for any finite value of z', as a meromorphic function of z. Its poles, which are necessarily zeros of $\mathfrak{D}(z)$, are independent of z'. Conversely, a classical argument[9] which can easily be extended here, shows that all the zeros of $\mathfrak{D}(z)$ are poles of $K_z(z')$.

1.3. Poles of $zV(z)$

Equations (1.1) and (1.6) enable us to get the analytic continuation of $V(r)$

$$V(z) = -2z^{-1}\{d[zK_z(z)]/dz\}. \tag{1.20}$$

Since the only singularities of $K_z(z)$ for finite z are poles, the only singularities of $zV(z)$ are poles of order at least equal to 2. Since the c_μ are real, if z_i is a pole, z_i^* is also a pole. Now let z_i be a pole of $K_z(z')$ $(z' \neq z)$. For $z = z_i$, the homogeneous integral equation derived from (1.7) has at least a nontrivial solution:

$$\chi_{z_i}(z') = -\int_0^{z_i} dz_1\, \chi_{z_i}(z_1)\varphi(z_1, z'). \tag{1.21}$$

This solution is an entire function of z'. The existence of such a nonvanishing solution could lead us to question the validity of Newton's method. Indeed, in this method the following step makes use of the vanishing of such a solution. Let us retrace the argument. $D_0(r)$ is the differential operator $r^2(\delta^2/\delta r^2 + 1)$.

From the partial differential equation verified by $f(r, r')$:

$$\begin{cases} [D_0(r) - D_0(r')]f(r, r') = 0, \\ f(0, r') = f(r, 0) = 0, \end{cases} \tag{1.22}$$

and from the integral equation (1.2), it is shown in the above method that the function,

$$\xi(r, r') = \{D(r) - D_0(r')\}K(r, r'), \tag{1.23}$$

where

$$D(r) \equiv D_0(r) - r^2 V(r), \tag{1.24}$$

is a solution of the homogeneous form of (1.2), which is equivalent to (1.21). The proof of this statement makes use of (1.2) and (1.22), together with straightforward but tedious differentiations and integrations by parts. The vanishing of $\xi(r, r')$ shows then that $K(r, r')$ is a solution of a partial differential equation, from which all the equations of the method follow. We see therefore that the

[7] F. G. Tricomi, *Integral Equations* (Interscience Publishers, Inc., New York, 1957).

[8] F. Riesz and B. Sz, Nagy Leçons d'analyse fonctionnelle (Academie des Sciences de Hongrie, 1955).

[9] Reference 8, p. 174.

method is not valid at the points $z = z_i$. However, this is not a difficulty, since these are isolated points in the z complex plane. For $z = z_i$, $\xi(z, z')$ is a nonvanishing function, so that the solution (1.21) is not equal to zero in L_2. But in the whole z complex plane (except at these isolated points), $\xi(z, z')$ is zero, so that there is one and only one *analytic solution* of (1.2), and the existence of $\xi(z, z')$ at z_i is only another aspect of the nonanalyticity of $K(z, z')$ at these points. Therefore, all the equations of Newton's method hold provided that there is no pole on the real axis. Besides, it is clear that if such a pole, which would be of order 2 or more, existed, the scattering problem might have, strictly speaking, no physical meaning. It is remarkable that the formula (1.3) and those relating the c_l to the δ_l enable us to define the scattering problem even in this case, if we choose for all the integrals a contour which does not cross any pole, and provided that there is no pole at $+\infty$. This may give a way of studying some special classes of singular potentials.

The absence of a pole at $+\infty$ is certainly true in general, i.e., $\mathfrak{D}(r)$ can be zero for $r = +\infty$ only for particular sets of c_l. On the other hand, there are certainly two cases for which there is no pole at $+\infty$.

(1) When all the c_l are small enough, the Neumann series converges for the solutions of (1.2) [or (1.7)] so that there is no pole on the real axis.

(2) When all the l for which $c_l \neq 0$ are of the same parity, it is possible to write down the solution $g(r, r')$ of the equation obtained from (1.2) by replacing the upper bound r of the integral by $+\infty$. This is due to the orthogonality of the sets u_{2l} or u_{2l+1}. One obtains

$$g(r, r') = \sum_0^\infty g_{2l} u_{2l}(r) u_{2l}(r') \tag{1.25}$$

or

$$g(r, r') = \sum_0^\infty g_{2l+1} u_{2l+1}(r) u_{2l+1}(r'), \tag{1.26}$$

where

$$g_n = \left[1 + \frac{\pi}{2(2n+1)} c_n\right]^{-1} c_n. \tag{1.27}$$

We showed in a previous paper[1] that the solution $K(r, r')$ of (1.2) is also a solution of

$$K(r, r') = g(r, r') + \int_r^\infty \rho^{-2}\, d\rho\, K(r, \rho) g(\rho, r'). \tag{1.28}$$

If the coefficients g_n are bounded in such a way that $g(r, r')$ belongs to one of the three classes studied in I, $|g(r, r')|$ is bounded by

$$C(rr')^{\frac{1}{2}} \times [1 + |r - r'|]^{-\frac{1}{2}}.$$

For large values of r, the Neumann series converges and defines $K(r, r')$ from $g(r, r')$. $K(r, r')$ is equivalent to $g(r, r')$ for large values of r and r', so that, in the same conditions, $|K_r(r')|$ is bounded. It follows that the solution of (1.2) cannot have a pole at $+\infty$ in these conditions. As we see in the following example, there can be a pole at $+\infty$ if these conditions are not fulfilled, for example, if one of the g_n is not finite.

1.4. A Simple Example

Newton has considered a very simple example where all the phase shifts are easy to calculate exactly. We investigate the properties of the potentials involved in this case. It is assumed that only one c_l is nonvanishing. The following formulas are straightforwardly derived

$$f(z, z') = c_{l_0} u_{l_0}(z) u_{l_0}(z'), \tag{1.29}$$

$$\phi_{l_0}(z) = [1 + c_{l_0} L^{l_0}_{l_0}(z)]^{-1} u_{l_0}(z), \tag{1.30}$$

$$K(z, z') = c_{l_0} \phi_{l_0}(z) u_{l_0}(z'), \tag{1.31}$$

$$\phi_l(z) = u_l(z) - \frac{c_{l_0} u_{l_0}(z) L^l_{l_0}(z)}{1 + c_{l_0} L^{l_0}_{l_0}(z)}, \tag{1.32}$$

where

$$L^{l'}_l(z) = \int_0^z u_l(\rho) u_{l'}(\rho) \rho^{-2}\, d\rho. \tag{1.33}$$

The potential $V(z)$ is equal to

$$V(z) = -2z^{-1} c_{l_0} \frac{d}{dz}\left\{z^{-1} \frac{[u_{l_0}(z)]^2}{1 + c_{l_0} L^{l_0}_{l_0}(z)}\right\}. \tag{1.34}$$

The phase shifts are given exactly by the following formulas:

$$\begin{cases} \tan \delta_l = 0 & \text{for } |l - l_0| \text{ even} \\ \tan \delta_l = c_{l_0}(l_0 - l)^{-1}(l + l_0 + 1)^{-1} & \\ \qquad \times \{1 + \frac{1}{2}\pi[c_{l_0}/(2l_0 + 1)]\}^{-1}, & \\ \qquad \text{for } |l - l_0| \text{ odd.} & \end{cases} \tag{1.35}$$

Let us now study the analytic properties of the potential (1.34). It is analytic for $z = 0$ if $l_0 \geq 1$. For l_0 equal to zero, $zV(z)$ is analytic in a nonvanishing circle, with its center at the origin. The function $L^{l_0}_{l_0}(r)$ is odd, increasing monotonously from $-\frac{1}{2}\pi/(2l_0 + 1)$ to $+\frac{1}{2}\pi/(2l_0 + 1)$ when r increases from $-\infty$ to $+\infty$.

Therefore, the potential (1.34) has no pole on the real axis only if

$$|c_{l_0}| < 2\pi^{-1}(2l_0 + 1). \tag{1.36}$$

If c_{l_0} is not bounded as in (1.36), and according to its sign, there is one pole either on the negative or on the positive real axis.

Poles in the Right Upper Half-Plane

We prove now that the potential (1.34) exhibits an infinity of poles in the first quarter of the z complex plane. For this, let us write the denominator of (1.34) in the following form:

$$D = 1 + \tfrac{1}{2}\pi \frac{c_{l_0}}{2l_0 + 1} - c_{l_0} \int_z^\infty d\rho\, \rho^{-2}[u_{l_0}(\rho)]^2. \quad (1.37)$$

Well-known formulas[10] for spherical Bessel functions lead to

$$u_{l_0}(z) = \sin(z - l_0\tfrac{1}{2}\pi)[1 + \epsilon_1(z)] + \epsilon_2(z)\cos(z - l_0\tfrac{1}{2}\pi), \quad (1.38)$$

where $z\epsilon_1(z)$ and $z\epsilon_2(z)$ are polynomials of z^{-1}.

Suppose now that both Re z and Im z go to $+\infty$. The formula (1.38) is equivalent to

$$u_{l_0}(z) = \exp[-i(z - l_0\tfrac{1}{2}\pi)][1 + \epsilon_0(z)],$$

$\epsilon_0(z)$ is analytic for $z \neq 0$, whereas $z\epsilon_0(z)$, $z^2\epsilon_0'(z)$, $z^3\epsilon_0''(z)$ remain bounded as Re z and Im $z \to \infty$. Integrating twice by parts and evaluating the rest on a parallel to the real axis, we can write D in the following form:

$$D = 1 + \frac{\frac{1}{2}\pi c_{l_0}}{2l_0 + 1} - \tfrac{1}{4}c_{l_0}\frac{e^{-2iz}(-1)^{l_0}}{z^2}[1 + \epsilon(z)], \quad (1.39)$$

where $z\epsilon(z)$ is analytic, and remains bounded as both Re z and Im z tend to $+\infty$. Let us now call z_0 a zero of

$$D_0 = 1 + \tfrac{1}{2}\pi \frac{c_{l_0}}{2l_0 + 1} - \tfrac{1}{4}c_{l_0}(-1)^{l_0}\frac{e^{-2iz}}{z^2} = \alpha(e^{\beta + i\gamma} - z^{-2}e^{-2iz}), \quad (1.40)$$

where α, β, γ are convenient constants. It is easy to see that D_0 exhibits an infinity of zeros in the first quarter, located on the curve:

$$2\,\text{Im}\, z = \beta + \text{Log}\,[(\text{Re}\, z)^2 + (\text{Im}\, z)^2]$$

at the points

$$\text{Re}\, z = -\tfrac{1}{2}\gamma - \text{Arc tan}\frac{\text{Im}\, z}{\text{Re}\, z} + k\pi. \quad (1.41)$$

This curve and these points go asymptotically to the following ones:

$$\begin{cases} \text{Im}\, z \sim \text{Log Re}\, z, \\ \text{Re}\, z \sim -\tfrac{1}{2}\gamma + k\pi. \end{cases} \quad (1.42)$$

We now show that if $\bar{\epsilon}(z)$ and $\bar{\epsilon}'(z)$ are bounded by $|\epsilon|$ in a circle $|z - z_0| < C\epsilon$, $D - D_0(z_0)$ has a zero in this circle. It is sufficient to show that the property holds for

$$\{z_0^{-2}e^{-2iz_0} - z^{-2}e^{-2iz}[1 + \bar{\epsilon}(z)]\},$$

or for

$$f(z) = 1 - e^{-2i(z - z_0)}[1 + \eta(z - z_0)],$$

where $|\eta|$ and $|d\eta/dz|$ are bounded by ϵ in the circle $|z - z_0| < 2\epsilon$, and analytic in this circle. To show this point, we have only to evaluate the following integral:

$$\frac{1}{2\pi i}\int_{|z - z_0| = 2\epsilon} \frac{f'(z)}{f(z)}\,dz,$$

which is equal to the number of zeros of $f(z)$ in the circle ($|z - z_0| \neq 2\epsilon$). Since we have on the boundary of this circle, up to order ϵ^2,

$$\begin{aligned} 3\epsilon &< |f(z)| < 5\epsilon, \\ 1 - 2\epsilon &< \tfrac{1}{2}|f'(z)| < 1 + 2\epsilon, \end{aligned} \qquad dz = 2\epsilon\, d\theta,$$

it is clear that there is one (and only one) zero in the given circle. Now, it is easy to see that the properties of $\epsilon(z)$ enable us to obtain majorizations of this kind, so that the zeros of $D(z)$ go asymptotically to the zeros of $D_0(z)$. Therefore, $D(z)$ exhibits an infinity of zeros, which go asymptotically to the points (1.42), and which are poles of the potential.

Asymptotic Behavior of the Potential

The formula (1.34) shows readily that $V(r)$ decreases as r^{-2} when $r \to \infty$, provided only c_{l_0} be different from $-2(2l_0 + 1)/\pi$.

The Case of a Pole at $+\infty$

If c_{l_0} is equal to $-2(2l_0 + 1)/\pi$, $V(r)$ behaves like $Cr^{-1}\sin 2r$ for large r, and this is not sufficient to give sense to the scattering problem. It is not surprising to see that, in this case, $\tan\delta_l$ and A^l are infinite for odd $(l - l_0)$.

Other Examples

Similar examples can be found if we take a finite number of nonvanishing c_l. These examples lead to Pincherle–Goursat kernels in the fundamental Fredholm equation, so that it is easy to obtain an exact solution of it.

[10] *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw–Hill Book Company, Inc., New York, 1953), Eq. 7.11 (1). Hereafter referred to as HTF, followed by the number of the formula.

2. DEFINITION AND ANALYTIC PROPERTIES OF JOST FUNCTIONS

In this section, we first show that the regular solutions of the Schrödinger equation for any value ν of l in the complex half-plane (Re $\nu > -1$) can be obtained from $K(r, r')$ through a formula which readily interpolates the formula giving the $u_l(r)$. We then use the asymptotic form of this relation to get the Jost functions. These can be continued in the whole ν plane as meromorphic functions with poles at the negative integers, and the expansions we obtain are their Mittag–Leffler expansions.[11] The result is used to show the unitarity of the interpolation of the scattering amplitude. We give then formulas which enable one to study the properties of Jost functions either from the asymptotic properties of the coefficients c_l or from the asymptotic behavior of the phase shifts.

These formulas enable us to get at the location of Regge poles for large values of $|\nu|$. We show that, for large $|\nu|$, the zeros of $f_2(\nu)$ in the right half-plane are located on a curve of zero slope in the first quarter. This curve is on the boundary of a domain described by Martin,[4] which cannot contain any Regge pole. On the other hand, it is known, from a work of Bessis,[12] that, if the potential is holomorphic and bounded by a decreasing exponential in the right r half-plane, there cannot be any pole in the ν plane with Re $\nu \to +\infty$. The existence of poles in the present case shows that the potentials built through Newton's method cannot fulfill these interesting conditions. Besides, the example we give exhibits an infinity of poles in the right r half-plane.

2.1. Definition of the Jost Functions

Let us now recall that $u_\nu(r)$ is an eigenfunction of $D_0(r)$

$$D_0(r)u_\nu(r) = \nu(\nu + 1)u_\nu(r). \tag{2.1}$$

From $K(r, \rho)$, we define, for any value of ν such that Re $\nu > -1$, the function

$$\phi_\nu(r) = u_\nu(r) - \int_0^r K(r, \rho)u_\nu(\rho)\rho^{-2}\, d\rho. \tag{2.2}$$

Applying the differential operator $D(z)$ to (2.2), performing two integrations by parts, and using (2.1) and (1.23), show that, for $z \neq z_i$, $\phi_\nu(z)$ is an eigenfunction of $D(z)$

$$D(z)\phi_\nu(z) = \nu(\nu + 1)\phi_\nu(z). \tag{2.3}$$

[11] C. Caratheodory, *Theory of Functions of a Complex Variable* (Chelsea Publishing Company, New York, 1958).
[12] D. Bessis, Nuovo Cimento **23**, 797 (1964).

Furthermore, it follows from (2.2), (1.2), and the well-known properties of $u_\nu(r)$ that $\phi_\nu(r)$ is the regular solution of (2.3) if Re $\nu > -1$, since

$$\begin{aligned} u_\nu(r) &\sim r^{\nu+1}, \\ \phi_\nu(r) &\sim r^{\nu+1} + O(r^{\nu+2}). \end{aligned} \tag{2.4}$$

Let us now recall the expansion of $K(r, r')$ (in Refs. 1, 2)

$$K(r, r') = \sum_{l=0}^{\infty} c_l\phi_l(r)u_l(r'). \tag{2.5}$$

Substituting this result in (2.2), we obtain

$$\phi_\nu(r) = u_\nu(r) - \sum_{l'} L_\nu^{l'}(r)c_{l'}\phi_{l'}(r), \tag{2.6}$$

where

$$L_\nu^{l'}(r) = \int_0^r d\rho\ \rho^{-2}u_\nu(\rho)u_{l'}(\rho). \tag{2.7}$$

In order to obtain the Jost functions, we let $r \to +\infty$ in (2.6). Then[13]

$$\begin{aligned} u_\nu(r) &\sim (2i)^{-1}[e^{i(r-\nu\frac{1}{2}\pi)} - e^{-i(r-\nu\frac{1}{2}\pi)}], \\ \phi_\nu(r) &\sim (2i)^{-1}[f_1(\nu)e^{ir} - f_2(\nu)e^{-ir}]. \end{aligned} \tag{2.8}$$

Formula (2.6) is valid for any finite value of r, modulo the Assumption I of Sec. 1. To replace the functions by their asymptotic expansion in (2.6), it is *necessary* that the series (2.6) converge for any real positive value of r, including $+\infty$.

Now, the well-known results yield

$$L_\nu^{l'}(r) = \frac{u_\nu(r)\dfrac{d}{dr}u_{l'}(r) - u_{l'}(r)\dfrac{d}{dr}u_\nu(r)}{l'(l'+1) - \nu(\nu+1)}. \tag{2.9}$$

We showed in I that $|u_l(r)|$ and $|u_l'(r)|$ are bounded, for any real positive value of r, by $Cl^{\frac{1}{2}}$. If there is no pole on the real axis, so that $K_\nu(r')$ is bounded for any real positive value of r and r', it follows from (2.2) that $|\phi_\nu(r)|$ is bounded by $C\nu^{\frac{1}{2}}$ for any real positive value of r, and, with some algebra, it is easy to get the same bound for $|\phi_\nu'(r)|$. Consequently, the series (2.6) and the differentiated series are convergent for any real positive value of r if

$$|c_l| < l^{\frac{1}{2}-\epsilon}. \tag{2.10}$$

Modulo this important condition, we can replace the functions in (2.6) by their asymptotic behavior (2.8), and obtain

[13] E. J. Squires, *Complex Angular Momenta and Particle Physics* (W. A. Benjamin, Inc., New York, 1963).

$$\begin{cases} f_1(\nu) = e^{-\frac{1}{2}i\pi\nu} - \sum_{l'} \dfrac{\sin(\nu - l')\frac{1}{2}\pi}{\nu(\nu+1) - l'(l'+1)} \\ \qquad \times a_{l'}(1 + i\tan\delta_{l'})e^{-\frac{1}{2}il'\pi}, \qquad (2.11a) \\ f_2(\nu) = e^{\frac{1}{2}i\pi\nu} - \sum_{l'} \dfrac{\sin(\nu - l')\frac{1}{2}\pi}{\nu(\nu+1) - l'(l'+1)} \\ \qquad \times a_{l'}(1 - i\tan\delta_{l'})e^{+\frac{1}{2}il'\pi}, \qquad (2.11b) \end{cases}$$

where we put, as in I,

$$a_l = c_l A_l \cos\delta_l. \qquad (2.12)$$

It follows from the bounds of the functions ϕ that A_l is bounded by $Cl^{\frac{1}{2}}$. Furthermore, it follows from the analysis of I [I, (3.2)] that A_l approaches 1 as $l \to \infty$, if, for instance, the phase shifts go to zero faster than $l^{-3-\epsilon}$. Since the a_l are bounded in the same conditions, the condition (2.10) is obviously fulfilled and Newton's method is self-consistent.

2.2. Properties of the Jost Functions and the Scattering Amplitude

2.2.1. Unitarity

We first remark that the unitarity condition

$$S(\nu)[S^*(\nu^*)] = 1 \qquad (2.13)$$

holds for any value of ν in the complex plane, since

$$S(\nu) = \frac{e^{\frac{1}{2}i\pi\nu}f_1(\nu)}{e^{-\frac{1}{2}i\pi\nu}f_2(\nu)}, \qquad (2.14)$$

and since it is obvious in (2.11) that

$$[f_1(\nu^*)]^* = f_2(\nu). \qquad (2.15)$$

2.2.2. Analyticity of $f_{1,2}(\nu)$

We see readily from (2.11) that $f_1(\nu)$ and $f_2(\nu)$ can be continued for any value of ν in the complex ν plane, except for negative integers $-l - 1$, for which $f_1(\nu)$ and $f_2(\nu)$ have simple poles, with residues $a_l(1 \pm i\tan\delta_l)(\pm i)^l/(-2l - 1)$. $f_1(\nu)$ and $f_2(\nu)$ are therefore meromorphic functions, with fixed simple poles, and the expansion (2.11) is their Mittag–Leffler expansion.

2.2.3. Another Expansion of the Jost Functions

Several properties of the functions which we are to study are related to the asymptotic behavior of the expansion coefficients in (2.11). Since the phase shifts δ_l go to zero as $l \to \infty$, the asymptotic behavior of the expansion coefficients in (2.11) is simply related to that of coefficients c_l, i.e., to $V(r)$. It would be interesting to have a formula similar to (2.11), where the asymptotic behavior of the expansion coefficients would be readily related to that of the phase shifts. It is not difficult to find such a formula when $|\delta_l|$ is bounded by $Cl^{-3-\epsilon}$ as $l \to \infty$. We show in Appendix AI that it is possible to obtain in this case the formula

$$\mathbf{a} = \mathbf{v} + M^{-1}\mathbf{\Delta}', \qquad (2.16)$$

where the components of $\mathbf{\Delta}'$ have the same asymptotic behavior as the phase shifts δ_l, i.e., the ratio of the lth component of $\mathbf{\Delta}'$ to δ_l is bounded.

Substituting this result in (2.11) and using the formulas (AII7) and (AII16) from Appendix II, we obtain for $f_2(\nu)$

$$\begin{aligned} f_2(\nu) &= e^{\frac{1}{2}i\pi\nu} - \alpha\sum_{l'} L_\nu^{l'} v_{l'} e^{-il'\frac{1}{2}\pi} \\ &+ i\sum_{l'} L_\nu^{l'} a_{l'}\tan\delta_{l'}\, e^{il'\frac{1}{2}\pi} - \sum_{l'} C_\nu^{l'}\,\Delta'_{l'}, \end{aligned} \qquad (2.17)$$

where the $C_\nu^{l'}$ are given by

$$\begin{aligned} C_\nu^{2p+1} &= -(-1)^p 4\pi^{-1}(2p + \tfrac{3}{2}) \\ &\quad \times \left[\frac{\Gamma(p + \frac{3}{2})\Gamma(\frac{1}{2}\nu + \frac{1}{2})}{\Gamma(p+1)\Gamma(\frac{1}{2}\nu + 1)}\right]^2 L_\nu^{2p+1}, \\ C_\nu^{2p} &= -i(-1)^p 4\pi^{-1}(2p + \tfrac{1}{2}) \\ &\quad \times \left[\frac{\Gamma(p + \frac{1}{2})\Gamma(\frac{1}{2}\nu + 1)}{\Gamma(p+1)\Gamma(\frac{1}{2}\nu + \frac{1}{2})}\right]^2 L_\nu^{2p}, \end{aligned} \qquad (2.18)$$

and

$$L_\nu^\mu = [(\nu + \tfrac{1}{2})^2 - (\mu + \tfrac{1}{2})^2]^{-1}\sin(\nu - \mu)\tfrac{1}{2}\pi. \qquad (2.19)$$

As we see later, the second term in the right-hand side of (2.17) can be calculated exactly. The expansion of $f_2(\nu)$ can be obtained from that of $f_1(\nu)$ through the formula (2.15). It is clear that the expansion (2.17) is very similar to the expansion (2.11), but once we have computed the term which could be done so exactly, the expansion coefficients of the remainder decrease like δ_l. We can therefore study the properties of $f_2(\nu)$ by using either one of the expansions (2.11) and (2.19). For the sake of simplicity, we limit ourselves here to (2.11).

2.2.4. Asymptotic Location of Regge Poles

The Regge poles are given by the zeros of $f_2(\nu)$. We try to find their positions in the ν complex plane for large values of $|\nu|$. The coefficients of $L_\nu^{l'}$ in (2.11b) are written as follows:

$$\begin{aligned} a_{2n}(1 - i\tan\delta_{2n}) &= \alpha\bar{\mu}_{2n} + \bar{a}_{2n}, \\ a_{2n+1}(1 - i\tan\delta_{2n+1}) &= \beta\bar{\mu}_{2n+1} + \bar{a}_{2n+1}, \end{aligned} \qquad (2.20)$$

and the $\bar{a}_n$ are assumed to go to zero as $n \to \infty$ faster than n^{-1}. According to an analysis of I [Sec. I, (3.2)], this condition is related to a bound $Cl^{-7/3-\epsilon}$ for

the δ_l. Insertion in (2.11b) of the formulas (2.20) leads us to consider (2.11b) as the sum of five terms, $T_1, T_2, \cdots, T_5$, where T_1 is $e^{\frac{1}{2}i\pi\nu}$, and the others are now considered separately. T_2 and T_4, which contain, respectively, the terms $\bar{\mu}_{2n}$ and $\bar{\mu}_{2n+1}$ of (2.11), can be calculated exactly, using formulas for the $\bar{\mu}_n$ given in I and well-known formulas,

$$\begin{aligned} -\alpha^{-1}T_2 &= \sum_n L_\mu^{2n} i^n \bar{\mu}_{2n} \\ &= \tfrac{1}{2}\pi \int_0^\infty J_{\nu+\frac{1}{2}}(\rho) \sum_n \bar{\mu}_{2n} J_{2n+\frac{1}{2}}(\rho)(-1)^n \\ &= (\tfrac{1}{2}\pi)^{\frac{1}{2}} \int_0^\infty J_{\nu+\frac{1}{2}}(\rho) J_0(\rho) \rho^{-\frac{1}{2}}\, d\rho \\ &= \frac{1}{2} \frac{\Gamma(\frac{1}{2}\nu + \frac{1}{2})\Gamma(\frac{1}{2}\nu + \frac{1}{2})}{\Gamma(\frac{1}{2}\nu + 1)\Gamma(\frac{1}{2}\nu + 1)} \cos \nu\tfrac{1}{2}\pi, \end{aligned} \tag{2.21}$$

so that, as $|\nu| \to \infty$, with $|\operatorname{Arg} \nu| < \pi$

$$T_2 \sim -\nu^{-1}\alpha \cos \nu\tfrac{1}{2}\pi \exp\left[-\tfrac{1}{2}\nu^{-1} + \tfrac{1}{6}\nu^{-2}\right]. \tag{2.22}$$

In the same way

$$\begin{aligned} T_4 &= -i\beta \sum_p L_\nu^{2p+1} \bar{\mu}_{2p+1} i^{2p} \\ &= -\tfrac{1}{2} i\beta \frac{\Gamma(\frac{1}{2}\nu + 1)\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu + \frac{3}{2})\Gamma(\frac{1}{2}\nu + \frac{1}{2})} \sin \nu\tfrac{1}{2}\pi, \end{aligned} \tag{2.23}$$

so that, as $|\nu| \to \infty$, with $|\operatorname{Arg} \nu| < \pi$

$$T_4 \sim -i\nu^{-1}\beta \sin \nu\tfrac{1}{2}\pi \exp\left[-\tfrac{1}{2}\nu^{-1} - \tfrac{1}{8}\nu^{-2}\right]. \tag{2.24}$$

Let us now study the terms T_3 and T_5 which contain the $\bar{a}_n$, and which can be put together into the form

$$\sum_{l'} \bar{a}_{l'} e^{\frac{1}{2}il'\pi} \frac{\sin(\nu - l')\frac{1}{2}\pi}{(\nu + \frac{1}{2})^2 - (l' + \frac{1}{2})^2}. \tag{2.25}$$

In order to avoid possible difficulties coming from the real axis, we first exclude from our study a strip $|\operatorname{Im} \nu| < \epsilon_0$, where ϵ_0 is a fixed number. Let us write (2.25) in the form

$$\begin{aligned} &e^{\frac{1}{2}i\nu\pi} \sum_{l'} \frac{\bar{a}_{l'}}{2i[(\nu + \frac{1}{2})^2 - (l' + \frac{1}{2})^2]} \\ &\quad - e^{-\frac{1}{2}i\nu\pi} \sum_{l'} \frac{\bar{a}_{l'}(-1)^{l'}}{2i[(\nu + \frac{1}{2})^2 - (l' + \frac{1}{2})^2]}. \end{aligned} \tag{2.26}$$

Suppose now the $\bar{a}_n$ are bounded in the following way:

$$|\bar{a}_n| < Cn^{-k} \qquad (k > 1), \tag{2.27}$$

a well-known identity enables us to write

$$\begin{aligned} &\sum_{l'} \frac{\bar{a}_{l'}}{(\nu + \frac{1}{2})^2 - (l' + \frac{1}{2})^2} \\ &= \sum_{l'} \frac{\bar{a}_{l'}}{(\nu + \frac{1}{2})^2} \left\{1 + \frac{(l' + \frac{1}{2})^2}{(\nu + \frac{1}{2})^2} + \cdots + \left[\frac{l' + \frac{1}{2}}{\nu + \frac{1}{2}}\right]^{2p-2}\right\} \\ &\quad + \sum_{l'} \bar{a}_{l'} \frac{[l' + \frac{1}{2}]^{2p}}{(\nu + \frac{1}{2})^{2p}[(\nu + \frac{1}{2})^2 - (l' + \frac{1}{2})^2]}. \end{aligned} \tag{2.28}$$

Let us make p equal to $E[\frac{1}{2}(k + 1 - \epsilon)]$, in order that the following condition should hold:

$$-1 \leq 2p - k < 1 - \epsilon. \tag{2.29}$$

The condition (2.29) and the results of Appendix AIII show readily that the remainder in (2.28) is of the order of $|\nu|^{-k-1+\epsilon}$ as $|\nu| \to \infty$ (with $\epsilon_0\nu \to \infty$), so that the last term in the brace of (2.28) is larger than the remainder. If the same work is done on the second term of (2.26), we get for $(T_3 + T_5)$ the expansion

$$\begin{aligned} T_3 + T_5 = e^{\frac{1}{2}i\nu\pi} &\left\{\frac{T_+^1}{(\nu + \frac{1}{2})^2} + \frac{T_+^2}{(\nu + \frac{1}{2})^4} + \cdots \right. \\ &\left. + \frac{T_+^p}{(\nu + \frac{1}{2})^{2p}} + O(\nu^{-k-1+\epsilon})\right\} \\ - e^{-\frac{1}{2}i\nu\pi} &\left\{\frac{T_-^1}{(\nu + \frac{1}{2})^2} + \frac{T_-^2}{(\nu + \frac{1}{2})^4} + \cdots \right. \\ &\left. + \frac{T_-^p}{(\nu + \frac{1}{2})^{2p}} + O(\nu^{-k-1+\epsilon})\right\}, \end{aligned} \tag{2.30}$$

where the coefficients T_+^q and T_-^q are linear combinations of the $\bar{a}_l$. Besides, the terms T_2 and T_4 can be combined to give

$$\begin{aligned} T_2 + T_4 &= -\tfrac{1}{2}\nu^{-1}e^{-\frac{1}{2}\nu^{-1}} \\ &\times \{e^{\frac{1}{2}i\nu\pi}[(\alpha + \beta) + \tfrac{1}{4}(\alpha - \beta)\nu^{-2} + O(\nu^{-3})] \\ &+ e^{-\frac{1}{2}i\nu\pi}[(\alpha - \beta) + \tfrac{1}{4}(\alpha + \beta)\nu^{-2} + O(\nu^{-3})]. \end{aligned} \tag{2.31}$$

If we compare $T_1 (= e^{\frac{1}{2}i\nu\pi})$ to the other terms for large $|\nu|$, we see that T_1 is larger in the lower half-plane, so that there cannot be zeros of $f_2(\nu)$ for large $|\nu|$ in the lower half-plane. When ν lies in the upper half-plane and $|\nu|$ is large, it is necessary that $\operatorname{Im} \nu$ and $\operatorname{Log} |\operatorname{Re} \nu|$ be of the same order so that $(T_2 + T_3 + T_4 + T_5)$ is of the same order as T_1. Therefore, we may limit our study to the large values of $\operatorname{Re} \nu$ and $\operatorname{Im} \nu$. According to the values of the δ_l and α, we may encounter one of the following cases:

(1) $\alpha \neq \beta$ $\quad f_2(\nu)$ can be put in the form

$$f_2(\nu) = e^{\frac{1}{2}i\nu\pi} - \tfrac{1}{2}\nu^{-1}(\alpha - \beta)e^{-\frac{1}{2}i\nu\pi}[1 + \epsilon(\nu)], \tag{2.32}$$

(2) $\alpha = \beta$ $\quad f_2(\nu)$ can be put in the form

$$f_2(\nu) = e^{\frac{1}{2}i\nu\pi} - (\nu + \tfrac{1}{2})^{-2} T e^{-\frac{1}{2}i\nu\pi}[1 + \epsilon_2(\nu)], \tag{2.33}$$

where T is obtained from T_-^1 and the coefficient of ν^{-2} in the terms $T_2 + T_4$. T may happen to vanish. This means that the c_l, or, equivalently, the phase shifts, are related by a definite constraint, so that the potential is restricted to a subclass of the class we study in general. The terms $(T_2 +$

$T_3 + T_4 + T_5$) are then of the order of $(\alpha + \beta)\nu^{-3}$ which is zero if (and only if) $\beta = 0$ (since $\alpha = \beta$). This method is applicable up to terms of the order of ν^{-2p}, if we limit ourselves to more and more restricted subclasses of potentials, or of phase shifts. If k is ∞, p is $+\infty$. If we limit our study to the cases where there is a nonvanishing coefficient of index q lower than k, the function $f_2(\nu)$ can be studied starting from the form

$$f_2(\nu) = e^{\frac{1}{2}i\pi\nu} - \nu^{-q}T^{(q)}e^{-\frac{1}{2}i\pi\nu}[1 + \epsilon_2(\nu)]. \tag{2.34}$$

This form is similar to the form (1.40) of Sec. 1, and can be studied in a similar way. If $\tau^{(q)} = e^{\sigma+i\tau}$, we find that $f_2(\nu)$ has an infinity of zeros in the upper half-plane, which, for large $|\nu|$, approach the points ($\mathrm{Re}\ \nu = \tau + 2k$) on the curve

$$\mathrm{Im}\ \nu = \pi^{-1}q\ \mathrm{Log}\ |\mathrm{Re}\ \nu|. \tag{2.35}$$

This curve is the boundary of a domain described by Martin which should not contain any Regge pole. It is not surprising to find Regge poles at $+\infty$ in the right half-plane. Only in the case of a potential which is holomorphic and bounded by a decreasing exponential in the right half of the r complex plane *has Bessis shown* that there could be *no pole* in the right half of the ν complex plane. Their existence in our problem proves that the conditions imposed on the potential are not fulfilled. As for the strip $|\mathrm{Im}\ \nu| < \epsilon_0$, which we did not study up to now, it is easy to show, with the help of Appendix III, that it does not contain any zero of $f_2(\nu)$ for large $|\nu|$ if $\mathrm{Re}\ \nu > 0$. It is possible that there are zeros in this strip for $\mathrm{Re}\ \nu < 0$, which are not interesting.

2.2.5. Analytic Properties of the Scattering Amplitude

From the above study and from formula (2.14), we see that $S(\nu)$ is a meromorphic function of ν. As $\nu \to -i\infty$, $S(\nu)$ behaves as $C\nu^{-q}e^{i\pi\nu}$. This behavior, and the existence of poles for $\mathrm{Re}\ \nu = \infty$, exclude any possibility of applying the Carlson theorem. We now show that Newton's potentials cannot include "good" superpositions of Yukawa potentials. Indeed, a "good" superposition of Yukawa potentials exhibits two important features

(1) δ_l decreases exponentially as $l \to \infty$.
(2) There is no Regge pole as $\mathrm{Re}\ \nu \to +\infty$.

Now, in the formula (2.17), if the δ_l decrease exponentially, it is possible to impose to the phase shifts an infinity of constraints in order to avoid a behavior of the Jost functions like (2.34). More precisely, we have to impose the vanishing of all the linear combinations

$$\sum_0^\infty \Delta(l + \tfrac{1}{2})(l + \tfrac{1}{2})^{2q} = 0, \tag{2.36}$$

where $q = 0, 1, \cdots, \infty$, and the coefficients $\Delta(l + \frac{1}{2})$ behave asymptotically like the δ_l. If δ_l decrease exponentially, we can write

$$\Delta(l + \tfrac{1}{2}) < Ce^{-\Gamma l}. \tag{2.37}$$

Let us now introduce a generating function $g(x)$ for $\Delta(l + \frac{1}{2})$

$$g(x) = \pi^{-1}\sum_0^\infty \Delta(l + \tfrac{1}{2})\cos(l + \tfrac{1}{2})x,$$
$$\Delta(l + \tfrac{1}{2}) = \int_0^\pi \cos(l + \tfrac{1}{2})x g(x)\,dx. \tag{2.38}$$

It is clear that $g(x)$ is analytic in the strip $(\mathrm{Im}\ x) < \Gamma$. Furthermore, $g(x)$ is an even function, and it follows from (2.36) that all its even-order derivatives are equal to zero. $g(x)$ is therefore equal to zero and so is $\Delta(l + \frac{1}{2})$. This means that, at least from $l = l_0$, all the δ_l are equal to zero, and that, even for $l_0 \neq 0$, $f_2(\nu)$ is equal to $e^{\frac{1}{2}i\pi\nu}$, so that the only potential consistent with these properties is $V(r) = 0$.

Remark: One can show that if there are several equivalent potentials, and if $S(\nu)$ cannot increase faster than $e^{|\tau\nu|}$, it is necessary that, for all these potentials, except maybe one of them, $S(\nu)$ should have an infinite number of poles and zeros in the right half-plane. Indeed, suppose that $S(\nu)$ is holomorphic and different from zero for $\mathrm{Re}\ \nu > A$. It would then be possible to calculate the function $[S(\nu)]^{\frac{1}{2}}$ in this domain, which would fulfill the applicability conditions of Carlson's theorem, and this can happen for only one of the equivalent potentials.

APPENDIX I

Our aim is to obtain for the asymptotic behavior of the coefficients a_l more precise formulas than those derived in I. We start with the fundamental formula

$$\tan \Delta\mathbf{e} = M\mathbf{a} + \tan\Delta\, M \tan\Delta\, \mathbf{a}, \tag{AI1}$$

where M is the matrix:

$$M_l^{l'} = \begin{cases} [l'(l'+1) - l(l+1)]^{-1}, & \text{if } l' - l \text{ is odd,} \\ 0, & \text{if } l' - l \text{ is even.} \end{cases} \tag{AI2}$$

The inverse matrix M^{-1} of M is equal to

$$M^{-1} = \mu M \mu, \tag{AI3}$$

where μ is a diagonal matrix, whose elements are equal to

$$\begin{aligned}\bar{\mu}_{2n+1} &= 2\pi^{-1}(2n+\tfrac{3}{2})[\Gamma(n+\tfrac{3}{2})/\Gamma(n+1)]^2 \\ &\sim 4\pi^{-1}n^2, \quad \text{for} \quad n\to\infty, \\ \bar{\mu}_{2n} &\sim 2\pi^{-1}(2n+\tfrac{1}{2})[\Gamma(n+\tfrac{1}{2})/\Gamma(n+1)]^2 \\ &\sim 4\pi^{-1}, \quad \text{for} \quad n\to\infty.\end{aligned} \tag{AI4}$$

In order to transform (AI1), we should first study the conditions in which the products $MM^{-1}\mathbf{x}$ and $M^{-1}M\mathbf{x}$ are associative, that is to say, are equal to $\mathbf{x}$. Using (AI3), (AI4), and bounds given in the Sec. 1.2 of I, it is easy to show that all the double series in $M(M^{-1}\mathbf{x})$ are bounded by summable positive series, and therefore summable, if

$$\text{(a)} \quad x_{l,\text{even}} < Cl_2^{-\epsilon}, \tag{AI5a}$$

$$\text{(b)} \quad x_{l,\text{odd}} < Cl_2^{1-\epsilon}. \tag{AI5b}$$

The Fubini theorem leads us straightforwardly to the result

$$M(M^{-1}\mathbf{x}) = MM^{-1}\mathbf{x} = \mathbf{x}. \tag{AI6}$$

It should be noticed that the condition (AI5b) is necessary to allow the application of M^{-1} to $\mathbf{x}$.

It is easy to see in the same way that $M^{-1}M\mathbf{x}$ is associative if

$$\begin{aligned}x_{l,\text{even}} &< Cl_2^{-\epsilon}, \\ x_{l,\text{odd}} &< Cl_2^{1-\epsilon}.\end{aligned} \tag{AI7}$$

Let us now return to Eq. (AI1). The summability conditions in this equation are

$$|a_l| < Cl^{1-\epsilon}, \qquad |a_l \tan\delta_l| < Cl^{1-\epsilon}. \tag{AI8}$$

If, more precisely, we suppose that $|l^{1+\epsilon}\delta_l|$ is bounded, the operator M^{-1} may be applied to all the terms of (AI1). According to (AI5) and (AI6), recalling that the vector $\mathbf{v}$ is annihilated by M, we deduce from (AI1) the equation

$$\begin{aligned}M^{-1}\tan\Delta\,\mathbf{e} + \alpha\mathbf{v} &= (1 + M^{-1}\tan\Delta\, M\tan\Delta)\mathbf{a} \\ &= (1+R)\mathbf{a}.\end{aligned} \tag{AI9}$$

This equation was given by Newton,[2] and studied in detail by the author,[1] but the validity conditions had not been fully investigated. If $l^{4/3+\epsilon}\delta_l$ is bounded, the analysis given in I shows that it is possible to get a set of a_l which are bounded. We see here that this analysis is consistent with the above conditions. It is assumed hereafter that

$$l^{4/3+s}\,\delta_l < C; \qquad s > 0. \tag{AI10}$$

Equation (AI9) is equivalent to the following set of equations:

$$\begin{aligned}&\text{(a)} \quad \alpha\mathbf{v} = (1+R)(\alpha\mathbf{v} + \mathbf{c}), \\ &\text{(b)} \quad M^{-1}\tan\Delta\mathbf{e} = (1+R)\mathbf{b}, \\ &\text{(c)} \quad \mathbf{a} = \mathbf{b} + \mathbf{c} + \alpha\mathbf{v}.\end{aligned} \tag{AI11}$$

According to (AI10) and the results in (3.66) of I, and since R connects only components of the same parity, the following orders hold:

$$\begin{aligned}a_{2r} &= \alpha v_{2r} + O(r^{-\epsilon}), \\ c_{2r} &= O(r^{-\epsilon}), \qquad c_{2r+1} = 0, \\ |b_{2r}| &< C(r^{-\epsilon}), \qquad |b_{2r+1}| < C.\end{aligned} \tag{AI12}$$

The last two inequalities enable us to write

$$\mathbf{b} = M^{-1}\tan\Delta\cot\Delta\, M\mathbf{b}. \tag{AI13}$$

Since $\tan\Delta$ and $\cot\Delta$ are diagonal matrices, it is clear that the associativity of the product in (AI13) is ensured by the associativity of MM^{-1}. Equation (AI11b) may therefore be written as

$$M^{-1}\tan\Delta\mathbf{e} = M^{-1}\tan\Delta(1+R^*)\cot\Delta\, M\mathbf{b}, \tag{AI14}$$

where

$$R^* = M\tan\Delta\, M^{-1}\tan\Delta. \tag{AI15}$$

If we notice that the vector $\mu^{-1}\mathbf{v}$ is annihilated by M^{-1}, we deduce from (AI14) the equation

$$\mathbf{e} - \gamma\cot\Delta\,\mu^{-1}\mathbf{v} = (1+R^*)\cot\Delta\, M\mathbf{b}. \tag{AI16}$$

A comparison between the asymptotic behaviors of the two sides in (AI16) shows that $\gamma = 0$. Proceeding carefully as above, we get from (AI16) a formula for $\mathbf{b}$

$$\mathbf{b} = M^{-1}\tan\Delta(1+R^*)^{-1}\mathbf{e}. \tag{AI17}$$

The same technique enables one to derive $\mathbf{c}$ through the following steps:

$$-R\alpha\mathbf{v} = (1+R)\mathbf{c}, \tag{AI18}$$

$$\mathbf{c} = M^{-1}\tan\Delta\cot\Delta\, M\mathbf{c}, \tag{AI19}$$

$$-\alpha M\tan\Delta\mathbf{v} = (1+R^*)\cot\Delta\, M\mathbf{c}. \tag{AI20}$$

From (AI11a), (AI17), and (AI20), we derive the value of $\mathbf{a}$

$$\mathbf{a} = \alpha\mathbf{v} + M^{-1}\tan\Delta(1+R^*)^{-1}[\mathbf{e} - \alpha M\tan\Delta\,\mathbf{v}]. \tag{AI21}$$

We now have to study the equation

$$\mathbf{y}_0 = \mathbf{x}_0 - R^*\mathbf{y}_0 = \mathbf{x}_0 - \mu^{-1}R\mu\mathbf{y}_0 \tag{AI22}$$

or

$$\mu\mathbf{y}_0 = \mu\mathbf{x}_0 - R\mu\mathbf{y}_0. \tag{AI23}$$

We see that this problem is equivalent to the problem of the inversion of $(1 + R)$, which has been treated in I, but that the application of the operator $(1 + R)^{-1}$ may involve some new difficulties, since $(\mu\mathbf{x})_l$ behaves asymptotically as l^2. From I we know that the solution of

$$\mathbf{y} = \mathbf{x} - R\mathbf{y} \tag{AI24}$$

is

$$\begin{aligned} y_{2r} &= x_{2r} + \sum_p V_{2r}^{2p} \tan \delta_{2p} x_{2p} (-1)^p, \\ y_{2r+1} &= \sum_p U_{2r+1}^{2p+1} (x_{2p+1} - \beta_1 \bar{\mu}_{2p+1}), \end{aligned} \tag{AI25}$$

where

$$U_{2r+1}^{2p+1} = \delta_{2r+1}^{2p+1} + \bar{V}_{2r+1}^{2p+1} \tan \delta_{2p+1} (-1)^p. \tag{AI26}$$

V_{2r}^{2p} is defined, in function of the operators γ_0 and γ_0^r defined in I, by

$$V_{2r}^{2p} = \mathfrak{D}^{-1}(\gamma_0) \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \int \cdots \int \begin{vmatrix} \int \gamma_0^r(\theta') \bar{u}_2(\theta')\, d\theta'; & \gamma_0^r(\theta_1) \cdots \gamma_0^r(\theta_m) \\ \int \gamma_0(\theta_1, \theta') \bar{u}_{2p}(\theta')\, d\theta'; & \gamma_0(\theta_1, \theta_1) \cdots \\ \vdots & \\ \int \gamma_0(\theta_m, \theta') \bar{u}_{2p}(\theta')\, d\theta'; & \gamma_0(\theta_m, \theta_1) \cdots \gamma_0(\theta_m, \theta_m) \end{vmatrix} d\theta_1 \cdots d\theta_m, \tag{AI27}$$

where

$$\mathfrak{D}(\gamma) = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \int \cdots \int \begin{vmatrix} \gamma_0(\theta_1, \theta_1) \cdots \gamma_0(\theta_1, \theta_m) \\ \gamma_0(\theta_m, \theta_1) \cdots \gamma_0(\theta_m, \theta_m) \end{vmatrix} d\theta_1 \cdots d\theta_m. \tag{AI28}$$

All the integrations are taken between 0 and $\frac{1}{2}\pi$. The functions γ_0 and γ_0^r are bounded for any value in the domain of integration if $l^{4/3+\epsilon}\delta_l$ is bounded. $\bar{V}_{2r+1}^{2p+1}$ is given by (AI27), provided we replace γ_0 by γ_1, γ_0^r by γ_1^r everywhere. If $\mathfrak{D}$ happens to be equal to zero, either for $\gamma = \gamma_0$ or for $\gamma = \gamma_1$, the inversion of $(1 + R)$ is not possible. As in I, we discard the particular sets of phase shifts for which this happens. β_1 is given by

$$\beta_1 = \frac{\sum_r \sigma_{2r+1} \sum_p U_{2r+1}^{2p+1} x_{2p+1}}{1 + \sum_r \sigma_{2r+1} \sum_p U_{2r+1}^{2p+1} \bar{\mu}_{2p+1}}, \tag{AI29}$$

where

$$\sigma_{2r+1} = \tan \delta_{2r+1} \sum_q M_{2q}^{2r+1} \bar{\mu}_{2q} \tan \delta_{2q}. \tag{AI30}$$

Now, if $l^{4/3+\epsilon}$ is bounded, it is clear that V_{2r}^{2p} or V_{2r+1}^{2p+1} are bounded by $Cp^{1/3}$. According to (AI25), (AI26), (AI29), and (AI30), a condition sufficient to ensure the convergence of all the series is that $p^{1/3+\epsilon} |\tan \delta_p x_p|$ be bounded as $p \to \infty$. If we return to (AI22), we see therefore that this equation can be solved if $l^{10/3+\epsilon}\delta_l$ is bounded. If this condition is valid, then (AI21) holds and we can write this equation in the equivalent form

$$\mathbf{a} = \alpha\mathbf{v} + M^{-1}\Delta', \tag{AI31}$$

where Δ' is a vector the components of which behave asymptotically as $\tan \delta_l$:

$$\Delta_l' \sim \tan \delta_l [1 + O(l^{-2})], \quad \text{as} \quad l \to \infty. \tag{AI32}$$

APPENDIX II

The matrices $L_\nu^{l'}$ and M^{-1} being defined by (2.7) and (AI3), we compute the elements of

$$C_\nu^l = \sum_{l'} L_\nu^{l'} e^{\frac{1}{2}il'\pi} \gamma_{l'}^l. \tag{AII1}$$

For $l = 2p + 1$, C_ν^l is equal, according to (I, 3.27), (I, 3.28), and (I, 2.3), to

$$C_\nu^{2p+1} = -(-1)^p \tfrac{1}{2}\pi \int_0^\infty J_{\nu+\frac{1}{2}}(\rho) \rho^{-1} A(\rho)\, d\rho, \tag{AII2}$$

where

$$\begin{aligned} A(\rho) = 2\pi\bar{\mu}_{2p+1}(p + \tfrac{1}{2})(p + 1) &\int_0^\infty J_{2p+\frac{1}{2}}(\rho_1) \rho_1^{-1}\, d\rho_1 \\ &\times \sum_n \bar{\mu}_{2n} J_{2n+\frac{1}{2}}(\rho_1) J_{2n+\frac{1}{2}}(\rho). \end{aligned} \tag{AII3}$$

Using for the series in the right-hand side of (AII3) the form given in I [I, (AII3)], we can write $A(\rho)$ as

$$\begin{aligned} A(\rho) = 4\bar{\mu}_{2p+1}(p + \tfrac{1}{2})(p + 1)\rho^{\frac{1}{2}} &\\ \times \int_0^{\pi/2} J_0(\rho \sin \alpha) \sin \alpha\, d\alpha &\\ \times \int_0^\infty J_0(\rho_1 \sin \alpha) J_{2p+\frac{1}{2}}(\rho_1) \rho_1^{-\frac{1}{2}}\, d\rho_1. & \end{aligned} \tag{AII4}$$

Using now the well-known formulas,[14] we can compute the integral in (AII4) and obtain

$$A(\rho) = 4\bar{\mu}_{2p+1}\frac{\Gamma(p+2)}{\Gamma(p+\frac{1}{2})}(\tfrac{1}{2}\rho)^{\frac{1}{2}} \times \int_0^{\pi/2} J_0(\rho\sin\alpha)P_{2p+1}(\cos\alpha)\sin\alpha\,d\alpha. \quad \text{(AII5)}$$

Inserting this result in (AII2) and performing the integral over ρ with the help of the well-known formulas,[15] we obtain

$$C_\nu^{2p+1} = -(-1)^p 2(2p+\tfrac{3}{2})\frac{\Gamma(p+\frac{3}{2})}{\Gamma(p+1)}\cdot\frac{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}{\Gamma(\frac{1}{2}\nu+1)} \times \int_0^{\pi/2} P_{2p+1}(\cos\alpha)P_\nu(\cos\alpha)\sin\alpha\,d\alpha. \quad \text{(AII6)}$$

A well-known formula[16] yields then the result

$$C_\nu^{2p+1} = -(-1)^p 2(2p+\tfrac{3}{2})\left[\frac{\Gamma(p+\frac{3}{2})}{\Gamma(p+1)}\cdot\frac{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}{\Gamma(\frac{1}{2}\nu+1)}\right]^2 \times \frac{\sin\frac{1}{2}\pi(\nu-2p-1)}{\frac{1}{2}\pi(\nu-2p-1)(\nu+2p+2)}. \quad \text{(AII7)}$$

For $l = 2p$, C_ν^l can be written, with the help of formulas (3.27), (3.28), and (2.3) of I, as

$$C_\nu^{2p} = \tfrac{1}{2}i\pi\int_0^\infty J_{\nu+\frac{1}{2}}(\rho)\,d\rho\,\rho^{-1}B(\rho), \quad \text{(AII8)}$$

where

$$B(\rho) = \bar{\mu}_{2p}\sum_r \bar{\mu}_{2r+1}J_{2r+\frac{1}{2}}(\rho)(-1)^r + 2\pi\bar{\mu}_{2p}p(p+\tfrac{1}{2})(-1)^p\int_0^\infty J_{2p+\frac{1}{2}}(\rho') \times \sum_r \bar{\mu}_{2r+1}J_{2r+\frac{1}{2}}(\rho)J_{2r+\frac{1}{2}}(\rho'). \quad \text{(AII9)}$$

Using for the series in the right-hand side of (AII9) integral expressions given in I [see (2.8) with $\alpha = \frac{1}{2}$, and (AII7)], we obtain for $B(\rho)$ the formula

$$B(\rho) = \bar{\mu}_{2p}(2/\pi)^{\frac{1}{2}}\rho^{\frac{1}{2}}J_1(\rho) + 4p(p+\tfrac{1}{2})(-1)^p\bar{\mu}_{2p}\rho^{\frac{1}{2}} \times \int_0^{\pi/2} J_1(\rho\sin\alpha)\sin\alpha\,d\alpha \times \int_0^\infty J_1(\rho'\sin\alpha)J_{2p+\frac{1}{2}}(\rho')\rho'^{-\frac{1}{2}}\,d\rho'. \quad \text{(AII10)}$$

We denote, respectively, by B_1 and B_2 the first and the second term in (AII10). The last integral in $B_2(\rho)$ can be transformed with the help of the well-known formulas[17] and $B_2(\rho)$ takes the form

$$B_2(\rho) = 2\bar{\mu}_{2p}(\tfrac{1}{2}p)^{\frac{1}{2}}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})} \times \int_0^{\pi/2} J_1(\rho\sin\alpha)\sin^2\alpha\,d\alpha\frac{d}{d\cos\alpha}P_{2p}(\cos\alpha). \quad \text{(AII11)}$$

After inserting this result in (AII8), and computing the integral over ρ in the same way as above, we get for the contribution of $B_2(\rho)$ to C_ν^{2p}:

$$\tfrac{1}{4}i\pi(-1)^p\bar{\mu}_{2p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}\frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+\frac{3}{2})} \times \int_0^1 (1-x^2)\frac{d}{dx}P_\nu(x)\frac{d}{dx}P_{2p}(x)\,dx. \quad \text{(AII12)}$$

Integrating by parts and using the Legendre differential equation readily yields the following formula, in which we have replaced $\bar{\mu}_{2p}$ by its expression in [I, (3.28)]:

$$2i(-1)^p(2p+\tfrac{1}{2})\frac{\Gamma(p+\frac{3}{2})}{\Gamma(p)}\frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+\frac{3}{2})} \times \int_0^1 P_\nu(x)P_{2p}(x)\,dx. \quad \text{(AII13)}$$

With the help of a well-known formula,[18] we get the result corresponding to $B_2(\rho)$:

$$-\tfrac{1}{2}i\bar{\mu}_{2p}\frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+\frac{3}{2})}\frac{\Gamma(1+\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+\frac{1}{2})} \times \frac{2p(2p+1)}{(2p-\nu)(2p+\nu+1)}\sin\tfrac{1}{2}\nu\pi. \quad \text{(AII14)}$$

The contribution of $B_1(\rho)$ is much easier to compute: a known formula[19] readily yields the result

$$\tfrac{1}{2}i\bar{\mu}_{2p}\frac{\Gamma(\frac{1}{2}\nu+1)}{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}\frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+\frac{3}{2})}\sin\tfrac{1}{2}\pi\nu. \quad \text{(AII15)}$$

If we add (AII14) and (AII15), we find the expression of C_ν^{2p}

$$C_\nu^{2p} = -i(-1)^p 2(2p+\tfrac{1}{2})\left[\frac{\Gamma(p+\frac{1}{2})}{\Gamma(p+1)}\frac{\Gamma(\frac{1}{2}\nu+1)}{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}\right]^2 \times \frac{\sin\frac{1}{2}\pi(\nu-2p)}{\frac{1}{2}\pi(\nu-2p)(\nu+2p+1)}. \quad \text{(AII16)}$$

All the convergence proofs, and, in general, the validity of the mathematical derivation result readily from the mathematical studies given in I.

APPENDIX III

We study here the asymptotic behavior for large values of $|\nu|$ of the series

$$S = \sum_{q=0}^\infty \frac{\gamma_q}{(\nu-q)(\nu+q+\alpha)}, \quad \alpha > 0, \quad \text{(AIII1)}$$

[14] HTF 7.7 (29).
[15] HTF 7.7 (29) and HTF 3.2 (20).
[16] HTF 3.12 (15).
[17] HTF 7.7 (29), HTF 2.11 (10), HTF 2.8 (20), and HTF 3.2 (14).
[18] HTF 3.12 (15).
[19] HTF 7.7 (30).

where the $|\gamma_q|$ are bounded by $Cq^{-\beta}$ ($\beta > -1$) and where ν takes its values in the complex plane, except the neighborhood of the real axis. For the sake of simplicity, we more precisely exclude the strip $|\mathrm{Im}\ \nu| < \epsilon_0$, where ϵ_0 is a fixed number. Let λ be the real part and μ be the imaginary part of $(\nu + \alpha)$. The modulus of S can be bounded as follows:

$$|S| \leq \sum_{q=0}^{\infty} |\gamma_q| \{[\lambda^2 + \mu^2 - (q+\alpha)^2]^2 + 4\mu^2(q+\alpha)^2\}^{-\frac{1}{2}} \quad \text{(AIII2)}$$

or

$$|S| \leq \sum_{q=0}^{\infty}{}' |\gamma_q| [(\lambda^2 + \mu^2) - (q+\alpha)^2]^{-1} + C\,|\nu|^{-1-\beta}, \quad \text{(AIII3)}$$

where the prime on the sum means that we excluded from the summation the two values of q nearest to $q + \alpha = (\lambda^2 + \mu^2)^{\frac{1}{2}}$.

According to a study given in I [I, (AI3)], $|S|$ is therefore bounded by $(C\,|\nu|^{-1-\beta+\epsilon} + C\,|\nu|^{-2})$. From the present study, it is easy to deduce the asymptotic behavior of the following series for large values of ν in the complex plane (including the *positive* real axis)

$$\begin{cases} S_1 = \sin \frac{1}{2}\nu\pi \sum_{n=0}^{\infty} \dfrac{\gamma_{2n}}{(\nu - 2n)(\nu + 2n + \alpha)}, \\ S_2 = \cos \frac{1}{2}\nu\pi \sum_{n=0}^{\infty} \dfrac{\gamma_{2n+1}}{(\nu - 2n - 1)(\nu + 2n + 1 + \alpha)}. \end{cases} \quad \text{(AIII4)}$$

Proceeding exactly as above, and excluding from the sums the terms with index nearest to $|\frac{1}{2}\nu|$, we find, for example, for S_1

$$\begin{cases} |S_1| < |\sin \frac{1}{2}\nu\pi|\,[C\,|\nu|^{-\beta-1+\epsilon} + C\,|\nu|^{-2}], \\ \qquad\qquad \text{for} \quad |\mathrm{Im}\ \nu| > \epsilon_0, \\ |S_1| < C\,|\nu|^{-\beta-1+\epsilon} + C\,|\nu|^{-2}, \\ \qquad\qquad \text{for} \quad |\mathrm{Im}\ \nu| < \epsilon_0. \end{cases} \quad \text{(AIII5)}$$

APPENDIX IV

It is remarkable that, for the special example given in (1.4), the scattering amplitude (and the cross section) can be given an exact closed form. To show this point, let us first introduce the following notations:

$$a = c_{l_0}(1 + \tfrac{1}{2}\pi c_{l_0}/2l_0 + 1)^{-1}, \quad \text{(AIV1)}$$

$$b = [(l_0 + \tfrac{1}{2})^2 - ia]^{\frac{1}{2}}; \qquad \mathrm{Re}\ b > 0. \quad \text{(AIV2)}$$

We can write the scattering amplitude as

$$f(\theta) = (ik)^{-1} \sum_{0}^{\infty} (2l+1)(1 + i\cot\delta_l)^{-1} P_l(\cos\theta), \quad \text{(AIV3)}$$

where, according to (1.35),

$$\cot\delta_l = a^{-1}[(l_0 + \tfrac{1}{2})^2 - (l + \tfrac{1}{2})^2] \quad \text{(AIV4)}$$

for $|l - l_0|$ odd, ∞ otherwise.

1. l_0 odd

From (AIV3), (AIV4), and the well-known results,[20] we derive readily the following relation:

$$f(\theta) = \frac{-\pi a k^{-1}}{\sin(b - \frac{1}{2})\frac{1}{2}\pi} \int_0^{\infty} J_b(x) x^{-1}\,dx \times \sum_{0}^{\infty} (2l + \tfrac{1}{2})(-1)^l J_{2l+\frac{1}{2}}(x) P_{2l}(\cos\theta) \quad \text{(AIV5)}$$

$$= \frac{-\pi a k^{-1}}{\sin(b - \frac{1}{2})\frac{1}{2}\pi (2\pi)^{\frac{1}{2}}} \times \int_0^{\infty} J_b(x) \cos(x\,|\cos\theta|) x^{-\frac{1}{2}}\,dx. \quad \text{(AIV6)}$$

This Weber–Schafheitlein integral can be calculated[14] as

$$f(\theta) = \frac{-\pi a k^{-1} \Gamma(\frac{1}{4} + \frac{1}{2}b)}{2\Gamma(\frac{1}{2})\Gamma(\frac{3}{4} + \frac{1}{2}b)\sin(b - \frac{1}{2})\frac{1}{2}\pi} \times {}_2F_1(\tfrac{1}{4} + \tfrac{1}{2}b, \tfrac{1}{4} - \tfrac{1}{2}b; \tfrac{1}{2}; \cos^2\theta). \quad \text{(AIV7)}$$

It is possible to express the result (AIV7) in terms of Legendre functions[21]

$$f(\theta) = (2\cos b\pi)^{-1} \pi a k^{-1} \{{}_2F_1(\tfrac{1}{2} + b, \tfrac{1}{2} - b; 1; \sin^2\tfrac{1}{2}\theta) + {}_2F_1(\tfrac{1}{2} + b, \tfrac{1}{2} - b; 1; \cos^2\tfrac{1}{2}\theta)\}$$

or

$$f(\theta) = (2\cos b\pi)^{-1} \pi a k^{-1} \{P_{(b-\frac{1}{2})}(\cos\theta) + P_{(b-\frac{1}{2})}[\cos(\pi - \theta)]\}. \quad \text{(AIV8)}$$

2. l_0 even

With the help of the well-known formulas,[22] we obtain, in the same way as above,

$$f(\theta) = \frac{-\pi a k^{-1} \cos\theta\, \Gamma(\frac{1}{2}b + \frac{3}{4})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}b + \frac{1}{4})\sin(b - \frac{3}{2})\frac{1}{2}\pi} \times {}_2F_1(\tfrac{3}{4} + \tfrac{1}{2}b, \tfrac{3}{4} - \tfrac{1}{2}b; \tfrac{3}{2}; \cos^2\theta). \quad \text{(AIV9)}$$

A quadratic transformation[21] enables us to express this result in terms of the Legendre functions[22]

$$f(\theta) = (2\cos b\pi)^{-1} \pi a k^{-1} \{P_{(b-\frac{1}{2})}(\cos\theta) - P_{(b-\frac{1}{2})}[\cos(\pi - \theta)]\}. \quad \text{(AIV10)}$$

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. K. Chadan and Professor A. Martin for interesting discussions.

[20] HTF 7.14 (32) and HTF 7.15 (44).
[21] HTF 2.11 (7) and HTF 3.2 (14).
[22] HTF 7.14 (32) and HTF 7.15 (45).

Radiative Transfer. II*

C. E. SIEWERT† AND P. F. ZWEIFEL‡
Department of Theoretical Physics, Middle East Technical University, Ankara, Turkey
(Received 6 January 1966)

This paper presents an exact solution to the equations of radiative transfer for a generalization of the Uniform-Picket-Fence model discussed in a previous work. Here the absorption coefficient is allowed to take N different values over the frequency spectrum. Case's method is used to construct the normal mode solutions to the set of N coupled integral equations. Then half-range completeness and orthogonality theorems are proved that enable one to solve typical half-space problems. Explicitly, the asymptotic solution to the Milne problem is developed, including the extrapolated end point, while implicitly the complete solution is available.

I. INTRODUCTION

In a previous paper[1] (hereafter referred to as I) we presented exact solutions for the equations of radiative transfer with local thermodynamic equilibrium for a particular model of the absorption coefficient K_ν (uniform picket-fence model). In this model, K_ν was assumed to take only two values K_1 and K_2. In the present paper, we generalize the results of I to the case where K_ν can take on N values. The derivation of the basic (matrix) transport equation follows similar lines of reasoning as that of I. However, we briefly give the cogent points in the derivation below with a slight generalization of the case treated in I; namely, we include a scattering term in the transport equation which we neglect in order to obtain an explicit solution.

We begin, then, with the equation for radiative transfer

$$\mu \frac{\partial \psi_\nu(z, \mu)}{\partial z} + \rho(z)(K_\nu + S_\nu)\psi_\nu(z, \mu) = \rho(z)K_\nu\beta_\nu[T(z)] + \rho(z)S_\nu \int \psi_\nu(z, \mu')f_\nu(\Omega'\cdot\Omega)\, d\Omega'. \quad (1)$$

Here, as in I, we assume plane symmetry (coordinate z), where μ is the cosine of the angle between the photon velocity vector and the z axis; ψ_ν is the angular energy density of frequency ν; $\rho(z)$ is the material density, and

$$\beta_\nu[T(z)] = (2h\nu^3/c^2)[\exp h\nu/kT(z) - 1]^{-1}$$

is the Planck black body function for the "local temperature" $T(z)$. This equation is identical to Eq. (1) of I except that here a monochromatic scattering term is included; S_ν is the scattering coefficient, and

$$\int f(\Omega'\cdot\Omega)\, d\Omega' = \int f(\Omega'\cdot\Omega)\, d\Omega = 1. \quad (2)$$

We now assume that the frequency spectrum can be divided into ranges $\Delta\nu_i$ in each of which K_ν and S_ν take on the same N different values ($K_1, K_2, \cdots, K_N$; $S_1, S_2, \cdots, S_N$) and that the fractional width $w_n(i)$ of $\Delta\nu_i$ over which K_ν and S_ν have the same values is the same for all $\Delta\nu_i$.[2] Further, we must assume that $B_\nu[T(z)]$ can be taken independent of ν over the range $\Delta\nu_i$. The meaning of these assumptions may be clarified by examining Fig. 1. If these assumptions are not reasonable *in detail* (uniform model) they may be so on *the average* (random model).

Keeping in mind these assumptions, we integrate Eq. (1) over the frequency range Δ_n in which K_ν and S_ν have values K_n and S_n, respectively (this includes contributions from all $\Delta\nu_i$), to obtain

$$\mu \frac{\partial}{\partial z}\psi_n(z, \mu) + \rho(z)(K_n + S_n)\psi_n = \rho(z)K_nB_n[T(z)] + \rho(z)S_n \int \psi_n(z, \mu')f(\Omega'\cdot\Omega)\, d\Omega', \quad n = 1, 2, \cdots, N, \quad (3)$$

where

$$\psi_n(z, \mu) \triangleq \int_{\Delta_n} d\nu\, \psi_\nu(z, \mu), \quad (4)$$

and $B_n[T(z)]$ has a similar definition. The Schwarzs-

* This work is based on a Ph.D. thesis submitted by one of the authors (C.E.S.) to the University of Michigan.

† United States Atomic Energy Commission Pre-doctoral Fellow with a leave of absence from the Department of Nuclear Engineering, the University of Michigan. Present Address: Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina.

‡ On Sabbatical leave, academic year, 1964–1965, from the University of Michigan.

[1] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N. Y.) **36**, 61 (1966).

[2] We have assumed that the steps for K_ν and S_ν always occur at the same value of ν. Also when K_ν has the value K_n, S_ν has the value S_n.

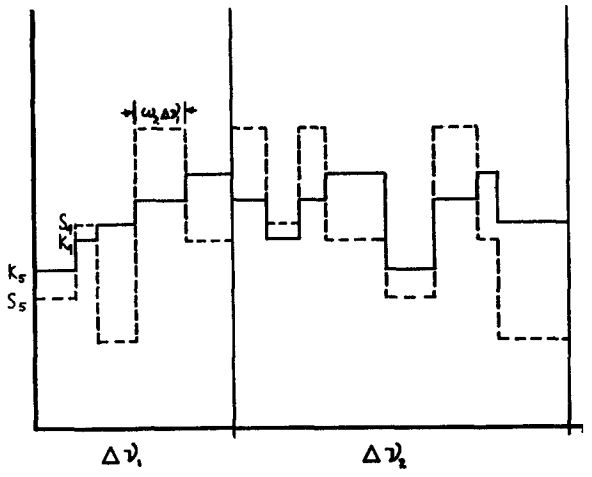


FIG. 1. Generalized picket-fence model, $N = 5$. We see that K_ν (solid lines) and S_ν (dashed lines) take on 5 different values in $\Delta\nu_i$ and that the frequency ranges over which K_ν and S_ν have constant values are the same. Further, the fractional width $w_n(i)$ of $\Delta\nu_i$ covered by K_ν and S_ν is the same for all i.

child condition,[1] which states local energy conservation, is

$$\int_0^\infty K_\nu B_\nu[T(z)]\,d\nu = \tfrac{1}{2}\int_0^\infty K_\nu\,d\nu\int_{-1}^1 \psi_\nu(z,\mu')\,d\mu'. \tag{5a}$$

In the present model, it takes the form

$$\sum_{n=1}^N K_n B_n[T(z)] = \tfrac{1}{2}\sum_{n=1}^N K_n\int_{-1}^1 \psi_n(z,\mu')\,d\mu'. \tag{5b}$$

Our uniformity assumptions (cf. Fig. 1) easily lead to the result

$$B_n[T(z)] = w_n\int_0^\infty d\nu\, B_\nu[T(z)] \tag{6a}$$

$$= (w_n\sigma/\pi)T^4(z), \tag{6b}$$

(σ is the Stefan–Boltzmann constant) so that Eq. (5b) can be written in the form

$$\frac{\sigma T^4(z)}{\pi}\sum_{n=1}^N K_n w_n = \tfrac{1}{2}\sum_{n=1}^N K_n\int_{-1}^1 \psi_n(z,\mu')\,d\mu'. \tag{7}$$

We can now eliminate $B_n[T(z)]$ among Eqs. (3), (6), and (7), and obtain ($dx \triangleq \rho(z)K_N dz$; $K_N \le K_i$, $i < N$):

$$\mu\frac{\partial}{\partial x}\Psi(x,\mu) + \mathbf{\Sigma}'\Psi(x,\mu) = \mathbf{C}\int_{-1}^1 \Psi(x,\mu')\,d\mu' + \mathbf{C}'\int \Psi(x,\mu')f(\mathbf{\Omega}'\cdot\mathbf{\Omega})\,d\mathbf{\Omega}', \tag{8}$$

where

$$(\mathbf{\Sigma}')_{ij} = (\sigma_i + S_i/K_N)\,\delta_{ij}, \tag{9a}$$

$$\sigma_i \triangleq K_i/K_N, \qquad (\sigma_N = 1), \tag{9b}$$

$$(\mathbf{C})_{ij} = \sigma_i\sigma_j w_i \Big/ 2\sum_{\alpha=1}^N \sigma_\alpha w_\alpha, \tag{10}$$

and

$$(\mathbf{C}')_{ij} = (S_i/K_N)\,\delta_{ij}. \tag{11}$$

Even for isotropic scattering [$f(\mathbf{\Omega}'\cdot\mathbf{\Omega}) = (4\pi)^{-1}$] we do not know how to obtain explicit solutions to this set of equations [$\det(\mathbf{C}+\mathbf{C}') \neq 0$]. However, if the scattering term may be neglected, we obtain a transport equation of the form

$$\mu\frac{\partial}{\partial x}\Psi(x,\mu) + \mathbf{\Sigma}\Psi(x,\mu) = \mathbf{C}\int_{-1}^1 \Psi(x,\mu')\,d\mu', \tag{12}$$

where

$$(\mathbf{\Sigma})_{ij} = \sigma_i\,\delta_{ij}. \tag{13}$$

We note

$$\det\mathbf{C} = 0 \tag{14a}$$

and, in fact,

$$\det\mathbf{M} = 0, \tag{14b}$$

where $\mathbf{M}$ is any minor of any rank >1 of $\mathbf{C}$. Equation (12) can be solved explicitly by an extension of the technique described in I. The procedure is described in the subsequent sections.

II. THE EIGENVALUES AND EIGENSOLUTIONS

As in I, translational invariance suggests we seek solutions of Eq. (12) in the form[3]

$$\Psi(\eta, x, \mu) = e^{-x/\eta}\mathbf{F}(\eta,\mu). \tag{15}$$

Substituting this ansatz into Eq. (12), we obtain an equation for

$$\mathbf{F}(\eta,\mu),\ (\mathbf{\Sigma} - \mu/\eta\,\mathbf{E})\mathbf{F}(\eta,\mu) = \mathbf{C}\int_{-1}^1 \mathbf{F}(\eta,\mu')\,d\mu', \tag{16}$$

where $\mathbf{E}$ is the unit matrix. First, we consider the continuum solutions, i.e., $\eta \in [-1, 1]$. In I it was necessary to divide this range into two subranges. Here, as we shall see, there are N such subranges

Region 1: $\eta \in [-1/\sigma_1, 1/\sigma_1]$; (17a)

Region 2: $\eta \in [-1/\sigma_2, -1/\sigma_1]$ and $[1/\sigma_1, 1/\sigma_2]$; (17b)

thus, in general,

Region i: $\eta \in [-1/\sigma_i, -1/\sigma_{i-1}]$ and $[1/\sigma_{i-1}, 1/\sigma_i]$, $i > 1$. (17c)

[3] K. M. Case, Ann. Phys. (N. Y.) 9, 1 (1960).

The eigensolutions for the ith-region take the form

$$\mathbf{F}^{(i)}(\eta, \mu) = (\mathbf{P}_i + \mathbf{\Lambda}_i)\mathbf{C} \int_{-1}^{1} \mathbf{F}^{(i)}(\eta, \mu')\, d\mu', \tag{18}$$

where

$$(\mathbf{P}_i)_{jk} = \eta P\left(\frac{1}{\sigma_j\eta - \mu}\right) \delta_{jk} \tag{19a}$$

and

$$(\mathbf{\Lambda}_i)_{jk} = \lambda_j^{(i)}(\eta)\, \delta(\sigma_j\eta - \mu)\, \delta_{ij}. \tag{19b}$$

Here the $\lambda_j^{(i)}(\eta)$ are unspecified functions that must be selected so that Eq. (18) is consistent. We note that for $j < i$ the symbol "P", denoting the Cauchy principal value in Eq. (19a), is superfluous because the denominator can never vanish. Similarly $\lambda_j^{(i)}(\eta)$ may be taken to be zero for $j < i$ since the argument of the delta function never vanishes. Thus, we see that $\mathbf{F}^{(i)}(\eta, \mu)$ contains $(N + 1 - i)$ unknown functions $\lambda_j^{(i)}(\eta)$. In addition, there are N unknown functions $A_j^{(i)}(\eta)$ in Eq. (16), defined by

$$\int_{-1}^{1} \mathbf{F}^{(i)}(\eta, \mu')\, d\mu' = \mathbf{A}^{(i)}(\eta) = \begin{bmatrix} a_1^{(i)}(\eta) \\ \vdots \\ a_N^{(i)}(\eta) \end{bmatrix}. \tag{20}$$

Thus the solutions $\mathbf{F}^{(i)}(\eta, \mu)$ are $(N + 1 - i)$-fold degenerate. There are $(N + 1 - i)$ linearly independent eigensolutions in region i, which are denoted by

$$\mathbf{F}_\alpha^{(i)}(\eta, \mu), \qquad \alpha = i, i + 1, \cdots, N. \tag{21}$$

(For notational convenience we run α from i to N rather than from 1 to $N + i - 1$.)

It is a straightforward matter to obtain the explicit form of the $F_\alpha^{(i)}(\eta, \mu)$. We find [$\tau(x) \triangleq \tanh^{-1}(x)$]

$$\mathbf{F}_i^{(i)}(\eta, \mu) = \begin{bmatrix} (\sigma_1\eta - \mu)^{-1}\eta C_{1\alpha} \\ \vdots \\ (\sigma_{i-1}\eta - \mu)^{-1}\eta C_{i-1,\alpha} \\ P(\sigma_i\eta - \mu)^{-1}\eta C_{i\alpha} - 2\eta C_{i\alpha}\tau(\sigma_i\eta)\delta(\sigma_i\eta - \mu) \\ \vdots \\ P(\sigma_\alpha\eta - \mu)^{-1}\eta C_{\alpha\alpha} + \left[1 - 2\eta C_{\alpha\alpha}\tau(\sigma_\alpha\eta) - 2\eta \sum_{\beta=1}^{i-1} C_{\beta\beta}\tau\left(\frac{1}{\sigma_\beta\eta}\right)\right]\delta(\sigma_\alpha\eta - \mu) \\ P(\sigma_{\alpha+1}\eta - \mu)^{-1}\eta C_{\alpha+1,\alpha} - 2\eta C_{\alpha+1,\alpha}\tau(\sigma_{\alpha+1}\eta - \mu) \\ \vdots \end{bmatrix}. \tag{22}$$

Although the derivation of Eq. (22) is tedious, it is easy to verify that it is a solution. In doing so, the relation

$$C_{j\alpha}C_{\alpha k} = C_{jk}C_{\alpha\alpha} \tag{23}$$

must be kept in mind.

Next, we consider the discrete spectrum, i.e., $\eta \notin [-1, 1]$. Thus from Eq. (16) we obtain

$$\mathbf{F}(\eta, \mu) = \mathbf{DC} \int_{-1}^{1} \mathbf{F}(\eta, \mu')\, d\mu', \tag{24}$$

where

$$(\mathbf{D})_{ij} = (\sigma_i\eta - \mu)^{-1}\eta\, \delta_{ij}. \tag{25}$$

The eigenvalues are obtained by integrating Eq. (24) and noting that nontrivial solutions exist only if

$$\Omega(z) \triangleq \det(\mathbf{E} - \boldsymbol{\tau}\mathbf{C}) = 0, \tag{26}$$

where

$$(\boldsymbol{\tau})_{jk} = 2z\tau\left(\frac{1}{\sigma_j z}\right)\delta_{jk}. \tag{27}$$

To evaluate Eq. (26), we write[4]

$$\det(\lambda\mathbf{E} - \boldsymbol{\tau}\mathbf{C}) = \lambda^N - \mathrm{Tr}\,(\boldsymbol{\tau}\mathbf{C})\lambda^{N-1} + a_2\lambda^{N-2} + \cdots + a_N. \tag{28}$$

Here, the coefficients of λ^{N-k}, $k = 2, 3, \cdots, N$ are defined as the sum of all the k by k minor determinants that can be formed using k of the diagonal elements of $\boldsymbol{\tau}\mathbf{C}$ (there are $N!/k!(N - k)!$ such determinants). One easily verifies that all such minor determinants of $\boldsymbol{\tau}\mathbf{C}$ vanish. Thus, setting $\lambda = 1$ in Eq. (28), we obtain

$$\Omega(z) = 1 - 2z \sum_{\beta=1}^{N} C_{\beta\beta}\tau\left(\frac{1}{\sigma_\beta z}\right). \tag{29}$$

In Appendix A we show that $\Omega(z)$ has only two zeros which, from Eqs. (29) and (10), are $\eta_0 = \pm\infty$. Thus, the discrete eigenvalues are identical with those obtained in I. The discrete eigensolutions are similar; we find

[4] J. H. M. Wedderburn, *Lectures on Matrices* (American Mathematical Society, New York, 1934), Chap. II.

$$\Psi_+(x, \mu) = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \triangleq \Phi_+ \tag{30a}$$

and

$$\Psi_-(x, \mu) = \begin{bmatrix} w_1(x - \mu/\sigma_1) \\ w_2(x - \mu/\sigma_2) \\ \vdots \\ w_N(x - \mu/\sigma_N) \end{bmatrix}. \tag{30b}$$

As in I, we choose to work mostly with certain linear combinations of the $\mathbf{F}_\alpha^{(i)}(\eta, \mu)$ which we call $\mathbf{\Phi}_\alpha^{(i)}(\eta, \mu)$. These are defined by

$$\left\{1 - 2\eta \sum_{\beta=1}^{i-1} C_{\beta\beta\tau}\left(\frac{1}{\sigma_\beta \eta}\right)\right\}\mathbf{\Phi}_\alpha^{(i)}(\eta, \mu) = \frac{1}{C_{i\alpha}}\mathbf{F}_\alpha^{(i)}(\eta, \mu) - \frac{1}{C_{i,\alpha+1}}\mathbf{F}_{\alpha+1}^{(i)}(\eta, \mu), \tag{31a}$$

where $\alpha = i, i + 1, \cdots, N - 1$. Thus

$$\mathbf{\Phi}_\alpha^{(i)}(\eta, \mu) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ (C_{i\alpha})^{-1}\delta(\sigma_\alpha \eta - \mu) \\ (-C_{i,\alpha+1})^{-1}\delta(\sigma_{\alpha+1}\eta - \mu) \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \qquad \alpha = i, \cdots, N - 1. \tag{31b}$$

Also, taking

$$\mathbf{\Phi}_N^{(i)}(\eta, \mu) = \mathbf{F}_N^{(i)}(\eta, \mu), \tag{32}$$

we see that, in each region, there are $(N - i)$ eigensolutions of the form (31b), containing only two nonzero elements, which involve only delta functions and one eigensolution of the form (32).

Aside from the simple form taken by the $\mathbf{\Phi}_\alpha^{(i)}(\eta, \mu)$, we see that these solutions take the same form in different regions; thus, we can recombine the $\mathbf{\Phi}_\alpha^{(i)}(\eta, \mu)$ in the following way. First, for $K < N$,

$$\mathbf{\Phi}_K^{(i)}(\eta, \mu) = \mathbf{\Phi}_K^{(j)}(\eta, \mu), \qquad j = i, i + 1, \cdots, K. \tag{33}$$

This suggests defining[5]

$$\mathbf{\Phi}_K(\eta, \mu) = \begin{cases} \mathbf{\Phi}_K^{(j)}(\eta, \mu), & j \le K < N, \quad (34a) \\ 0, & K < j. \quad (34b) \end{cases}$$

There are $(N - 1)$ eigensolutions of this form; since there are N of the $\mathbf{\Phi}_N^{(i)}(\eta, \mu)$, we are now dealing with only $(2N - 1)$ different eigensolutions instead of $[\frac{1}{2}N(N + 1)]$. If we, in fact, also define

$$\mathbf{\Phi}_N(\eta, \mu) = \sum_{i=1}^{N} \mathbf{\Phi}_N^{(i)}(\eta, \mu) \oplus_i (\eta), \tag{35}$$

where

$$\oplus_i (\eta) = \begin{cases} 1, & \eta \in \text{region } i, \\ 0, & \text{otherwise}, \end{cases} \tag{36}$$

we need only consider N eigensolutions, $\mathbf{\Omega}_1(\eta, \mu)$, $\cdots$, $\mathbf{\Phi}_N(\eta, \mu)$. To recapitulate, we have $(N - 1)$ continuum eigensolutions of the forms

$$\mathbf{\Phi}_1(\eta, \mu) = \begin{bmatrix} (C_{11})^{-1}\delta(\sigma_1\eta - \mu) \\ (-C_{12})^{-1}\delta(\sigma_2\eta - \mu) \\ 0 \\ 0 \\ \vdots \end{bmatrix} \oplus_1 (\eta), \tag{37a}$$

$$\mathbf{\Phi}_2(\eta, \mu) = \begin{bmatrix} 0 \\ (C_{12})^{-1}\delta(\sigma_2\eta - \mu) \\ (-C_{13})^{-1}\delta(\sigma_3\eta - \mu) \\ 0 \\ 0 \\ \vdots \end{bmatrix} \times \{\oplus_1(\eta) + \oplus_2(\eta)\}, \tag{37b}$$

or, in general,

$$\mathbf{\Phi}_i(\eta, \mu) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ (C_{ii})^{-1}\delta(\sigma_i\eta - \mu) \\ (-C_{i,i+1})^{-1}\delta(\sigma_{i+1}\eta - \mu) \\ 0 \\ 0 \\ \vdots \end{bmatrix} \times \sum_{\alpha=1}^{i} \oplus_\alpha (\eta). \tag{37c}$$

[5] This method of attack was suggested by J. Mika (private communication).

In addition, we have one continuum eigensolution [Eq. (35)] as well as two discrete eigenmodes [Eqs. (30)].

III. COMPLETENESS

Theorem I: The functions $\mathbf{\Phi}_i(\eta, \mu)$, $i = 1 \cdots N$, $\eta \geq 0$ and $\mathbf{\Phi}_+$ are complete for arbitrary N-vector functions, $\mathbf{\Psi}(\mu)$, defined on the "half-range," $0 \leq \mu \leq 1$.

This theorem means that an N-component function $\mathbf{\Psi}(\mu)$ can be expanded in the form

$$\mathbf{\Psi}(\mu) = A_+\mathbf{\Phi}_+ + \sum_{\beta=1}^{N} \int_0^{1/\sigma_\beta} \alpha_\beta(\eta)\mathbf{\Phi}_\beta(\eta, \mu)\, d\eta, \qquad \mu \in [0, 1], \tag{38}$$

where A_+ and $\alpha_\beta(\eta)$ are "scalar" functions.

Proof: We proceed as in I, i.e., we attempt an expansion in terms of the continuum modes alone:

$$\mathbf{\Psi}(\mu) = \sum_{\beta=1}^{N-1} \int_0^{1/\sigma_\beta} \alpha_\beta(\eta)\mathbf{\Phi}_\beta(\eta, \mu)\, d\eta + \int_0^1 \alpha_N(\eta)\mathbf{\Phi}_N(\eta, \mu)\, d\eta. \tag{39}$$

Here, the last term has been split off from the sum because the first $(N - 1)$ terms have the simple form given in Eq. (37c). These integrals can all be carried out explicitly,

$$\int_0^{1/\sigma_\beta} \alpha_\beta(\eta)\mathbf{\Phi}_\beta(\eta, \mu)\, d\eta = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ (\sigma_\beta C_{1,\beta})^{-1}\alpha_\beta(\mu/\sigma_\beta) \\ -(\sigma_{\beta+1}C_{1,\beta+1})^{-1}\alpha_\beta(\mu/\sigma_{\beta+1})S(\sigma_{\beta+1}/\sigma_\beta - \mu) \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \qquad \beta = 1, \cdots, N - 1, \tag{40}$$

where $S(X)$ is the unit step function

$$S(X) = 1, \quad X \geq 0, \qquad = 0, \quad \text{otherwise}. \tag{41}$$

When these results are substituted into Eq. (39), we obtain

$$\begin{aligned}\psi_i(\mu) = {} & -(C_{1,i}\sigma_i)^{-1}\alpha_{i-1}(\mu/\sigma_i)S(\sigma_i/\sigma_{i-1} - \mu) \\ & + (C_{1,i}\sigma_i)^{-1}\alpha_i(\mu/\sigma_i) \\ & - 2\mu C_{iN}\alpha_N(\mu/\sigma_i)\tau(\mu)\sigma_i^{-2} \\ & + C_{iN}P\int_0^1 \frac{\alpha_N(\eta)\eta\, d\eta}{\sigma_i\eta - \mu}, \\ & i = 1, 2, \cdots, N - 1, \end{aligned} \tag{42a}$$

and

$$\begin{aligned}\psi_N(\mu) = {} & -C_{1N}^{-1}\alpha_{N-1}(\mu)S(\sigma_{N-1}^{-1} - \mu) \\ & + C_{NN}P\int_0^1 \frac{\alpha_N(\eta)\eta\, d\eta}{\eta - \mu} \\ & + \alpha_N(\mu)[1 - 2\mu C_{NN}\tau(\mu)] \\ & - 2\mu\sum_{\beta=1}^{N-1} C_{\beta\beta}\tau\left(\frac{1}{\sigma_\beta\mu}\right)S(\mu - \sigma_\beta^{-1}). \end{aligned} \tag{42b}$$

Here $\psi_i(\mu)$ is the ith component of $\mathbf{\Psi}(\eta)$. [We recall in obtaining Eq. (42b) that $\sigma_N = 1$.]

We note from Eqs. (42) that the unknowns $\alpha_i(\eta)$, $i = 1, \cdots, N - 1$, can be eliminated successively starting with $i = 1(\alpha_0(\eta) = 0)$. In this way, Eqs. (42) can be converted into a singular integral equation for $\alpha_N(\eta)$. To carry this out, we make the change of variable in Eq. (42a),

$$\mu/\sigma_i \to \mu \tag{43}$$

and multiply the equation by σ_i^2. Then we add all N equations [i.e., including (42b)] to obtain the simple result

$$\begin{aligned}&\sum_{i=1}^{N} \sigma_i^2\psi_i(\sigma_i\mu)S(1/\sigma_i - \mu) \\ &= \sum_{i=1}^{N} C_{i,i}S(1/\sigma_i - \mu)P\int_0^1 \frac{\alpha_N(\eta)\eta\, d\eta}{\eta - \mu} + \alpha_N(\mu) \\ &\quad \times \left\{1 - 2\mu\sum_{i=1}^{N} C_{i,i}[\tau(1/\sigma_i\mu)\,\delta(\mu - 1/\sigma_i) \right. \\ &\quad \left. + \tau(\sigma_i\mu)S(1/\sigma_i - \mu)]\right\}. \end{aligned} \tag{44}$$

The various step functions were introduced by the variable change, since we must require the argument of $\psi_i(\eta)$, for example, to be less than or equal to unity; thus, under (43),

$$\psi_i(\mu) \to \psi_i(\sigma_i\mu)\{\oplus_1(\mu) + \cdots + \oplus_i(\mu)\}. \tag{45}$$

We note

$$\sum_{\alpha=1}^{i} \oplus_\alpha(\mu) = S\left(\frac{1}{\sigma_i} - \mu\right). \tag{46}$$

Equation (44) is now in canonical form, since, we

note that, from the definition of $\Omega(z)$ in Eq. (29),

$$\Omega^{\pm}(\mu) = 1 - 2\mu \sum_{\beta=1}^{N} C_{\beta\beta}\tau(1/\sigma_\beta\mu)S(\mu - 1/\sigma_\beta) - 2\mu \sum_{\beta=1}^{N} C_{\beta\beta}[\tau(\sigma_\beta\mu) \mp \tfrac{1}{2}\pi i]S\left(\frac{1}{\sigma_\beta} - \mu\right). \quad (47)$$

[$\Omega(z)$ is analytic in the complex z plane cut from -1 to $+1$ along the real line; $\Omega^{\pm}$ represents the boundary values above and below the branch cut.] Thus, Eq. (44) can be written as

$$\sum_{i=1}^{N} \sigma_i^2\psi_i(\sigma_i\mu)S\left(\frac{1}{\sigma_i} - \mu\right) = \frac{\Omega^+(\mu) - \Omega^-(\mu)}{2\pi i\mu} P\int_0^1 (\eta - \mu)^{-1}\alpha_N(\eta)\, d\eta + \tfrac{1}{2}(\Omega^+(\mu) - \Omega^-(\mu))\alpha_N(\mu). \quad (48)$$

This equation is in standard form, (c.f, I) and its solution is well known in terms of the X-function, which is analytic in the complex plane cut from 0 to 1 along the real line.

$$X(z) = (1 - z)^{-1} \times \exp\{\pi^{-1}\int_0^1 \arg\,\Omega^+(\mu)(\mu - z)^{-1}\, d\mu\}. \quad (49)$$

Once $\alpha_N(\eta)$ is known, the other $\alpha_i(\eta)$ may all be found from Eqs. (42). Also, the discrete mode is introduced, as in I, by the condition at infinity on the auxiliary function

$$N(z) = (2\pi i)^{-1}\int_0^1 (\eta - z)^{-1}\alpha_N(\eta)\eta\, d\eta. \quad (50)$$

Since the details are identical to those in I, we go no further than to note that the completeness theorem is hereby proved. The coefficients in the expansion, Eq. (38), can be found in principal from the above solution; but it is simpler to use the orthogonality relations derived in the next section.

IV. ORTHOGONALITY

Theorem II: The continuum functions $\Phi(\eta, \mu)$, $\eta \in [0, 1]$ and the discrete mode Φ_+ are orthogonal to the adjoint eigensolutions $\Phi^\dagger(\eta, \mu)$, $\eta \in [0, 1]$ and $\Phi_+^\dagger$ on the range $0 \leq \mu \leq 1$ with weight function $\mathbf{W}(\mu)$, where

$$[\mathbf{W}(\mu)]_{ij} = \sigma_i\gamma(\mu/\sigma_i)\delta_{ij}, \quad (51)$$

and $\gamma(\mu)$ is defined as

$$\gamma(\mu) = \mu X^+(\mu)/\Omega^+(\mu). \quad (52)$$

The proof of this theorem is a generalization of the one given in I for $N = 2$. There, we prove orthogonality of the $\mathbf{F}_\alpha^\dagger(\eta, \mu)$ rather than the $\Phi_K(\eta, \mu)$. However, this makes no difference, since the theorem states that[6] (the over tilde denotes transpose)

$$\int_0^1 \tilde{\Phi}^\dagger(\eta', \mu)\mathbf{W}(\mu)\Phi(\eta, \mu)\, d\mu = 0, \qquad \eta \neq \eta', \quad (53a)$$

and we *prove* that

$$\int_0^1 \tilde{\mathbf{F}}^\dagger(\eta', \mu)\mathbf{W}(\mu)\mathbf{F}(\eta, \mu)\, d\mu = 0, \qquad \eta = \eta'. \quad (53b)$$

Clearly, (53a) and (53b) imply each other. $\mathbf{F}^\dagger(\eta, \mu)$, obeying the adjoint equation given below, may be obtained from $\mathbf{F}(\eta, \mu)$ under the interchange $C_{ij} \Leftrightarrow C_{ji}$; $\Phi^\dagger(\eta, \mu)$ and $\Phi(\eta, \mu)$ are related in the same way. As in I, we note that care must be taken when this interchange is made. For example, Φ_+ is given by Eq. (30a). However, this form is a reduction of the form obtained if Eq. (16) with $\eta_0 = \pm\infty$ is solved directly; this is to say that C_{ij} no longer appear. One finds easily, however,

$$\Phi_+^\dagger = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (54)$$

Equations (22) for $\mathbf{F}_\alpha^\dagger(\eta, \mu)$,and Eqs. (31b) and (32) for $\Phi_\alpha(\eta, \mu)$ are in the form so that the interchange $C_{ij} \Leftrightarrow C_{ji}$ gives the adjoint solutions.

Proof: We turn now to the proof of Theorem II, i.e., Eq. (53b). We begin, as in I, with the equation obeyed by $\mathbf{F}(\eta, \mu)$ and the adjoint equation obeyed by $\mathbf{F}^\dagger(\eta', \mu)$. [To simplify notation, we allow $\mathbf{F}(\eta, \mu)$ and $\mathbf{F}^\dagger(\eta', \mu)$ to denote either discrete or continuum modes.] Thus,

$$\left(\frac{\mathbf{\Sigma}}{\mu} - \frac{\mathbf{C}}{\mu}\int_{-1}^1 d\mu'\right)\mathbf{F}(\eta, \mu) = \frac{1}{\eta}\mathbf{F}(\eta, \mu) \quad (55a)$$

and

$$\left(\frac{\mathbf{\Sigma}}{\mu} - \frac{\tilde{\mathbf{C}}}{\mu}\int_{-1}^1 d\mu'\right)\mathbf{F}^+(\eta', \mu) = \frac{1}{\eta}\mathbf{F}^\dagger(\eta', \mu). \quad (55b)$$

The method of proof is to multiply Eq. (55a) from the left by $\tilde{\mathbf{F}}^+(\eta', \mu)\mathbf{W}(\mu)$ and to multiply the transpose of Eq. (55b) from the right by $\mathbf{W}(\mu)\mathbf{F}(\eta, \mu)$, integrate both over μ from 0 to 1 and subtract. The right-hand side of the resulting equation becomes simply

$$\left(\frac{1}{\eta'} - \frac{1}{\eta}\right)\int_0^1 \tilde{\mathbf{F}}^\dagger(\eta', \mu)\mathbf{W}(\mu)\mathbf{F}(\eta, \mu)\, d\mu,$$

[6] The function Ψ_- is not included in the orthogonal set because it is not a solution of Eqs. (55) [it satisfies Eq. (12)].

and thus we wish to prove that the left-hand side vanishes if $\eta \neq \eta'$. By following a procedure identical to that outlined in I the proof can be reduced to showing that a quantity J_2 is a constant, where

$$J_2 = X(0) - \omega(\eta)\eta^{-1}\gamma(\eta) + \sum_{\alpha=1}^{N} C_{\alpha\alpha}\lambda^{\alpha}(\eta)\eta^{-1}\gamma(\eta), \quad (56)$$

with

$$\omega(\eta) = \tfrac{1}{2}[\Omega^{+}(\eta) - \Omega^{-}(\eta)]. \quad (57)$$

The proof of the theorem hinges on proving that J_2 is independent of η and symmetric in i, j. We showed this in I by direct substitution of the $\lambda^{\alpha}(\eta)$ (see note added in proof). Here, we prove it in general. We should clarify what is meant by the symbol $\lambda^{\alpha}(\eta)$. We note from Eq. (16) that any of the components $f_i(\eta, \mu)$ of $\mathbf{F}(\eta, \mu)$ can be written in the form

$$f_{\alpha}(\eta, \mu) = \{P\eta(\sigma_{\alpha}\eta - \mu)^{-1} + \lambda^{\alpha}(\eta)\delta(\sigma_{\alpha}\eta - \mu)\} \times \sum_{\beta=1}^{N} C_{\alpha\beta}a_{\beta}(\eta), \qquad \sigma_{\alpha}\eta \leq 1, \quad (58a)$$

or

$$f_i(\eta, \mu) = \eta(\sigma_i\eta - \mu)^{-1} \times \sum_{\beta=1}^{N} C_{i\beta}a_{\beta}(\eta), \qquad \sigma_i\eta \geq 1, \quad (58b)$$

where

$$a_{\beta}(\eta) = \int_{-1}^{1} f_{\beta}(\eta, \mu')\, d\mu'. \quad (59)$$

[Thus the $\lambda^{\alpha}(\eta)$ are the unknown coefficients in Eq. (19b) which one must find for explicit evaluation of the eigensolutions; thus they should be denoted by $\lambda_{ik}^{\alpha}(\eta)$ in order to distinguish the region and the degeneracy. However, we do not have to use their explicit form. The symbol $\lambda^{\alpha}(\eta)$ thus denotes *any* of the $\lambda_{ik}^{\alpha}(\eta)$.] Multiply Eq. (58a) by $C_{N\alpha}$ and sum over all α for which $\sigma_{\alpha}n \leq 1$; then multiply Eq. (58b) by C_{Ni} and sum over all j for which $\sigma_i n > 1$. We integrate the resulting two equations over μ from -1 to 1 and add them to obtain

$$\sum_{\alpha}^{s} C_{N\alpha}a_{\alpha}(\eta) + \sum_{i}^{ns} C_{Ni}a_i(\eta) = \sum_{\alpha}^{s} C_{N\alpha}[2\eta\tau(\sigma_{\alpha}\eta) + \lambda^{\alpha}(\eta)] \times \sum_{\beta=1}^{N} C_{\alpha\beta}a_{\beta}(\eta) + \sum_{i}^{ns} C_{Ni}2\eta\tau\left(\frac{1}{\sigma_i\eta}\right) \sum_{\beta=1}^{N} C_{i\beta}a_{\beta}(\eta). \quad (60)$$

Here the superscripts s and ns indicate that the sums are to be taken only over the singular [Eq. (58a)] and the nonsingular [Eq. (58b)] sets respectively. Using $C_{\alpha\beta}C_{N\alpha} = C_{\alpha\alpha}C_{N\beta}$ in the right-hand side of Eq. (60), the sums separate and a factor $\sum_{\beta=1}^{N} C_{N\beta}A_{\beta}(\eta)$ can therefore be canceled to give

$$1 - 2\eta \sum_{\alpha}^{s} C_{\alpha\alpha}\tau(\sigma_{\alpha}\eta) - \sum_{i}^{ns} 2\eta C_{ii}\tau\left(\frac{1}{\sigma_i\eta}\right) = \sum_{\alpha}^{s} C_{\alpha\alpha}\lambda^{\alpha}(\eta). \quad (61)$$

The left-hand side of Eq. (61) is exactly $\omega(\eta)$, thus

$$\sum_{\alpha}^{s} C_{\alpha\alpha}\lambda^{\alpha}(\eta) = \omega(\eta). \quad (62)$$

Substituting this result into Eq. (56) [remembering that $\lambda^{\alpha}(\eta)$ appears only when $f_{\alpha}(\eta, \mu)$ is singular] we find

$$J_2 \equiv 1, \quad (63)$$

the theorem is thus proved.

V. NORMALIZATION

The results of the previous two sections can be used to expand functions $\Psi(\mu)$ for $\mu \in [0, 1]$ and to obtain the expansion coefficients if

(i) the normalization integrals are known, and
(ii) the degenerate eigenfunctions are orthogonalized.

As in I, we introduce a new set of functions $\mathbf{x}_K(\eta, \mu)$, constructed so as to be orthogonal to all of the $\mathbf{\Phi}_K(\eta, \mu)$. We define our scalar product as

$$(\mathbf{U}, \mathbf{V}) \triangleq \int_0^1 \tilde{\mathbf{U}}^{\dagger}(\eta', \mu)\mathbf{W}(\mu)\mathbf{V}(\eta, \mu)\, d\mu. \quad (64)$$

Abbreviating

$$(\mathbf{\Phi}_i, \mathbf{\Phi}_j) = (i, j)\delta(\eta - \eta'), \quad (65)$$

we easily calculate

$$(i, j) = -\gamma(\eta)[(C_{1,i+1}C_{i+1,1})^{-1}\delta_{i+1}^{i} + (C_{1i}C_{i1})^{-1}\delta_{i-1}^{i} - [(C_{1i}C_{i1})^{-1} + (C_{1,i+1}C_{i+1,1})^{-1}]\delta_i^i\}, \quad i, j < N. \quad (66)$$

Also, keeping Theorem II in mind, we have

$$(\mathbf{\Phi}_i, \mathbf{\Phi}_N) = \sum_{\alpha=1}^{i} (\mathbf{\Phi}_i, \mathbf{\Phi}_N^{(\alpha)}). \quad (67)$$

The product $(\mathbf{\Phi}_i, \mathbf{\Phi}_N^{(\alpha)})$ is also easily calculated, and is seen to be independent of α. Thus the sum in Eq. (67) merely introduces a factor i, and we find

$$(i, N) = i\frac{C_{iN}}{C_{i1}}2\eta\gamma(\eta)\{\tau(\sigma_{i+1}\eta) - \tau(\sigma_i\eta)\}, \quad i < N - 1, \quad (68a)$$

$$(N, i) = (i, N)\frac{C_{Ni}}{C_{1i}}\frac{C_{i1}}{C_{iN}}, \quad i < N-1, \tag{68b}$$

$$= w_N/w_i(i, N), \quad i < N-1. \tag{68c}$$

Also,

$$(N-1, N) = -\gamma(\eta)\Big[(N-1)\frac{C_{N-1,N}}{C_{N-1,1}}2\eta\tau(\sigma_{N-1}\eta) + \frac{1}{C_{N1}}\sum_{\alpha=1}^{N-1}\Big(1 - 2\eta\sum_{\beta=1}^{\alpha-1}C_{\beta\beta}\tau(1/\sigma_\beta\eta) - 2\eta C_{NN}T(\eta)\Big)\Big]. \tag{69}$$

$(N, N-1)$ may be found from $(N-1, N)$ under the interchange $C_{ij} \to C_{ji}$. Finally we need

$$(\Phi_N, \Phi_N) = \sum_{i=1}^{N}(\Phi_N^{(i)}, \Phi_N^{(i)}), \tag{70}$$

where $(\Phi_N^{(i)}, \Phi_N^{(i)})$ may be calculated from the explicit form, Eq. (22), by making use of the X-function identities of Appendix B. We quote the result,

$$(N, N) = \sum_{i=1}^{N}\Big\{\sum_{\alpha=1}^{N}C_{NN}4\eta^2C_{\alpha\alpha}\tau^2(\sigma_\alpha\eta) + \Big[1 - 2\eta\sum_{\beta=1}^{i-1}C_{\beta\beta}\tau(1/\sigma_\beta\eta) - 2\eta C_{NN}\tau(\eta)\Big]^2 + \sum_{\alpha=i}^{N}C_{NN}C_{\alpha\alpha}\pi^2\eta^2\Big\}\gamma(\eta). \tag{71}$$

Equations (66), (68), (69), and (71) give all the necessary normalization integrals for the construction of the $\mathbf{x}_K$. We write, in general,

$$\mathbf{x}_i(\eta, \mu) = \sum_{\alpha=1}^{N}\mathrm{N}_\alpha^{(i)}\Phi_\alpha(\eta, \mu), \tag{72}$$

where $\mathrm{N}_\alpha^{(i)}$ are to be choosen such that

$$(\mathbf{x}_i, \Phi_j) = \delta_{ij}\,\delta(\eta - \eta'). \tag{73}$$

The $\mathrm{N}_\alpha^{(i)}$ are readily found from Eqs. (72) and (73). We take the scalar product of Eq. (72) from the right successively with the $\Phi_\beta(\eta, \mu)$, $\beta = 1, \cdots N$. This yields the following equations for the $\mathrm{N}_\alpha^{(i)}$:

$$\begin{aligned}
0 &= \mathrm{N}_1^{(i)}(1,1) + \mathrm{N}_2^{(i)}(2,1) + \mathrm{N}_3^{(i)}(3,1) + 0 + \cdots + 0 + \mathrm{N}_N^{(i)}(N,1),\\
0 &= \mathrm{N}_1^{(i)}(1,2) + \mathrm{N}_2^{(i)}(2,2) + \mathrm{N}_3^{(i)}(3,2) + 0 + \cdots + 0 + \mathrm{N}_N^{(i)}(N,2),\\
0 &= 0 + \mathrm{N}_2^{(i)}(2,3) + \mathrm{N}_3^{(i)}(3,3) + \mathrm{N}_4^{(i)}(4,3) + 0 + \cdots + 0 + \mathrm{N}_N^{(i)}(N,3),\\
&\vdots\\
1 &= 0 + \cdots \mathrm{N}_{i-1}^{(i)}(i-1,i) + \mathrm{N}_i^{(i)}(i,i) + \mathrm{N}_{i+1}^{(i)}(i+1,i) + 0 + \cdots \mathrm{N}_N^{(i)}(N,i),\\
&\vdots\\
0 &= \mathrm{N}_1^{(i)}(1,N) + \mathrm{N}_2^{(i)}(2,N) + \cdots \mathrm{N}_N^{(i)}(N,N).
\end{aligned} \tag{74}$$

We see that the first and the $(N-1)$th equations have only three nonvanishing coefficients, all the rest have four such coefficients except the Nth, which has N. The set of equations (74) is easily solved for the $\mathrm{N}_\alpha^{(i)}$.

We note that we have set $(\mathbf{x}_K, \Phi_K) = \delta(\eta - \eta')$. Also $(\mathbf{x}_K, \Phi_+) = 0$. The discrete (asymptotic) coefficient is found from the relation

$$(\Phi_+, \Phi_+) = \sum_{i=1}^{N}\int_0^1 w_i\sigma_i\gamma(\mu/\sigma_i)\,d\mu. \tag{75}$$

Changing variables, this can be written

$$(\Phi_+, \Phi_+) = \sum_{i=1}^{N}\int_0^1 w_i\sigma_i^2\int_0^{1/\sigma_i}\gamma(\mu)\,d\mu, = 2\sum_{K=1}^{N}w_K\sigma_K\sum_{i=1}^{N}C_{ii}\int_0^{1/\sigma_i}\gamma(\mu)\,d\mu. \tag{76}$$

This integral can be evaluated from Eq. (B1) (in Appendix B) in the limit $z \to \infty$,

$$-\lim_{z\to\infty} Z\,X\,(z) = \sum_{i=1}^{N}C_{ii}\int_0^{1/\sigma_i}\gamma(\mu)\,d\mu. \tag{77}$$

But from Eq. (49), we see that the limit is -1. Thus

$$(\Phi_+, \Phi_+) = 2\sum_{i=1}^{N}\sigma_i w_i. \tag{78}$$

In applying this result to obtain expansion coefficients, one might have an expansion of the form

$$\Psi(\mu) = A_+\Phi_+ + \sum_{i=1}^{N}\int_0^{1/\sigma_i}\alpha_i(\eta)\Phi_i(\eta, \mu)\,d\eta, \quad \mu \in [0, 1]. \tag{79}$$

Then, from Eq. (78), we find

$$A_+ = (\Phi_+, \Psi(\mu))\Big(2\sum_{i=1}^{N}\sigma_i w_i\Big)^{-1}, \tag{80}$$

while

$$\alpha_i(\eta) = (\mathbf{x}_i, \Psi(\mu)). \tag{81}$$

VI. THE MILNE PROBLEM

We seek the angular density, $\Psi_M(x, \mu)$, in the source-free half-space under the boundary conditions:

(a) $\Psi_M(x, \mu = 0, \mu \geq 0$ (zero re-entrant radiation),
(b) $\Psi_M(x, \mu) \sim \Psi_-(x, \mu)$ (for large x).

The second condition specifies that $\Psi_M(x, \mu)$ diverges no more rapidly than the slowest diverging mode $\Psi_-(x, \mu)$.

The solution can be constructed from the normal modes of the transport equation. Condition (b) requires that no $\Psi(\eta, x, \mu)$ be included for $\eta \in [-1, 0]$. Thus, we write

$$\Psi_M(x, \mu) = A_-\Psi_-(x, \mu) + A_+\Phi_+ + \sum_{i=1}^{N} \int_0^{1/\sigma_i} \alpha_i(\eta) e^{-x/\eta} \Phi_i(\eta, \mu)\, d\eta. \quad (82)$$

The coefficient A_- we leave arbitrary (it depends upon the normalization). The other coefficients are obtained from condition (a). Setting $x = 0$ in Eq. (82), we have

$$-A_-\Psi_-(0, \mu) = A_+\Phi_+ + \sum_{i=1}^{N} \int_0^{1/\sigma_i} \alpha_i(\eta)\Phi_i(\eta, \mu)\, d\eta, \qquad \mu \in [0, 1]. \quad (83)$$

Thus, the coefficients are just the half-range expansion coefficients for the function

$$-A_-\Psi_-(0, \mu) = A_- \begin{bmatrix} w_1/\sigma_1 \\ w_2/\sigma_2 \\ \vdots \\ w_N \end{bmatrix} \mu. \quad (84)$$

They are found immediately from the orthogonality relations once the $\mathbf{X}_i(\eta, \mu)$ are constructed. For the asymptotic solution (i.e., the part of Ψ_M involving Ψ_- and Φ_+), we have

$$\frac{A_+}{A_-} = \frac{-\int_0^1 \tilde{\Phi}_+^\dagger \mathbf{W}(\mu)\Psi_-(0, \mu)\, d\mu}{2\sum_{i=1}^{N} \sigma_i w_i}, \quad (85)$$

where the normalization integral, Eq. (78), is used. Expanding Eq. (85), we obtain

$$\frac{A_+}{A_-} = \frac{\sum_{i=1}^{N} w_i \int_0^1 \gamma(\mu/\sigma_i)\mu\, d\mu}{2\sum_{i=1}^{N} \sigma_i w_i}. \quad (86)$$

Changing variables and noting Eq. (10), (86) becomes

$$\frac{A_+}{A_-} = \sum_{i=1}^{N} C_{i1} \int_0^{1/\sigma_i} \gamma(\mu)\mu\, d\mu. \quad (87)$$

This expression can be put in terms of the X-function by use of Identity IV, Appendix B. We find

$$\frac{A_+}{A_-} = \frac{3}{2} \sum_{i=1}^{N} C_{i1} \int_0^{1/\sigma_i} \frac{\mu^2}{X(-\mu)}\, d\mu. \quad (88)$$

The continuum expansion coefficients can be found in just the same manner. [However, since, in general, one must solve the set of equations (74) and then use Eq. (72) to construct the $\mathbf{x}_i$'s we merely formally indicate the solution.]

$$\alpha_i(\eta)/A_- = -[\mathbf{x}_i, \Psi_-(0, \mu)]. \quad (89)$$

The customary normalization[1] is to set

$$-2\pi \int_{-1}^{1} \mu\, d\mu \int_0^\infty d\nu\, \psi_\nu(x, \mu) = \sigma T_e^4, \quad (90)$$

where T_e is the "effective temperature" and σ is the Stefan–Boltzmann constant. Equation (90) can be written as

$$\frac{-\sigma T_e^4}{2\pi} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}^\sim \int_{-1}^{1} \mu\Psi_M(x, \mu) \quad (91a)$$

or

$$\frac{-\sigma T_e^4}{2\pi} = \int_{-1}^{1} \mu\tilde{\Phi}_+^\dagger \Psi_M(x, \mu)\, d\mu. \quad (91b)$$

Evaluating Eq. (91b), we find

$$A_- = 3\sigma T_e^4 \left(4\pi \sum_{i=1}^{N} \frac{w_i}{\sigma_i}\right)^{-1}. \quad (92)$$

The expansion coefficients are now found (in principle) to solve the problem. We have

$$\Psi_M(x, \mu) = \left[3\sigma T_e^4 \left(4\pi \sum_{i=1}^{N} \frac{w_i}{\sigma_i}\right)^{-1}\right]\left[\Psi_-(x, \mu) + \frac{A_+}{A_-} \times \Phi_+ + \sum_{i=1}^{N} \int_0^{1/\sigma_i} \frac{\alpha_i(\eta)}{A_-} e^{-x/\eta}\Phi_i(\eta, \mu)\, d\eta\right]. \quad (93)$$

The energy density,

$$\mathcal{F}(x) \triangleq 2\pi \int_{-1}^{1} d\mu \int_0^\infty d\nu\, \psi_\nu(x, \mu) \quad (94)$$

is given by

$$\mathcal{F}(x) = 2\pi \int_{-1}^{1} d\mu\, \tilde{\Phi}_+^\dagger \Psi_M(x, \mu). \quad (95)$$

The extrapolated endpoint is defined in terms of the quantity

$$\mathfrak{F}_{asymp}(x) = 2\pi \int_{-1}^{1} d\mu\, \tilde{\Phi}_+^{\dagger} \Psi_{Masymp}(x, \mu). \quad (96)$$

This, in our model, becomes

$$\mathfrak{F}_{asymp}(x) = \left[3\sigma T_e^4 \left(\sum_{i=1}^{N} \frac{w_i}{\sigma_i}\right)^{-1}\right] \times \left[\sum_{i=1}^{N} w_i \left(x + \frac{A_+}{A_-}\right)\right]. \quad (97)$$

Thus the asymptotic energy density extrapolates to zero at $x = -x_0$, where

$$x_0 = A_+/A_- \quad (98)$$

which is already given by Eq. (88).

The temperature distribution in this model is given by

$$\frac{\sigma T^4(x)}{\pi} = \left(2\sum_{i=1}^{N} \sigma_i w_i\right)^{-1} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ 1 \end{bmatrix}^{\sim} \times \int_{-1}^{1} \Psi_M(x, \mu)\, d\mu. \quad (99)$$

The asymptotic temperature distribution is easily obtained as

$$\frac{T^4_{asymp}(x)}{T_*^4} = \left[\frac{3}{4}\left(\sum_{i=1}^{N} \frac{w_i}{\sigma_i}\right)^{-1}\right][x + x_0]. \quad (100)$$

Just as in I, the law of darkening [for the integrated quantities, $\Psi_M(0, \mu)$, $\mu < 0$] can be obtained. The fact that we restrict μ to be negative enables us to determine Ψ_M (for $x = 0$) explicitly without actually knowing any of the continuum expansion coefficients except $\alpha_N(\eta)$. The coefficient $\alpha_N(\eta)$ is expressed in terms of the N-function which then permits the evaluation of integrals involved. The procedure follows exactly as in I. We simply state the result,

$$\Psi_M(0, -\mu) = \left[3\sigma T_e^4 \left(4\pi \sum_{i=1}^{N} \frac{w_i}{\sigma_i}\right)^{-1}\right] \begin{bmatrix} \frac{w_1}{X(-\mu/\sigma_1)} \\ \frac{w_2}{X(-\mu/\sigma_2)} \\ \vdots \\ \frac{w_N}{X(-\mu)} \end{bmatrix}, \quad \mu \in [0, 1]. \quad (101)$$

It is clear how other half-range problems could be solved. For example, consider the albedo problem. Here we have a source-free half-space with incident distribution

$$\Psi_{inc}(\mu) = \begin{bmatrix} \delta(\mu - \mu_1) \\ \delta(\mu - \mu_2) \\ \vdots \\ \delta(\mu - \mu_N) \end{bmatrix}, \quad \mu_i, \mu \geq 0. \quad (102)$$

Here, the solution must not diverge at infinity, so we set

$$\Psi_a(x, \mu) = A_+\Phi_+ + \sum_{i=1}^{N} \int_0^{1/\sigma_i} \alpha_i(\eta) e^{-x/\eta} \Phi_i(\eta, \mu)\, d\eta. \quad (103)$$

Since

$$\Psi_a(0, \mu) = \Psi_{inc}(\mu), \quad \mu \geq 0, \quad (104)$$

the expansion coefficients are found as integrals of the adjoint functions times delta functions. As in the Milne problem, the determination of the solution is quite trivial once the set of x-functions has been constructed.

The construction of the half-space Green's function requires a special technique, this is discussed in I. The procedure here for the case of general N follows in exactly the same manner.

Note added in proof: In I we "proved" $J_2 = 1$. Actually, $J_2 = X(0) \neq 1$. However, since $X(0) =$ const, symmetric in i, j (cf. Identity II, Appendix B), the proof is still valid.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Professor Feza Gürsey for the aid given to the development of the dispersion relation in this paper. Both authors wish to thank the Middle East Technical University as well as the Çekmece Nuclear Center in Istanbul for their hospitality and one of the authors (P.F.Z.) is grateful to the Ford Foundation for partial financial support.

APPENDIX A. THE NUMBER OF DISCRETE EIGENVALUES

The discrete eigenvalues, $\eta_{i\pm}$, are defined as the zeros of the dispersion function $\Omega(z)$, Eq. (29). Here we verify that there are only two zeros for any N.

Since $\Omega(z)$ is analytic in the cut plane and vanishes at infinity, the number of zeros is $(2\pi)^{-1}$ times the change in the argument of $\Omega(z)$ as a contour en-

circling the cut is traversed.[7] Because $\Omega^+(\mu) = [\Omega^-(\mu)]^*$, and $\Omega(z) = \Omega(-z)$, the change in argument is four times the change in going from $0 + i\epsilon$ to $1 + i\epsilon$. Call this change $\Delta_+(0, 1)$. We have

$$\Delta_+(0, 1) = \sum_{i=0}^{N-1} \Delta_+\left(\frac{1}{\sigma_i}, \frac{1}{\sigma_{i+1}}\right), \tag{A1}$$

where we define $1/\sigma_0 = 0$.

From Eq. (47), we can write for the boundary value in region i

$$\Omega_i^+(\mu) = 1 - 2\mu \sum_{\beta=1}^{i-1} C_{\beta\beta} T\left(\frac{1}{\sigma_\beta \mu}\right) - 2\mu \sum_{\beta=\lambda}^{N} C_{\beta\beta}\left\{T(\sigma_\beta \mu) \pm \frac{\pi i}{2}\right\}. \tag{A2}$$

From (A2) it is easily verified that

$$\Delta_+(0, 1/\sigma_1) = \pi, \tag{A3}$$

$$\Delta_+\left(\frac{1}{\sigma_i}, \frac{1}{\sigma_{i+1}}\right) = 0, \qquad i = 1, 2, \cdots N - 1. \tag{A4}$$

Thus

$$\Delta_+(0, 1) = \pi, \tag{A5}$$

and the total change (for the encircled cut) is 4π. Thus $\Omega(z)$ has two zeros.

[7] R. V. Churchill, *Complex Variables and Applications* (McGraw-Hill Book Company, Inc., New York, 1960), Chap. 12.

APPENDIX B. X-FUNCTION IDENTITIES

The derivations of the X-function identities are trivial generalizations of the corresponding derivations in I (Appendix A), so we present them without proof.

Identity I:

$$X(z) = \sum_{i=1}^{N} C_{ii} \int_0^{1/\sigma_i} \frac{\gamma(\mu)\, d\mu}{\mu - z}. \tag{B1}$$

Identity II:

$$X(z)X(-z) = \left[\frac{3}{2}\left(\sum_{i=1}^{N} \frac{C_{ii}}{\sigma_i^3}\right)^{-1}\right]\Omega(z). \tag{B2}$$

By combining Identities I and II we get a nonlinear nonsingular integral equation for the numerical evaluation of $X(z)$. Thus we find

Identity III:

$$X(z) = \left[\frac{3}{2}\left(\sum_{i=1}^{N} \frac{C_{ii}}{\sigma_i^3}\right)^{-1}\right] \times \sum_{\alpha=1}^{N} C_{\alpha\alpha} \int_0^{1/\sigma_\alpha} \frac{\mu}{X(-\mu)} \frac{d\mu}{\mu - z}. \tag{B3}$$

Furthermore, Identity IV is the trivial result obtained by taking boundary values of Eq. (B2).

$$\gamma(\mu) = \left[\frac{3}{2}\left(\sum_{i=1}^{N} \frac{C_{ii}}{\sigma_i^3}\right)^{-1}\right] \frac{\mu}{X(-\mu)}. \tag{B4}$$

A Theorem on Peratization of Singular Potentials and Other Miscellanea

RICHARD M. SPECTOR
Physics Department, Wayne State University, Detroit, Michigan
(Received 3 January 1966)

It is proved for potentials of the form $V(r) = A(r)/r^n$ $(n > 3)$ [with $A(r)$ more singular at $r = 0$ than $r(dA/dr)$, or apporaching zero less rapidly] that the peratized scattering length can be written down immediately in known form. The distrinction between regularization and peratization is made clear, and a common conjecture about the g (coupling constant) behavior of the peratized scattering length is disproved by a counter example. Finally, various mistakes in the literature are corrected.

I. INTRODUCTION

IN the last few years, extensive effort has gone into the study of singular potential scattering.[1-7] Stimulated by a prescription for handling divergent field theories (called peratization, developed by Feinberg and Pais[8,9]) many workers have undertaken to study a similar process for the Schrödinger equation. The results in the literature are sometimes inaccurate and even contradictory. In this paper, we clarify some of these past results and present a theorem which enables one to peratize a certain general class of potentials to be specified later.

Briefly, peratization involves replacing a singular potential, $V(r)$, by a potential $U(r, \alpha)$ such that $\lim_{\alpha\to 0} U(r, \alpha) = V(r)$ and where $U(0, \alpha)$ exists. The Born series for the scattering length for $U(r, \alpha)$ is then calculated. Obviously every term diverges as $\alpha \to 0$. The leading singular term in each order in g (the coupling constant) is then retained and this sum is expressed in closed form, if possible, denoted as $a(\alpha)$. This is not in general the scattering length for $U(r, \alpha)$, which we denote as $a'(\alpha)$, because only the most divergent terms in the Born series are summed, thus neglecting many contributions to the scattering length for $U(r, \alpha)$.

The hope now is that $\lim_{\alpha\to 0} a(\alpha) = a(0)$ exists, and moreover, that $a(0) = a$, where a is the scattering length for $V(r)$. Notice that a is a function of the coupling constant, and also that it could well happen that $a(0) \neq a$. This is called peratization of the potential $V(r)$ and is usually done at zero energy.[2]

There is another process called regularization which is not the same as peratization, a fact not clearly understood by a number of authors. The distinction has recently been pointed out.[6] Regularization involves first solving an integral equation for the wavefunction and then using this in a certain integral relationship in order to obtain the scattering length. This process involves the scattering length $a'(\alpha)$ for $U(r, \alpha)$ which is *not* $a(\alpha)$ as defined above. Regularization involves the unjustified interchange of two limiting processes[7] and is not generally valid. The point here is that the validity of peratization and of regularization are two separate questions.

Another point in need of clarification is that, in the peratization procedure, the summation of leading singular terms is usually taken *in potential theory*, in order to find the exact scattering length for $V(r)$. *In field theory*, on the other hand, one only hopes to find the leading order for small g in the scattering amplitude. There is one example in potential theory[5] where peratization yields only a leading term in g. However, we will indicate below a counter example to the often found conjecture that the summation of leading singular terms in α will yield the leading term for small g in the scattering amplitude.

II. STATEMENT AND PROOF OF THE THEOREM

Let us assume a potential

$$V(r) = A(r)/r^n, \qquad n > 3, \tag{1}$$

where $A(r) > 0$ in some finite region $\epsilon > r > 0$ so that we have a repulsive potential near the origin. We also assume the usual conditions

$$\int_0^b r\,|V(r)|\,dr \quad \text{diverges for any} \quad b > 0, \tag{2}$$

$$\int_\epsilon^\infty r^2\,|V(r)|\,dr \quad \text{exists for any} \quad c > 0. \tag{3}$$

[1] G. Tiktopoulos and S. B. Treiman, Phys. Rev. **134**, B844 (1964). Equation (4) of this paper is incorrect; however, the result, Eq. (5), does follow from the correct infinite series.
[2] N. N. Khuri and A. Pais, Rev. Mod. Phys. **36**, 590 (1964).
[3] H. H. Aly, Riazuddin, and A. H. Zimerman, Phys. Rev. **136**, B1174 (1964).
[4] H. H. Aly, Riazuddin, and A. H. Zimerman, Nuovo Cimento **35**, 324 (1964).
[5] T. T. Wu, Phys. Rev. **136**, B1176 (1964).
[6] F. Calogero and M. Cassandro, Nuovo Cimento **37**, 760 (1965).
[7] F. Calogero, Phys. Rev. **139**, B602 (1965). This paper gives further references to work dealing with singular potentials.
[8] G. Feinberg and A. Pais, Phys. Rev. **131**, 2724 (1963).
[9] G. Feinberg and A. Pais, Phys. Rev. **133**, B477 (1964).

Then our theorem follows:

Theorem: If $A(r)$ is more singular at the origin than $r(dA/dr)$ (or approaches zero less quickly), then the peratized zero-energy scattering length (in the sense of Ref. 1 or 3) for the potential

$$U(r, \alpha) = V(r)\theta(r - \alpha), \qquad \alpha > 0,$$

is

$$a(\alpha) = -[A(\alpha)]^{\frac{1}{2}} \tanh ([A(\alpha)]^{\frac{1}{2}}/\alpha), \qquad n = 4, \tag{4}$$

or, in general,

$$a(\alpha) = -\frac{\pi\nu^{2\nu-1}}{[\Gamma(\nu)]^2 \sin \nu\pi} A^{\nu}(\alpha) \frac{I_{1-\nu}(\Omega)}{I_{\nu-1}(\Omega)} \tag{5}$$

with

$$\nu = (n - 2)^{-1}, \qquad \Omega = \frac{2\nu}{\alpha^{(n/2)-1}} A^{\frac{1}{2}}(\alpha).$$

Equation (4) is a special case of Eq. (5) with $n = 4$.

Proof: The zero-energy wavefunction for the potential $U(r, \alpha)$ can be written as[3]

$$\psi(r, \alpha) = r - gr \int_r^\infty V(r')\psi(r', \alpha)\, dr' - g \int_\alpha^r r' V(r')\psi(r', \alpha)\, dr' \tag{6}$$

and the exact scattering length $a'(\alpha)$ is

$$a'(\alpha) = -g \int_\alpha^\infty r' V(r')\psi(r', \alpha)\, dr'.$$

The function $a(\alpha)$ differs from $a'(\alpha)$ in that it contains only the most singular terms in α for each order of g. The lowest-order term in g of $a'(\alpha)$ is

$$a'(\alpha) = -g \int_\alpha^\infty r'^{2-n} A(r')\, dr' \tag{7}$$

$$= -\frac{g}{(3 - n)} \frac{A(r')}{r'^{n-3}}\bigg|_\alpha^\infty + \frac{g}{(3 - n)} \times \int_\alpha^\infty r'^{3-n} \left(\frac{dA}{dr}\right)_{r=r'} dr'$$

$$= -\frac{g}{(3 - n)} \frac{A(\alpha)}{\alpha^{n-3}} + \frac{g}{(3 - n)} \times \int_\alpha^\infty r'^{3-n} \left(\frac{dA}{dr}\right)_{r=r'} dr', \tag{8}$$

where we have integrated by parts. The infinite limit in the integrated part does not contribute because of condition (3).

If $r(dA/dr)$ is less singular at $r = 0$ than $A(r)$, or even approaches zero more rapidly than $A(r)$, then we need keep only the integrated term in expression (8) if we desire only the most singular term in α. Under such conditions, the integral (7) will be more singular than the integral term[10] in (8), so that the integrated term in (8) is more singular than the integral.

To find the second-order term in g of $a'(\alpha)$, we need the first-order g term for $\psi(r, \alpha)$. In general, the $(m + 1)$ order in g of $a'(\alpha)$, call it $a'_{m+1}(\alpha)$, is evaluated using $\psi_m(r, \alpha)$, the mth order in g of the wavefunction. We can easily see from Eq. (6) that $\psi_m(r, \alpha)$ will contain terms in r, terms in α, and cross terms involving both r and α of the form $j(\alpha)k(r)$. When we evaluate $a'_{m+1}(\alpha)$, we obviously need retain only the most singular r term in $\psi_m(r, \alpha)$, the most singular α term, and the most singular of each type of cross term. By this we mean that for the cross terms of the form $j(\alpha)k(r)$, we retain, for each $j(\alpha)$, only the term with the most singular $k(r)$. But if there are two or more identical $k(r)$ terms, we then keep only the term with the most singular $j(\alpha)$. All other terms in $\psi_m(r, \alpha)$ will give terms in $a'_{m+1}(\alpha)$ that are less singular in α than the ones we retain.

The first-order terms in g of $\psi(r, \alpha)$ can be integrated by parts, and the integrated terms will be of the forms $\beta_1 r^{-n+3} A(r)$ and $\beta_2 \alpha^{-n+3} A(\alpha)$, while the remaining integrals look like

$$r \int_r^\infty r'^{-n+2} \left(\frac{dA}{dr}\right)_{r=r'} dr' \quad \text{and} \quad \int_\alpha^\infty r'^{-n+3} \left(\frac{dA}{dr}\right)_{r=r'} dr',$$

which can be discarded if $r(dA/dr)$ again bears the previously specified relationship to $A(r)$ at $r = 0$. Notice that the constants β_1 and β_2 do not depend on the form of $A(r)$ in any way but only on n.

In the g^2 term of $a'(\alpha)$, we now find that an integration by parts yields integrated terms of the form $\gamma_2 \alpha^{5-2n} A^2(\alpha)$ while the remaining integrals contain $r'^{5-2n} A(r')(dA/dr)_{r=r'}$ and $r'^{2-n}(dA/dr)_{r=r'}$. To find only the leading term in α, we apply the same argument as used before and we keep only $\gamma_2 \alpha^{5-2n} A^2(\alpha)$, where γ_2 depends only on n and not on $A(r)$.

It is easy to see that the *leading* terms of the integral for $\psi_{m-1}(r, \alpha)$ will involve various terms of the form of a product of some power of r, some power of α, and $A^{\mu}(\alpha)A^{\nu}(r)$ with $\mu + \nu = m - 1$. When any of these terms is integrated by parts, the new integral will contain one more power of r, one less power of $A(r)$, and dA/dr. Therefore, by the same reasoning as before, we need keep only

[10] It takes only a simple geometrical argument to prove that if $G(x)$ is more singular at $x = 0$ than $H(x)$ then $\int_\epsilon^\infty G(x)dx$ is more singular as $\epsilon \to 0$ than $\int_\epsilon^\infty H(x)dx$ where there are no troubles at infinity.

the integrated term. Similarly, to find $a_m'(\alpha)$, we can see that we need retain only the integrated term of an integration by parts.

Thus, a summation of the most singular α terms in the Born series will give the form

$$a(\alpha) = \sum_{m=0}^{\infty} \gamma_m(n)[gA(\alpha)]^m, \tag{9}$$

where the $\gamma_m(n)$ do not depend on the form of $A(r)$.

This sum has been evaluated for $A(r) = 1$ and $n = 4$, and is[1]

$$a(\alpha) = -g^{\frac{1}{2}} \tanh (g^{\frac{1}{2}}/\alpha), \tag{10}$$

while, for $A(r) = -g \ln r$ and any $n > 3$, we have

$$a(\alpha) = -\frac{\pi\nu^{2\nu-1}}{[\Gamma(\nu)]^2 \sin \nu\pi} g^{\frac{1}{2}}(-\ln \alpha)^{\frac{1}{2}} \frac{I_{1-\nu}(\Omega)}{I_{\nu-1}(\Omega)}, \tag{11}$$

$$\nu = (n-2)^{-1}, \quad \Omega = 2\nu[\alpha^{\frac{1}{2}n-1}]^{-1} g^{\frac{1}{2}}(-\ln \alpha)^{\frac{1}{2}}.$$

We have corrected formula (E21) of Ref. 5 which makes use of (the incorrect equation) (E20). The correct formula, as may be easily checked in Ref. 12, p. 133, is

$$\int dz\, z^{-1}[I_{\nu-1}(z)]^{-2} = -\frac{\nu}{2 \sin \nu\pi} \frac{I_{1-\nu}(z)}{I_{\nu-1}(z)}.$$

The function $I_\lambda(x)$ is the commonly defined Bessel function of the third kind of real argument.[11] Since the form of these results is independent of $A(r)$, we can immediately conclude that Eqs. (4) and (5) hold and the theorem is proved.

III. APPLICATIONS AND DISCUSSION

We immediately see that we can get the results of Ref. 3 for $A(r) = g\,(\ln)^2\, r$ correctly as

$$a(\alpha) = -g^{\frac{1}{2}} \ln \alpha \tanh (g^{\frac{1}{2}} \ln \alpha/\alpha).$$

This is a useful check on the validity of the theorem.

One powerful result of our theorem is the following corollary.

Corollary: If $A(r)$ satisfies the condition of the previous theorem and $A(0)$ does not exist, then peratization fails—assuming the scattering length exists—for $V(r) = A(r)/r^n$, $n > 3$.

Proof: The limit as $\alpha \to 0$ of Eq. (5) yields

$$a(0) \propto g^\nu A^\nu(0), \quad \text{which does not exist.}$$

On the other hand, if $A(0)$ is a nonzero constant, then $a(0)$ exists but *may or may not* be the real scattering length a. It is not possible without other considerations to say whether peratization works in this case.

[11] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1958), 2nd ed.

The three examples of $A(r) = g$, $A(r) = g\,(\ln)^2\, r$, and $A(r) = -g \ln r$ have all been treated individually in the literature. However, there are other functions $A(r)$ that satisfy the requirements of the theorem. A few of these are[12]

(a) $A(r) = g(-\ln r^\mu)^\lambda$, $\quad \mu > 0$,

(b) $A(r) = ge^{\lambda r^\mu}$, $\quad \lambda > 0, \quad \mu > 0$;

(c) $A(r) = g \cos^\lambda r^\mu$, $\quad \mu > 0$.

We see examples here where very different potentials like (b) and (c) have the same summation to most singular terms in α, yet their scattering lengths are surely not the same.

Another important point is that, we see from Eq. (E23) of Ref. 5, that, for $A(r) = -g \ln r$, the summation of each singular series (i.e., the most singular terms, the next most singular, and so on) yields exactly the same form of $g^{\frac{1}{2}}\ (-\ln \alpha)^{\frac{1}{2}}$. Here is a clear counter example to the conjecture that the summation of leading singularities will give the leading term in the scatttering length for small g.

It is fairly obvious that this cannot in general be true because the general form

$$H_1\{f_1(\alpha), g\} = \sum_{i=0}^{\infty} C_i\{f_1(\alpha)g\}^i$$

will yield a closed form H_1, whose character is determined by all the C_i. A summation of second most singular terms

$$H_2(f_2(\alpha), g) = \sum_{i=0}^{\infty} D_i\{f_2(\alpha)g\}^i$$

will yield a form, H_2, determined only by the D_i which have only a very complex relationship to the C_i as can be seen by working with $\psi_{m-1}(r, \alpha)$ and $a_m(\alpha)$ for a little while. The limits

$$\lim_{\alpha\to 0} H_1 = h_1(g), \qquad \lim_{\alpha\to 0} H_2 = h_2(g)$$

depend not only on $f_1(\alpha)$ and $f_2(\alpha)$, but on H_1 and H_2 as well, i.e., on all the C_i's and D_i's. There is, in general, no possible way to determine $h_1(g)$ and $h_2(g)$ without knowing H_1 and H_2 (or knowing every C_i and D_i exactly).

[A series can be highly unstable. Recall that $e^{-1/x} = \sum_{n=0}^{\infty} C_n(1/x)^n$, $C_n = 1/n!$, which is zero as $x \to 0^+$. But if we alter just one C_n to $C_n + \epsilon$, then the sum diverges like ϵ/x^n as $x \to 0^+$.]

[12] If $\mu < 0$ in (a) then $A(r) = g(\ln r^\mu)^\lambda$. This case has been treated by H. Cornille, Nuovo Cimento **38**, 1243 (1965); **39**, 557 (1965); **43**, 786 (1966). In these papers, Cornille investigates, in great depth, the validity of various limiting procedures.

Indeed, one can produce simple examples of infinite series where h_2 is the leading g term even though f_1 is more singular at $\alpha = 0$ than f_2.

The fact that one must know every C_i exactly to discern the features of $h_1(g)$ is what hinders the investigation of peratization. We cannot in general tell if the second and further singular terms contribute to the infinite sum. In the rare case that

$$a(\alpha) = \sum_{i=0}^{\infty} C_i\{B(\alpha)g\}^i \tag{12}$$

exactly for all terms, we can write

$$B(\alpha) = B_1(\alpha) + B_r(\alpha), \tag{13}$$

where $B_1(\alpha)$ is the most singular term in α and $B_r(\alpha)$ contains everything else. Then, *if* peratization works,

$$a = a(0) = \lim_{\alpha\to 0} H(B(\alpha), g) = \lim_{\alpha\to 0} H(B_1(\alpha), g),$$

and we could have merely summed

$$H_1(B_1(\alpha), g) = \sum_{i=0}^{\infty} C_i\{B_1(\alpha), g\}^i.$$

However, the form (12) would be almost impossible to achieve due to the complicated equations defining $a(\alpha)$.

IV. CRITIQUE OF THE LITERATURE

As pointed out by Calogero,[6] Khuri and Pais[2] and Aly *et al.*[4] have not precisely distinguished between regularization and peratization. This confusion of terms appears again in a more recent work of Aly *et al.*[13]

In Refs. 4 and 6, the authors treat the potential

$$V(r) = ge^{2/r}/r^4 + g'(1/r^4), \tag{14}$$

but the Aly group deals only with $g' = \frac{1}{4}$. There is some disagreement among the two sets of authors as to the validity of peratization.

To clear up some of the confusion, we quote the results for the scattering length for Eq. (14) with $g' = \frac{1}{4}$ to be

$$a = -(g^{\frac{1}{2}} + \tfrac{1}{2}). \tag{15}$$

Since the scattering length for $\frac{1}{4}(1/r^4)$ alone is $a = -(\frac{1}{4})^{\frac{1}{2}} = -\frac{1}{2}$, it is assumed by Aly and his coworkers that the $-\frac{1}{2}$ in Eq. (15) comes from the second term in Eq. (14). The situation is more complicated than this. The formula given for $a'(\alpha)$ in Ref. 6 for the potential [Eq. (14)] is[14]

$$a'(\alpha) = g^{\frac{1}{2}}\frac{K_\sigma[g^{\frac{1}{2}}e^{1/\alpha}]I'_\sigma(g^{\frac{1}{2}}) - I_\sigma[g^{\frac{1}{2}}e^{1/\alpha}]K'_\sigma(g^{\frac{1}{2}})}{K_\sigma[g^{\frac{1}{2}}e^{1/\alpha}]I_\sigma(g^{\frac{1}{2}}) - I_\sigma[g^{\frac{1}{2}}e^{1/\alpha}]K_\sigma(g^{\frac{1}{2}})} \tag{16}$$

with $\sigma = (g')^{\frac{1}{2}}$. For $g' = \frac{1}{4}$ we have

$$a'(0) = -(g^{\frac{1}{2}} + \tfrac{1}{2}) \text{ exactly for all } g.$$

When $g = 0$, we have, for any g'

$$a'(0) = -\left(2^{1-2\sigma}\sigma\frac{\Gamma(1-\sigma)}{\Gamma(1+\sigma)}g^{\sigma} + \sigma\right). \tag{17}$$

When $g \gg g'$ and g is large we have

$$a'(0) = -\left(g^{\frac{1}{2}} + \frac{1}{2} + \frac{4\sigma^2 - 1}{8}g^{-\frac{1}{2}}\right). \tag{18}$$

We see from Eqs. (17) and (18) that the second term in Eq. (15) does not come simply from the second term in the potential (14), and we also see that the g' part of the potential affects also the first term of Eq. (15). Though the first term in (17) is some sort of a cross term between the g and g' terms, we see from (18) that for large g the first two terms of $a(0)$ are like (15) regardless of the value of g'! Thus Eq. (15) *is not* a linear combination of the $g' = \frac{1}{4}$ scattering length and a cross term between g and g'.

V. CONCLUSIONS

We have shown how a special class of singular potentials may be peratized and whether or not the peratization procedure is finite. To discuss, in general, whether peratization works or not is a task impossible without explicit calculation and comparison with the correct answer. It is difficult at the moment to state general criteria for the procedure to be successful, though important progress has been made by Cornille.[12] Significantly, it fails in many cases.

We have again cautioned against the confusion between regularization and peratization, and we have clarified some misconceptions prevalent in the literature about whether leading singular sums will yield leading terms in g. Finally, some other mistakes and misinterpretations in the literature have been illuminated.

ACKNOWLEDGMENTS

The author would like to thank Dr. H. Cornille, Dr. F. Calogero, and Dr. M. Cassandro for stimulating correspondence as well as several helpful suggestions.

[13] H. H. Aly, Riazuddin, and A. H. Zimerman, J. Math. Phys. **6**, 1115 (1965).

[14] The expression in Eq. (16) is lacking a term which is zero when $\alpha = 0$, and this gives the expression the wrong limit as $\alpha \to \infty$. According to a private communication from the authors of Ref. 6, this will be corrected in a future publication.

A Non-Hilbert-Space Formulation of Quantum Field Theory

EDUARD PRUGOVEČKI
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada
(Received 17 March 1966)

Based on earlier papers in which a non-Hilbert-space formalism for quantum physics, called the $\mathfrak{L}$-formalism, was proposed, a new formalism for quantum field theory is developed by introducing the field concept into the $\mathfrak{L}$-formalism. The present formalism is then compared with the Wightman approach and with the Haag–Araki approach to quantum field theory. The necessity of a global vacuum within the $\mathfrak{L}$-framework and the alternative concept of the local vacuum are discussed.

1. INTRODUCTION

IN a series of previous papers[1] (referred to throughout the present paper as [1], [2], [3], and [4]), we have formulated an axiomatic approach to the formalism of quantum mechanics. This formalism is partially motivated by a new approach to the theory of measurement in quantum mechanics (see [1]). However, disregarding the acceptability of the proposed theory of measurement, our formalism (called the $\mathfrak{L}$-formalism) is of independent interest as a new tool for formulating quantum theories.

For reasons which soon become clear, it is of no particular interest to apply this formalism to ordinary nonrelativistic quantum mechanics. Consequently, our task in this paper is to enrich the $\mathfrak{L}$-formalism with additional structure so that, out of this formalism, field theories may be formulated. Next, we investigate the relation between our axiomatic formulation of a quantum field theory and each of the two best known contemporary formulations of this kind [that of Wightman (Sec. 3) and that of Haag and Araki (Sec. 4)].

In order to understand the basically unique features of a quantum field theory formulated in terms of the $\mathfrak{L}$-formalism (we will refer to such a theory as $\mathfrak{L}$QFT) in contradistinction to that of a quantum field theory formulated in terms of Hilbert-space concepts ($\mathfrak{H}$QFT), we have to elucidate the relation between the $\mathfrak{L}$-formalism and the Hilbert-space formalism ($\mathfrak{H}$-formalism). In [3] and [4] we have proved that there exists a *physical* equivalence between the $\mathfrak{L}$-formalism and the $\mathfrak{H}$-formalism. The formal nature of this equivalence has been explained in Sec. 1 of [3] and in Sec. 3 of [4]. Its physical content is the following: In the $\mathfrak{L}$-formalism, the set $\mathfrak{L}$ of basic objects consists of (cf., Appendix) the set $\mathfrak{O}$ of all observables (on which no prerequired sturcture is imposed) and the set $\mathfrak{P}$ of all physical states. A physical state is defined as a mapping attaching to each n-tuple $\alpha = (\alpha_1, \cdots, \alpha_n)$ of observables $\alpha_1, \cdots, \alpha_n$ a complex probability measure P^α in such a way that certain axioms are fulfilled. If a quantum mechanical theory is given in terms of the $\mathfrak{H}$-formalism, then we immediately know the set $\mathfrak{O}$—an observable being represented by a self-adjoint operator. The complex probability measures corresponding to any physical state can also be written immediately [see Appendix, formula (A18)], and thus the theory is instantaneously recasted in the $\mathfrak{L}$-formalism.

On the other hand, if a theory is given in terms of the $\mathfrak{L}$-formalism, we prove (in [3] and [4]) the *existence* of a Hilbert space $\mathfrak{H}$ so that the theory can be recasted in terms of an $\mathfrak{H}$-formalism associated with that Hilbert space. We have to stress, however, that this theorem is essentially an *existence* theorem. In other words, the proof of the existence of this Hilbert space $\mathfrak{H}$ does not guarantee the possibility of the straightforward numerical construction of any particular physically interesting object of $\mathfrak{H}$ out of the given objects in $\mathfrak{L}$. This situation parallels the case of, say, existence theorems of solutions of systems of differential equations under given initial conditions, when the existence proof does not offer a really practical way of explicitly obtaining such solutions.

This suggests the fact that the practical problem of theoretical formulation and experimental computation in the $\mathfrak{L}$-formalism might be quite different from that of the $\mathfrak{H}$-formalism. This conjecture is supported further by the fact that, in any actual experiment, only a finite number of observables is

[1] These papers are: [1] E. Prugovečki, "On a Theory of Measurement of Incompatible Observables in Quantum Mechanics" (to be published). [2] E. Prugovečki, J. Math. Phys. 7, 1054 (1966); [3] *ibid.* 7, 1070 (1966); [4] *ibid.* 7, 1680 (1966). Familiarity with [1]–[4] is desirable but not absolutely essential for reading the present paper. The notation and the main ideas of the formalism developed in [2]–[4] have been incorporated in the Appendix, in a form best suited to the needs of this paper.

involved. Consequently, in the £-formalism in order to get information concerning such an experiment, it is necessary to compute only complex probability measures related only to the involved observables. This is possible because the £-formalism gives the explicit conditions which such measures have to fulfill in order to belong to a physical state. This situation is different from that of an $\mathfrak{C}$-formulated theory, where no finite complete set of observables exists—such as the case of field theory.

We have to remark that, in a £QFT, only *observable* fields can play a role. This is due to the fact that £ is assumed to contain only observable quantities, and the £-formalism has been formulated accordingly. We have shown, however, in a separate paper [Nuovo Cimento **45A**, 327 (1966)] that, in electrodynamics, the unobservable Dirac field can be replaced with observable tensor fields.

2. THE AXIOMS FOR QUANTUM FIELD THEORY

The axiomatic system presented in [2], [3], and [4] is meant to characterize the basic framework of any quantum theory. However, in order to formulate a framework for quantum field theories, additional axioms must be supplemented.

The main characteristic of a conventional field theory is the fact that one observable is assigned to each point in space–time from each of the *kinds* of observables belonging to a certain considered class $\mathfrak{K}$ of *categories* of observables. For example, each known field theory contains (as one of the class $\mathfrak{K}$ categories of observables) the category called "the energy or energy density at some point x of space–time". Any quantum field theory contains such a category as "the total angular momentum density of particles of the kind κ at a point of space–time", while it is the characteristic of electrodynamics (classical or quantum) that it contains the category "the electromagnetic field at some point of space–time". Each of these categories of observables is called a field, and $\mathfrak{K}$ represents the class of all fields.

Wightman has shown,[2] however, that there does not exist a nontrivial local *quantum* field theory with a unique global vacuum which could be formulated in terms of fields defined at points of space–time. Furthermore, general consideration of Bohr and Rosenfeld,[3] as well as that of Heisenberg,[4] seem to indicate that it might be necessary to consider exclusively "smeared" fields. Consequently, we formulate £QFT in terms of fields which are distributions over some space of test functions defined at all the points of the Minkowski four-dimensional space–time $\mathfrak{M}$.

To do this, we choose some adequate space of test functions. For the sake of being definite, let us take this space to be the $\mathfrak{D}$-space of Schwartz.[5] $\mathfrak{D}$ is the space of all infinitely many times differentiable functions $f(x)$ of compact support in the space–time, in which a suitable topology is introduced. We can still regard fields as categories of observables from a class $\mathfrak{K}$, but, instead of attaching for each such category $\varphi \in \mathfrak{K}$ an observable $\varphi(x)$ to each point x, we attach an "observable" $\varphi(f)$ to each real function $f = \bar{f} \in \mathfrak{D}$.

Axiom W1: The set $\mathfrak{O}$ of observables of a £QFT contains all the fields $\varphi(f)$ taken at all real points $f \in \mathfrak{D}$. For a complex $f \in \mathfrak{D}$,

$$f(x) = g(x) + ih(x), \quad g, h \in \mathfrak{D}, \quad x \in \mathfrak{M}, \tag{2.1}$$

$\varphi(f)$ is defined by

$$P^{\psi,\varphi(f)\psi}(B) = P^{\psi,\varphi(g)\psi}(B) + iP^{\psi,\varphi(h)\psi}(B). \tag{2.2}$$

The fields $\varphi(f)$ are distribution-valued fields on $\mathfrak{D}$ in the sense that

$$\int_{-\infty}^{+\infty} \lambda \, dP^{\psi,\varphi(af+bg)\psi}(I_\lambda) = a \int_{-\infty}^{+\infty} \lambda_1 \, dP^{\psi,\varphi(f)\psi}(I_{\lambda_1}) + b \int_{-\infty}^{+\infty} \lambda_2 \, dP^{\psi,\varphi(g)\psi}(I_{\lambda_2}),$$
$$f, g \in \mathfrak{D}, \quad a, b \in \mathbf{C}^1, \tag{2.3}$$

and

$$\int_{-\infty}^{+\infty} \lambda_n \, dP^{\psi,\varphi(f_n)\psi}(I_{\lambda_n}) \to \int_{-\infty}^{+\infty} \lambda \, dP^{\psi,\varphi(f)\psi}(I_\lambda), \quad f_n \to f, \tag{2.4}$$

if $f_1, f_2, \cdots, \in \mathfrak{D}$ is any sequence which converges to $f \in \mathfrak{D}$ in the topology of $\mathfrak{D}$. Any other observable is a function of some finite number of compatible field observables and global observables (to be defined later).

In the above relations we have employed the following notation (which will be used consistently from now on): For any $\lambda = (\lambda_1, \cdots, \lambda_n) \in \mathbf{R}^n$ the symbol I_λ or $I_{\lambda_1 \times \ldots \times \lambda_n}$ denotes the direct product of the sets $(-\infty, \lambda_1), \cdots, (-\infty, \lambda_n)$, i.e.,

[2] A. S. Wightman, Ann. Inst. Henri Poincaré **1**, 403 (1964).

[3] N. Bohr and L. Rosenfeld, Kgl. Danske Vid. Sels. Math-fys. Medd. **12**, No. 8 (1933); Phys. Rev. **78**, 794 (1950).

[4] W. Heisenberg, Verh. d. Sachs. Ak. Leipzig **83**, 3 (1931); **86**, 317 (1934).

[5] L. Schwartz, *Théorie des distributions* (Hermann & Cie., Paris, 1957), Vol. I.

$$I_\lambda = \{\lambda' = (\lambda_1', \cdots, \lambda_n'): -\infty < \lambda_k' < \lambda_k, k = 1, \cdots, n\}. \quad (2.5)$$

In formulating Axiom W1, we also used the expression *field observable*, by which we refer to a field φ taken at some $f = \bar{f} \in \mathfrak{D}$, or to a function $F[\varphi(f)]$ of $\varphi(f)$, defined by

$$P^{\nu_1 F[\varphi(f)]\nu_2}(B) = P^{\nu_1 \varphi(f)\nu_2}[F^{-1}(B)], \quad B \in \mathfrak{B}^1, \quad (2.6)$$

The definition (2.6) is naturally valid for any Borel function $F(\lambda)$ on $\mathbf{R}^1$.

We note that the set of all observables consists of fields taken at some real $f \in \mathfrak{D}$, as well as other observables. Furthermore, in general it is not desirable to impose the condition that, for any Borel function $F(\lambda)$, the field observables $F[\varphi(f)]$ define a field for all $f = \bar{f} \in \mathfrak{D}$. Namely, this would require that $F[\varphi(f)]$ obey (2.3), which is too stringent—as is more evident in Sec. 3, where we make the transition to the Hilbert-space formalism.

In Axiom W1 we required, however, that any function of compatible $\varphi_1(f_1), \cdots, \varphi_n(f_n)$, and, in particular, of compatible fields $\varphi_1, \cdots, \varphi_n$, is an observable. In this context the fields $\varphi_1, \cdots, \varphi_n$ are compatible if the *observables* $\varphi_1(f_1), \cdots, \varphi_n(f_n)$ are compatible (in the sense of Def. 1, Appendix) for any $f_1 = \bar{f}_1, \cdots, f_n = \bar{f}_n \in \mathfrak{D}$. Then for a Borel function $F(\lambda), \lambda \in R^{\varphi_1 \times \cdots \times \varphi_n}$, we define $F[\varphi_1(f_1), \cdots, \varphi_n(f_n).]$ as the observable for which

$$P^{\nu_1 F[\varphi_1(f_1), \cdots, \varphi_n(f_n)]\nu_2}(B) = P^{\nu_1 \varphi_1(f_1), \cdots, \varphi_n(f_n)\nu_2}[f^{-1}(B)], \quad P \in \mathfrak{S}, \quad f_k = \bar{f}_k \in \mathfrak{D}, \quad k = 1, \cdots, n. \quad (2.7)$$

Besides field observables, we also have, in the present quantum field theories, global observables. We denote by $\pi = (\pi^0, \pi^1, \pi^2, \pi^3)$ the four one-dimensional observables determining the total global 4-momentum (the term total refers to all kinds of considered "particles", while the term global means the 4-momentum on the entire space). In the conventional Hilbert-space quantum field theory (to which we refer by the shorthand $\mathfrak{H}$QFT), the total global four momentum is represented by selfadjoint operators $(P^0, P^1, P^2, P^3) = P$, which are the generators of a representation of the translation group in space–time, i.e., for each field operator $\Phi(x)$, we have

$$\Phi(x) = e^{iP\cdot(x-y)}\Phi(y)e^{-iP\cdot(x-y)}. \quad (2.8)$$

We have used here the notation (which is to be employed throughout) in which

$$\xi\eta = \xi_\nu\eta^\nu = \xi^0\eta^0 - \xi^1\eta^1 - \xi^2\eta^2 - \xi^3\eta^3 \quad (2.9)$$

for any two Minkowsky-space 4-vectors ξ, η. The spectrum of P is assumed to lie in the closure $\bar{V}_+$ of the forward light-cone V_+. All these requirements are inserted in the £QFT by means of the following axiom:

Axiom W2: The set $\mathfrak{O}$ contains the total global 4-momentum π, having a spectrum S^π lying within the forward light-cone $\bar{V}_+$.

A representation of the translation group in space–time is given by the condition that, for any $\varphi \in \mathfrak{K}$,

$$P^{\nu_1 \varphi(f_{\{x,1\}})\nu_2}(B) = \int e^{i(p_1 - p_2)\cdot x} \times dP^{V_1 \pi; \varphi(f); \pi V_2}(I_{p_1} \times B \times I_{p_2}), \quad P \in \mathfrak{S}. \quad (2.10)$$

For a given inhomogeneous Lorentz transformation

$$x \to x' = \Lambda x + a, \quad x, a \in \mathfrak{M}, \quad \Lambda \in L_+^\uparrow,$$

we define $f_{\{a,\Lambda\}}$ by

$$f_{\{a,\Lambda\}}(x) = f[\Lambda^{-1}(x - a)]; \quad (2.11)$$

we can easily see that $f_{\{a,\Lambda\}} \in \mathfrak{D}$ if $f \in \mathfrak{D}$.

In $\mathfrak{H}$ QFT the existence of a representation of the restricted Lorentz group $L_+^\uparrow$ is also assumed. The infinitesimal generators $M_{\kappa\lambda}$ are taken to represent the total global angular momentum $\mu_{\kappa\lambda}$. We arrive at that representation by assuming that the fields can be grouped together in finite sets

$$\varphi^{(1)} = \{\varphi_1^{(1)}, \varphi_2^{(1)}, \cdots, \varphi_{k_1}^{(1)}\},$$
$$\varphi^{(2)} = \{\varphi_1^{(2)}, \varphi_2^{(2)}, \cdots, \varphi_{k_2}^{(2)}\};$$

each element $\varphi_l^{(i)}$ of such a set $\varphi^{(i)}$ can now be called the lth-component of $\varphi^{(i)}$. These sets $\varphi^{(i)}$ then are such that we have, in terms of the corresponding field operators, the following representation of $L_+^\uparrow$:

$$\Phi_i^{(i)}(\Lambda\cdot x) = \sum_k S_{ik}(\Lambda)U(0, \Lambda)\Phi_k^{(i)}(x) \times U(0, \Lambda)^{-1}, \quad x \in \mathfrak{M}. \quad (2.12)$$

Here $S_{ik}(\Lambda)$ stands for a finite-dimensional irreducible tensor or spinor representation of $L_+^\uparrow$.

The case of £QFT is distinct in two respects. Our concept of a field is in a certain respect more general, because it refers to any kind of observable which is attached to each point of space–time. Thus it includes, e.g., the "field" which represents the probability that the energy density at each point of $\mathfrak{M}$ lies within [0, 1]. On the other hand, it is more restrictive, just because it refers *only* to observables, and subsequently the spinor representations of $L_+^\uparrow$ have to be ignored.

Axiom W3: The set $\mathfrak{O}$ includes the global observables $\mu_{\kappa\lambda}$ which stand for the components of the total angular momentum. There exists a family $\mathfrak{F}$ of "relativistic fields" obtained by grouping the fields of a certain subset $\mathfrak{F}_c$ of $\mathfrak{K}$ into finite sets $\varphi^{(1)} = \{\varphi_1^{(1)}, \cdots, \varphi_{k_1}^{(1)}\}, \varphi_2^{(2)} = \{\varphi_1^{(2)}, \cdots, \varphi_{k_2}^{(2)}\}, \cdots$ of compatible fields with identical spectrum. These sets are such that, for any Lorentz rotation $\Lambda \in L_+^\uparrow$ in any of the two-dimensional coordinate planes (κ, λ), and for any $P \in \mathfrak{S}$, $\varphi \in \mathfrak{F}$, $B \in \mathfrak{B}^\varphi$,

$$P^{\varphi_1\varphi_i(f_{[0,\Lambda]})\varphi_k}(B) = \sum_{k=1}^{n} S_{ik}(\Lambda) \int \exp\left[i(m' - m'')\Lambda^{\alpha\beta}\right] \times dP^{\varphi_1\mu_{\alpha\beta}\varphi_k(f)\mu_{\alpha\beta}\varphi_k}(I_{m'} \times B \times I_{m''}), \tag{2.13}$$

where

$$\mu_{\kappa\lambda} = -\mu_{\lambda\kappa}, \quad \kappa, \lambda = 0, \cdots, 3 \tag{2.14}$$

and the matrices $S_{ik}(\Lambda)$ belong to a finite-dimensional tensor representation of $L_+^\uparrow$. The family $\mathfrak{F}$ is such that all fields within $\mathfrak{F}_c$ are Borel-independent (Appendix, Definition 7) and any field in $\mathfrak{K}$ is a function of a finite set of compatible fields from $\mathfrak{F}_c$.

From now on we refer to $\varphi^{(1)}, \varphi^{(2)}, \cdots$ as relativistic fields and to $\varphi_1^{(1)}, \varphi_2^{(1)}, \cdots$ as the components of a relativistic field, reserving the word "field", as we have done until now, for any element of $\mathfrak{K}$. This is convenient because very often we are not concerned with the relativistic aspects of the theory.

We note that, since any Lorentz transformation can be obtained by performing three successive Lorentz transformations in the coordinate planes, (2.13) really implies a "representation" of the restricted Lorentz group. Furthermore, from (2.9) and (2.13), we can immediately deduce the form of any transformation corresponding to an element of the inhomogeneous Lorentz group.

Another of the basic assumptions of $\mathfrak{K}$QFT is local commutativity (sometimes referred to as microcausality) which requires that

$$[\varphi_1(x), \varphi_2(y)]_\pm = 0, \quad (x - y)^2 < 0 \tag{2.15}$$

for any two fields φ_1, φ_2—where the anticommutator is for half-integer spin and the commutator for integral spin. This assumption, based on the relativistic postulate that no signal spreads with a speed greater than that of light, is in fact completely *ad hoc* for unobservable fields. However, in £QFT it can be justified from the point of view of bounded signal velocity, and it is expressed in the following axiom.

Axiom W4: If $\varphi_1, \varphi_2 \in \mathfrak{K}$ are any two fields, and the supports of the real test-functions $f, g, \in \mathfrak{D}$ are spacelike separated with respect to one another, then the observables $\varphi_1(f)$ and $\varphi_2(g)$ are compatible.

3. COMPARISON WITH WIGHTMAN'S FORMULATION OF QUANTUM FIELD THEORY

When we compare the formulation of quantum field theory given in the preceding section with other formulations, one point should be obvious: we have to consider such versions of these formulations in which only observable fields are treated. Consequently, the theorem which will be formulated now should be understood accordingly.

Theorem 1: An £QFT obeying Axioms W1–W4 is physically equivalent to a quantum field theory formulated in terms of the Wightman axioms[6,7] on a Hilbert-space, which is not necessarily separable, and which does not necessarily contain a global vacuum, if and only if the following axiom is fulfilled.

Axiom W5: For any $\varphi \in \mathfrak{K}$ and for any real $f = \bar{f} \in \mathfrak{D}$,

$$\int_{-\infty}^{+\infty} \lambda^2 \, dP^{\varphi(f)}(I_\lambda) < +\infty, \qquad P \in \mathfrak{S}_0. \tag{3.1}$$

In addition, the total global and angular momentum are such that

$$\int_{-\infty}^{+\infty} \lambda^2 \, dP^{\nu\nu}(I_\lambda) < +\infty, \quad \nu = 0, \cdots, 3,$$
$$\int_{-\infty}^{+} \lambda^2 \, dP^{\mu_{\alpha\beta}}(I_\lambda) < +\infty, \quad \alpha, \beta = 0, \cdots, 3. \tag{3.2}$$

We demonstrate now that an £QFT obeying Axioms W1–W5 can be formulated in such terms as to satisfy the Wightman axioms [with the exception of the separability of the Hilbert space and the existence of a (unique) global vacuum]. In other words, we prove the "sufficiency part" of the above-stated theorem. The necessity of introducing Axiom W5 becomes obvious in the course of the proof.

From the theorems in [3] and [4] on the existence of a Hilbert-space representation[8] of the £-formalism, we extract the following: There exists a Hilbert

[6] A. S. Wightman, *Quelques Problèmes Mathématique de la théorié quantique relativiste* in *Les problèmes mathématique de la théorie quantique des champes* (Centre Nationale de la Recherche Scientifique, Lille, 1959).

[7] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964).

[8] Cf. [3], Sec. 1.

space $\mathfrak{K}$ and an injective mapping of the set $\mathcal{S}_0$ of all pure physical states into a subset of all rays

$$\{a\Psi, a \in \mathbf{C}^1\}, \quad \Psi \in \mathfrak{K}, \tag{3.3}$$

defined on $\mathfrak{K}$. If we denote by H_0 the subset of all vectors Ψ in $\mathfrak{K}$ belonging to all the rays which are images of pure physical states, then the linear manifold D spanned by H_0 is everywhere dense in $\mathfrak{K}$ (in the norm topology).

There exists also an injective mapping of the set $\mathcal{O}_b$ of all bounded observables (i.e., observable with a bounded spectrum) into the C^*-algebra $\mathfrak{B}(\mathfrak{K})$ of all bounded linear operators on $\mathfrak{K}$. This mapping is such that, given any observables $\alpha_1, \cdots, \alpha_n$ and Borel sets $B_1 \in \mathfrak{B}^{\alpha_1}, \cdots, B_n \in \mathfrak{B}^{\alpha_n}$,[9] we can attach to them projection operators $E_{\alpha_1}(B_1), \cdots, E_{\alpha_n}(B_n)$ in a manner such that

$$P^{\alpha_1;\cdots;\alpha_n}(B_1 \times \cdots \times B_n) = \langle \Psi_P | E_{\alpha_1}(B_1) \cdots E_{\alpha_n}(B_n) | \Psi_P \rangle, \quad P \in \mathcal{S}_0, \quad \Psi_P \in H_0. \tag{3.4}$$

In (3.4) Ψ_P is any normalized vector, $\langle \Psi_P, \Psi_P \rangle = 1$, belonging to the ray which is the image of the pure physical state P.

If α is a bounded observable, the set

$$\{E_\alpha(B), B \in \mathfrak{B}^\alpha\} \tag{3.5}$$

of projection operators on $\mathfrak{K}$ can be regarded as the spectral decomposition of a bounded self-adjoint operator

$$A_\alpha = \int_{-\infty}^{+\infty} \lambda \, dE_\alpha(I_\lambda). \tag{3.6}$$

In fact, the mapping of $\mathcal{O}_b$ into $\mathfrak{K}$ is such that A_α is just the image of α.

On the basis of (3.4), we have

$$\langle \Psi_P | A_{\alpha_1} \cdots A_{\alpha_n} | \Psi_P \rangle = \int \lambda_1 \cdots \lambda_n \, dP^{\alpha_1;\cdots;\alpha_n}(I_{\lambda_1} \times \cdots \times I_{\lambda_n}), \quad P \in \mathcal{P}_0, \quad \Psi_P \in H_0, \quad \langle \Psi_P | \Psi_P \rangle = 1, \tag{3.7}$$

if the observables $\alpha_1, \cdots, \alpha_n$ are bounded.

If α is an unbounded observable, the set (3.5) of projection operators can still be regarded to determine an (unbounded) self-adjoint operator, provided that the set D_α of all vectors $\Psi \in \mathfrak{K}$ for which

$$\int \lambda^2 \, d\langle \Psi | E_\alpha(I_\lambda) | \Psi \rangle < +\infty \tag{3.8}$$

is everywhere dense in $\mathfrak{K}$. For the field observables $\varphi(f)$, $f \in \mathfrak{D}$, $f = \bar{f}$, the global total momentum π_μ, and the angular momentum $\mu_{\lambda\nu}$, condition (3.8) is satisfied by all $\Psi \in H_0$ due to Axiom W5; it is consequently satisfied also by all $\Psi \in D$. Since D is everywhere dense in $\mathfrak{K}$, the operators

$$A_{\varphi(f)} = \int_{-\infty}^{+\infty} \lambda \, dE_{\varphi(f)}(I_\lambda), \quad f \in \mathfrak{D}, \quad f = \bar{f}, \tag{3.9}$$

$$P_\mu = \int_{-\infty}^{+\infty} \lambda \, dE_{\pi_\mu}(I_\lambda), \quad \mu = 0, \cdots, 3, \tag{3.10}$$

$$M_{\kappa\nu} = \int \lambda \, dE_{\mu_{\kappa\nu}}(I_\lambda), \quad \kappa, \nu = 0, \cdots, 3 \tag{3.11}$$

are defined as self-adjoint operators with domains of definition containing D. For a complex $f \in \mathfrak{D}$, we can naturally define $A_{\varphi(f)}$ as

$$A_{\varphi(f)} = A_{\varphi(g)} + iA_{\varphi(h)}, \quad f(x) = g(x) + ih(x), \quad g(x) = \overline{g(x)}, \quad h(x) = \overline{h(x)}. \tag{3.12}$$

In order to prove that the Wightman Axiom I (in chap. 3 of Ref. 7) is fulfilled, we can take our D to coincide with the set D introduced in that axiom. Due to Axiom W1 we can easily see that, for any $\varphi \in \mathfrak{K}$, $\Psi_1, \Psi_2 \in \mathfrak{D}$, the functionals

$$\langle \Psi_1 | \varphi(f) | \Psi_2 \rangle \tag{3.13}$$

are distributions on $\mathfrak{D}$. The validity of Wightman's Axioms 0 and I[10] in our framework would be essentially established (with the exception of the existence of a global vacuum state Ψ_0) if we prove that

$$P_\mu D \subset D, \quad M_{\mu\nu} D \subset D, \quad \varphi(f) D \subset D, \quad \mu, \nu = 0, \cdots, 3, \quad \varphi \in \mathfrak{K}, \quad f \in \mathfrak{D}. \tag{3.14}$$

Take any $\Psi_P \in H_0$ corresponding to a pure physical state P, and any observable $\alpha \in \mathcal{O}$. If, for a Borel set $B \in \mathfrak{B}^\alpha$ the vector $E_\alpha(B)\Psi_P$ is different from zero, i.e.,

$$\langle \Psi_P | E_\alpha(B) | \Psi_P \rangle = \langle \Psi_P | \Psi_P \rangle P^\alpha(B) \neq 0, \tag{3.15}$$

then, according to Axiom IV in the Appendix, it corresponds to a physical state P_1 defined by

$$P_1^\beta(B') = [P^\alpha(B)]^{-1} P^{\alpha;\beta;\alpha}(B \times B' \times B), \quad \beta \in \hat{\mathcal{O}}, \quad B' \in \mathfrak{B}^\beta, \tag{3.16}$$

i.e., we have $E_\alpha(B)\Psi_P \in H_0$.

Due to the same axiom, we can prove that, if

$$E_\alpha(B_1), \cdots, E_\alpha(B_n) \neq 0, \quad B_1, \cdots, B_n \in \mathfrak{B}^\alpha, \tag{3.17}$$

[9] An ordered pair consisting of one observable α and a $B \in \mathfrak{B}$ defines a special kind of observable called a question. A question has its spectrum concentrated at the points 0 and 1 (cf. [2], Sec. 3).

[10] Cf. Ref. 7, pp. 97–98.

then, for any $\lambda_1, \cdots, \lambda_n \in \mathbf{C}^1$, either

$$(\lambda_1 E_\alpha(B_1) + \cdots + \lambda_n E_\alpha(B_n))\Psi_P \in H_0 \tag{3.18}$$

or

$$(\lambda_1 E_\alpha(B_1) + \cdots + \lambda_n E_\alpha(B_n))\Psi_P = \mathbf{O}. \tag{3.19}$$

Assume now that A_α is defined as a self-adjoint operator on $\mathfrak{H}$, and that

$$0 < \int_{-\infty}^{+\infty} \lambda^2 \, d\langle \Psi_P | \, E_\alpha(I_\lambda) \, | \Psi_P \rangle < +\infty. \tag{3.20}$$

Choose a sequence $\Pi_1, \Pi_2, \cdots$ of finer and finer partitions

$$\Pi_k = \{ \cdots [\lambda_{-1}^{(k)}, \lambda_0^{(k)}), [\lambda_0^{(k)}, \lambda_1^{(k)}), \times [\lambda_1^{(k)}, \lambda_2^{(k)}), \cdots \}, \quad k = 1, 2, \cdots, \tag{3.21}$$

of $\mathbf{R}^1 = (-\infty, +\infty)$ in which the maximum length of the finite intervals in the partition tends to zero when $k \to \infty$. For sufficiently fine partitions, i.e., for sufficiently high $k \geq r$, we have [due to (3.20)]

$$0 < \sum_j |\lambda_j^{(k)}|^2 \langle \Psi_P | \, E_\alpha([\lambda_j^{(k)}, \lambda_{j+1}^{(k)})) \, | \Psi_P \rangle, \tag{3.22}$$

where the summation in j is only over finite intervals.

From the previous considerations we know that, given a k for which (3.22) is satisfied, there exists a physical state P_k (which is necessarily a pure state) defined as

$$P_k^\beta(B) = [\sum_j |\lambda_j^{(k)}|^2 P^\alpha([\lambda_j^{(k)}, \lambda_{j+1}^{(k)}))]^{-1} \times \sum_{i,j} \lambda_i^{(k)} \lambda_j^{(k)} P^{\alpha;\beta;\alpha}([\lambda_i^{(k)}, \lambda_{i+1}^{(k)}) \times B \times [\lambda_j^{(k)}, \lambda_{j+1}^{(k)})), \quad \beta = (\beta_1, \cdots, \beta_s) \in \hat{\mathfrak{O}}, \quad B = B_1' \times \cdots \times B_s' \in \mathfrak{B}^{\beta_1} \times \cdots \times \mathfrak{B}^{\beta_s}, \tag{3.23}$$

where

$$P^{\alpha;\beta;\alpha}(B_1 \times B \times B_2) = \langle \Psi_P | E_\alpha(B_1) F_\beta(B) E_\alpha(B_2) \, | \Psi_P \rangle, \quad B_1, B_2 \in \mathfrak{B}^\alpha, \quad F_\beta(B) = E_{\beta_1}(B_1') \cdots E_{\beta_s}(B_s'), \tag{3.24}$$

having the property that

$$\Psi_k = \sum_j \lambda_j^{(k)} E_\alpha([\lambda_j^{(k)}, \lambda_{j+1}^{(k)}))\Psi_P \tag{3.25}$$

belongs to the image-ray of P_k, i.e., $\Psi_k \in H_0$. Now, from (3.20), $A_\alpha \Psi_P$ is defined and

$$\lim_{k \to \infty} ||\Psi_k - A_\alpha \Psi_P|| = 0. \tag{3.26}$$

On the other hand, according to Axiom II B in the Appendix, the set $\mathcal{S}$ of all physical states is closed in the weak topology. This means that if we prove that $P_{r+1}, P_{r+2}, \cdots$ is a Cauchy sequence in the weak topology of $\mathcal{S}$, i.e., if for any $\epsilon > 0$ and any $\beta \in \hat{\mathfrak{O}}$, $B \in \mathring{\mathfrak{B}}^\beta$ [11] we can find an $N(\epsilon, \beta, B)$ such that

$$|P_k^\beta(B) - P_m^\beta(B)| < \epsilon, \; k, m > N(\epsilon, \beta, B), \tag{3.27}$$

then there exists a physical state P_∞ such that

$$P_\infty^\beta(B) = \lim_{k \to \infty} P_k^\beta(B), \quad \beta \in \hat{\mathfrak{O}}, \quad B \in \mathring{\mathfrak{B}}^\beta. \tag{3.28}$$

But (3.27) is true for the sequence $P_k^\beta(B)$ in (3.23). Namely, the sequence

$$\sum_j |\lambda_j^{(k)}|^2 P^\alpha(I_j^{(k)}), \quad k = r+1, r+2, \cdots, \tag{3.29}$$

where

$$I_j^{(k)} = [\lambda_j^{(k)}, \lambda_{j+1}^{(k)}) \tag{3.30}$$

converges towards

$$0 < \int_{-\infty}^{+\infty} \lambda^2 \, dP^\alpha(I_\lambda) = \int_{-\infty}^{+\infty} \lambda^2 \, dP\langle \Psi_P | \, E_\alpha(I_\lambda) \, | \Psi_P \rangle < +\infty, \tag{3.31}$$

and the sequence

$$\sum_{i,j} \lambda_i^{(k)} \lambda_j^{(k)} P^{\alpha;\beta;\alpha}(I_i^{(k)} \times B \times I_j^{(k)}), \quad k = r+1, r+2, \cdots \tag{3.32}$$

is also convergent. We establish this last statement by writing

$$\sum_{i,j} \lambda_i^{(k)} \lambda_j^{(k)} P^{\alpha;\beta;\alpha}(I_i^{(k)} \times B \times I_j^{(k)}) = \sum_{i,j} \lambda_i^{(k)} \lambda_j^{(k)} \langle E_\alpha(I_i^{(k)}) \Psi_P | \, F_\beta(B) \, | E_\alpha(I_j^{(k)}) \Psi_P \rangle = \langle \Psi_k | \, F_\beta(B) \, | \Psi_k \rangle. \tag{3.33}$$

Due to (3.26),

$$\lim_{k \to \infty} \langle \Psi_k | \, F_\beta(B) \, | \Psi_k \rangle = \langle A_\alpha \Psi_P | \, F_\beta(B) \, | A_\alpha \Psi_P \rangle < +\infty, \tag{3.34}$$

and the existence of P_∞ is established.

Thus we have simultaneously established that, due to (3.34), we can write

$$P_\infty^{\beta_1;\cdots;\beta_m}(B_1 \times \cdots \times B_m) = ||A_\alpha \Psi_P||^{-2} \times \langle A_\alpha \Psi_P | \, E_{\beta_1}(B_1) \cdots E_{\beta_m}(B_m) \, | A_\alpha \Psi_P \rangle, \quad \beta_1, \cdots \beta_m \in \mathfrak{O}, \quad B_1 \in \mathfrak{B}^{\beta_1}, \cdots, B_m \in \mathfrak{B}^{\beta_m}, \tag{3.35}$$

and consequently $A_\alpha \Psi_P$ belongs to the image ray of P_∞, i.e., $A_\alpha \Psi_P \in H_0$.

---

[11] If $\beta(\beta_1, \ldots, \beta_s) = \epsilon \hat{\mathfrak{O}}$, then we prove the statement, strictly speaking, only for $B \in \mathfrak{B}^{\beta_1} \times \ldots \times \beta^{\beta_s}$. However, it is trivial to extend this to $\mathring{\mathfrak{B}}^\beta$, because $\mathring{\mathfrak{B}}^\beta$ is the Boolean algebra generated by $\mathfrak{B}^{\beta_1} \times \ldots \times \mathfrak{B}^{\beta_s}$.

Finally, if, instead of (3.20), we had

$$\int_{-\infty}^{+\infty} \lambda^2 \, d\langle \Psi_P | \, E_\alpha(I_\lambda) \, | \Psi_P \rangle = 0, \tag{3.36}$$

then

$$A_\alpha \Psi_P = \mathbf{O} \in D. \tag{3.37}$$

Thus, we have proved that, to any observable α for which

$$\int_{-\infty}^{+\infty} \lambda^2 \langle \Psi | \, E_\alpha(I_\lambda) \, | \Psi \rangle < +\infty, \quad \Psi \in D, \tag{3.38}$$

there corresponds a unique, (in general) unbounded, and self-adjoined operator A_α which leaves D invariant, i.e., $A_\alpha D \subset D$. Thus, due to Axiom W5, (3.14) is proved.

Since the operators P_μ are self-adjoint, the operators

$$U(x, 1) = \exp iP_\mu x^\mu \tag{3.39}$$

exist, are unitary, and are defined on the entire $\mathfrak{H}$. They provide a representation of the translation group which is continuous in the strong topology of $\mathfrak{B}(\mathfrak{H})$:

$$\lim_{x \to y} ||(U(x, 1) - U(y, 1))\Psi|| = 0, \quad \Psi \in \mathfrak{H}. \tag{3.40}$$

In order to prove that, for any $\varphi \in \mathfrak{K}$, $f \in \mathfrak{D}$,

$$A_{\varphi(f_{(x,1)})}\Psi = e^{iP \cdot x} A_{\varphi(f)} e^{-iP \cdot x}\Psi, \quad \Psi \in D, \tag{3.41}$$

we make use of Axiom W2. Namely, inserting in (2.13) for v_1 an arbitrary $\alpha = (\alpha_1, \cdots, \alpha_m) \in \hat{\mathfrak{O}}$ and $B' = B_1' \times \cdots \times B_m' \in \mathring{\mathfrak{B}}^\alpha$, and for v_2 any $\beta = (\beta_1, \cdots, \alpha_n) \in \hat{\mathfrak{O}}$ and $B'' = B_1'' \times \cdots \times B_n'' \in \mathring{\mathfrak{B}}^\beta$, we can write, due to (3.7), for any $\Psi_P \in H_0$:

$$\begin{aligned}
&\langle \Psi_P | \, F_\alpha(B') E_{\varphi(f_{(x,1)})}(B) F_\beta(B'') \, | \Psi_P \rangle \\
&= P^{\alpha;\varphi(f_{(x,1)});\beta}(B' \times B \times B'') \\
&= \int e^{i(p'-p'') \cdot x} \, dP^{\alpha;\pi;\varphi(f);\pi;\beta} \\
&\qquad \times (B' \times I_{p'} \times B \times I_{p''} \times B'') \\
&= \int e^{i(p'-p'') \cdot x} \, d\langle \Psi_P | \, F_\alpha(B') E_\pi(I_{p'}) \\
&\qquad \times E_{\varphi(f)}(B) E_\pi(I_{p''}) F_\beta(B'') \, | \Psi_P \rangle \\
&= \langle \Psi_P | F_\alpha(B') e^{iP \cdot x} E_{\varphi(f)}(B) e^{-iP \cdot x} F_\beta(B'') \, | \Psi_P \rangle,
\end{aligned} \tag{3.42}$$

where for $F_\alpha(B')$ and $F_\beta(B'')$ we used the notation introduced in (3.24), while

$$\begin{aligned}
E_\pi(I_p) &= E_{\pi^0}(I_{p^0}) \cdots E_{\pi^3}(I_{p^3}), \\
p &= (p^0, \cdots, p^3).
\end{aligned} \tag{3.43}$$

Now, from the way in which the space $\mathfrak{H}$ is constructed,[12] it follows that the linear manifold, spanned by all the vectors

$$F_\alpha(B)\Psi_P, \quad \alpha \in \hat{\mathfrak{O}}, \quad B \in \mathring{\mathfrak{B}}^\alpha \tag{3.44}$$

is everywhere dense in $\mathfrak{H}^P$ in the norm-topology. Consequently, (3.42) can be true if and only if

$$E_{\varphi(f_{(x,1)})}(B) = e^{iP \cdot x} E_{\varphi(f)}(B) e^{-iP \cdot x}. \tag{3.45}$$

For arbitrary $\Psi_1, \Psi_2 \in D$ the integrals

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \lambda \, d\langle \Psi_1 | \, E_{\varphi(f_{(x,1)})}(I_\lambda) \, | \Psi_2 \rangle \\
&\quad = \int_{-\infty}^{+\infty} \lambda \, d\langle \Psi_1 | \, e^{iP \cdot x} E_{\varphi(f)}(I_\lambda) e^{-iP \cdot x} \, | \Psi_2 \rangle
\end{aligned} \tag{3.46}$$

exist. Relation (3.46) is equivalent to (3.41). Thus the fact is established that all the conditions of Wightman's Axiom O (with the exception of the existence of the global vacuum) are fulfilled; namely, the last requirement of this axiom that has not yet been checked is the spectral condition. This condition is obviously fulfilled because, naturally, the spectrum S^π of π and the joint spectrum of $P^0, \cdots, P^3$, defined as the complement of the set

$$\begin{aligned}
&\bigcup (B_0 \times B_1 \times B_2 \times B_3), \quad B_0, \cdots, B_3 \in \mathfrak{B}^1, \\
&\qquad E_{\pi^0}(B_0) \cdots E_{\pi^3}(B_3) = \mathbf{O}
\end{aligned} \tag{3.47}$$

are identical, and $S^\pi \subset V_+$ according to Axiom W2.

In the same manner as for the translation group, we can show that, due to Axiom W3, if any rotation $\Lambda \in L_+^\uparrow$ in the (μ, ν) plane is given, then we have[13]

$$\begin{aligned}
A_{\varphi_j(f_{(0,\Lambda)})}\Psi &= \sum_k S_{jk}(\Lambda) \exp\left(-\tfrac{1}{2} i M_{\mu\nu} \Lambda^{\mu\nu}\right) \\
&\quad \times A_{\varphi_k(f)} \exp\left(\tfrac{1}{2} i M_{\mu\nu} \Lambda^{\mu\nu}\right)\Psi, \quad \Psi \in D,
\end{aligned} \tag{3.48}$$

and, of course, $\exp\left(\frac{1}{2} i M_{\mu\nu} \Lambda^{\mu\nu}\right) D \subset D$.

Due to the fact that any Lorentz transformation $\Lambda \in L_+^\uparrow$ can be reduced to three consecutive rotations of the above type, we can define, on the basis of (3.48), a representation $U(0, \Lambda)$ of $L_+^\uparrow$ having the property

$$U(0, \Lambda) \, D \subset D. \tag{3.49}$$

Thus Wightman's Axiom II[14] is also fulfilled.

---

[12] Consult the proof of Theorem 3 in [3], Sec. 3.3.

[13] We do not have in $\exp\left(\frac{1}{2} i M_{\mu\nu} \Lambda^{\mu\nu}\right)$ any problems with the noncommutativity of the $M_{\mu\nu}$, because, for $\kappa \neq \mu$, $\lambda \neq \nu$, $\Lambda^\kappa{}_\lambda = \delta_{\kappa\lambda}$ if Λ is a rotation in the (μ, ν) plane. As $M_{\mu\nu} = -M_{\mu\nu}$, the sum $\frac{1}{2} M_{\mu\nu} \Lambda^{\mu\nu}$ reduces to $M_{\mu\nu} \Lambda^{\mu\nu}$ (no summation over μ and ν).

[14] Cf. Ref. 7, p. 99.

Finally, we can deduce from Axiom W4 that, if

$$f(x)g(y) = 0, \quad x, y \in \mathfrak{M}, \quad f, g \in \mathfrak{D}, \tag{3.50}$$

then, for arbitrary $\alpha, \beta \in \hat{\mathfrak{O}}$, $B' \in \mathring{\mathfrak{B}}^{\alpha}$, $B'' \in \mathring{\mathfrak{B}}^{\beta}$ and arbitrary $\Psi_P \in H_0$,

$$\begin{aligned}
&\langle \Psi_P |\, F_\alpha(B')E_{\varphi_1(f)}(B_1)E_{\varphi_2(g)}(B_2)F_\beta(B'')\,|\Psi_P\rangle \\
&= P^{\alpha;\varphi_1(f);\varphi_2(g);\beta}(B' \times B_1 \times B_2 \times B'') \\
&= P^{\alpha;\varphi_2(g);\varphi_1(f);\beta}(B' \times B_2 \times B_1 \times B'') \\
&= \langle \Psi_P |\, F_\alpha(B')E_{\varphi_2(g)}(B_2)E_{\varphi_1(f)}(B_1)F_\beta(B'')\,|\Psi_P\rangle, \\
&\qquad \varphi_1, \varphi_2 \in \mathfrak{K}, \quad B_1 \in \mathfrak{B}^{\varphi_1}, \quad B_2 \in \mathfrak{B}^{\varphi_2}.
\end{aligned} \tag{3.51}$$

From the earlier mentioned fact that (3.44) is everywhere dense in $\mathfrak{H}^P$ in the norm-topology, we conclude from (3.51) that

$$E_{\varphi_1(f)}E_{\varphi_2(g)} = E_{\varphi_2(g)}E_{\varphi_1(f)} \tag{3.52}$$

and thus

$$[A_{\varphi_1(f)}A_{\varphi_2(g)} - A_{\varphi_2(g)}A_{\varphi_1(f)}]\Psi = \mathbf{O}, \quad \Psi \in D. \tag{3.53}$$

4. COMPARISON WITH THE HAAG–ARAKI ALGEBRAIC FRAMEWORK

The main concept of the Haag–Araki[15] approach to quantum field theory is an algebra of observables in which some suitable topology is introduced. We consider here the case of the algebra of "quasi-local observables" introduced by Haag and Kastler,[16] in which the uniform topology is adopted. We refer to the six postulates introduced in the Introduction of Ref. 16 as the Axioms H_1–H_6. We intend to prove the following theorem.

Theorem 2: In an ℒQFT obeying Axioms W1–W4, a B^*-algebra generated by local observables can be associated with any finite region in space–time. These algebras fulfill the conditions of Axioms H_1–H_5 in Ref. 16.

We denote by Δ open sets with compact (in the Euclidean metric) closure in the Minkowski space $\mathfrak{M}$. In the Haag–Araki formalism observables are attached to each such set Δ.

If we adopt Axioms W1–W4, then we denote by $\mathcal{O}(\Delta)$ the set consisting of all the fields $\varphi \in \mathfrak{K}$ taken at all $\bar{f} \in \mathfrak{D}$, $f = \bar{f}$, with support lying within Δ, and of all the observables which are functions of any finite number of such compatible field observables, i.e.,

$$\begin{aligned}
\mathcal{O}(\Delta) = \{&F[\varphi_1(f_1), \cdots \varphi_n(f_n)]: \\
&f_1 = \bar{f}_1, \cdots f_n = \bar{f}_n \in \mathfrak{D}, \\
&\text{supp } f_1, \cdots, \text{supp } f_n \subset \Delta, \\
&\{\varphi_1(f_1), \cdots \varphi_n(f_n)\} = C, \quad F \in \mathcal{A}\},
\end{aligned} \tag{4.1}$$

where $\mathcal{A}$ stands for the set of all Borel functions defined on finite-dimensional Euclidean spaces.

According to Theorem 6 in Sec. 2.5 of [3], and Theorem 1 in Sec. 4, of [4], there exists a B^*-algebra $\mathfrak{A}(\mathcal{O})$ into which the set $\mathcal{O}$ of all observables can be embedded. Denote by $\mathfrak{A}(\Delta)$ the smallest B^*-algebra containing the image in $\mathfrak{A}(\mathcal{O})$ of the set (4.1). These algebras $\mathfrak{A}(\Delta)$ can be taken to be the algebras introduced by Axiom H_1.

Due to the fact that $\mathcal{O}(\Delta_1) \subset \mathcal{O}(\Delta_2)$ if $\Delta_1 \subset \Delta_2$, it follows immediately that the isotony property,

$$\mathfrak{A}(\Delta_1) \subset \mathfrak{A}(\Delta_2), \quad \Delta_1 \subset \Delta_2, \tag{4.2}$$

is satisfied. It is also clear that all the algebras $\mathfrak{A}(\Delta)$ contain the unit element—which is the image of all trivial questions like

$$F[\varphi(f)], \quad F(\lambda) \equiv 1, \quad \lambda \in \mathbf{R}^1. \tag{4.3}$$

Therefore Axiom H_2 is also satisfied.

If Δ_1 and Δ_2 are spacelike with respect to each other, then, due to Axiom W4, each element of $\mathcal{O}(\Delta_1)$ is compatible with any element of $\mathcal{O}(\Delta_2)$. According to the definition of the concept of embedding[17] of the set $\mathcal{O}$ into $\mathfrak{A}(\mathcal{O})$, this compatibility is reflected in the commutativity of the images in $\mathfrak{A}(\mathcal{O})$ of the elements of $\mathcal{O}(\Delta_1)$ with the images in $\mathfrak{A}(\mathcal{O})$ of the elements of $\mathcal{O}(\Delta_2)$. This implies the commutativity of the B^*-algebras $\mathfrak{A}(\Delta_1)$ and $\mathfrak{A}(\Delta_2)$; we derive such a statement from the fact that any element of an algebra $\mathfrak{A}(\Delta)$ is obtained by taking finite sums and products (in any order) of a finite number of images in $\mathfrak{A}(\mathcal{O})$ of elements from $\mathcal{O}(\Delta)$, or by taking uniform limits of such. Consequently, Axiom H_3 is fulfilled.

Denote by $\mathfrak{A}$ the smallest B^*-subalgebra of $\mathfrak{A}(\mathcal{O})$ containing the image of the set

$$\begin{aligned}
\mathcal{O}_\varphi = \{&F[\varphi_1(f_1), \cdots, \varphi_n(f_n))\,;\, F \in \mathcal{A}, \varphi_1, \cdots, \\
&\varphi_n \in \mathfrak{K}, f_1 = \bar{f}_1, \cdots, f_n = \bar{f}_n \in \mathfrak{D}\}.
\end{aligned} \tag{4.4}$$

This algebra does not coincide with the B^*-algebra $\mathfrak{A}(\mathcal{O})$, because all the global observables are not included in (4.4)—corresponding to the fact that these global quantities do not belong to the category of "observables" in the sense in which Haag uses the word.

[15] H. Araki, Progr. Theoret. Phys. (Kyoto) **32**, 844 (1964).
[16] J. R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[17] Cf. [3], Sec. 1, and [4], Sec. 3, Def. 4.

It is obvious that

$$\mathcal{O}_\varphi = \bigcup_\Delta \mathcal{O}(\Delta). \tag{4.5}$$

Consequently we have

$$\mathfrak{A} \supset \bigcup_\Delta \mathfrak{A}(\Delta). \tag{4.6}$$

If we denote by $\mathfrak{A}_1(\Delta)$ and $\mathfrak{A}_1$ the algebras generated (in the ordinary algebraic sense) by the images of $\mathcal{O}(\Delta)$ and $\mathcal{O}_\varphi$, respectively, in $\mathfrak{A}(\mathcal{O})$, than it is easy to prove that

$$\mathfrak{A}_1 = \bigcup_\Delta \mathfrak{A}_1(\Delta). \tag{4.7}$$

In general, however, we cannot extend (4.7) to transform (4.6) into an identity.

The B^*-algebra $\mathfrak{A}$ is obviously the smallest B^*-sub-algebra of $\mathfrak{A}(\mathcal{O})$ which contains all the algebras $\mathfrak{A}(\Delta)$. It is the algebra of "quasi-local observables" postulated in Axiom H_4.

From Axioms W2 and W3 ,we can immediately infer that, under an inhomogeneous Lorentz transformation (a, Λ), $\Lambda \in L_+^\uparrow$, the set $\mathcal{O}(\Delta)$ goes over into

$$\mathcal{O}_{(a,\Lambda)}(\Delta) = \mathcal{O}(a + \Lambda \cdot \Delta). \tag{4.8}$$

This implies the transformation

$$\mathfrak{A}(\Delta) \to \mathfrak{A}_{(a,\Lambda)}(\Delta) = \mathfrak{A}(a + \Lambda \cdot \Delta), \tag{4.9}$$

and consequently the existence of an automorphism of $\mathfrak{A}$—due to the fact that the set $\bigcup_\Delta \mathfrak{A}(\Delta)$ is everywhere dense (in the uniform topology) in $\mathfrak{A}$. This is precisely what is required in Axiom H_5.

Finally, we know that the £-formalism admits a Hilbert space representation.[18] This means that $\mathfrak{A}(\mathcal{O})$, and consequently $\mathfrak{A}$, possess a faithful representation in a Hilbert space. We do not know, however, whether this representation of $\mathfrak{A}$ is algebraically irreducible, i.e., whether $\mathfrak{A}$ is primitive.

5. THE CONCEPT OF LOCAL VACUUM

In the conventional quantum field theory as well as in the Wightman axiomatic approach, the existence of a (unique) ground state, called the vacuum, is assumed. The introduction of such a state of the global total energy-momentum is, at once, convenient mathematically, and seemingly justified from the physical point of view. Namely, the existence of such a mathematical object seems to be necessitated by the *imaginable* "physical" situation in which a physical vacuum is realized in the entire universe.

However, the existence of such a vacuum state, which we call from now on a *global vacuum*, seems much less justified once it is realized that *any* actual experiment is and *can be carried out only* in a finite region of space–time. If that is the case, the *ad hoc* introduction of a global vacuum is not physically justified and can impose a very serious restriction on the structure of a theory. Therefore, the desirable thing would be either to *derive* the existence of such a vacuum from more plausible assumptions (as has been carried out under specific circumstances by Borchers[19]), or to replace this stringent requirement of the global vacuum with some more physical assumption.

We do not know whether a result like Borcher's can be derived in the formulated £QFT. Namely, although (see Sec. 4) the £QFT can be reformulated in terms of B^*-algebras of "quasi-local observables" and Haag's concept of physical equivalence applied to it, we do not know whether the weak additivity property used by Borchers in his proof is valid for the resulting von Neumann algebras. On the other hand, in the £QFT formalism the assumption of the existence of a global vacuum would introduce no remarkable computational or formal simplification—such as in ℜQFT. Furthermore, the transition to a physically equivalent theory with a vacuum (if such a theory existed) might be quite difficult in practice. Therefore, it is desirable to have an alternative to the global vacuum.

Definition 1: A physical state P_0 is said to represent a local vacuum over a space–time domain Δ, which is bounded in the Euclidean metric,

$$x_0^2 + x_1^2 + x_2^2 + x_3^2, \tag{5.1}$$

if and only if

$$P_0^{\overline{\varphi^*(f)}} = \int \lambda^2 \, dP_0^{\varphi(f)}(I_\lambda) = 0, \tag{5.2}$$

for all $f = \bar{f} \in \mathfrak{D}$, supp $f \subset \Delta$, and for all field components $\varphi \in \mathfrak{F}_c$ from the family $\mathfrak{F}$ of relativistic fields (mentioned in Axiom F3, Sec. 2).

In (5.2) the convenient notation of a bar over an observable in a physical state indicates the fact that the mean value has been taken over that observable for that state; i.e., for any $\beta, \gamma \in \hat{\mathcal{O}}$, $B' \in \mathfrak{B}^\beta$, $B'' \in \mathfrak{B}^\gamma$,

$$P^{\beta;\bar{\alpha};\gamma}(B' \times B'') = \int_{-\infty}^{+\infty} \lambda \, dP^{\beta;\alpha;\gamma}(B' \times I_\lambda \times B'') \tag{5.3}$$

for any observable α.

[18] Cf. Theorem 3 in [3], Sec. 3.3, or Theorem 3 in [4], Sec. 5.

[19] H. J. Borchers, Commun. Math. Phys. 1, 57 (1965).

It is easy to see that (5.2) is equivalent to the requirement:

$$P^{\varphi(f)}(B) = 0, \quad B \in \mathfrak{B}^{\varphi}, \quad \{0\} \cap B = \varnothing. \tag{5.4}$$

Due to Axiom IV we can prove the following:

Theorem 3:[20] If $P_0 \in \mathcal{S}$ is a local vacuum over Δ, then

$$P_0^{\varphi_1(f_1),\cdots,\varphi_n(f_n)}(B) = 0,$$
$$B \in \mathfrak{B}^{\varphi_1\times\cdots\times\varphi_n}, \quad \{0\} \cap B = \varnothing, \tag{5.5}$$

for any $\varphi_1, \cdots, \varphi_n \in \mathfrak{F}_c$, and any $f_k = \bar{f}_k \in \mathfrak{D}$, supp $f_k \subset \Delta (k = 1, \cdots, n)$.

Proof: In order to prove (5.5), we first assume that $\varphi_1, \cdots, \varphi_n$ are bounded fields, i.e., that $S^{\varphi_1}, \cdots, S^{\varphi_n}$ are bounded sets in $\mathbf{R}^1$. Then we can employ the fundamental theorem of [3] according to which there exists a Banach *-algebra $\mathfrak{A}(\mathcal{O})$, self-adjoint indempotent elements $E_{\varphi_1(f_1)}(B_1), \cdots, E_{\varphi_n(f_n)}(B_n)$ of $\mathfrak{A}(\mathcal{O})$ and a positive-definite linear functional $< x >_{P_0}$, $x \in \mathfrak{A}(\mathcal{O})$, on $\mathfrak{A}(\mathcal{O})$, which are such that

$$P_0^{\varphi_1(f_1),\cdots;\varphi_n(f_n)}(B_1 \times \cdots \times B_n)$$
$$= \langle E_{\varphi_1(f_1)}(B_1) \cdots E_{\varphi_n(f_n)}(B_n)\rangle_{P_0},$$
$$B_1 \in \mathfrak{B}^{\varphi_1}, \cdots, B_n \in \mathfrak{B}^{\varphi_n}. \tag{5.6}$$

By applying on the positive functional $< x >_{P_0}, x \in \mathfrak{A}(\mathcal{O})$, the Schwartz–Cauchy inequality, we get

$$|\langle E_{\varphi_1(f_1)}(B_1) \cdots E_{\varphi_n(f_n)}(B_n)\rangle_{P_0}|^2$$
$$\leq \langle E_{\varphi_1(f_1)}(B_1)E_{\varphi_1(f_1)}(B_1)\rangle_{P_0}\langle E_{\varphi_n(f_n)}(B_n) \cdots$$
$$\times E_{\varphi_2(f_2)}(B_2)E_{\varphi_2(f_2)}(B_2) \cdots E_{\varphi_n(f_n)}(B_n)\rangle_{P_0}. \tag{5.7}$$

Now, by employing again the results of [3] and afterwards Axiom IB, we can write

$$\langle E_{\varphi_1(f_1)}(B_1)E_{\varphi_1(f_1)}(B_1)\rangle_{P_0}$$
$$= P_0^{\varphi_1(f_1);\varphi_1(f_1)}(B_1 \times B_1) = P_0^{\varphi_1(f_1)}(B_1). \tag{5.8}$$

According to (5.4), which is equivalent to (5.2),we have

$$P^{\varphi_1(f_1)}(B_1) = 0, \quad \{0\} \cap B_1 = \varnothing, \quad B_1 \in \mathfrak{B}^{\varphi_1}. \tag{5.9}$$

Thus, due to (5.6), (5.7), and (5.8), the proposition is proved for $B \in \mathfrak{B}^{\varphi_1\times\cdots\times\varphi_n}$ of the form $B_1 \times \cdots \times B_n$, $\{0\} \cap B_1 = \varnothing$, and $\varphi_1, \cdots, \varphi_n$ bounded.

In the case B_1 contains the zero point of the spectrum of φ_1, (i.e., $\{0\} \subset B_1$), we have on the basis of the above and of Axiom IA that

$$P_0^{\varphi_1(f_1);\cdots;\varphi_n(f_n)}(B_1 \times \cdots \times B_n)$$
$$= P_0^{\varphi_1(f_1);\varphi_2(f_2);\cdots;\varphi_n(f_n)}(\mathbf{R}^1 \times B_2 \times \cdots \times B_n)$$
$$= P_0^{\varphi_2(f_2);\cdots;\varphi_n(f_n)}(B_2 \times \cdots \times B_n). \tag{5.10}$$

Hence, we have reduced this case to the previous case, where we deal with $n - 1$ fields.

An unbounded field $\varphi_k(f)$ can be approximated by a bounded field $\varphi_k^{(N)}(f)$, where $\varphi_k^{(N)}$ is defined as the function of φ_k in the following way:

$$\varphi_k^{(N)} = F_N[\varphi_k],$$
$$F_N(\lambda) = \begin{cases}\lambda, & -N \leq \lambda \leq +N,\\ 0, & |\lambda| > N.\end{cases} \tag{5.11}$$

We then get

$$P_0^{\varphi_1(f_1);\cdots;\varphi_n(f_n)}(B_1 \times \cdots \times B_n) = 0,$$
$$B_1 \times \cdots \times B_n \in \mathfrak{B}^{\varphi_1\times\cdots\times\varphi_n}, \tag{5.12}$$
$$\{0\} \cap (B_1 \times \cdots \times B_n) = \varnothing$$

for unbounded fields $\varphi_1, \cdots, \varphi_n$ by going in the above procedure to the limit $N \to +\infty$.

The fact that (5.12) is valid not only for Borel sets of the form $B_1 \times \cdots \times B_n$, but also for arbitrary $B \in \mathfrak{B}^{\varphi_1\times\cdots\times\varphi_n}$ follows from basic theorems of measure theory.[21] Q.E.D.

Naturally, the above theorem implies that

$$P_0^{\overline{\varphi_1(f_1)};\cdots;\overline{\varphi_n(f_n)}} = 0, \tag{5.13}$$

$$P_0^{\overline{[\varphi_1(f_1)]^2};\cdots;\overline{[\varphi_n(f_n)]^2}} = 0 \tag{5.14}$$

for arbitrary $\varphi_1, \cdots, \varphi_n \in \mathfrak{F}_c$, $f_k = \bar{f}_k \in \mathfrak{D}$, supp $f_k \subset \Delta$, $k = 1, \cdots, n$. However, due to the fact that the measures $P_0^{\varphi_1(f_1),\cdots,\varphi_n(f_n)}(B)$ do not, in general, have to be positive definite, the opposite is not true; (5.5) is a stronger result than (5.13) and (5.14).

We can now introduce:

Postulate 1: Given any finite (in the Euclidean metric) region of space–time, there is at least one local vacuum over that region.

However, due to the fact that one always makes measurements only with a nonvanishing error, and, consequently, that one can determine only weak neighborhoods of physical states (see [1] and the Appendix), the above postulate is still more stringent then the existence of experimentally realizable vacuum states would require. Namely, it would be

[20] Our main interest in this theorem is due to the possibility that a physical interpretation of complex probability measures (as was proposed in [1]) assigned to incompatible observables might prove feasible and desirable.

[21] P. R. Halmos, *Measure Theory* (D. Van Nostrand, Inc., Princeton, New Jersey, 1950), Sec. 13, Theorem A, and Secs. 33 and 35.

sufficient to introduce the following weaker version of Postulate 1.

Postulate 2: Given any finite region of space–time, there exists, for any $\epsilon > 0$, and ϵ-approximation of a local vacuum over that region.

In the above axiom we used the following concept:

Definition 2: Given any finite region Δ in space–time, we say that a physical state P_ϵ provides an ϵ-approximation of a local vacuum over Δ if we have

$$|P_\epsilon^{\overline{\varphi^*(f)}}| < \epsilon, \qquad |f(x)| \leq 1, \tag{5.15}$$

for all $\varphi \in \mathfrak{F}_c$, $f = \bar{f} \in \mathfrak{D}$, supp $f \subset \Delta$.

From (5.15) we can derive the following estimate

$$P_\epsilon^{\varphi(f)}(B_\epsilon) < \epsilon, \quad B_\epsilon = (-\infty, \epsilon] \cup [\epsilon, +\infty) \tag{5.16}$$

for any f of the above type.

In general, we can write that, if $f_k = \bar{f}_k \in \mathfrak{D}$, supp $f_k \subset \Delta (k = 1, \cdots n)$,

$$|P_\epsilon^{\varphi_1(f_1);\cdots;\varphi_n(f_n)}(B_\epsilon \times \cdots \times B_\epsilon)| \leq \epsilon,$$
$$|f_k(x)| \leq 1, \tag{5.17}$$

for any $\varphi_1, \cdots, \varphi_n \in \mathfrak{F}_c$. This inequality might be considered as a version of the earlier proved theorem for the local vacuum. We can prove (5.17) by first taking bounded fields and employing (5.7) [first on $E_{\varphi_1(f_1)}(B_\epsilon)$ and then on $E_{\varphi_n(f_n)}(B_\epsilon)$], always remembering that

$$|(E_{\alpha_1}(B_1) \cdots E_{\alpha_n}(B_n))_P| \leq 1, \quad \alpha_1, \cdots, \alpha_n \in \mathfrak{O},$$
$$B_1 \in \mathfrak{B}^{\alpha_1}, \cdots, B_n \in \mathfrak{B}^{\alpha_n}, \quad P \in \mathfrak{S}. \tag{5.18}$$

Afterwards we can generalize (5.17) to the case of unbounded fields by the method used in the proof of Theorem 3.

As for the concept of local vacuum, we note that there is no guarantee, and no compelling reason, why a state which is a local vacuum in some region should be a pure physical state, and consequently representable in 3CQFT by a Hilbert vector. Namely, the concept of local vacuum has been introduced in order to cover the case when a specific field theory is believed to refer also to the case when an actual experimental vacuum (i.e., a region of space–time not containing any systems described by that field theory) is realizable in some (finite) region of space–time. As we have argued, in such a case it is not necessary to introduce a global vacuum. Furthermore, it is not even necessary to adopt Postulate 1. It would be sufficient, from the physical point of view, to have Postulate 2 valid.

We would like to remark that, if a specific field theory possesses a global vacuum, then it seems to us that, if the theory has been correctly built and if that vacuum state is more than a mathematical convenience, such a global vacuum should be a local vacuum, in the sense of Definition 1, over any finite region Δ of space–time (this is certainly true, at least in the case of that mathematical abstraction known as "free fields"). However, we expect that the converse is not true; there might exist interesting quantum field theories which satisfy Postulate 2, or even the more restricted Postulate 1, without possessing a global vacuum.

ACKNOWLEDGMENT

The author would like to express his appreciation to the Theoretical Physics Institute of the University of Alberta for awarding him a Postdoctoral Fellowship.

APPENDIX

For the purpose of this paper, it is desirable to give a short summary of the most important concepts and results of the previous papers [1]–[4]. Some of these concepts and results are somewhat modified here to suit the needs of this paper.

The general formalism with which we are concerned (and to which we sometimes refer as the £-formalism) is given in the form of four axioms (see [4]). The basic concepts in terms of which these axioms are formulated are the set of observables $\mathfrak{O}$ and the set of physical states $\mathfrak{S}$. We do not impose any artificial *a priori* structure on these sets (e.g., we do not require that $\mathfrak{O}$ be an algebra or that all pure physical states constitute a Hilbert space). All the required structure is contained in the four axioms which, with the exception of the last one, represent a set of requirements which occur in a natural manner when a certain probabilistic interpretation of the basic concepts of the given formalism (see [1]) is adopted. The last axiom is introduced in order to make the formalism physically equivalent to the Hilbert-space formalism.

The axioms, as we present them here, are essentially those given in [4]. However, we do not formulate them in the most general form by means of probability functionals; rather we adopt the more specialized formalism in terms of complex probability measures, which intuitively is more transparent and quite adequate for our present needs.

In order to facilitate the notation, we introduce the following conventions:

If a set S is given, we denote by $\hat{S}$ the family of all ordered finite sets of elements belonging to the set

S. To any n-tuple $\alpha = (\alpha_1, \cdots \alpha_n) \in \hat{\mathfrak{O}}$ of one-dimensional observables $\alpha_1, \cdots, \alpha_n$ we assign an n-dimensional Euclidean space $\mathbf{R}^\alpha$ and denote with $\mathfrak{B}^\alpha$ the family of all Borel sets on $\mathbf{R}^\alpha$, i.e., the Boolean *σ-algebra* generated by $\mathfrak{B}^{\alpha_1} \times \cdots \times \mathfrak{B}^{\alpha_n}$, while by $\mathring{\mathfrak{B}}^\alpha$ we denote the Boolean *algebra* generated by $\mathfrak{B}^{\alpha_1} \times \cdots \times \mathfrak{B}^{\alpha_n}$.[22]

In the $\mathfrak{L}$-formalism, we assign to each n-tuple $\alpha \in \hat{\mathfrak{O}}$ of observables, and to each physical state $P \in \mathfrak{S}$ a complex probability measure $P^\alpha(B)$, $B \in \mathfrak{B}^\alpha$, on $\mathbf{R}^\alpha$. We define a complex probability measure $P^\alpha(B)$ as a normalized [i.e., $P^\alpha(\mathbf{R}^\alpha) = 1$] complex measure for which $|P^\alpha(B)| < +\infty$ for all $B \in \mathring{\mathfrak{B}}^\alpha$.[23] If we have $B_1 \in \mathfrak{B}^{\alpha_1 \times \cdots \times \alpha_m}$, $B_2 \in \mathfrak{B}^{\beta_1 \times \cdots \times \beta_n}$, $\alpha_1, \cdots, \beta_n \in \mathfrak{O}$, in order to denote that, in $P^{\alpha_1, \cdots, \alpha_m, \beta_1 \cdots, \beta_n}(B_1 \times B_2)$, B_1 refers to the first m observables, we write $P^{\alpha_1, \cdots, \alpha_m; \beta_1, \cdots, \beta_n}(B_1 \times B_2)$. In general, in an expression of the form $P^{\alpha;\beta;\cdots}(B_1 \times B_2 \times \cdots)$, $\alpha, \beta, \cdots \in \hat{\mathfrak{O}}$, the lth factor B_l in the direct product $B_1 \times B_2 \times \cdots$ refers to the lth ordered set α_l of observables contained between the $(l-1)$th and lth *semicolons*.

If a relation is true for all values $\alpha \in \hat{\mathfrak{O}}$ and for all $B \in \mathfrak{B}^\alpha$, it is very useful to replace α with the "variable" symbol v. For example, the so abbreviated relation (A2) below stands for the following more detailed written expression

$$P^{\beta;\alpha;\gamma}(B_1 \times \mathbf{R}^\alpha \times B_2) = P^{\beta;\gamma}(B_1 \times B_2)$$
$$\beta, \gamma \in \hat{\mathfrak{O}}, \quad B_1 \in \mathfrak{B}^{\alpha_1}, \quad B_2 \in \mathfrak{B}^{\alpha_2}. \tag{A1}$$

Axiom I:[24] For a given set $\mathfrak{O}$ of observables, a physical state P is defined by assigning to each finite ordered set $\alpha \in \hat{\mathfrak{O}}$ of observables a complex probability measure $P^\alpha(B)$, $B \in \mathfrak{B}^\alpha$. Each physical state P has to satisfy the following relations:

(A). For any observable $\alpha \in \mathfrak{O}$

$$P^{\nu_1 \alpha \nu_2}(\mathbf{R}^\alpha) = P^{\nu_1 \nu_2}. \tag{A2}$$

(B) If the symbols $\alpha_1, \cdots, \alpha_n$ stand for the same observable α, then

$$P^{\nu_1 \alpha_1; \cdots; \alpha_n \nu_2}(B_1 \times \cdots \times B_n) = P^{\nu_1 \alpha \nu_2}(B_1 \cap \cdots \cap B_n) \tag{A3}$$

for any $B_1, \cdots, B_n \in \mathfrak{B}^\alpha$.

(C) For any observable $\alpha \in \mathfrak{O}$ and any $B \in \mathfrak{B}^\alpha$

$$P^\alpha(B) \geq 0, \quad P^\alpha(\mathbf{R}^\alpha) = 1. \tag{A4}$$

[22] Cf. [2], Sec. 2. 1, and [3], Sec. 2. 5, Lemma 10.

[23] Note the differences between the present definition and that in [2], Sec. 2.2, where it was required that $[P^\alpha(B)] < +\infty$ for all $B \in \mathfrak{B}^\alpha$. We will mention in Ref. 27 the effect of this change.

[24] This axiom can be derived from Axiom IV and Axiom IB.

Definition 1: A finite set $\{\alpha_1, \cdots, \alpha_n\}$ of observables is called a set of compatible observables, symbolically written $\{\alpha_1, \cdots, \alpha_n\} = C$, if and only if

$$P^{\alpha_1, \cdots, \alpha_n}(B) \geq 0, \quad B \in \mathfrak{B}^{\alpha_1 \times \cdots \times \alpha_n},$$
$$P^{\nu_1 \alpha_1; \cdots; \alpha_n \nu_2}(B_1 \times \cdots \times B_n) = P^{\nu_1 \alpha_{k_1}; \cdots; \alpha_{k_n} \nu_2}(B_{k_1} \times \cdots \times B_{k_n}) \tag{A5}$$
$$B_1 \in \mathfrak{B}^{\alpha_1}, \cdots, B_n \in \mathfrak{B}^{\alpha_n},$$

for any physical state P and any permutation $k_1, \cdots, k_n$ of the indices, $1, \cdots, n$.

Definition 2: Two physical states P_1 and P_2 are called equal if and only if

$$P_1^\alpha(B) = P_2^\alpha(B) \tag{A6}$$

for all $\alpha \in \hat{\mathfrak{O}}$, $B \in \mathfrak{B}^\alpha$.

Definition 3: Two observables α and β are called equal if and only if

$$P^{\nu_1 \alpha \nu_2}(B) = P^{\nu_1 \beta \nu_2}(B) \tag{A7}$$

for all $B \in \mathfrak{B}^\alpha \equiv \mathfrak{B}^\beta$.

Definition 4: The weak topology[25] in the set $\mathfrak{S}$ of all physical states is the topology in which the neighborhood basis of a physical state $P_0 \in \mathfrak{S}$ consists of all the states

$$W(P_0; \alpha; B_1, \cdots, B_n; \epsilon) = \{P: |P^\alpha(B_1) - P_0^\alpha(B_1)| < \epsilon, \cdots |P^\alpha(B_n) - P_0^\alpha(B_n)| < \epsilon\}, \tag{A8}$$

corresponding to all choices of $\alpha \in \hat{\mathfrak{O}}$, $B_1, \cdots, B_n \in \mathring{\mathfrak{B}}^\alpha$, $\epsilon > 0$.

Axiom II:

(A) If P_1 and P_2 are any two physical states and $0 \leq t \leq 1$, then the family of all complex probability measures

$$tP_1^\alpha(B) + (1-t)P_2^\alpha(B), \quad B \in \mathfrak{B}^\alpha, \tag{A9}$$

corresponding to all $\alpha \in \hat{\mathfrak{O}}$, determines a (unique) physical state P, i.e., $\mathfrak{S}$ is a convex set.

(B) $\mathfrak{S}$ is closed in the weak topology.

As we can notice from Axiom IIA above, our concept of physical state refers not only to pure states but also to mixed states—which, in the Hilbert space formalism, are represented by statistical (density) operators.

Definition 5: A point $\lambda \in \mathbf{R}^\alpha$ belongs to the spec-

[25] Note the difference between the present definition in which only $B_1, \ldots, B_n \in \mathring{\mathfrak{B}}^\alpha$ are considered, and that employed in [2], where $B_1, \ldots, B_n \in \mathfrak{B}^\alpha$ are considered.

trum S^α of $\alpha \in \hat{\mathfrak{O}}$ if and only if, for any interval $I \subset R^\alpha$ containing λ, there can be found at least one physical state P for which $P^\alpha(\lambda) \neq 0$.

Axiom III:

(A) If $\alpha_1, \cdots, \alpha_n$ are any n observables ($n = 1, 2, \cdots$), then

$$S^{\alpha_1 \times \cdots \times \alpha_n} \subset S^{\alpha_1} \times \cdots \times S^{\alpha_n}. \quad \text{(A10)}$$

(B) If $\{\alpha\} = C$, $\alpha = (\alpha_1, \cdots, \alpha_n) \in \hat{\mathfrak{O}}$, and B is a Borel set in $\mathbf{R}^\alpha$ containing a point of the spectrum S^α of α, then there exists at least one physical state P for which $P^\alpha(B) = 1$.

Finally, we have an axiom whose necessity is not immediately obvious. It is introduced because it ensures the validity of a generalization of the superposition principle—valid even when superselection rules are present. Namely, for pure physical states it guarantees, as it is shown in [3] (Sec. 2.3) that (expressed in the Hilbert space language), if the normalized Hilbert vector Ψ corresponds to a physical state and A is any element of *-algebra generated by all the observables whose spectral decompositions belong to the same coherent subspace, then $A\Psi$ corresponds to a physical state for $A\Psi \neq \mathbf{O}$.

Axiom IV:

For any $P \in \mathcal{P}$ and any $\beta = (\beta_1 \cdots \beta_n) \in \hat{\mathfrak{O}}$, the expression

$$M = \sum_{i,j=1}^{n} \bar{a}_i a_j P^{\beta^*;\beta}(B_i^* \times B_j),$$

$$\beta^* = (\beta_n, \cdots, \beta_1), \quad B^* = (B_n' \times \cdots \times B_1')$$

$$\text{if} \quad B = (B_1' \times \cdots \times B_n') \quad \text{(A11)}$$

is nonnegative for arbitrary[26] $B_1, \cdots, B_n \in \mathring{\mathfrak{B}}^\beta$ and arbitrary complex numbers $a_1, \cdots, a_n (n = 1, 2, \cdots)$. If (A11) is positive for some choice of $B_1, \cdots, B_n \in \mathring{\mathfrak{B}}^\beta$ and $a_1, \cdots, a_n \in \mathbf{C}^1$, then there exists a physical state P_1 for which

$$P_1^\alpha(B) = M^{-1} \sum_{i,j=1}^{n} \bar{a}_i a_j P^{\beta^*;\alpha;\beta}(B_i^* \times B \times B_j) \quad \text{(A12)}$$

for any $\alpha \in \hat{\mathfrak{O}}$, $B \in \mathfrak{B}^\alpha$.

An additional concept, which will prove to be very useful, is that of a function of an observable.

Definition 6: An observable β is said to be a function of a set $\{\alpha_1, \cdots, \alpha_n\}$ of compatible observables if and only if there exists such a real valued Borel function $f(\lambda)$, $\lambda \in R^{\alpha_1 \times, \cdots, \times \alpha_n}$, defined on $\mathbf{R}^{\alpha_1 \times, \cdots, \times \alpha_n}$, so that we have

$$P^{\nu_1 \beta \nu_2}(B) = P^{\nu_1 \alpha_1, \cdots, \alpha_n \nu_2}[f^{-1}(B)] \quad \text{(A13)}$$

for all $P \in \mathcal{S}$. We then write $\beta = f(\alpha_1, \cdots, \alpha_n)$.

We note that, if the set $\mathfrak{O}$ of all observables does not already include all the functions of compatible observables, we can enlarge it by attaching to it all such functions *defined* by means of (A13).

Definition 7: A set of observables is said to be a set of *independant* compatible observables if no observable in that set is a function of a finite number of observables belonging to that same set.

We have shown in [3] and [4] that the above formalism is physically equivalent to the Hilbert space formalism.[27] Of course, in order to be able to speak of physical equivalence we must have in mind a certain (experimental) interpretation or a class of interpretations of the above formalism. In other words, we must have available some *correspondence rules* relating at least some of the above introduced theoretical constructs to laboratory procedures. In [1], after carrying out a thorough analysis of the theory of measurement in quantum mechanics we propose a generalization of Born's interpretation of quantum mechanics or, as we prefer to call it, of Born's correspondence rule. This generalization is based especially on an analysis of the concept of "simultaneous" measurement of incompatible observables.

As we show in [1], we can express the outcome of any measurement or state preparation on an n-tuple α of observables in terms of a principal histogram $h = [\Pi, p(B)]$, consisting of a partition Π of $\mathbf{R}^\alpha$, a set function

$$0 \leq p(B) \leq 1, \quad B \in \pi, \quad \text{(A14)}$$

and an experimental error function $\epsilon(B)$, $B \in \Pi$. We call such a totality of mathematical objects an empirical state. We propose then the following:

The generalized Born's correspondence rule: To a given empirical state corresponding to a state preparation or a measurement on $\alpha \in \mathfrak{O}$, and determined by a principal histogram $h = [\Pi, p(B)]$ and an

[26] Again, unlike in [2], only $B_1, \ldots, B_n \in \mathring{\mathfrak{B}}^\beta$ instead of $B_1, \ldots, B_n \in \mathfrak{B}^\beta$ are considered.

[27] This physical equivalence is not affected by the changes which have been underlined by Refs. 23–25. The proof in [3] of this equivalence, however, has to undergo some minor changes and becomes in its nature more like the case treated in [4]. These changes include the elimination of the necessity of introducing the normed*-algebra $\mathfrak{A}_4(\mathfrak{O})$ in [3], Sec. 2.4. They also solve the difficulties mentioned in Refs. 20 and 21 of [3].

experimental error function $\epsilon(B)$, $B \in \Pi$, corresponds any physical state P which satisfies the relations

$$p(B) - \epsilon(B) \leq \operatorname{Re} P^{\alpha}(B) \leq p(B) + \epsilon(B), \tag{A15}$$

$$|P^{\alpha}(B)| \leq p(B) + \epsilon(B), \tag{A16}$$

for all $B \in \Pi \cap \mathfrak{B}^{\alpha}$.

This correspondence rule contains as a *special case* the conventional Born's correspondence rule, which is valid only for measurements or state-preparations on finite sets of *compatible* observables. Regardless of whether we adopt the original Born's correspondence rule or the above generalization, one feature is worth noticing: when we make measurements or state preparations we do not, generally speaking, determine a single physical state but rather weak neighborhoods of such states! This fact is worth remembering when we are considering the physical equivalence of different formalisms as well as when we are investigating which features in a theory have a direct empirical significance.

Now, speaking simply, a theory $\mathfrak{I}_1$ formulated in terms of the £-formalism is physically equivalent to a theory $\mathfrak{I}_2$ formulated in the Hilbert space language in the following sense: Given $\mathfrak{I}_1$ we can always construct (as is shown in [3] and [4]) a theory $\mathfrak{I}_2$ and, once the above interpretation of the £-formalism (*regardless* of whether we accept the generalized Born's correspondence rule or *only* the conventional form) and the conventional (Hilbert space) interpretation of $\mathfrak{I}_2$ are adopted, we can describe *any realistic experiments* equally well in terms of $\mathfrak{I}_1$ as well as in terms of $\mathfrak{I}_2$.

Naturally, given a Hilbert-space theory $\mathfrak{I}_2$ it is very easy to construct the £-formalism theory $\mathfrak{I}_1$ (see [2]) in the following way:

Once $\mathfrak{I}_2$ is given, the set $\mathfrak{O}$ of all observables is also automatically given. These observables are represented by certain self-adjoint operators.

Take any physical state (pure of mixed) in $\mathfrak{I}_2$, generally described by a statistical (density) operator ρ. We assign to ρ and to any n-tuple $(\alpha_1, \cdots, \alpha_n)$ of observables represented by the self-adjoint operators

$$A_1 = \int \lambda_1 \, dE^{(1)}_{\lambda_1}, \cdots, A_n = \int \lambda_n \, dE^{(n)}_{\lambda_n}, \tag{A17}$$

$$E_\lambda = E(I_\lambda), \quad I_\lambda = (-\infty, \lambda),$$

the complex measure $P^{\alpha_1, \cdots, \alpha_n}(B)$, $B \in \mathfrak{B}^n$, which is such that

$$P^{\alpha_1, \cdots, \alpha_n}_{\rho}(B_1 \times \cdots \times B_n) = \operatorname{Tr}[\rho E^{(1)}(B_1) \cdots E^{(n)}(B_n)],$$

$$B_1 \in \mathfrak{B}^{\alpha_1}, \cdots, B_n \in \mathfrak{B}^{\alpha_n}.$$

It is easy to establish that $P^{\alpha_1, \cdots, \alpha_n}(B)$, $B \in \mathfrak{B}^n$, is a complex probability measure. The set

$$P_\rho = \{P^{\alpha_1, \cdots, \alpha_n}_{\rho}(B), (\alpha_1, \cdots, \alpha_n) \in \hat{\mathfrak{O}}\} \tag{A18}$$

of complex probability measures assigned to the given ρ for all n-tuples $(\alpha_1, \cdots, \alpha_n)$ of observables defines a physical state in the sense of the £-formalism. It can be checked (see [2], Sec. 2.3) that the resulting formalism is an £-formalism.