JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Generating Functions for the Exact Solution of the Transport Equation.* II. Time-Dependent with Anisotropic Scattering

I. K. ABU-SHUMAYS AND E. H. BAREISS
Argonne National Laboratory, Argonne, Illinois

(Received 1 January 1968)

Part I of this series [J. Math. Phys. **9**, 1722 (1968)] introduced in detail the general method of transforming *integrodifferential* transport equations to *partial differential* equations. The treatment there is restricted to isotropic transport in slab geometry. This part extends the method to time-dependent anisotropic transport for slab geometry. Generating functions are used as an analytic tool to define appropriate transformations whose inverses are known. The general solution of the transport equation considered are expressed in terms of expansion modes. The expansion coefficients are determined by a combination of Fourier transforms and orthogonality relations. Fourier transforms in the time variable are used instead of the usual Laplace transforms. The solutions of the initial-value problem and its analog with the role of space and time interchanged are given.

1. INTRODUCTION

Our approach is based on applying the generating-function techniques to the spherical-harmonics method. The time-dependent isotropic transport and the stationary anisotropic transport treated by Case, Mika, and others[1,2] will be included as special cases of the treatment below. Therefore this work can also be regarded as a natural supplement to the work of Case and others.

In this paper generating functions are used as analytic tools for the transformation of the integro-differential *time-dependent transport equation for slab geometry with anisotropic scattering* into a partial differential equation. The new equation is solved to yield analytic expressions for the solution of the corresponding transport equation.

In the first part of this series,[3] which will henceforth be referred to as I, our approach was applied to the stationary isotropic transport equation in general slab geometry. The notation of I will be followed here.

2. TIME-DEPENDENT ANISOTROPIC TRANSPORT IN SLAB GEOMETRY WITH AXIAL SYMMETRY

The axially symmetric transport equation under consideration takes the form

$$\frac{\partial}{\partial t}\psi + \mu\frac{\partial}{\partial z}\psi + \psi = \frac{c}{2}\int_{-1}^{1} f(\mu, \mu')\psi(z, \mu', t)\, d\mu'. \tag{2.1}$$

Equation (2.1) has been normalized so that the constant velocity v is 1 and the total cross section σ is 1. The scattering function $f(\mu, \mu')$ has the form

$$f(\mu, \mu') = \sum_{k=1}^{N} a_k P_k(\mu) P_k(\mu'), \quad a_0 = 1. \tag{2.2}$$

The angular distribution $\psi(z, \mu, t)$ is expanded in Legendre polynomials so that

$$\psi(z, \mu, t) = \sum_{n=0}^{\infty} \frac{2n+1}{2} f_n(z, t) P_n(\mu), \tag{2.3}$$

where

$$f_n(z, t) = \begin{cases} \int_{-1}^{1} \psi(z, \mu, t) P_n(\mu)\, d\mu, & \text{for } n \geq 0, \\ 0, & \text{for } n < 0. \end{cases} \tag{2.4}$$

Application of the relation

$$(2n+1)P_n(\mu) = (n+1)P_{n+1}(\mu) + nP_{n-1}(\mu) \tag{2.5}$$

to Eqs. (2.1)–(2.3) yields the recurrence relation

$$(2n+1)\frac{\partial}{\partial t}f_n + (n+1)\frac{\partial}{\partial z}f_{n+1} + n\frac{\partial}{\partial z}f_{n-1} + (2n+1)f_n - ca_n f_n = 0, \tag{2.6}$$

where $a_n = 0$ for $n > N$.

The generating function for the expansion coefficients $f_n(z, t)$ is defined by

$$\chi(z, t; \zeta) = \sum_{n=0}^{\infty} \zeta^n f_n(z, t) \quad (\zeta = \xi + i\eta). \tag{2.7}$$

Substituting (2.4) into (2.7) gives the following transformation of the angular distribution $\psi(z, \mu, t)$ to the generating function $\chi(z, t; \zeta)$:

$$\chi(z, t; \zeta) = \int_{-1}^{1} \psi(z, \mu, t) \sum_{n=0}^{\infty} \zeta^n P_n(\mu)\, d\mu = \int_{-1}^{1} \frac{\psi(z, \mu, t)\, d\mu}{[1 - 2\mu\zeta + \zeta^2]^{\frac{1}{2}}}, \quad |\zeta| < 1. \tag{2.8}$$

* Major work performed under the auspices of the U.S. Atomic Energy Commission; one author was supported in part by the Office of Naval Research under Contract Nonr 1866(34) with Harvard University.

[1] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1967), Chap. 4.

[2] J. R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).

[3] I. K. Abu-Shumays and E. H. Bareiss, J. Math. Phys. **9**, 1722 (1968).

Applying the generating function Eq. (2.7) to the recurrence relations of Eq. (2.6), or, equivalently, applying the integral transform (2.8) to (2.1), gives

$$(1+\zeta^2)\frac{\partial^2}{\partial z\partial\zeta}\chi + 2\zeta\frac{\partial^2}{\partial t\partial\zeta}\chi + \zeta\frac{\partial}{\partial z}\chi + 2\zeta\frac{\partial}{\partial\zeta}\chi + \frac{\partial}{\partial t}\chi + \chi - c\sum_{k=0}^{N}\frac{a_k}{k!}\chi^{(k)}(z,t;0)\zeta^k = 0, \quad (2.9)$$

where $\chi^{(k)}(z, t, 0)$ is the kth derivative of χ with respect to ζ evaluated at $\zeta = 0$. From Eq. (2.8) follows as in Sec. 2 of I:

Theorem 2.1: Only the solutions of Eq. (2.9) which are analytic in ζ for $|\zeta| < 1$ are generating functions (2.7) for the Fourier coefficients $f_n(z, t)$ of (2.3).

Equation (2.9) can be solved by the method of separation of variables if we let

$$\chi_{\lambda,\omega}(z,t;\zeta) = e^{-\lambda z}e^{-\omega t}G_{\lambda,\omega}(\zeta), \quad (2.10)$$

where $G_{\lambda,\omega}(\zeta)$ satisfies

$$[2\zeta(1-\omega) - \lambda(1+\zeta^2)]\frac{d}{d\zeta}G_{\lambda,\omega}(\zeta) + (1-\omega-\lambda\zeta)G_{\lambda,\omega}(\zeta) - c\sum_{k=0}^{N}a_k g_k\zeta^k = 0. \quad (2.11)$$

For the moment λ, ω are arbitrary constants and

$$g_k = \frac{1}{k!}G^{(k)}_{\lambda,\omega}(0). \quad (2.12a)$$

An appropriate normalization is to require $G_{\lambda,\omega}(0) = 1$, i.e.,

$$g_0 = 1. \quad (2.12b)$$

For $\zeta = 0$, Eq. (2.11) becomes $-\lambda g_1 + 1 - \omega - c = 0$. Thus

$$g_1 = (1-\omega-c)/\lambda \quad \text{for} \quad \lambda \neq 0. \quad (2.12c)$$

By successive differentiation of (2.11), the remaining g_k's can likewise be determined. Hence the term in (2.11) involving the sum may be regarded as known if c, a_k, λ, ω, and ζ are known. Equation (2.11) can be solved explicitly for both $\lambda = 0$ and $\lambda \neq 0$.

(a) $\lambda = 0$: Equation (2.11) becomes

$$(1-\omega)\left[2\zeta\frac{d}{d\zeta}G_{0,\omega} + G_{0,\omega}\right] = c\sum_{k=0}^{N}a_k g_k\zeta^k.$$

Multiplying by the integrating factor $\zeta^{-\frac{1}{2}}$ yields

$$(1-\omega)\frac{d}{d\zeta}[2\zeta^{\frac{1}{2}}G_{0,\omega}] = c\sum_{k=1}^{N}a_k g_k\zeta^{k-\frac{1}{2}}.$$

Then integrating from 0 to ζ and simplifying yields

$$(1-\omega)G_{0,\omega}(\zeta) = c\sum_{k=0}^{N}\frac{a_k}{2k+1}g_k\zeta^k. \quad (2.13a)$$

Because of Eq. (2.12a), this equation is satisfied if and only if

$$1-\omega = c,\ a_k = 2k+1 \quad \text{for} \quad k = 1,\cdots,N. \quad (2.13b)$$

The g_k's for $k = 1, \cdots, N$ are arbitrary. From (2.3), (2.7), and (2.10) it follows that the solution of the transport equation (2.1) corresponding to (2.13) is

$$\psi(z,\mu,t) = p(\mu)e^{-(1-c)t}, \quad (2.14)$$

where $p(\mu)$ is an arbitrary polynomial of degree not greater than N. Unless all conditions of Eqs. (2.13b) are satisfied, no solution for (2.11) exists for $\lambda = 0$.

(b) $\lambda \neq 0$: Define the variable ν by

$$\nu = (1-\omega)/\lambda. \quad (2.15)$$

Dividing Eq. (2.11) by $-\lambda$ yields

$$[1-2\nu\zeta+\zeta^2]\frac{d}{d\zeta}G_{\lambda,\omega} + (\zeta-\nu)G_{\lambda,\omega} + \frac{c}{\lambda}\sum_{k=0}^{N}a_k g_k\zeta^k = 0. \quad (2.16)$$

If $[1-2\nu\zeta+\zeta^2] = 0$, $G_{\lambda,\omega}(\zeta)$ is determined from the remaining terms of (2.16), since a_k and g_k are known. If $[1-2\nu\zeta+\zeta^2] \neq 0$, multiplying (2.16) by the integrating factor $[1-2\nu\zeta+\zeta^2]^{-\frac{1}{2}}$ yields

$$\frac{d}{d\zeta}\{[1-2\nu\zeta+\zeta^2]^{\frac{1}{2}}G_{\lambda,\omega}\} = -\frac{c}{\lambda}\sum_{k=0}^{N}a_k g_k\frac{\zeta^k}{[1-2\nu\zeta+\zeta^2]^{\frac{1}{2}}}. \quad (2.17)$$

Integration from $\zeta = 0$ to ζ, which is carried out in the Appendix, yields $G_{\lambda,\omega}(\zeta)$. It is convenient to distinguish between two cases for the expression of $G_{\lambda,\omega}(\zeta)$, corresponding to $\nu \in [-1, 1]$ and $\nu \notin [-1, 1]$.

(i) For $\nu \in [-1, 1]$ we have

$$G_{\lambda,\omega}(\xi) = \sum_{l=0}^{N-1}K_l(\nu)\zeta^l + [1-2\nu\zeta+\zeta^2]^{-\frac{1}{2}}\left\{1 - \frac{c}{\lambda}\sum_{k=0}^{N}a_k g_k Q_k(\nu) + \frac{c}{\lambda}\sum_{k=0}^{N}a_k g_k P_k(\nu)\log\frac{\nu-\zeta+[1-2\nu\zeta+\zeta^2]^{\frac{1}{2}}}{[1-\nu^2]^{\frac{1}{2}}}\right\}, \quad (2.18)$$

where $P_k(\nu)$ and $Q_k(\nu)$ are the Legendre polynomials of the first and second kind, respectively, and $K_l(\nu)$ is a polynomial in ν and depends on c, λ, and $a_1, \cdots, a_n$

as parameters. Some explicit forms of $K_l(\nu)$ are given in the Appendix. Since, for all $\nu \in [-1, 1]$, $[1 - 2\nu\zeta + \zeta^2]$ has no zeros in the region $|\zeta| < 1$, $G_{\lambda,\omega}(\zeta)$ of (2.18) is analytic in ζ for all $|\zeta| < 1$ and hence by Theorem 2.1 is a factor of the generating function $\chi_{\lambda,\omega}$. [Remember that $\lambda \neq 0$ and ω are related to ν by (2.15).]

(ii) For $\nu \notin [-1, 1]$, it is convenient to write the solution in the form

$$G_{\lambda,\omega}(\zeta) = \sum_{l=0}^{N-1} K_l(\nu)\zeta^l + [1 - 2\nu\zeta + \zeta^2]^{-\frac{1}{2}}\left\{1 - \frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k Q_k(\nu) + \frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k P_k(\nu) \log \frac{\nu - \zeta + [1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}}{\nu - \zeta - [1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}}\right\}. \tag{2.19}$$

The requirement that $G_{\lambda,\omega}(\zeta)$ be analytic for $|\zeta| < 1$ to satisfy Theorem 2.1 and the fact that $[1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}$ vanishes inside $|\zeta| < 1$ at

$$\zeta_- = \nu - (\nu - 1)[(\nu + 1)/(\nu - 1)]^{\frac{1}{2}} \tag{2.20}$$

requires that

$$\lim_{\zeta \to \zeta_-} [1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}} G_{\lambda,\omega}(\zeta) = 0,$$

or, equivalently, from Eq. (2.19)

$$1 - \frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k Q_k(\nu) = 0. \tag{2.21}$$

We call Eq. (2.21) the characteristic equation and its roots for a given λ will be denoted by $\nu_{\lambda,i}$. There is also the option to give ω instead of λ. Recalling (2.15), the characteristic equation for a given ω can be written in the form

$$1 - \frac{c}{1 - \omega}\sum_{k=0}^{N} a_k g_k Q_k(\nu) = 0. \tag{2.22}$$

To evaluate the number of roots of (2.21) and (2.22), define $\Lambda(z)$ by

$$\Lambda_\lambda(z) = 1 - \frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k(z) Q_k(z) \tag{2.23}$$

or

$$\Lambda_\omega(z) = 1 - \frac{cz}{1 - \omega}\sum_{k=0}^{N} a_k g_k(z) Q_k(z). \tag{2.24}$$

From explicit expressions for the functions $g_k(z)$ and $Q_k(z)$, it follows that the function $\Lambda(z)$ is holomorphic in the plane cut on the real axis from -1 to $+1$. Hence, using the argument principle, the number of roots in the cut plane can be expressed for a given λ or a given ω as

$$N_\lambda = \frac{1}{2\pi}\Delta_c \arg \Lambda_\lambda(z), \quad N_\omega = \frac{1}{2\pi}\Delta_c \arg \Lambda_\omega(z), \tag{2.25}$$

where the contour c encircles the cut on the real axis from -1 to $+1$ clockwise.

The set of all admissible ν for a fixed λ or ω will be called the spectral set S_λ or S_ω, respectively:

$$\begin{aligned} S_\lambda &= S_{\lambda p} U S_{\lambda c}, \\ S_{\lambda p} &= \{\nu : \Lambda_\lambda(\nu) = 0\}, \\ S_{\lambda c} &= \{\nu : 1 \geq \nu \geq -1\}. \end{aligned} \tag{2.26}$$

A similar definition holds for S_ω. Further, let $\int_{(S)} d\nu$ denote a Stieltjes-type integral operator over the sum of the discrete terms for all possible $\nu_{\lambda,i}$ (or $\nu_{\omega,i}$) plus over the integral $\int_{-1}^{1} d\nu$. Then the general solution of Eq. (2.9) for the generating function corresponding to a given λ is

$$\chi_\lambda(z, t; \zeta) = \int_{(S_\lambda)} d\nu A_\lambda(\nu) G_{\lambda,1-\lambda\nu}(\zeta) e^{-\lambda z} e^{-(1-\lambda\nu)t}, \tag{2.27}$$

where $A_\lambda(\nu)$ is the expansion coefficient function, and the generating function $\chi_\omega(z, t, \zeta)$ corresponding to a given ω is

$$\chi_\omega(z, t; \zeta) = \int_{(S_\omega)} d\nu A_\omega(\nu) G_{(1-\omega)/\nu,\omega}(\zeta) e^{-(1-\omega)z/\nu} e^{-\omega t}. \tag{2.28}$$

The scalar flux or particle density $\rho(z)$ is given by $\chi(z, t; \zeta)$ at $\zeta = 0$, as follows from Eq. (2.8). Since we normalized $G_{\lambda,\omega}(0) \equiv 1$, this is

$$\rho_\lambda(z, t) = \int_{(S_\lambda)} d\nu A_\lambda(\nu) e^{-\lambda z} e^{-(1-\lambda\nu)t} \tag{2.29}$$

and

$$\rho_\omega(z, t) = \int_{(S_\omega)} d\nu A_\omega(\nu) e^{-(1-\omega)z/\nu} e^{-\omega t}. \tag{2.30}$$

In particular, for $\omega = 0$, $\rho_{\omega=0}(z, t)$ is the known result for the stationary anisotropic transport equation.[1,2]

Thus, letting $\omega = 0$ is physically meaningful. Letting ω be pure imaginary also leads to physically meaningful results in the study of pulsed neutron experiments or excitations of neutron waves.[4] In general, for applications, only a partial range for λ or ω is needed and the corresponding equations (2.27) and (2.29) or (2.28) and (2.30) are integrated over this range to yield the solution of the generating function and the particle density.

[4] N. Corngold, *Proceedings of the Symposium on Pulsed Neutron Research* (International Atomic Energy Agency, Karlsruhe, 1965).

3. INVERSION OF THE GENERATING FUNCTION AND THE GENERAL SOLUTION OF (2.1)

The method of inverting the transform (2.8) to obtain the solution of the angular distribution $\psi(z, \mu, t)$ of (2.1) from the generating function $\chi(z, t; \zeta)$ is identical to that given in Sec. 3 of I: For λ and ω given so that $\nu \in S$, we have

$$\begin{aligned}\psi_{\lambda,\omega}(z, \mu, t) &= \sum_{n=0}^{\infty} \frac{2n+1}{2} \frac{1}{n!} \chi^{(n)}_{\lambda,\omega}(z, t; 0) P_n(\mu) \\ &= e^{-\lambda z} e^{-\omega t} \sum_{n=0}^{\infty} \frac{2n+1}{2} \frac{1}{n!} G^{(k)}_{\lambda,\omega}(0) P_n(\mu) \\ &= e^{-\lambda z} e^{-\omega t} \Phi_{\lambda,\omega}(\mu) \quad (\nu \in S). \end{aligned} \tag{3.1}$$

It follows, by inserting (3.1) into (2.1), that $\Phi_{\lambda,\omega}(\mu)$ satisfies the equation

$$(1 - \omega - \lambda\mu)\Phi_{\lambda,\omega}(\mu) = \frac{c}{2} \int_{-1}^{1} f(\mu, \mu') \Phi_{\lambda,\omega}(\mu')\, d\mu'. \tag{3.2a}$$

Because of (2.15),

$$(1 - \omega - \lambda\mu) \equiv \lambda(\nu - \mu) \equiv (1 - \omega)(\nu - \mu)/\nu. \tag{3.2b}$$

Given λ(or ω), $\Phi_{\lambda,\omega}$ will be called an eigenfunction of (3.2) belonging to the eigenvalue $\nu = (1 - \omega)/\lambda$, provided $\nu \in S$. [Using the middle term of (3.2b), Eq. (3.2a) reduces to the eigenvalue problem

$$(\nu - \mu)\phi(\mu) = \frac{c'}{2} \int_{-1}^{1} f(\mu, \mu']\phi(\mu')\, d\mu',$$

where $c' = c/\lambda$ is now a complex number. The isotropic case of this equation is presented in Ref. 1.]

Following the steps of Sec. 3 of I, but using (3.1) and the results of the previous section, we obtain

$$\begin{aligned}\Phi_{\lambda,\omega}(\mu) = \frac{c}{2\lambda} \frac{1}{\nu - \mu} \sum_{k=0}^{N} a_k g_k P_k(\nu) \\ + \sum_{l=0}^{N-1} \frac{2l+1}{2} K_l(\nu) P_l(\mu), \\ \nu = \frac{1 - \omega}{\lambda}, \quad \nu \notin [-1, 1], \quad \Lambda(\nu) = 0, \end{aligned} \tag{3.3}$$

$$\begin{aligned}\Phi_{\lambda,\omega}(\mu) = \frac{c}{2\lambda} \mathrm{P} \frac{1}{\nu - \mu} \sum_{k=0}^{N} a_k g_k P_k(\nu) \\ + \sum_{l=0}^{N-1} \frac{2l+1}{2} K_l(\nu) P_l(\mu) + \mathrm{P}\Lambda(\nu)\delta(\nu - \mu), \\ \nu \in [-1, 1], \end{aligned} \tag{3.4a}$$

where P denotes the principal value

$$\mathrm{P}\Lambda(\nu) = \tfrac{1}{2}\{\Lambda(\nu + i0) + \Lambda(\nu - i0)\}, \tag{3.4b}$$

$$\mathrm{P} \frac{1}{\nu - \mu} = \sum_{n=0}^{\infty} (2n+1) Q_n(\nu) P_n(\mu), \tag{3.4c}$$

and

$$\delta(\nu - \mu) = \sum_{n=0}^{\infty} \frac{2n+1}{2} P_n(\nu) P_n(\mu). \tag{3.4d}$$

The eigenfunctions obey the orthogonality relations (3.5) and (3.6) given below. For the continuous part of the spectrum $\nu \in [-1, 1]$, for λ fixed and $\omega = 1 - \lambda\nu$, we have

$$\int_{-1}^{1} \Phi_{\lambda,1-\lambda\nu}(\mu) \Phi_{\lambda,1-\lambda\nu'}(\mu)\, d\mu = L(\lambda, \nu)\delta(\nu - \nu'), \tag{3.5a}$$

where

$$L(\lambda, \nu) = \Lambda^{+}_{\lambda}(\nu)\Lambda^{-}_{\lambda}(\nu), \tag{3.5b}$$

and, for ω fixed and $\lambda = (1 - \omega)/\nu$, we have

$$\int_{-1}^{1} \Phi_{(1-\omega)/\nu,\omega}(\mu) \Phi_{(1-\omega)/\nu',\omega}(\mu)\mu\, d\mu = M(\omega, \nu)\delta(\nu - \nu'), \tag{3.6a}$$

where

$$M(\omega, \nu) = \nu\Lambda^{+}_{\omega}(\nu)\Lambda^{-}_{\omega}(\nu), \tag{3.6b}$$

and Λ^{+} and Λ^{-} are the limits of $\Lambda(z)$ as z approaches $\nu \in [-1, 1]$ from the positive or negative sides of the cut from -1 to 1. It will be assumed here, as in Ref. 2, and in the rest of this treatment that $\Lambda^{+}(\nu)\Lambda^{-}(\nu) \neq 0$ for $\nu \in [-1, 1]$. For the case $\Lambda^{+}(\nu)\Lambda^{-}(\nu) = 0$ we refer to Mika.[2] The orthogonality relations for the discrete part of the spectrum $\nu \notin [-1, 1]$, $\nu \in S$, are

$$\int_{-1}^{1} \Phi_{\lambda,1-\lambda\nu}(\mu) \Phi_{\lambda,1-\lambda\nu'}(\mu)\, d\mu = L(\lambda, \nu)\delta_{\nu,\nu'}, \tag{3.5c}$$

where, from Eq. (3.3),

$$\begin{aligned} L(\lambda, \nu) &= \int_{-1}^{1} \Phi^2_{\lambda,1-\lambda\nu}(\mu)\, d\mu \\ &= \frac{2s^2}{\nu^2 - 1} + \sum_{l=0}^{N-1} (2l+1) K_l(\nu)[2sQ_l(\nu) + \tfrac{1}{2}K_l(\nu)], \end{aligned} \tag{3.5d}$$

with

$$s = \frac{c}{2\lambda} \sum_{k=0}^{N} a_k g_k(\nu) P_k(\nu) \quad \{\nu: \Lambda(\nu) = 0\} \tag{3.5e}$$

and

$$\int_{-1}^{1} \Phi_{(1-\omega)/\nu,\omega}(\mu) \Phi_{(1-\omega)/\nu',\omega}(\mu)\mu\, d\mu = M(\omega, \nu)\delta_{\nu,\nu'}. \tag{3.6c}$$

Here

$$\begin{aligned} M(\omega, \nu) &= \int_{-1}^{1} \Phi^2_{(1-\omega)/\nu,\omega}(\mu)\mu\, d\mu \\ &= \frac{2\nu s^2}{\nu^2 - 1} + s^2 \log \frac{\nu - 1}{\nu + 1} \\ &\quad + \sum_{l=0}^{N-1} \{2(2l+1) s K_l(\nu) Q_l(\nu)\nu \\ &\quad + \frac{2l+1}{2} K_l(\nu) K_{l+1}(\nu) + \frac{l}{2} K_l(\nu) K_{l-1}(\nu), \\ &\quad \{\nu: \Lambda(\nu)\} = 0. \end{aligned} \tag{3.6d}$$

The solutions of the transport equation (2.1), for given λ or ω corresponding to (2.27) and (2.28) for the generating function $\chi(z, t; \zeta)$, take the form

$$\psi_\lambda(z, \mu, t) = \int_{(S_\lambda)} d\nu A_\lambda(\nu) e^{-\lambda z} e^{-(1-\lambda\nu)t} \Phi_{\lambda, 1-\lambda\nu}(\mu), \quad (3.7a)$$

$$\psi_\omega(z, \mu, t) = \int_{(S_\omega)} d\nu A_\omega(\nu) e^{-(1-\omega)z/\nu} e^{-\omega t} \Phi_{(1-\omega)/\nu, \omega}(\mu), \quad (3.8a)$$

where, in particular at $z = 0$, $t = 0$,

$$\psi_\lambda(0, \mu, 0) = \int_{(S_\lambda)} d\nu A_\lambda(\nu) \Phi_{\lambda, 1-\lambda\nu}(\mu), \quad (3.7b)$$

$$\psi_\omega(0, \mu, 0) = \int_{(S_\omega)} d\nu A_\omega(\nu) \Phi_{(1-\omega)/\nu, \omega}(\mu). \quad (3.8b)$$

For an infinite medium the orthogonality relations (3.5) and (3.6) can be used to evaluate the expansion coefficients of the representations (3.7) and (3.8). These representations are possible because of full- and partial-range completeness theorems for functions of μ. The full-range completeness theorem is:

Theorem 3.1: Any arbitrary function $\psi(\mu)$ which obeys the H^* condition for all $\mu \in [-1, 1]$, [i.e.,

$$|\psi(\mu_2) - \psi(\mu_1)| \leq A|\mu_2 - \mu_1|^\gamma, \quad (3.9a)$$

where A and γ are positive numbers and, in the neighborhood of ± 1,

$$\psi(\mu) = \frac{\psi^*(\mu)}{|\mu \pm 1|^\delta}, \quad 0 \leq \delta \leq 1, \quad (3.9b)$$

where $\psi^*(\mu)$ satisfies (3.9a)] can be represented in the form (3.7b) or (3.8b) for any given λ or ω.

However, we look for the general representation of solutions to the transport equation (2.1) and not merely for representations of arbitrary functions of μ. At a first glance, the general solution of the transport equation (2.1) may be expected to be the double integral of (3.7) [or (3.8)] over all $\lambda = \lambda_1 + i\lambda_2 \neq 0$ (or all ω). This combination, however, overdetermines the solution; a line integral suffices, as for the initial-value problem presented below.

Consider an initial angular distribution $\psi(z, \mu, 0)$ which obeys the H^* condition in μ and possesses a Fourier transform in z. By Theorem 3.1 the Fourier transform of $\psi(z, \mu, 0)$ can always be represented in the form

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} dz\, e^{i\lambda z} \psi(z, \mu, 0) = \int_{(S_{i\lambda})} d\nu\, A_{i\lambda}(\nu) \Phi_{i\lambda, 1-i\lambda\nu}(\mu), \quad (3.10)$$

where, using the orthogonality relations (3.5),

$$A_{i\lambda}(\nu) = \frac{1}{2\pi L(i\lambda, \nu)} \int_{-\infty}^{+\infty} dz\, e^{i\lambda z} \times \int_{-1}^{1} d\mu\, \psi(z, \mu, 0) \Phi_{i\lambda, 1-i\lambda\nu}(\mu). \quad (3.11)$$

The inverse of the Fourier transform (3.10) is

$$\psi(z, \mu, 0) = \int_{-\infty}^{+\infty} d\lambda\, e^{-i\lambda z} \int_{(S_{i\lambda})} d\nu\, A_{i\lambda}(\nu) \Phi_{i\lambda, 1-i\lambda\nu}(\mu). \quad (3.12)$$

Thus a solution of the transport equation (2.1) which has the initial angular distribution $\psi(z, \mu, 0)$ given above is [see (3.1)]

$$\psi(z, \mu, t) = \int_{-\infty}^{+\infty} d\lambda\, e^{-i\lambda z} \times \int_{(S_{i\lambda})} d\nu\, A_{i\lambda}(\nu) e^{-(1-i\lambda\nu)t} \Phi_{i\lambda, 1-i\lambda\nu}(\mu). \quad (3.13)$$

Equation (3.13) is a representation for solutions of the transport equation (2.1) and, for isotropic scattering, agrees with that given in Ref. 1. In the infinite medium, the expansion coefficients $A_{i\lambda}(\nu)$ are given by (3.11). This representation also leads to an expression for the initial-value infinite-medium Green's function and hence[1,3] to the solution of the transport equation for arbitrary source distributions. The expression is similar to that given by Case and Zweifel (Ref. 1, Chap. 7) for the special case of isotropic scattering.

The representation (3.13), however, is not appropriate for the Milne problem (Ref. 1, p. 183). This is not surprising since the Fourier transform approach for the stationary isotropic transport equation leads only to an asymptotic solution and does not account for the expansion modes belonging to the continuum.

Case and Zweifel (Ref. 1, Chap. 7) also give a representation of the angular distribution for the time-dependent isotropic transport equation in terms of Laplace transforms with respect to time. A generalization of their work to the anisotropic case is straightforward and will be omitted here. Instead, a representation based on Fourier transforms in the time variable will be discussed. As mentioned above, the representation (3.8) is complete and physically meaningful for $\omega = 0$ (stationary case) or ω pure imaginary (pulsed reactors, fixed neutron waves[4]). Thus a linear combination of (3.8) for pure imaginary ω, i.e., a Fourier transform in the time variable, is expected to yield a representation of solutions to the

transport equation (2.1):

$$\psi(z,\mu,t)=\int_{-\infty}^{+\infty}d\omega\,\psi_{i\omega}(z,\mu,t)$$
$$=\int_{-\infty}^{+\infty}d\omega\,e^{-i\omega t}\int_{(S_{i\omega})}d\nu\,A_{i\omega}(\nu)$$
$$\times\exp\left(-\frac{1-i\omega}{\nu}z\right)\Phi_{(1-i\omega)/\nu,i\omega}(\mu), \tag{3.14}$$

where

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}dt\,e^{i\omega t}\psi(z,\mu,t)$$
$$=\int_{(S_{i\omega})}d\nu\,A_{i\omega}(\nu)\exp\left(-\frac{1-i\omega}{\nu}z\right)\Phi_{(1-i\omega)/\nu,i\omega}(\mu), \tag{3.15}$$

and, using the orthogonality relations (3.6),

$$A_{i\omega}(\nu)=\frac{1}{2\pi M(\omega,\nu)}\int_{-\infty}^{+\infty}dt\,e^{i\omega t}$$
$$\times\int_{-1}^{1}d\mu\,\mu\psi(0,\mu,t)\Phi_{(1-i\omega)/\nu,i\omega}(\mu). \tag{3.16}$$

If the angular distribution $\psi(z,\mu,t)$ (a) obeys the H^* condition in μ, (b) if it is known at a fixed point in space, say $z=0$, for all times t, and (c) if it has a Fourier transform in time, then the general angular distribution will be given by (3.14) with the expansion coefficients defined by (3.16). This result is analogous to the initial-value problem treated above, i.e., from the time history of ψ at a fixed point (the measuring point) we can calculate ψ for any other point (μ, z, t).

As an application of the above representations the case of purely absorbing media will be shown to be a limit case for anisotropic scattering.

The restriction of the scattering function $f(\mu,\mu')$ to be nonnegative and $\int f(\mu,\mu')\,d\mu'=1$ require[1,2] that $a_k\leq 2k+1$. In the limit case $N\to\infty$ and $a_k=2k+1$, $\frac{1}{2}f(\mu,\mu')$ takes the form

$$\tfrac{1}{2}f(\mu,\mu')=\sum_{k=0}^{\infty}\frac{2k+1}{2}P_k(\mu)P_k(\mu')=\delta(\mu-\mu'). \tag{3.17}$$

Then the transport equation (2.1) reduces to

$$\frac{\partial}{\partial t}\psi(z,\mu,t)+\mu\frac{\partial}{\partial z}\psi(z,\mu,t)+\psi(z,\mu,t)$$
$$=c\psi(z,\mu,t). \tag{3.18}$$

This equation is equivalent to the case of a purely absorbing medium treated in Ref. 1. The transformation (renormalization) $z\to z/(1-c)$, $t\to t/(1-c)$ reduces Eq. (3.18) to the transport equation (2.1) with $c=0$. Hence the representation of the general solution $\psi(z,\mu,t)$ of (3.18), as deduced from Eq. (3.13), is

$$\psi(z,\mu,t)=\int_{-\infty}^{+\infty}d\lambda\,e^{-i\lambda z}\int_{-1}^{1}d\nu\,A_{i\lambda}(\nu)e^{-(1-c-i\lambda\nu)t}\delta(\nu-\mu)$$
$$=\int_{-\infty}^{+\infty}d\lambda\,e^{-i\lambda z}A_{i\lambda}(\mu)e^{-(1-c-i\lambda\mu)t}, \tag{3.19}$$

where

$$A_{i\lambda}(\mu)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dz\,e^{i\lambda z}\psi(z,\mu,0). \tag{3.20}$$

The representation as deduced from (3.14) is

$$\psi(z,\mu,t)=\int_{-\infty}^{+\infty}d\omega\,e^{-i\omega t}\int_{-1}^{1}d\nu A_{i\omega}(\nu)$$
$$\times\exp\left(-\frac{1-c-i\omega}{\nu}z\right)\delta(\nu-\mu)$$
$$=\int_{-\infty}^{+\infty}d\omega\,e^{-i\omega t}A_{i\omega}(\mu)\exp\left(-\frac{1-c-i\omega}{\nu}z\right), \tag{3.21}$$

where

$$A_{i\omega}(\mu)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}d\omega\,e^{i\omega t}\psi(0,\mu,t). \tag{3.22}$$

Since here $\Phi_{\lambda,\omega}(\mu)=\delta(\nu-\mu)$, $\nu=(1-c-\omega)/\lambda\in[-1,1]$, Theorem 3.1 is satisfied for arbitrarily prescribed λ or ω. Further, since the functions $\{e^{-i\lambda z}\colon \lambda\in(-\infty,\infty)\}$ [or $\{e^{-i\omega t}\colon \omega\in(-\infty,\infty)\}$] form a complete basis for functions of z (or of t) which possess Fourier transforms (this includes the δ function), the general solutions of the transport equation (3.18) have the above representations (3.19) or (3.21). This simple example should be a reminder that scattering functions $f(\mu,\mu')$, for which the expansion (2.2) has $(N=\infty)$-many terms, need individual attention.

4. TIME-DEPENDENT ISOTROPIC TRANSPORT IN GENERAL SLAB GEOMETRY

Time-dependent isotropic transport with axially symmetric slab geometry is treated above (the special case of transport with general geometry and $a_k=0$ for $k\geq 1$). Stationary isotropic transport in general slab geometry was treated in I. Here we combine these two problems and solve the time-dependent transport equation for isotropic scattering in general slab geometry

$$\frac{\partial}{\partial t}\psi+\mu\frac{\partial}{\partial z}\psi+\psi=\frac{c}{4\pi}\int_{0}^{2\pi}d\phi\int_{-1}^{1}d\mu\,\psi(z,\mu,\phi,t). \tag{4.1}$$

As in I, the angular distribution is expanded in

general spherical harmonics so that

$$\psi(z, \mu, \phi, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} f_{nm}(z, t) Y_n^m(\mu, \phi). \quad (4.2)$$

The expansion coefficients are given by

$$f_{nm}(z, t) = \int_0^{2\pi} d\phi \int_{-1}^{1} d\mu \psi(z, \mu, \phi, t) Y_n^{m*}(\mu, \phi). \quad (4.3)$$

The restriction of the angular distribution ψ to be real implies[3] that

$$f_{n,-m} = (-1)^m \frac{(n-m)!}{(n+m)!} f_{nm}^*. \quad (4.4)$$

Substituting Eq. (4.2) into (4.1) and using the known properties of the spherical harmonics[3,5] yields

$$(2n+1)\frac{\partial}{\partial t} f_{nm} + (n+1-m)\frac{\partial}{\partial z} f_{n+1,m} + (n+m)\frac{\partial}{\partial z} f_{n-1,m} + (2n+1) f_{nm} - c f_{nm} \delta_{n0} \delta_{m0} = 0. \quad (4.5)$$

Since the second subscript m of f_{nm} is the same for all terms in (4.5), this set of equations can be solved separately for each value of m. Equation (4.4) makes it sufficient to solve (4.5) for $m \geq 0$.

For $m = 0$, the system (4.5) is a special case of (2.6) with $a_k = 0$ for $k \geq 1$ and has been solved. For each m fixed where $m > 0$, we define a generating function for the Fourier coefficients f_{nm} of the angular distribution ψ by

$$\chi_m(z, t; \zeta) \equiv \sum_{n=m}^{\infty} \zeta^n f_{nm}(z, t), \quad m > 0. \quad (4.6)$$

As in I, χ_m can be shown to be a transform of the angular distribution $\psi(z, \mu, \phi, t)$ and to be analytic in ζ for $|\zeta| < 1$. Applying this transform to Eq. (4.1) or, equivalently, multiplying Eq. (4.5) by ζ^n, summing over n, and using the definition of χ_m above yields

$$(1+\zeta^2)\frac{\partial^2}{\partial z \partial \zeta} \chi_m + 2\zeta \frac{\partial^2}{\partial t \partial \zeta} \chi_m + \left[(m+1)\zeta - \frac{m}{\zeta}\right]\frac{\partial}{\partial z} \chi_m + \frac{\partial}{\partial t} \chi_m + 2\zeta \frac{\partial}{\partial \zeta} \chi_m + \chi_m = 0. \quad (4.7)$$

Here again (see Theorem 2.1), only the solutions of Eq. (4.7) which are analytic in ζ for $|\zeta| < 1$ are generating functions for the Fourier coefficients f_{nm} of (4.2).

[5] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, London, 1931).

Equation (4.7) can be solved by the method of separation of variables if we let

$$\chi_{m,\lambda,\omega} = e^{-\lambda z} e^{-\omega t} G_{m,\lambda,\omega}(\zeta), \quad (4.8)$$

where $G_{m,\lambda,\omega}(\zeta)$ satisfies

$$[2(1-\omega)\zeta - \lambda(1+\zeta^2)]\frac{d}{d\zeta} G_{m,\lambda,\omega}(\zeta) + \left[1 - \omega - (m+1)\lambda\zeta + \frac{m\lambda}{\zeta}\right] G_{m,\lambda,\omega}(\zeta) = 0. \quad (4.9)$$

For $\lambda = 0$, multiplying Eq. (4.9) by the integrating factor $\zeta^{-\frac{1}{2}}$ yields

$$(1-\omega)\frac{d}{d\zeta}[2\zeta^{\frac{1}{2}} G_{m,0,\omega}(\zeta)] = 0.$$

By integrating the last equation from $\zeta = 0$ to ζ we get

$$(1-\omega)2\zeta^{\frac{1}{2}} G_{m,0,\omega}(\zeta) - [(1-\omega)2\zeta^{\frac{1}{2}} G_{m,0,\omega}(\zeta)]_{\zeta=0} = 0. \quad (4.10)$$

Hence $G_{m,0,\omega}(\zeta) \equiv 0$ unless $\omega = 1$, in which case $G_{m,0,1}(\zeta)$ is an arbitrary function of ζ. The corresponding solution of the transport equation (4.1) is given by

$$\psi_{m,0,1} = e^{-t} e^{im\phi} \gamma(\mu), \quad m \neq 0, \quad \gamma(\mu) \text{ arbitrary}. \quad (4.11)$$

For $\lambda \neq 0$, dividing by $-\lambda$ and defining $\nu = (1-\omega)/\lambda$ reduces Eq. (4.9) to Eq. (4.16) of I. Thus the corresponding elementary solutions of the transport equation (4.1) are

$$\psi_{m,\lambda,\omega} = e^{-\lambda z} e^{-\omega t} e^{im\phi} \delta(\mu - \nu), \quad \nu = \frac{1-\omega}{\lambda} \in [-1, 1]. \quad (4.12a)$$

Therefore

$$\Phi_{m,\lambda,\omega} = e^{im\phi} \delta(\mu - \nu), \quad \nu = (1-\omega)/\lambda \in [-1, 1] \quad (4.12b)$$

form a complete set for arbitrary functions of μ. Equations (4.11) and (4.12), which were derived for $m > 0$, hold for all integers $m \neq 0$ because of (4.4).

The general solution of the transport equation (4.1) corresponding to (3.13), but with $a_k = 0$ for $k \geq 1$, is

$$\begin{aligned}\psi(z, \mu, \phi, t) &= \int_{-\infty}^{+\infty} d\lambda\, e^{-i\lambda z} \left\{ \int_{(S_{i\lambda})} d\nu\, A_{i\lambda}(\nu) e^{-(1-i\lambda\nu)t} \Phi_{i\lambda,1-i\lambda\nu}(\mu) + \sum_{\substack{m=-\infty \\ m\neq 0}}^{+\infty} \int_{-1}^{1} d\nu A_{m,i\lambda}(\nu) e^{-(1-i\lambda\nu)t} \Phi_{m,i\lambda,1-i\lambda\nu}(\mu) \right\} \\ &= \int_{-\infty}^{+\infty} d\lambda\, e^{-i\lambda z} \left\{ \int_{(S_{i\lambda})} d\nu\, A_{i\lambda}(\nu) e^{-(1-i\lambda\nu)t} \Phi_{i\lambda,1-i\lambda\nu}(\mu) + \sum_{\substack{m=-\infty \\ m\neq 0}}^{+\infty} A_{m,i\lambda}(\mu) e^{-(1-i\lambda\mu)t} e^{im\phi} \right\}. \end{aligned} \quad (4.13)$$

The expansion coefficients $A_{i\lambda}(\nu)$, $A_{m,i\lambda}(\mu)$ can be determined as in Sec. 4 of I and Sec. 3 above. For example, if the initial distribution $\psi(z, \mu, \phi, 0)$ is given, obeys the H^* condition in μ, and possesses a Fourier transform in z, then (4.13) is the solution for the initial-value problem when we let

$$A_{m,i\lambda}(\mu) = \frac{1}{4\pi^2}\int_0^{2\pi} d\phi \int_{-\infty}^{+\infty} dz\, e^{i\lambda z}\psi(z, \mu, \phi, 0)e^{-im\phi}, \qquad m \neq 0, \quad (4.14)$$

and [using the orthogonality relation (3.5) with $a_k = 0$ for $k \geq 1$]

$$A_{i\lambda}(\nu) = \frac{1}{4\pi^2 L(i\lambda, \nu)}\int_0^{2\pi} d\phi \int_{-1}^{1} d\mu \times \int_{-\infty}^{+\infty} dz\, e^{i\lambda z}\psi(z, \mu, \phi, 0)\Phi_{i\lambda,1-i\lambda\nu}(\mu). \quad (4.15)$$

In particular, let the initial distribution $\psi(z, \mu, \phi, 0)$ be a product of δ functions:

$$\psi(z, \mu, \phi, 0; z_0, \mu_0, \phi_0) = \frac{1}{2\pi}\,\delta(z - z_0)\delta(\mu - \mu_0)\delta(\phi - \phi_0). \quad (4.16)$$

Then Eqs. (4.14) and (4.15) yield the following simple expansion coefficients corresponding to (4.16):

$$A_{m,i\lambda}(\mu) = \frac{1}{8\pi^3}\, e^{i\lambda z_0}\exp(-im\phi_0)\delta(\mu - \mu_0), \quad (4.17)$$

$$A_{i\lambda}(\nu) = \frac{1}{8\pi^3 L(i\lambda, \nu)}\exp(i\lambda z_0)\Phi_{i\lambda,1-i\lambda\nu}(\mu_0). \quad (4.18)$$

Hence, from Eqs. (4.13), (4.17), (4.18), and (5.9) of I, the general solution of the transport equation (4.1) having the initial distribution (4.16) is

$$\psi(z, \mu, \phi, t; z_0, \mu_0, \phi_0) = \frac{1}{8\pi^3}\int_{-\infty}^{+\infty} d\lambda\, e^{-i\lambda(z-z_0)} \times \left\{\int_{(S_{i\lambda})} d\nu\, \frac{e^{-(1-i\lambda\nu)t}}{L(i\lambda, \nu)}\,\Phi_{i\lambda,1-i\lambda\nu}(\mu_0)\Phi_{i\lambda,1-i\lambda\nu}(\mu) + e^{-(1-i\lambda\mu)t}\delta(\mu - \mu_0)[2\pi\delta(\phi - \phi_0) - 1]\right\}. \quad (4.19)$$

As expected, the angular distribution (4.19) indeed reduces to (4.16) when $t = 0$ (see Ref. 7 of I). Expressions similar to (4.13)–(4.19) hold for the representation corresponding to (3.14) based on Fourier transforms in the time variable.

ACKNOWLEDGMENTS

The authors wish to express their sincere thanks to Garrett Birkhoff for valuable comments. R. F. King edited this report and Doris Haight typed the manuscript.

APPENDIX: INTEGRATION OF (2.17) FROM $\zeta = 0$ TO ξ

The integrals involved are of the type

$$I_k \equiv \int_0^{\zeta} \frac{\zeta^k}{(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}, \quad k \text{ a positive integer.} \quad (A1)$$

From Refs. 3 and 6:

$$I_0 = \log\left\{\frac{\zeta - \nu + (1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}{1 - \nu}\right\} = -\log\left\{\frac{[\nu - \zeta + (1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}}{1 + \nu}\right\}, \quad (A2)$$

$$I_1 = [1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}} - 1 + \nu I_0. \quad (A3)$$

Further, for $k > 1$,

$$I_k = [1 - 2\nu\zeta + \zeta^2]^{\frac{1}{2}}\frac{\zeta^{k-1}}{k} + \frac{(2k-1)\nu}{k} I_{k-1} - \frac{k-1}{k} I_{k-2}. \quad (A4)$$

Equations (A2)–(A4) are sufficient to determine I_k for all positive integers k. A more convenient expression for I_k can be deduced from the above by mathematical induction and is

$$I_k(\zeta) = (1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}\{u_{k-1}(\zeta, \nu) + W_{k-1}(\nu)\} - W_{k-1}(\nu) + P_k(\nu)I_0, \quad (A5)$$

where u_{k-1} is a polynomial in ζ and ν of degree $k - 1$ in ζ and $k - 2$ in ν,

$$u_0 = 0, \quad u_1 = \zeta/2, \quad (A6a)$$

and the rest of the u_{k-1} are determined from the recurrence relation

$$u_{k-1}(\zeta, \nu) = \frac{\zeta^{k-1}}{k} + \frac{(2k-1)\nu}{k} u_{k-2} - \frac{k-1}{k} u_{k-3}. \quad (A6b)$$

Clearly,

$$u_{k-1}(0, \nu) = 0. \quad (A6c)$$

Further, $W_{k-1}(\nu)$ is a polynomial of degree $k - 1$ in ν defined in terms of Legendre polynomials by

$$W_{k-1}(\nu) = P_k(\nu)Q_0(\nu) - Q_k(\nu). \quad (A7)$$

Integrating (2.17) from $\zeta = 0$ to ζ and using (A5)

[6] G. Petit Bois, *Tables of Indefinite Integrals* (Dover Publications, Inc., New York, 1961).

yields

$$(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}G_{\lambda,\omega}(\zeta) - 1$$
$$= -\frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k I_k$$
$$= -\frac{c}{\lambda}(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}\sum_{k=1}^{N} a_k g_k\{u_{k-1}(\zeta, \nu) + W_{k-1}(\nu)\}$$
$$+ \frac{c}{\lambda}\sum_{k=1}^{N} a_k g_k W_{k-1}(\nu) - \frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k P_k(\nu) I_0. \quad \text{(A8)}$$

Since $u_{k-1}(\zeta, \nu)$ is a polynomial in ζ of degree $k - 1$, the right-hand side of (A8) can be written in the form

$$-\frac{c}{\lambda}\sum_{k=1}^{N} a_k g_k[u_{k-1}(\zeta, \nu) + W_{k-1}(\nu)] \equiv \sum_{k=0}^{N-1} K_l(\nu)\zeta^l, \quad \text{(A9)}$$

where $K_l(\nu)$ is a polynomial in ν and depends on c, λ, and the parameters $a_1, \cdots, a_N$. In particular, using (A6c), it follows that

$$K_0(\nu) = -\frac{c}{\lambda}\sum_{k=1}^{N} a_k g_k W_{k-1}(\nu) \quad \text{(A10a)}$$

and, using (A6b),

$$K_{N-1}(\nu) = -\frac{c}{N\lambda}\, a_N g_N, \quad \text{(A10b)}$$

$$K_{N-2}(\nu) = -\frac{c}{\lambda}\left[\frac{2N-1}{N(N-1)}\, a_N g_N \nu + \frac{1}{N-1}\, a_{N-1} g_{N-1}\right], \quad \text{(A10c)}$$

and so on. The expressions (2.18) to (2.21) for $G_{\lambda,\omega}(\zeta)$ and the characteristic equation follow directly from (A8), (A9), (A2), and (A7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Inverse Wave Propagator*

ÉAMON LALOR
Department of Physics and Astronomy, University of Rochester, Rochester, New York

(Received 16 February 1968)

In a recent publication, Wolf and Shewell gave a formal solution to the inverse diffraction problem, i.e., to finding the field distribution in the plane $z = 0$ from the knowledge of the field in an arbitrary plane $z = z_1 > 0$ in the half-space into which the field is propagated. The solution involved the use of a singular kernel. In the present paper the inverse diffraction problem is treated in a rigorous manner. Our method makes use of the representation of the field as an angular spectrum of plane waves and demonstrates the usefulness of this type of representation. It is shown that by the use of a suitable truncation procedure one may avoid the use of a singular kernel or the generalized function theory.

1. INTRODUCTION

In a recent note, Wolf and Shewell[1] gave the solution to the inverse diffraction problem. The problem may be formulated as follows. Consider a monochromatic wave field

$$V(\mathbf{R}, t) = U(\mathbf{R}) \exp(-i\omega t) \quad (1.1)$$

which is well behaved in the half-space $z > 0$. The space-dependent part $U(\mathbf{R})$ obeys the Helmholtz equation

$$(\nabla^2 + k^2)U(\mathbf{R}) = 0 \quad (k = \omega/c), \quad (1.2)$$

where c is the velocity of light. It is assumed that U obeys the Sommerfeld radiation condition at infinity in the half-space $z > 0$. If the field distribution $U(\mathbf{R}_2)$ in the plane $z = z_2$ is given, is it possible to derive an expression for the field $U(\mathbf{R}_1)$ at any point $\mathbf{R} = \mathbf{R}_1$ in any plane $z = z_1$ such that $0 \leq z_1 \leq z_2$?

The solution, which may be regarded as the inverse of one of Rayleigh's diffraction formulas,[2] was given as follows:

$$U(\mathbf{R}_1) = \iint U(\mathbf{R}_2) K(\mathbf{R}_1, \mathbf{R}_2)\, dS_2, \quad (1.3)$$

where the integration extends over the plane $z = z_2$ and the kernel $K(\mathbf{R}_1, \mathbf{R}_2)$ may be expressed in the form

$$K(\mathbf{R}_1, \mathbf{R}_2) = -\frac{1}{2\pi}\frac{\partial}{\partial z_2}[\exp(-ikr)/r] + S(\mathbf{R}_1, \mathbf{R}_2), \quad (1.4)$$

* Research supported by the Air Force Office of Scientific Research.

[1] E. Wolf and J. R. Shewell, Phys. Letters **25A** (1967),417; **26A** (1967) 104.

[2] The Rayleigh formula may be written in the form

$$U(\mathbf{R}_2) = \iint U(\mathbf{R}_1) K(\mathbf{R}_2, \mathbf{R}_1)\, dS_1,$$

where

$$K(\mathbf{R}_2, \mathbf{R}_1) = -\frac{1}{2\pi}\frac{\partial}{\partial z_2}\left(\frac{\exp(ik|\mathbf{R}_1 - \mathbf{R}_2|)}{|\mathbf{R}_1 - \mathbf{R}_2|}\right)$$

and the integration is carried out over the plane $z = z_1$.

yields

$$(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}G_{\lambda,\omega}(\zeta) - 1$$
$$= -\frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k I_k$$
$$= -\frac{c}{\lambda}(1 - 2\nu\zeta + \zeta^2)^{\frac{1}{2}}\sum_{k=1}^{N} a_k g_k\{u_{k-1}(\zeta, \nu) + W_{k-1}(\nu)\}$$
$$+ \frac{c}{\lambda}\sum_{k=1}^{N} a_k g_k W_{k-1}(\nu) - \frac{c}{\lambda}\sum_{k=0}^{N} a_k g_k P_k(\nu) I_0. \quad \text{(A8)}$$

Since $u_{k-1}(\zeta, \nu)$ is a polynomial in ζ of degree $k - 1$, the right-hand side of (A8) can be written in the form

$$-\frac{c}{\lambda}\sum_{k=1}^{N} a_k g_k[u_{k-1}(\zeta, \nu) + W_{k-1}(\nu)] \equiv \sum_{k=0}^{N-1} K_l(\nu)\zeta^l, \quad \text{(A9)}$$

where $K_l(\nu)$ is a polynomial in ν and depends on c, λ, and the parameters $a_1, \cdots, a_N$. In particular, using (A6c), it follows that

$$K_0(\nu) = -\frac{c}{\lambda}\sum_{k=1}^{N} a_k g_k W_{k-1}(\nu) \quad \text{(A10a)}$$

and, using (A6b),

$$K_{N-1}(\nu) = -\frac{c}{N\lambda}\, a_N g_N, \quad \text{(A10b)}$$

$$K_{N-2}(\nu) = -\frac{c}{\lambda}\left[\frac{2N-1}{N(N-1)}\, a_N g_N \nu + \frac{1}{N-1}\, a_{N-1} g_{N-1}\right], \quad \text{(A10c)}$$

and so on. The expressions (2.18) to (2.21) for $G_{\lambda,\omega}(\zeta)$ and the characteristic equation follow directly from (A8), (A9), (A2), and (A7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Inverse Wave Propagator*

ÉAMON LALOR
Department of Physics and Astronomy, University of Rochester, Rochester, New York

(Received 16 February 1968)

In a recent publication, Wolf and Shewell gave a formal solution to the inverse diffraction problem, i.e., to finding the field distribution in the plane $z = 0$ from the knowledge of the field in an arbitrary plane $z = z_1 > 0$ in the half-space into which the field is propagated. The solution involved the use of a singular kernel. In the present paper the inverse diffraction problem is treated in a rigorous manner. Our method makes use of the representation of the field as an angular spectrum of plane waves and demonstrates the usefulness of this type of representation. It is shown that by the use of a suitable truncation procedure one may avoid the use of a singular kernel or the generalized function theory.

1. INTRODUCTION

In a recent note, Wolf and Shewell[1] gave the solution to the inverse diffraction problem. The problem may be formulated as follows. Consider a monochromatic wave field

$$V(\mathbf{R}, t) = U(\mathbf{R}) \exp(-i\omega t) \quad (1.1)$$

which is well behaved in the half-space $z > 0$. The space-dependent part $U(\mathbf{R})$ obeys the Helmholtz equation

$$(\nabla^2 + k^2)U(\mathbf{R}) = 0 \quad (k = \omega/c), \quad (1.2)$$

where c is the velocity of light. It is assumed that U obeys the Sommerfeld radiation condition at infinity in the half-space $z > 0$. If the field distribution $U(\mathbf{R}_2)$ in the plane $z = z_2$ is given, is it possible to derive an expression for the field $U(\mathbf{R}_1)$ at any point $\mathbf{R} = \mathbf{R}_1$ in any plane $z = z_1$ such that $0 \leq z_1 \leq z_2$?

The solution, which may be regarded as the inverse of one of Rayleigh's diffraction formulas,[2] was given as follows:

$$U(\mathbf{R}_1) = \iint U(\mathbf{R}_2) K(\mathbf{R}_1, \mathbf{R}_2)\, dS_2, \quad (1.3)$$

where the integration extends over the plane $z = z_2$ and the kernel $K(\mathbf{R}_1, \mathbf{R}_2)$ may be expressed in the form

$$K(\mathbf{R}_1, \mathbf{R}_2) = -\frac{1}{2\pi}\frac{\partial}{\partial z_2}[\exp(-ikr)/r] + S(\mathbf{R}_1, \mathbf{R}_2), \quad (1.4)$$

* Research supported by the Air Force Office of Scientific Research.

[1] E. Wolf and J. R. Shewell, Phys. Letters **25A** (1967),417; **26A** (1967) 104.

[2] The Rayleigh formula may be written in the form

$$U(\mathbf{R}_2) = \iint U(\mathbf{R}_1) K(\mathbf{R}_2, \mathbf{R}_1)\, dS_1,$$

where

$$K(\mathbf{R}_2, \mathbf{R}_1) = -\frac{1}{2\pi}\frac{\partial}{\partial z_2}\left(\frac{\exp(ik|\mathbf{R}_1 - \mathbf{R}_2|)}{|\mathbf{R}_1 - \mathbf{R}_2|}\right)$$

and the integration is carried out over the plane $z = z_1$.

$r = |\mathbf{R}_1 - \mathbf{R}_2|$, and $S(\mathbf{R}_1, \mathbf{R}_2)$ is a singular integral defined formally by the equation

$$S(\mathbf{R}_1, \mathbf{R}_2) = \frac{k^2}{\pi}\int_1^\infty \sinh\,[k(\rho^2 - 1)^{\frac{1}{2}}(z_2 - z_1)]J_0(kr_1\rho)\rho\, d\rho. \tag{1.5}$$

Here $r_1 = [(x_2 - x_1)^2 + (y_2 - y_1)^2]^{\frac{1}{2}}$,

$$\mathbf{R}_1 \equiv (x_1, y_1, z_1), \quad \mathbf{R}_2 \equiv (x_2, y_2, z_2),$$

and J_0 is the Bessel function of the first kind and zero order.

It is the aim of the present paper to derive a solution to this problem in a mathematically rigorous manner and to demonstrate that it is possible to avoid difficulties associated with the use of a singular kernel in Eq. (1.3) while at the same time expressing the solution in such a way that it may be used in the treatment of physical problems.

2. MATHEMATICAL PRELIMINARIES[3]

Let us consider a scalar wave field $U(x, y, z)$ which satisfies the following conditions:

(a) it is a solution of the Helmholtz equation

$$(\nabla^2 + k^2)U(x, y, z) = 0 \quad \text{when} \quad z > 0;$$

(b) it assumes given boundary values

$$U(x, y, 0) = f(x, y),$$

where $f(x, y)$ has the following properties:

(i) it is sectionally continuous in the xy plane;

(ii) outside a certain circle of radius R_0, it is continuous and has continuous derivatives such that

$$|f(x, y)| < B/R, \quad |f_x(x, y)| < B/R, \quad |f_y(x, y)| < B/R,$$

where $R = (x^2 + y^2)^{\frac{1}{2}}$ and B is a constant independent of x and y;

(iii) it is square-integrable, i.e.,

$$\iint_{-\infty}^{\infty} |f(x, y)|^2\, dx\, dy < \infty;$$

(c) it is regular for $z > 0$ and satisfies the following conditions:

(i) in the domain $R = [x^2 + y^2 + z^2]^{\frac{1}{2}} > R_0$ of the half-space $z > 0$ there exists a constant C such that U and the derivative $\partial U/\partial R$ satisfy the inequalities

$$|U| < \frac{C}{R}, \quad \left|\frac{\partial U}{\partial R}\right| < \frac{C}{R};$$

(ii) in any solid sector $-\pi/2 + \delta < \theta < \pi/2 - \delta$ of the domain $R > R_0$, $z > 0$ there exists a constant $D(\delta)$ such that for all points (x, y, z) of the sector

$$\left|\frac{\partial U}{\partial R} - ikU\right| < \frac{D}{R^2}.$$

Conditions (a), (b), and (c) ensure that the solution to the boundary-value problem is unique and that we are dealing with physically reasonable fields.[4] One may show[5] that the solution $U(x, y, z)$ which satisfies conditions (a), (b), and (c) may be represented as an angular spectrum of plane waves, i.e., in the form

$$U(x, y, z) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} A(p, q) \times \exp\,[ik(px + qy + mz)]\, dp\, dq, \tag{2.1}$$

where

$$m = [1 - p^2 - q^2]^{\frac{1}{2}}, \quad \text{when} \quad p^2 + q^2 \leq 1,$$
$$= +i[p^2 + q^2 - 1]^{\frac{1}{2}}, \quad \text{when} \quad p^2 + q^2 > 1, \tag{2.2}$$

for all x and y and for $z \geq 0$. In (2.1), $A(p, q)$ is defined as

$$A(p, q) = \iint_{-\infty}^{\infty} U(x, y, 0) \exp\,[-ik(px + qy)]\, dx\, dy; \tag{2.3}$$

however, in Ref. 5 it is shown that

$$A(p, q) = \iint_{-\infty}^{\infty} U(x, y, z) \times \exp\,[-ik(px + qy + mz)]\, dx\, dy \quad \text{for all} \quad z \geq 0. \tag{2.4}$$

If we substitute from (2.4) for $A(p, q)$ (with $x = x_2$, $y = y_2$, $z = z_2$) into (2.1) (with $x = x_1$, $y = y_1$, $z = z_1$) we obtain

$$U(x_1, y_1, z_1) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} dp\, dq \iint_{-\infty}^{\infty} dx_2\, dy_2 U(x_2, y_2, z_2) \times \exp\{-ik[p(x_2 - x_1) + q(y_2 - y_1) + m(z_2 - z_1)]\}. \tag{2.5}$$

Equation (2.5) already expresses the field distribution $U(\mathbf{R}_1)$ at any point (x_1, y_1, z_1) in terms of the field distribution $U(\mathbf{R}_2)$ in the plane $z = z_2$. It should be noted that Eq. (2.5) is valid irrespective of the relative values of z_1 and z_2, i.e., irrespective of whether

[3] Reference 5, which was written after this paper was submitted, considers some of the mathematical points of this section in more detail and with greater precision.

[4] R. K. Luneburg, *Mathematical Theory of Optics* (University of California Press, Berkeley, 1966), p. 311. Note, however, that the conditions satisfied by $U(x, y, z)$ in this reference are weaker than those in the present paper [condition (b.iii) of our Sec. 2 is an additional condition, not made by Luneburg].

[5] É. Lalor, J. Opt. Soc. Am. **58**, 1235 (1968).

$z_1 \gtrless z_2$ provided only that $z_1 \geq 0$, $z_2 \geq 0$. This equation is not, however, very useful for practical calculations since it involves fourfold integrations.

Let us now consider (2.5) with $z_1 < z_2$ (solution of the inverse diffraction problem) and let us split the domain of integration into two parts:

$$\begin{aligned} U(x_1, y_1, z_1) &= \left(\frac{k}{2\pi}\right)^2 \iint_{D_1} dp\, dq \int_{-\infty}^{\infty}\int dx_2\, dy_2 U(x_2, y_2, z_2) \\ &\quad \times \exp\{-ik[p(x_2 - x_1) + q(y_2 - y_1) + m(z_2 - z_1)]\} \\ &\quad + \left(\frac{k}{2\pi}\right)^2 \iint_{D_2} dp\, dq \int_{-\infty}^{\infty}\int dx_2\, dy_2 U(x_2, y_2, z_2) \\ &\quad \times \exp\{-ik[p(x_2 - x_1) \\ &\quad + q(y_2 - y_1) + m(z_2 - z_1)]\}, \end{aligned} \tag{2.6}$$

where D_1 is the domain $p^2 + q^2 \leq 1$ and D_2 is the domain $p^2 + q^2 > 1$. It will be useful to add to the first term on the right of (2.6) the integral (shown in Appendix A to be convergent)

$$\begin{aligned} I = \left(\frac{k}{2\pi}\right)^2 &\iint_{D_2} dp\, dq \int_{-\infty}^{\infty}\int dx_2\, dy_2 U(x_2, y_2, z_2) \\ &\times \exp\{-ik[p(x_2 - x_1) + q(y_2 - y_1)]\} \\ &\times \exp[-k(p^2 + q^2 - 1)^{\frac{1}{2}}(z_2 - z_1)] \end{aligned} \tag{2.7}$$

and to subtract the same integral from the second term on the right of (2.6). Then (2.6) becomes

$$U(x_1, y_1, z_1) = U^{(1)}(x_1, y_1, z_1) + U^{(2)}(x_1, y_1, z_1), \tag{2.8}$$

where

$$\begin{aligned} U^{(1)}(x_1, y_1, z_1) &= \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} dp\, dq \int_{-\infty}^{\infty}\int dx_2\, dy_2 U(x_2, y_2, z_2) \\ &\quad \times \exp\{-ik[p(x_2 - x_1) \\ &\quad + q(y_2 - y_1) + m_1(z_2 - z_1)]\}, \end{aligned} \tag{2.9}$$

with

$$\begin{aligned} m_1 &= (1 - p^2 - q^2)^{\frac{1}{2}}, \quad \text{when} \quad p^2 + q^2 \leq 1, \\ &= -i(p^2 + q^2 - 1)^{\frac{1}{2}}, \quad \text{when} \quad p^2 + q^2 > 1, \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} U^{(2)}(x_1, y_1, z_1) &= \frac{k^2}{2\pi^2} \iint_{D_2} dp\, dq \int_{-\infty}^{\infty}\int dx_2\, dy_2 U(x_2, y_2, z_2) \\ &\quad \times \exp\{-ik[p(x_2 - x_1) + q(y_2 - y_1)]\} \\ &\quad \times \sinh[k(p^2 + q^2 - 1)^{\frac{1}{2}}(z_2 - z_1)]. \end{aligned} \tag{2.11}$$

We will now examine separately the behavior of $U^{(1)}$ and $U^{(2)}$.

3. THE BEHAVIOR OF $U^{(1)}(x_1, y_1, z_1)$

The integral (2.9) for $U^{(1)}(x_1, y_1, z_1)$ may be rewritten as follows:

$$\begin{aligned} U^{(1)}(x_1, y_1, z_1) &= \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} dp\, dq \Big[\int_{-\infty}^{\infty}\int U(x_2, y_2, z_2) \\ &\quad \times \exp[-ik(px_2 + qy_2)]\, dx_2\, dy_2\Big] \\ &\quad \times (\exp\{-ik[px_1 + qy_1 + m(z_1 - z_2)]\})^*, \end{aligned} \tag{3.1}$$

where the asterisk denotes complex conjugation. We note from (2.2) and (2.10) that m_1^* is equal to m.

We now make use of the following well-known result (Parseval's theorem) of Fourier transform theory (Ref. 6, p. 48). If $f(x)$ and $g(x)$ are square-integrable, i.e., are such that

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} |g(x)|^2\, dx < \infty,$$

then

$$\int_{-\infty}^{\infty} \hat{f}(y)\hat{g}(y)^*\, dy = 2\pi \int_{-\infty}^{\infty} f(x)g(x)^*\, dx, \tag{3.2}$$

where $\hat{f}(y)$ and $\hat{g}(y)$ are the Fourier transform of $f(x)$ and $g(x)$, respectively:

$$\hat{f}(y) = \int_{-\infty}^{\infty} f(x)e^{-ixy}\, dx, \quad \hat{g}(y) = \int_{-\infty}^{\infty} g(y)e^{-ixy}\, dx.$$

Now, let us define $\hat{U}(p, q, z_2)$ as the Fourier transform of $U(x_2, y_2, z_2)$ with respect to the first two variables:

$$\begin{aligned} \hat{U}(p, q, z_2) = \int_{-\infty}^{\infty}\int &U(x_2, y_2, z_2) \\ &\times \exp[-ik(px_2 + qy_2)]\, dx_2\, dy_2. \end{aligned} \tag{3.3}$$

Moreover, as is shown in Appendix B,

$$\exp\{-ik[px_1 + qy_1 + m(z_1 - z_2)]\} = \hat{V}(p, q, z_2), \tag{3.4}$$

where

$$V(x_2, y_2, z_2) = -\frac{1}{2\pi}\frac{\partial}{\partial z_2}[\exp(ikr)/r]. \tag{3.5}$$

In (3.5) the dependence of V upon x_1, y_1, and z_1 is suppressed and $r = |\mathbf{R}_1 - \mathbf{R}_2|$. Therefore using (3.3), (3.4), (3.5), and Parseval's theorem (3.2), $U^{(1)}$, given

by (3.1), may be expressed in the form

$$U^{(1)}(x_1, y_1, z_1) = -\frac{1}{2\pi}\iint_{-\infty}^{\infty} U(x_2, y_2, z_2) \times \frac{\partial}{\partial z_2}\left[\frac{\exp(-ik|\mathbf{R}_1 - \mathbf{R}_2|)}{|\mathbf{R}_1 - \mathbf{R}_2|}\right] dx_2\, dy_2. \quad (3.6)$$

The use of Parseval's theorem is justified since conditions (a), (b), and (c) of Sec. 2 imply that $U(x_2, y_2, z_2)$ is square-integrable,[5] i.e., that

$$\iint_{-\infty}^{\infty} |U(x_2, y_2, z_2)|^2\, dx_2\, dy_2 < \infty, \quad \text{if} \quad z_2 \geq 0, \quad (3.7)$$

and, as is easily verified,

$$\iint_{-\infty}^{\infty} |V(x_2, y_2, z_2)|^2\, dx_2\, dy_2 < \infty \quad \text{when} \quad z_2 > 0.$$

4. THE BEHAVIOR OF $U^{(2)}(x_1, y_1, z_1)$

It is convenient to rewrite Eq. (2.11) in the form

$$U^{(2)}(x_1, y_1, z_1) = \lim_{T\to\infty} \frac{k^2}{2\pi^2}\iint_{D_{2T}} dp\, dq \Big[\iint_{-\infty}^{\infty} U(x_2, y_2, z_2) \times \exp[-ik(px_2 + qy_2)]\, dx_2\, dy_2\Big] \times \exp[ik(px_1 + qy_1)] \times \sinh[k(p^2 + q^2 - 1)^{\frac{1}{2}}(z_2 - z_1)], \quad (4.1)$$

where D_{2T} is defined as the (p, q) domain such that $1 \leq (p^2 + q^2) \leq T$. Or, in a more compact form,

$$U^{(2)}(x_1, y_1, z_1) = \lim_{T\to\infty} \frac{k^2}{2\pi^2}\iint_{-\infty}^{\infty} f(p, q)\hat{U}(p, q, z_2)\, dp\, dq, \quad (4.2)$$

where

$$f(p, q) = \exp[ik(px_1 + qy_1)] \times \sinh[k(p^2 + q^2 - 1)^{\frac{1}{2}}(z_2 - z_1)] \quad \text{if} \quad 1 \leq (p^2 + q^2)^{\frac{1}{2}} \leq T,$$
$$= 0 \quad \text{otherwise.}$$

There is a well-known theorem in Fourier analysis (Ref. 6, p. 49) which states that if $f(x)$ and $g(x)$ are square-integrable, then

$$\int_{-\infty}^{\infty} f(x)\hat{g}(x)\, dx = \int_{-\infty}^{\infty} \hat{f}(y)g(y)\, dy. \quad (4.3)$$

Thus we see that since $f(p, q)$ is square-integrable {because of the truncation and since $f(p, q)$ is bounded by $\sinh[k(T^2 - 1)(z_2 - z_1)]$} and since $U(x_2, y_2, z_2)$ is also square-integrable [cf. Eq. (3.7)] we may rewrite Eq. (4.2) as

$$U^{(2)}(x_1, y_1, z_1) = \lim_{T\to\infty} \frac{k^2}{2\pi^2}\iint_{-\infty}^{\infty} U(x_2, y_2, z_2)\hat{f}(x_2, y_2)\, dx_2\, dy_2 \quad (4.4)$$

or

$$U^{(2)}(x_1, y_1, z_1) = \lim_{T\to\infty} \frac{k^2}{2\pi^2}\iint_{-\infty}^{\infty} dx_2\, dy_2 U(x_2, y_2, z_2) \times \iint_{D_{2T}} \sinh[k(p^2 + q^2 - 1)^{\frac{1}{2}}(z_2 - z_1)] \times \exp[ik(px_1 + qy_1)] \exp[-ik(px_2 + qy_2)]\, dp\, dq. \quad (4.5)$$

It is now convenient to change to polar coordinates:

$$p = r\cos\theta, \qquad q = r\sin\theta,$$
$$x_1 - x_2 = R_1\cos\varphi, \quad y_1 - y_2 = R_1\sin\varphi.$$

Equation (4.5) becomes

$$U^{(2)}(x_1, y_1, z_1) = \lim_{T\to\infty} \frac{k^2}{2\pi^2}\iint_{-\infty}^{\infty} dx_2\, dy_2 U(x_2, y_2, z_2) \times \int_1^T r\, dr \int_0^{2\pi} d\theta \sinh[k(r - 1)^{\frac{1}{2}}(z_2 - z_2)] \times \exp[ikrR_1\cos(\theta - \varphi)]. \quad (4.6)$$

The integral

$$I = \frac{1}{2\pi}\int_0^{2\pi} d\theta \exp[ikrR_1\cos(\theta - \varphi)]$$

is the well-known integral representation of[7] $J_0(krR_1)$. Using this fact, (4.6) may be rewritten as

$$U^{(2)}(x_1, y_1, z_1) = \lim_{T\to\infty} \frac{k^2}{\pi}\iint_{-\infty}^{\infty} dx_2\, dy_2 U(x_2, y_2, z_2) S_T(\mathbf{R}_1, \mathbf{R}_2), \quad (4.7)$$

where

$$S_T(\mathbf{R}_1, \mathbf{R}_2) \equiv \int_1^T J_0(krR_1)\sinh[k(r^2 - 1)^{\frac{1}{2}}(z_2 - z_1)]r\, dr. \quad (4.8)$$

Combining (3.6) with (4.7) according to (2.8), we finally obtain the following expression for the field in the plane $z = z_1$ in terms of the field in the plane

[6] R. R. Goldberg, *Fourier Transforms* (Cambridge University Press, Cambridge, England, 1961).

[7] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 621.

$z = z_2$:

$$U(x_1, y_1, z_1) = -\frac{1}{2\pi}\iint_{-\infty}^{\infty} U(x_2, y_2, z_2) \times \frac{\partial}{\partial z_2}\left[\frac{\exp[-ik|\mathbf{R}_1 - \mathbf{R}_2|]}{|\mathbf{R}_1 - \mathbf{R}_2|}\right] dx_2\, dy_2 + \lim_{T\to\infty} \frac{k^2}{\pi}\iint_{-\infty}^{\infty} U(x_2, y_2, z_2) S_T(\mathbf{R}_1, \mathbf{R}_2)\, dx_2\, dy_2, \tag{4.9}$$

where $S_T(\mathbf{R}_1, \mathbf{R}_2)$ is defined by (4.8).

The form (4.7) avoids the use of a singular kernel and gives a rigorous way of operating with the singular kernel introduced by Wolf and Shewell.[1] Examination of the form of $U^{(2)}(x_1, y_1, z_1)$ reveals that it is composed entirely of inhomogeneous waves, which, as is well known, carry information about spatial periodicities that are smaller than a wavelength.

Though the integral for $U^{(2)}(x_1, y_1, z_1)$ in Eq. (4.7) converges, it is evident that even a small perturbation of $U(x_2, x_2, z_2)$ away from its exact value may cause it to diverge. This instability is a mathematical consequence of the type of boundary-value problem considered. This fact would seem to limit the practical value of calculating with this term since experimental measurement of $U(x_2, y_2, z_2)$ will never be precise.

However, if information about spatial periodicities that are smaller than a wavelength is not required, as, for instance, in most questions of practical interest in optics, the term $U^{(2)}$ could simply be neglected. It should also be noted that the response of any physical detector is frequency-dependent and in fact any real detector is unable to resolve frequencies higher than some finite value. This physical cutoff ensures that in practical calculations one is never required to proceed to the limit in Eq. (4.7).

ACKNOWLEDGMENT

It is a pleasure to acknowledge the constant encouragement and assistance of Professor Emil Wolf during the course of this work.

APPENDIX A: THE CONVERGENCE OF THE INTEGRAL I IN EQUATION (2.7)

The convergence of the integral defined in (2.7),

$$I = \left(\frac{k}{2\pi}\right)^2 \iint_{D_2} dp\, dq \iint dx_2\, dy_2 U(x_2, y_2, z_2) \times \exp\{-ik[p(x_2 - x_1) + q(y_2 - y_1)]\} \times \exp[-k(p^2 + q^2 - 1)^{\frac{1}{2}}(z_2 - z_1)],$$

may be shown by the following argument. On carrying out the integrations over the x_2y_2 plane, it takes the form

$$I = \left(\frac{k}{2\pi}\right)^2 \iint_{D_2} dp\, dq A(p, q) \exp[ik(px_1 + qy_1)] \times \exp[-k(p^2 + q^2 - 1)^{\frac{1}{2}}(2z_2 - z_1)]. \tag{A1}$$

Since $2z_2 - z_1$ is greater than zero, we see from (2.1) that

$$U(x_1, y_1, 2z_2 - z_1) = \left(\frac{k}{2\pi}\right)^2 \iint_{D_1} dp\, dq A(p, q) \times \exp\{ik[px_1 + qy_1 + m(2z_2 - z_1)]\} + I. \tag{A2}$$

Since the left-hand side of (A2) is finite and since, as we shall show, the first term on the right-hand side of (A2) converges, I must be finite. To show that the first term

$$\left(\frac{k}{2\pi}\right)^2 \iint_{D_1} dp\, dq A(p, q) \times \exp[ik(px_1 + qy_1 + m(2z_2 - z_1))] \equiv I_1 \tag{A3}$$

on the right of (A2) is finite we note first that if I_1 diverges then certainly the integral

$$\iint_{D_1} |A(p, q)|\, dp\, dq \quad \text{diverges.} \tag{A4}$$

Let us now assume that I_1 does not converge. We shall show that this assumption leads to a contradiction. We now divide the domain of integration D_1 into two domains $D_{1>}$ and $D_{1<}$, where $D_{1>}$ consists of the set of points in D_1 for which $|A(p, q)| \geq 1$ and $D_{1<}$ consists of the set of points in D_1 for which $|A(p, q)| < 1$. Obviously $D_1 = D_{1>} + D_{1<}$. The integral

$$\iint_{D_{1<}} |A(p, q)|\, dp\, dq$$

is evidently finite; therefore our assumption implies that the integral

$$\iint_{D_{1>}} |A(p, q)|\, dp\, dq \quad \text{diverges.}$$

In the domain $D_{1>}$,

$$|A(p, q)|^2 \geq |A(p, q)| \tag{A5}$$

and therefore the integral

$$\iint_{D_1>} |A(p, q)|^2 \, dp \, dq \quad \text{diverges.}$$

This implies that the integral

$$\int_{-\infty}^{\infty}\int |A(p, q)|^2 \, dp \, dq \quad \text{diverges.}$$

However, we know from condition (b.iii) of Sec. 2, Eq. (2.3), and Parseval's theorem, that

$$\int_{-\infty}^{\infty}\int |A(p, q)|^2 \, dp \, dq < \infty. \tag{A6}$$

Thus we have a contradiction. Hence I_1 converges. This completes the proof.

APPENDIX B: PROOF OF RELATION (3.4)

We shall now show that the Fourier transform of

$$V(x_2, y_2, z_2) = -\frac{1}{2\pi}\frac{\partial}{\partial z_2}[\exp(ikr)/r] \tag{B1}$$

is $\exp\{-ik[px_1 + qy_1 + m(z_1 - z_2)]\}$. We start with the well-known formula due to Weyl[8]:

$$\frac{\exp(ik|\mathbf{R}_1 - \mathbf{R}_2|)}{|\mathbf{R}_1 - \mathbf{R}_2|} = \frac{ik}{2\pi}\iint \exp\{ik[p(x_2 - x_1) + q(y_2 - y_1) + m(z_2 - z_1)]\}\frac{dp\,dq}{m}, \tag{B2}$$

where m is defined by (2.2). On differentiating both sides of (B2) with respect to z_2 and interchanging the order of differentiation and integration on the right, we obtain

$$-\frac{1}{2\pi}\frac{\partial}{\partial z_2}\left[\frac{\exp(ik|\mathbf{R}_1 - \mathbf{R}_2|)}{|\mathbf{R}_1 - \mathbf{R}_2|}\right] = \left(\frac{k}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int dp\,dq \exp\{ik[p(x_2 - x_1) + q(y_2 - y_1) + m(z_2 - z_1)]\}. \tag{B3}$$

The inversion of the order of differentiation and integration is justified since the integral on the right of (B3) is uniformly convergent for $z_2 - z_1 > \delta$, where δ is any positive number. This may be seen by applying the Weierstrass M test[9] because the integrand is dominated by $|\exp(ikm\delta)|$ and since this exponential is absolutely integrable over the p, q plane. Now since $V(x_2, y_2, z_2)$ is square-integrable, i.e., since

$$\int_{-\infty}^{\infty}\int |V(x_2, y_2, z_2)|^2 \, dx_2 \, dy_2 < \infty,$$

it has a Fourier representation,

$$V(x_2, y_2, z_2) = \left(\frac{k}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int \hat{V}(p, q, z_2)\exp[ik(px_2 + qy_2)]\,dp\,dq. \tag{B4}$$

Since $V(x_2, y_2, z_2)$ is given by (B1), comparison of (B4) and (B3) shows that

$$\hat{V}(p, q, z_2) = \exp\{-ik[px_1 + qy_1 + m(z_1 - z_2)]\}.$$

This completes the proof.

[8] H. Weyl, Ann. Physik **60**, 481 (1919). Equation (B2) is a straightforward modification of the result given in this paper.

[9] E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford University Press, London, 1935), p. 111.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Response of a Many-Particle System to Quasistatic Changes in Volume*

N. S. GILLIS
Sandia Laboratory, Albuquerque, New Mexico

(Received 21 May 1968)

A study is made of the linear dissipative processes associated with viscous flow in a many-particle system. Central to this study is a direct examination of the linear-response properties of the system when it is influenced by a dynamical perturbation which induces a time-dependent change in the size and shape of the containing volume of the system. By examining the linear response of the pressure–tensor operator to such a perturbation, we derive correlation-function expressions for the coefficients of viscosity η and ζ and for the shear and bulk moduli G_0 and K_0. Essential to this discussion is the careful examination of the behavior of certain spectral functions at *finite* volume. It is found that in order to carry through the analysis in a consistent fashion, one must require that these spectral functions exhibit a special singular behavior. In particular, it is found that the static moduli G_0 and K_0 are related to the singular parts of the relevant spectral functions, whereas the viscosity coefficients are related to the nonsingular parts of these same spectral functions. The expressions obtained are compared with the familiar Kubo–Mori expressions for the coefficients of viscosity η and ζ.

I. INTRODUCTION

In a previous paper[1] we carried out a detailed study of the response of a macroscopic system to an external perturbation. A major part of that analysis was devoted to an examination of the properties of the time-dependent correlation functions from which the transport properties could be derived. The present study again concerns itself with the response of a macroscopic system, but from a somewhat more restricted point of view. In particular, we concern ourselves with a detailed study of the equilibrium and nonequilibrium phenomena associated with the visco–elastic processes which may be set up in the system. The analysis is carried out within the context of a linear-response calculation, employing a dynamical perturbation which induces a time-dependent change in the size and shape of the containing volume of the system. The dynamical perturbation chosen is that of a "box potential" $U(\mathbf{r})$ which becomes infinite at the spatial boundaries of the system. In order to eliminate the explicit presence of $U(\mathbf{r})$ in the dynamical perturbation, a canonical transformation is employed which induces a scale transformation on the field operators. This then allows us to take the large-volume limit at an appropriate point in the calculation.

Because the volume of the system is in effect constrained by the dynamical perturbation employed, one must examine carefully the behavior of the relevant correlation functions at *finite* volume. As in the infinite-volume case, certain frequency-dependent spectral functions derived from these correlation functions exhibit $\delta(\omega)$ singularities. Only by taking proper account of their singular nature and carrying out the large volume limit correctly can the calculations be carried through in a consistent fashion. In the course of our analysis, we obtain correlation-function expressions for the coefficients of viscosity ζ and η and for the compressional and shear moduli K_0 and G_0. We find that K_0 and G_0 are related to the singular parts of certain spectral functions, whereas ζ and η are obtained from the nonsingular parts of these same spectral functions. In the large-volume limit the expressions arrived at are consistent with those found previously for the infinite-volume system.[1]

The present work is an extension of an earlier calculation of Montroll.[2] In that calculation, Montroll restricted himself to a consideration of the shear viscosity alone and did not choose to consider those processes for which volume changes could occur. The present work attempts to be more general and, at the same time, allows for the singular nature of the relevant spectral functions.

In the following section we concern ourselves with a brief discussion of the phenomenological equations of viscosity, preparatory to the full-fledged microscopic linear-response theory to be presented in Sec. III.

II. PHENOMENOLOGICAL EQUATIONS

In this section we will concern ourselves with a brief discussion of the constitutive equations of hydrodynamics. Our primary goal here will be to recast the equations in a form which explicitly exhibits the parameters directly related to the external mechanical

* This work was supported by the U.S. Atomic Energy Commission.

[1] R. D. Puff and N. S. Gillis, Ann. Phys. (N.Y.) **46**, 364 (1968). Hereafter in the text this will be referred to as (PG).

[2] E. W. Montroll, *Rendiconti della Scuola Internazionale di Fisica, Corso* 10 (Edizione Nicola Zanichelli, Bologna, Italy, 1959), p. 242.

perturbation which alters the shape and size of the containing volume.

The constitutive equations which we will be concerned with are, in linearized form,

$$T_{ij}(\mathbf{r}t) = \delta_{ij}P(\mathbf{r}t) - \eta(\nabla_i v_j(\mathbf{r}t) + \nabla_j v_i(\mathbf{r}t)) - \delta_{ij}(\zeta - \tfrac{2}{3}\eta)(\boldsymbol{\nabla}\cdot\mathbf{v}(\mathbf{r}t)), \quad (1)$$

$$\mathbf{J}^{\varepsilon}(\mathbf{r}t) = (P_0 + \mathcal{E}_0)\mathbf{v}(\mathbf{r}t) - \kappa\boldsymbol{\nabla}T(\mathbf{r}t). \quad (2)$$

T_{ij} and $\mathbf{J}^{\varepsilon}$ are the momentum–flux tensor and energy current, respectively, $\mathbf{v}(\mathbf{r}t)$ and $T(\mathbf{r}t)$ represent the local velocity and temperature of a fluid element, P_0 and $\mathcal{E}_0$ are the equilibrium pressure and energy density, η and ζ are the shear and bulk coefficients of viscosity, and κ is the thermal conductivity. Finally, we define the local pressure $P(\mathbf{r}t)$ as follows: If $P_0(\rho_0, T_0)$ represents the equilibrium pressure as a function of the equilibrium–particle density and temperature ρ_0 and T_0, then $P(\mathbf{r}t) \equiv P_0(\rho(\mathbf{r}t), T(\mathbf{r}t))$. Equations (1) and (2), together with the conservation laws for particle density, momentum density, and energy density, provide a complete hydrodynamic description of a single-component fluid.

In what follows, we will be interested primarily in momentum–relaxation processes which occur in a fluid exhibiting both shear and bulk viscosities. In this respect it is well to note that, for *adiabatic* processes, (1) and (2) will not be coupled by the conservation laws. To see this, we remember first of all that adiabaticity requires that the entropy per unit mass remains constant. In terms of the entropy per unit volume $s(\mathbf{r}t)$, this implies that

$$s(\mathbf{r}t) - s_0 = [(\rho(\mathbf{r}t) - \rho_0)/\rho_0]s_0. \quad (3)$$

Thus, considering $P(\mathbf{r}t)$ as a function of $\rho(\mathbf{r}t)$ and $s(\mathbf{r}t)$, for small deviations from equilibrium we have

$$P(\rho(\mathbf{r}t), s(\mathbf{r}t)) = P(\rho_0, s_0) + (\partial P/\partial\rho)_s(\rho(\mathbf{r}t) - \rho_0) + (\partial P/\partial s)_\rho(s(\mathbf{r}t) - s_0). \quad (4)$$

Substituting (3) into (4) results in

$$P(\rho(\mathbf{r}t), s(\mathbf{r}t)) = P(\rho_0, s_0) + K_0[(\rho(\mathbf{r}t) - \rho_0)/\rho_0], \quad (5)$$

where $K_0 \equiv -V(\partial P/\partial V)_{SN}$ is the static compressional modulus. Equation (5), together with (1) and the momentum- and particle-density conservation laws, provides a closed set of equations which can be solved independently of the equations describing thermal relaxation processes.

For our purposes it will be useful to consider a somewhat more general situation than that described by (1) by formally examining the case of a single-component fluid with a nonzero static shear modulus. Now, for a strained elastic solid, the stress tensor takes the form

$$T_{ij}(\mathbf{r}t) = \delta_{ij}P_0 - G_0(\nabla_i\sigma_j(\mathbf{r}t) + \nabla_j\sigma_i(\mathbf{r}t)) - \delta_{ij}(K_0 - \tfrac{2}{3}G_0)\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\mathbf{r}t), \quad (6)$$

where $\boldsymbol{\sigma}(\mathbf{r}t)$ represents the vector displacement of an element of material from its equilibrium position and G_0 is the static shear modulus. We can represent both the hydrodynamic limit and the elastic-solid limit, and also any intermediate case (visco–elastic fluid), if we write the time–Fourier transform of

$$\delta T_{ij}(\mathbf{r}t) \equiv T_{ij}(\mathbf{r}t) - \delta_{ij}P_0$$

as

$$\delta T_{ij}(\mathbf{r}\omega) = -G(\omega)(\nabla_i\sigma_j(\mathbf{r}\omega) + \nabla_j\sigma_i(\mathbf{r}\omega)) - \delta_{ij}(K(\omega) - \tfrac{2}{3}G(\omega))\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\mathbf{r}\omega). \quad (7)$$

The parameters $G(\omega)$ and $K(\omega)$ are, in general, frequency-dependent and complex; they exhibit dispersion of the form

$$G(\omega) = G_0 - i\omega\eta(\omega), \qquad K(\omega) = K_0 - i\omega\zeta(\omega). \quad (8)$$

The extreme hydrodynamic limit is characterized by $G_0 = 0$ with η and ζ real constants (in the absence of dispersion). K_0 will, in general, be nonzero unless the fluid is incompressible. On the other hand, the elastic limit is characterized by $\eta, \zeta = 0$; G_0, K_0 nonzero.

We now rewrite (7) as

$$\delta T_{ij}(\mathbf{r}\omega) = -G_0(\nabla_i\sigma_j(\mathbf{r}\omega) + \nabla_j\sigma_i(\mathbf{r}\omega)) - \delta_{ij}(K_0 - \tfrac{2}{3}G_0)\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\mathbf{r}\omega) - \eta(\nabla_i v_j(\mathbf{r}\omega) + \nabla_j v_i(\mathbf{r}\omega)) - \delta_{ij}(\zeta - \tfrac{2}{3}\eta)\boldsymbol{\nabla}\cdot\mathbf{v}(\mathbf{r}\omega), \quad (9)$$

where $\mathbf{v}(\mathbf{r}\omega) = -i\omega\boldsymbol{\sigma}(\mathbf{r}\omega)$.

The expression (9) represents the response of the stress tensor when the system is taken from some equilibrium situation to a nonequilibrium state in which the "driving forces" $\nabla_i\sigma_j$ and $\nabla_i v_j$ are nonzero. To accomplish this latter situation, we expose the system to a well-defined external mechanical disturbance. Further, it will be desirable to put (9) into a more convenient form by eliminating the "driving forces" $\nabla_i\sigma_j$, $\nabla_i v_j$ in terms of parameters directly related to the external perturbation.

Let the original equilibrium situation be that in which the system is enclosed in a cubical box whose sides are defined by the planes

$$X = 0, \quad X = L_0, \qquad Y = 0, \quad Y = L_0, \qquad Z = 0, \quad Z = L_0. \quad (10)$$

The nonequilibrium state is to be simulated by allowing the sides of the box to move in time, i.e., at time t we have

$$\begin{aligned} X &= 0, \quad X = L_1(t), \\ Y &= 0, \quad Y = L_2(t), \\ Z &= 0, \quad Z = L_3(t), \end{aligned} \tag{11}$$

where

$$L_l(t) = L_0 e^{\epsilon_l(t)} = L_0 + \epsilon_l(t) L_0 + O(\epsilon^2),$$
$$\epsilon_l(t) = \epsilon_l e^{-i\omega t}; \quad l = 1, 2, 3.$$

Consider the expression (4). If we integrate (4) over the volume $V(t) = L_1(t)L_2(t)L_3(t)$, we obtain

$$\int_V d^3\mathbf{r} P(\mathbf{r}t) = P_0 V(t) + \left(\frac{\partial P}{\partial \rho}\right)_s [N(t) - \rho_0 V(t)] + \left(\frac{\partial P}{\partial s}\right)_\rho [S(t) - s_0 V(t)],$$

where

$$N(t) = \int_V d^3\mathbf{r}\rho(\mathbf{r}t),$$
$$S(t) = \int_V d^3\mathbf{r} s(\mathbf{r}t).$$

Clearly, the total number of particles remains constant, so that

$$N(t) = N_0 = \rho_0 V_0.$$

Also,

$$V(t) = V_0\left(1 + \sum_l \epsilon_l(t) + O(\epsilon^2)\right).$$

It is important to note at this point that we are considering a *thermally isolated* system; i.e., a system influenced only by a change in an external field. For such a system and for small deviations from equilibrium, one can show that[3]

$$S(t) = s_0 V_0 + O(\epsilon\dot{\epsilon}).$$

Thus, for a quasistatic change in the volume of the system,

$$\int_V d^3\mathbf{r}[P(\mathbf{r}t) - P_0] = -V_0 K_0 \sum_l \epsilon_l(t) + O(\epsilon^2).$$

The above relation is valid under the assumption that the change in the total entropy is of second order. This is somewhat less restrictive than the adiabatic assumption, which would require statements about the local entropy density. The point here is that, as long as we are concerned only with the response of the spatially integrated local pressure, we may neglect all terms due to changes in the entropy.

We can apply considerations similar to the above to the general expression (7). Integrating over the volume $V(t)$ results in

$$\int_V d^3\mathbf{r}\delta T_{ij}(\mathbf{r}\omega) = -G(\omega)\int_\Omega [d\Omega_i \sigma_j(\omega) + d\Omega_j \sigma_i(\omega)] - (K(\omega) - \tfrac{2}{3}G(\omega))\int_\Omega d\boldsymbol{\Omega}\cdot\boldsymbol{\sigma}(\omega), \tag{12}$$

where we have converted integrals over the volume $V(t)$ to integrals over the surface of the containing volume $\Omega(t)$. By imposing the condition that the material adhere to the walls, we arrive at the following physical boundary conditions:

$$\begin{aligned} \sigma_\perp(t) &= \epsilon_i(t) L_0 + O(\epsilon^2), \quad \text{at wall surface } i, \\ \sigma_\parallel(t) &= 0, \quad \text{at all wall surfaces.} \end{aligned} \tag{13}$$

$\sigma_\perp$ and $\sigma_\parallel$ are the normal and tangential components (respectively) of σ at the wall surface. Equation (13) in conjunction with (12) yields the result

$$\int_V d^3\mathbf{r}\delta T_{ij}(\mathbf{r}\omega) = -2G(\omega)V_0\epsilon_i\delta_{ij} - (K(\omega) - \tfrac{2}{3}G(\omega))V_0\delta_{ij}\sum_l \epsilon_l + O(\epsilon^2). \tag{14}$$

In Eq. (14) we have a constitutive equation which expresses the linear response of the integrated stress tensor in terms of the parameters ϵ_i defining the external perturbation. The form of (14) is particularly useful, since it will permit direct comparison with the exact linear response calculation to be presented in the following section.

III. LINEAR-RESPONSE CONSIDERATIONS

Our goal in this section will be to examine in detail the change in the average value of the pressure-tensor operator[4] $T_{ij}(\mathbf{r}t)$ when the system is influenced by an external time-dependent perturbation which alters the shape and size of the containing volume of the system.

As is well known, the general quantum-statistical average of an operator representing some dynamical quantity $O(\mathbf{r}t)$ is defined by

$$\langle O(\mathbf{r})\rangle(t) \equiv \mathrm{Tr}\,[w(t)O(\mathbf{r})]. \tag{15}$$

Here, $O(\mathbf{r})$ is the operator in the Schrödinger picture, and $w(t)$ is the density-matrix operator satisfying the equation of motion

$$i\frac{\partial}{\partial t} w(t) = [H(t), w(t)].$$

We suppose that the Hamiltonian $H(t)$ of our system may be written as

$$H(t) = \begin{cases} H_0 + H_1(t), & t > t_0, \\ H_0, & t < t_0. \end{cases}$$

[3] See, e.g., L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1958), pp. 36 ff.

[4] That there exists a well-defined operator $T_{ij}(\mathbf{r}t)$ has been shown in (PG). See Eq. (2.3) of Ref. 1.

$H_1(t)$ is to be regarded as a time-dependent perturbation (as yet unspecified) on the unperturbed system Hamiltonian H_0.

Now, we wish to examine the linear response of our system to a time-dependent change in volume. In order to do this, we confine our system in equilibrium to a finite volume V_0 by introducing a static "box potential" $U(\mathbf{r})$ which becomes infinite at the spatial boundaries of the system. We then vary the shape and size of the containing volume by performing a time-dependent scale transformation on $U(\mathbf{r})$, i.e., we make the replacement

$$U(x, y, z) \to U(xe^{-\epsilon_1(t)}, ye^{-\epsilon_2(t)}, ze^{-\epsilon_3(t)}). \quad (16)$$

Since we are ultimately concerned with the large-volume limit of the system, the geometry of the containing volume is irrelevant. However, for definiteness we assume that $U(\mathbf{r})$ represents the region defined by (10), whereas the scale-transformed U represents the region defined by (11). The Hamiltonian of the system containing the "box potential" we denote by $H_\epsilon(t)$. Thus,

$$\begin{aligned} H_\epsilon(t) = {} & \frac{\hbar^2}{2M}\int d^3\mathbf{r}(\nabla\psi^\dagger(\mathbf{r}))\cdot(\nabla\psi(\mathbf{r})) \\ & + \tfrac{1}{2}\int d^3\mathbf{r}d^3\mathbf{r}'V(|\mathbf{r}-\mathbf{r}'|)\psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r}) \\ & + \int d^3\mathbf{r}'\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')U(x'e^{-\epsilon_1(t)}, y'e^{-\epsilon_2(t)}, z'e^{-\epsilon_3(t)}). \end{aligned} \quad (17)$$

Clearly, if we expand $H_\epsilon(t)$ to terms linear in $\epsilon(t)$, our choice for $H_1(t)$ must involve $U(\mathbf{r})$ explicitly. In order to avoid the explicit presence of $U(\mathbf{r})$ in our perturbation expansion, we perform a canonical transformation on the field operators. Introducing the unitary transformation

$$R \equiv \exp\left[-i\sum_l \epsilon_l(t)S_{ll}\right],$$

discussed in the Appendix, we rewrite (15) as

$$\begin{aligned} \langle O(\mathbf{r})\rangle(t) &= \mathrm{Tr}\,[R^\dagger w(t)O(\mathbf{r})R] \\ &= \mathrm{Tr}\,[\bar{w}(t)\bar{O}(\mathbf{r})], \end{aligned} \quad (18)$$

where

$$\bar{w} \equiv R^\dagger w R,$$
$$\bar{O} = R^\dagger O R.$$

$\bar{w}_\epsilon(t)$ satisfies the equation of motion

$$i\frac{\partial}{\partial t}\bar{w}(t) = \left[\bar{H}_\epsilon(t) - \sum_l \dot{\epsilon}_l(t)S_{ll}, \bar{w}(t)\right], \quad (19)$$

where

$$\bar{H}_\epsilon(t) \equiv R^\dagger H_\epsilon(t) R.$$

We can interpret (19) by observing that $\bar{w}(t)$ develops in time according to the effective Hamiltonian

$$H_{\mathrm{eff}}(t) \equiv \bar{H}_\epsilon(t) - \sum_l \dot{\epsilon}_l(t)S_{ll}. \quad (20)$$

Up to terms linear in ϵ, (20) may be written

$$H_{\mathrm{eff}}(t) = H_0 + \sum_l \left[\epsilon_l(t)\left(\frac{\partial\bar{H}_\epsilon}{\partial\epsilon_l}\right)_{\epsilon=0} - \dot{\epsilon}_l(t)S_{ll}\right],$$

where

$$H_0 \equiv (\bar{H}_\epsilon(t))_{\epsilon=0} = (H_\epsilon(t))_{\epsilon=0}.$$

It is to be noted that $(\partial\bar{H}_\epsilon/\partial\epsilon_l)_0$ does not contain $U(\mathbf{r})$ explicitly. Indeed, we easily find that

$$\left(\frac{\partial\bar{H}_\epsilon}{\partial\epsilon_l}\right)_{\epsilon=0} = -\int d^3\mathbf{r}T_{ll}(\mathbf{r}).$$

The usual linear-response analysis can now be carried through using the dynamical perturbation

$$H_1(t) = -\sum_l\left[\epsilon_l(t)\int d^3\mathbf{r}T_{ll}(\mathbf{r}) + \dot{\epsilon}_l(t)S_{ll}\right]. \quad (21)$$

Thus, the linear-response statement derived from (15) reads

$$\begin{aligned} \langle O(\mathbf{r})\rangle(t) = \langle\bar{O}(\mathbf{r}t)\rangle_0 - i\sum_l\int_{t_0}^t dt'\Big\langle\Big[\epsilon_l(t')\int d^3\mathbf{r}'T_{ll}(\mathbf{r}'t') \\ + \epsilon_l(t')S_{ll}(t'), \bar{O}(\mathbf{r}t)\Big]\Big\rangle_0. \end{aligned} \quad (22)$$

All operators on the right-hand side of (22) are to be treated as Heisenberg operators developing in time according to H_0. $\bar{O}(\mathbf{r}t)$ to terms linear in ϵ may be written

$$\bar{O}(\mathbf{r}t) = O(\mathbf{r}t) + i\sum_l \epsilon_l(t)[S_{ll}(t), O(\mathbf{r}t)].$$

The above expression, in conjunction with (22), yields

$$\begin{aligned} \delta\langle O(\mathbf{r})\rangle(t) = {} & -i\sum_l\int_{t_0}^t dt'\epsilon_l(t')\langle[T_{ll}(t'), O(\mathbf{r}t)]\rangle_0 \\ & + i\sum_l\int_{t_0}^t dt'\epsilon_l(t')\langle[\dot{S}_{ll}(t'), O(\mathbf{r}t)]\rangle_0, \end{aligned} \quad (23)$$

where $T_{ij}(t)$ has been used to denote the spatially integrated stress tensor. In obtaining the final expression (23), we have integrated once by parts in (22) and noted that $\epsilon_l(t_0) = 0$. The equilibrium average $\langle\cdots\rangle_0$ in (23) is carried out with respect to the equilibrium-density matrix

$$w_{\mathrm{eq}} = \bar{w}(t_0) = \frac{\exp\beta(\mu N - H_0)}{\mathrm{Tr}\,[\exp\beta(\mu N - H_0)]}.$$

It is important to notice at this point that the Hamiltonian H_0, which appears in $w_\epsilon(t_0)$ and which determines the time dependence of the field operators, contains the static box potential $U(\mathbf{r})$. In general, then, neither the total momentum operator nor the total angular-momentum operator commutes with the Hamiltonian of the system. Hence, equilibrium correlation functions of the general form

$$\langle A(\mathbf{r}t)B(\mathbf{r}'0)\rangle_0 - \langle A(\mathbf{r})\rangle_0\langle B(\mathbf{r}')\rangle_0$$

will be functions of $\mathbf{r}$ and $\mathbf{r}'$, separately, and not merely functions of the magnitude of $\mathbf{r} - \mathbf{r}'$. Only in the large-volume limit will we be able to invoke space-translational and rotational invariance. Time-translational invariance still holds, of course, and the usual arguments involving time reversal remain unchanged.

The expression (23) is useful in that it does not exhibit the "box potential" explicitly. In principle, then, one could at this point calculate all relevant quantities by quantizing the field operators in a finite cubical volume, thus taking into account the implicit presence of the "box potential" by imposing on the fields the requirement that they vanish at the spatial boundaries of the system. This, of course, can be accomplished by expanding the field operators in terms of the appropriate complete set of eigenfunctions. Indeed, the results of the following calculations will serve to illustrate some special features of the equilibrium correlation functions for a finite system. In particular, we will see explicitly that general statements about the large-volume limit require careful discussion.

Let us now apply (23) to an actual physical situation. We suppose that at time $t = -\infty$ the system occupies a volume V_0. At this initial time we begin to increase the volume quasistatically, so that for $t > -\infty$ the system occupies a volume $V(t) > V_0$. In (23) we set $t_0 = -\infty$ and choose $\epsilon_l(t)$ to be independent of l and of the special form

$$\epsilon_l(t) = \begin{cases} \epsilon_0 \exp(\alpha t), & -\infty \leq t < 0, \\ \epsilon_0, & t \geq 0. \end{cases} \tag{24}$$

This form corresponds to a *quasistatic* increase in volume, since we take α to be a small positive real constant. *The adiabatic* limit $\alpha \to 0^+$ for our volume change will be taken at the end of calculations involving the time dependence, but it should be emphasized that we are always working with *finite* volume. If a final thermodynamic limit $V_0 \to \infty$ is required, we will take this limit last. The discussion below will clarify the distinction between the two limiting processes $V_0 \to \infty$, $\alpha \to 0^+$ and $\alpha \to 0^+$, $V_0 \to \infty$, and will emphasize the difference between the physical processes involved in the two cases.

The choice (24) for $\epsilon_l(t)$ implies that for any time $t > 0$ the system will have undergone a volume change

$$\begin{aligned} \delta V &= (V_0 \exp(3\epsilon_0) - V_0) \\ &= 3\epsilon_0 V_0 + O(\epsilon^2). \end{aligned}$$

We now define time–Fourier transforms (spectral functions) of the equilibrium commutator functions as follows:

$$\langle [A(\mathbf{r}t), B(\mathbf{r}'t')]\rangle_0 \equiv \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega(t-t')} \langle [A(\mathbf{r}), B(\mathbf{r}')]\rangle_\omega . \tag{25}$$

Equations (24) and (25) allow us to rewrite (23) in the convenient form

$$\begin{aligned} &\delta\langle O(\mathbf{r})\rangle(t) \\ &= -i\epsilon_0\langle [O(\mathbf{r}t), \Gamma(t)]\rangle_0 \\ &\quad - 3\epsilon_0 \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t}\left(1 - \frac{\omega}{\omega - i\alpha}\right)\frac{\langle [O(\mathbf{r}), P]\rangle_\omega}{\omega} \\ &\quad + 3\epsilon_0 \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{\langle [O(\mathbf{r}), P]\rangle_\omega}{\omega} \\ &\quad + i\epsilon_0 \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t}\left(1 - \frac{\omega}{\omega - i\alpha}\right)\langle [O(\mathbf{r}), \Gamma]\rangle_\omega . \end{aligned} \tag{26}$$

In the above we have introduced the definition P for the spatially integrated pressure operator ($= \frac{1}{3}\operatorname{Tr} T_{ij}$). Also,

$$\Gamma(t) \equiv \sum_l S_{ll}(t).$$

It is of interest to apply (26) to the case where $O(\mathbf{r}t) = \rho(\mathbf{r}t)$, since this simple example will provide insight into the formalism to be employed in the later discussion of a more complex situation. In this respect it is imperative that we outline some general considerations on the properties of the spectral functions appearing in (26).

Suppose for the moment that our system exhibits space-translational invariance. Then, as was shown in detail in (PG), we would have

$$\int d^3\mathbf{r}' \frac{\langle [\rho(\mathbf{r}), P(\mathbf{r}')]\rangle_\omega}{\omega} = 2\pi\delta(\omega)\langle\rho\rangle_0 . \tag{27}$$

That the sole contribution to (27) in the space-translational case is of the form $\delta(\omega)$ follows from the fact that $\rho(\mathbf{r}t)$ satisfies a differential conservation law.

We now destroy the space-translational invariance of the system, either by introducing a "box potential" into the Hamiltonian or by imposing appropriate boundary conditions on the field operators. One can then ask the question whether or not there exists a statement analogous to (27) for the finite-volume

system. Indeed, the appropriate generalization of (27) is

$$\int_V d^3\mathbf{r}' \frac{\langle[\rho(\mathbf{r}), P(\mathbf{r}')]\rangle_\omega}{\omega} = 2\pi\delta(\omega)C_1(\mathbf{r}) + C_2(\mathbf{r}\omega). \quad (28)$$

Here, $C_2(\mathbf{r}\omega)$ is a smooth function of ω for $\omega \sim 0$ and

$$\lim_{V\to\infty} C_1(\mathbf{r}) = \langle\rho\rangle_0,$$
$$\lim_{V\to\infty} C_2(\mathbf{r}\omega) = 0. \quad (28')$$

The essential point to note here is that the spectral function (27) or (28) exhibits an $\omega\delta(\omega)$ singularity, irrespective of whether the system is constrained to finite volume or not. More generally, one may make the following statement concerning a spectral function of the form $\langle[A(\mathbf{r}), B(\mathbf{r}')]\rangle_\omega$, where A and B are two intensive operators possessing the same parity under time reversal (this last implies that the spectral function is an odd function of ω): If the $k = 0$ component of the translationally invariant $\langle[A(\mathbf{r}), B(\mathbf{r}')]\rangle_\omega$ exhibits a contribution of the form $\omega\delta(\omega)$, then a similar contribution will exist in the finite-volume case. The crux of the argument supporting this assertion may be put forth as follows: Suppose the spatial dependence of the spectral function $\langle[A(\mathbf{r}), B(\mathbf{r}')]\rangle_\omega$ is expanded in a multiple Fourier sum in wave-vector space with respect to the complete set of eigenfunctions

$$\phi_k(r) \equiv (8/V)^{\frac{1}{2}} \sin k_1x \sin k_2y \sin k_3y,$$
$$k_i = \pi n_i/L, \quad n_i = 0, \pm1, \pm2, \cdots,$$

which vanish at the boundaries of the cubical volume (10). It is found that the $\omega\delta(\omega)$ contribution arises from a sum over the subset of Fourier components with $\mathbf{k}_1 = \mathbf{k}_2$, $\mathbf{K}_1 = \mathbf{K}_2$, where $(\mathbf{k}_1, \mathbf{k}_2)$ and $(\mathbf{K}_1, \mathbf{K}_2)$ are the relative and center-of-mass wave vectors of two pairs of particles, respectively. A spatial integration over $\mathbf{r}'$, followed by the large-volume limit, yields the same $\omega\delta(\omega)$ contribution as we would obtain if we allowed the volume to become infinite first and then integrated over all space (thus yielding the $k = 0$ component of the space-translational case). In conclusion, we can assert that in all situations to which (26) has been applied, in order to carry through the calculations in a consistent manner, one must assume that the presence of an $\omega\delta(\omega)$ singularity in the spectral function $\langle[A(\mathbf{r}), B(\mathbf{r}')]\rangle_\omega$ is independent of any volume constraint imposed on the system.

The above behavior of $\langle[A(\mathbf{r}), B(\mathbf{r}')]\rangle_\omega$ as a function of ω is in marked contrast to the behavior of the other spectral function of interest $\langle[\rho(\mathbf{r}), \Gamma]\rangle_\omega$. It is a property of the structure of the operator S that this function is well behaved at $\omega = 0$ for the case of the system constrained to a finite volume. Since $\langle[\rho(\mathbf{r}), \Gamma]\rangle_\omega$ is an even function of ω, one might anticipate a $\delta(\omega)$ singularity. However, the subset of Fourier components such that $\mathbf{k}_1 = \mathbf{k}_2$, $\mathbf{K}_1 = \mathbf{K}_2$ is zero for this function, and hence the $\delta(\omega)$ contribution, does not appear. We can also make this argument from a somewhat different point of view. It follows from (A2) that

$$\langle[\rho(\mathbf{r}), \Gamma]\rangle_\omega = i\omega\langle[\rho(\mathbf{r}), Q]\rangle_\omega, \quad Q \equiv \sum_l Q_{ll}.$$

From general time-reversal arguments, we know that $\langle[\rho(\mathbf{r}), Q]\rangle_\omega$ is odd in ω. If, now, for finite volume we can show that this function is well behaved for $\omega \sim 0$, then it will follow that

$$\langle[\rho(\mathbf{r}), \Gamma]\rangle_\omega \sim O(\omega^2), \quad \omega \to 0.$$

To support the above assertion we consider the following easily derived sum rule:

$$\frac{1}{V}\int_V d^3\mathbf{r}\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{\langle[\rho(\mathbf{r}), Q]\rangle_\omega}{\omega} = \frac{1}{V}\left[\frac{\partial}{\partial\mu}\langle Q\rangle_0\right]_{\beta V}.$$

Clearly, for finite V the right-hand side of the above expression is well defined. This, in turn, implies that $\langle[\rho(\mathbf{r}), Q]\rangle_\omega$ is well defined at $\omega = 0$.

The above discussion is sufficient to provide us with all the information we need in order to evaluate the linear response (26) with $O(\mathbf{r}t) = \rho(\mathbf{r}t)$. Substituting (28) into (26) and allowing $\alpha \to 0^+$ yields

$$\delta\langle\rho(\mathbf{r})\rangle(t > 0) = -i\epsilon_0\langle[\rho(\mathbf{r}t), \Gamma(t)]\rangle_0 + 3\epsilon_0\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} C_2(\mathbf{r}\omega). \quad (29)$$

We easily evaluate the equal-time commutator in (29):

$$i\langle[\rho(\mathbf{r}t), \Gamma]\rangle_0 = \nabla \cdot [\mathbf{r}\langle\rho(\mathbf{r})\rangle_0].$$

The subsequent large-volume limit then yields

$$\delta\langle\rho(\mathbf{r})\rangle(t > 0) = -3\epsilon_0\langle\rho\rangle_0 = -\frac{\delta V}{V_0}\langle\rho\rangle_0, \quad (30)$$

as a consequence of (28'). This result is equivalent to the thermodynamic statement

$$\left(\frac{\delta\rho}{\delta V}\right)_{SN} = -\frac{1}{V}\rho.$$

At first glance it appears that only the smooth part of the spectral function (28) contributes to the response (29). However, this is misleading, since $C_1(\mathbf{r})$, $C_2(\mathbf{r}\omega)$, and $\langle[\rho(\mathbf{r}t), \Gamma(t)]\rangle_0$ are connected by the exact sum

rule [see Eq. (2.21) of (PG)]:

$$C_1(\mathbf{r}) + \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} C_2(\mathbf{r}\omega) = \tfrac{1}{3} i\langle[\rho(\mathbf{r}t), \Gamma(t)]\rangle_0 + V\left[\frac{\partial}{\partial V}\langle\rho(\mathbf{r})\rangle_0\right]_{\beta\mu}.$$

Thus, (29) may be rewritten as

$$\delta\langle\rho(\mathbf{r})\rangle(t > 0) = -3\epsilon_0 C_1(\mathbf{r}) - 3\epsilon_0 V\left[\frac{\partial}{\partial V}\langle\rho(\mathbf{r})\rangle_0\right]_{\beta\mu} \tag{31}$$

in terms of the $\delta(\omega)$ contribution $C_1(\mathbf{r})$. In the large-volume limit

$$C_1(\mathbf{r}) \to \langle\rho\rangle_0,$$
$$V\left[\frac{\partial}{\partial V}\langle\rho(\mathbf{r})\rangle_0\right]_{\beta\mu} \to 0,$$

and we again recapture the result (30).

One can carry out a calculation completely analogous to the above for the energy-density operator $\mathcal{E}(\mathbf{r})$. The statement corresponding to (29) in this case would read

$$\delta\langle\mathcal{E}(\mathbf{r})\rangle(t > 0) = -i\epsilon_0\langle[\mathcal{E}(\mathbf{r}t), \Gamma(t)]\rangle_0 + O\left(\frac{1}{V}\right).$$

Evaluating the equal-time commutator and employing the familiar virial-theorem expression for the equilibrium pressure [see Eq. (2.12) of (PG)] yields

$$\delta\langle\mathcal{E}(\mathbf{r})\rangle(t > 0) = -3\epsilon_0[\langle\mathcal{E}\rangle_0 + \langle P\rangle_0],$$

which is equivalent to the thermodynamic result

$$\left[\frac{\delta}{\delta V}\frac{E}{V}\right]_{SN} = -\frac{1}{V}\left[\frac{E}{V} + P\right].$$

In conclusion to this preliminary discussion, one should note an additional interesting point with respect to the linear-response analysis of $\rho(\mathbf{r}t)$ and $\mathcal{E}(\mathbf{r}t)$. *If* we had taken the large-volume limit $V \to \infty$ *before* taking the adiabatic limit $\alpha \to 0^+$, then it is easy to show that we would have obtained *zero* response in both cases, corresponding to the thermodynamic results

$$\left(\frac{\delta\rho}{\partial V}\right)_{\beta\mu} = 0,$$
$$\left(\frac{\partial\mathcal{E}}{\partial V}\right)_{\beta\mu} = 0.$$

The preceding introductory considerations now allow us to proceed to the primary goal of this section; i.e., to examine the linear response of the stress-tensor operator under the influence of a time-dependent change in the volume of the system. In so doing, we will be able to make a direct comparison with the phenomenological considerations of Sec. II and thus obtain correlation-function expressions for the parameters K_0, G_0, ζ, and η. Instead of the choice (24) for $\epsilon_l(t)$, we now consider a somewhat more general situation in which the $\epsilon_l(t)$ have the form

$$\epsilon_l(t) = \epsilon_{0l}\exp(\alpha t - i\omega t), \quad -\infty \leq t \leq 0.$$

If we agree to look at the system at time $t = 0$, we will have that

$$\epsilon_l(0) = \epsilon_{0l},$$
$$\dot{\epsilon}_l(0) = -i\omega\epsilon_{0l}.$$

The desired linear-response statement can now be obtained by a slight modificiation of (26): we set $t = 0$, make the replacement $\alpha \to \alpha - i\omega$, and take account of the fact that the $\epsilon_l(t)$ are no longer independent of l. Then, with $O(\mathbf{r}t) = T_{ij}(\mathbf{r}t)$, we have

$$\begin{aligned}\delta\langle T_{ij}(\mathbf{r})\rangle(\omega) = \sum_l \epsilon_{0l}\Bigg\{&\int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\frac{\langle[T_{ij}(\mathbf{r}), T_{ll}]\rangle_{\omega'}}{\omega'} - i\langle[T_{ij}(\mathbf{r}t), S_{ll}(t)]\rangle_0 \\ &-\int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\frac{\langle[T_{ij}(\mathbf{r}), T_{ll}]\rangle_{\omega'}}{\omega'}\left(1 - \frac{\omega'}{\omega' - \omega - i\alpha}\right) \\ &+ i\int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\langle[T_{ij}(\mathbf{r}), S_{ll}]\rangle_{\omega'}\left(1 - \frac{\omega'}{\omega' - \omega - i\alpha}\right)\Bigg\}.\end{aligned} \tag{32}$$

It is convenient at this point to eliminate the first two terms in (32) by employing the exact sum rule (A11). Thus,

$$\begin{aligned}\delta\langle T_{ij}(\mathbf{r})\rangle(\omega) = \sum_l \epsilon_{0l}\Bigg\{&\left[\frac{\partial}{\partial\epsilon_l}\langle T_{ij}(\mathbf{r})\rangle_\epsilon\right]_{\epsilon=0} \\ &-\int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\frac{\langle[T_{ij}(\mathbf{r}), T_{ll}]\rangle_{\omega'}}{\omega'}\left(1 - \frac{\omega'}{\omega' - \omega - i\alpha}\right) \\ &+ i\int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\langle[T_{ij}(\mathbf{r}), S_{ll}]\rangle_{\omega'}\left(1 - \frac{\omega'}{\omega' - \omega - i\alpha}\right)\Bigg\}.\end{aligned} \tag{33}$$

Before proceeding further with (33), we examine the properties of the spectral function $\langle[T_{ij}(\mathbf{r}), T_{ll'}]\rangle_\omega/\omega$. It was shown in (PG) that, for the spatially infinite system, this function has both a $\delta(\omega)$ contribution and a contribution regular at $\omega = 0$, i.e.,

$$\frac{\langle[T_{ij}(\mathbf{r}), T_{lm}]\rangle_\omega}{\omega} = 2\pi\delta(\omega)b_{ijlm} + B_{ijlm}(\omega). \tag{34}$$

$B_{ijlm}(\omega)$, in general, is nonzero for the translationally invariant system, since $T_{ij}(\mathbf{r}t)$ is a nonconserving

operator. Now, within the spirit of the discussion following (26), we generalize (34) to read

$$\frac{\langle [T_{ij}(\mathbf{r}), T_{lm}]\rangle_\omega}{\omega} = 2\pi\delta(\omega)b_{ijlm}(\mathbf{r}) + B_{ijlm}(\mathbf{r}\omega), \quad (35)$$

for the finite-volume system. In the large-volume limit, of course,

$$b_{ijlm}(\mathbf{r}) \to b_{ijlm},$$
$$B_{ijlm}(\mathbf{r}\omega) \to B_{ijlm}(\omega). \quad (36)$$

The previous arguments concerning the spectral function $\langle [\rho(\mathbf{r}), \Gamma]\rangle_\omega$ hinge on the structure of the operators S_{ll} and hence remain essentially unchanged for $\langle [T_{ij}(\mathbf{r}), S_{ll}]\rangle_\omega$. Thus, we conclude that

$$\langle [T_{ij}(\mathbf{r}), S_{ll}]\rangle_\omega \sim O(\omega^2), \quad \omega \to 0,$$

for a system constrained to a finite volume. Keeping the property (36) in mind, we can now substitute (35) into (33) and take the limit $\alpha \to 0^+$. This yields

$$\delta\langle T_{ij}(\mathbf{r})\rangle(\omega) = \sum_l \epsilon_{0l}\left\{\left[\frac{\partial}{\partial\epsilon_l}\langle T_{ij}(\mathbf{r})\rangle_\epsilon\right]_{\epsilon=0} - b_{ijll}(\mathbf{r})\right\} - i\omega\sum_l \epsilon_{0l}\tfrac{1}{2}[B_{ijll}(\mathbf{r}_0)] + O(\omega^2). \quad (37)$$

In the large-volume limit,

$$\left[\frac{\partial}{\partial\epsilon_l}\langle T_{ij}(\mathbf{r})\rangle_\epsilon\right]_{\epsilon=0} \to \delta_{ij}V\left[\frac{\partial}{\partial V}\langle P\rangle_0\right]_{\beta\mu} = O\left(\frac{1}{V}\right).$$

Furthermore, symmetry considerations imply that we may write

$$b_{ijlm}(\mathbf{r}) \to \delta_{ij}\delta_{lm}b^{(1)} + (\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl})b^{(2)},$$
$$B_{ijlm}(\mathbf{r}\omega) \to \delta_{ij}\delta_{lm}B^{(1)}(\omega) + (\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl})B^{(2)}(\omega). \quad (38)$$

Hence, in the $V \to \infty$ limit, (37) becomes

$$\delta\langle T_{ij}(\mathbf{r})\rangle(\omega) = \delta_{ij}\left\{-2\epsilon_j b^{(2)} - \left(\sum_l \epsilon_l\right)b^{(1)} - 2\dot{\epsilon}_j(\tfrac{1}{2}B^{(2)}(0)) - \left(\sum_l \dot{\epsilon}_l\right)(\tfrac{1}{2}B^{(1)}(0))\right\}. \quad (39)$$

Comparison with the constitutive equations of Sec. II allows us to make the identifications

$$K_0 - \tfrac{2}{3}G_0 = b^{(1)},$$
$$G_0 = b^{(2)},$$
$$\zeta - \tfrac{2}{3}\eta = \tfrac{1}{2}B^{(1)}(0),$$
$$\eta = \tfrac{1}{2}B^{(2)}(0). \quad (40)$$

In order to exhibit more clearly the connection between $\langle [T_{ij}(\mathbf{r}), T_{lm}(\mathbf{r}')]\rangle_\omega$ and the macroscopic parameters ζ, η, K_0, and G_0, we recall (34) as well as (38). These relations allow us to write

$$\int d^3\mathbf{r}' \frac{\langle [P(\mathbf{r}), P(\mathbf{r}')]\rangle_\omega}{\omega} = 2\pi\delta(\omega)K_0 + (B^{(1)}(\omega) + \tfrac{2}{3}B^{(2)}(\omega)),$$

$$\int d^3\mathbf{r}' \frac{\langle [T_{lm}(\mathbf{r}), T_{lm}(\mathbf{r}')]\rangle_\omega}{\omega} = 2\pi\delta(\omega)G_0 + B^{(2)}(\omega), \quad l \neq m. \quad (41)$$

We see, then, that the weights of the $\delta(\omega)$ parts of the relevant correlation functions are directly related to the static parameters K_0 and G_0, whereas the zero-frequency components of the nonsingular parts of these same correlation functions provide us with the dynamic coefficients ζ and η.

IV. CONCLUSION

In the preceding sections we adopted a microscopic viewpoint in order to examine in detail the relaxation processes which occur in a many-particle system subjected to a time-dependent change in volume. Basic to this discussion was the realization that certain frequency-dependent spectral functions possessed singular contributions of the form $\delta(\omega)$. We approached the problem within the framework of a linear response analysis, employing a time-dependent dynamical perturbation which altered the shape and size of the containing volume of the system. As a test of the method, we analyzed two cases in which the answer was known in advance; i.e., the linear response of the particle density $\rho(\mathbf{r}t)$ and the energy density $\mathcal{E}(\mathbf{r}t)$ to an adiabatic change in volume. We subsequently applied the method to the stress-tensor operator and succeeded in relating the macroscopic parameters η and ζ to spectral functions defined in terms of the stress-tensor operator. A consistent formulation of the problem was found to require that these spectral functions exhibit a singular behavior of the form $\delta(\omega)$; further, it was shown that the static parameters G_0 and K_0 were intimately related to this singular behavior. To be explicit, we found that

$$\lim_{\omega\to 0}\left[\int d^3\mathbf{r}' \frac{\langle [T_{ij}(\mathbf{r}), T_{lm}(\mathbf{r}')]\rangle_\omega}{\omega} - \delta_{ij}\delta_{lm}2\pi\delta(\omega)(K_0 - \tfrac{2}{3}G_0) - (\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl})2\pi\delta(\omega)G_0\right] = \delta_{ij}\delta_{lm}2(\zeta - \tfrac{2}{3}\eta) + (\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl})2\eta. \quad (42)$$

Now, it should be noted that (42) is identical to the usual Kubo–Mori formulas[5] for ζ and η *if* we ignore the singular contributions from the spectral function

$$\int d^3\mathbf{r}' \frac{\langle [T_{ij}(\mathbf{r}), T_{lm}(\mathbf{r}')]\rangle_\omega}{\omega}.$$

In not adequately taking into account the singular behavior of this function, the Kubo–Mori formulation does not provide us with an insight into the relation which exists between the static parameters K_0 and G_0 and the microscopic theory. Further, the original Kubo–Mori formulation required the use of a local equilibrium–density matrix, as well as coarse-grained time averages. Neither the local equilibrium assumption nor the coarse-graining procedure is needed in the present formulation.

In general, the function $\langle [T_{ij}(\mathbf{r}), T_{lm}(\mathbf{r}')]\rangle_\omega$ is extremely complicated. However, as we saw in (PG), in the infinite-volume limit the $k \to 0$ limit of this function can be related to a simpler spectral function constructed from the current–current commutator. Thus, within the framework of the present linear-response treatment, all of the information relating to visco-elastic processes is contained in the spectral function of the current–current commutator.

ACKNOWLEDGMENT

I would like to thank Professor Robert D. Puff for valuable discussions.

APPENDIX: SOME FORMAL OPERATOR RELATIONS

In this Appendix we consider the properties of a set of Hermitian operators S_{lm} $(l, m = 1, 2, 3)$ defined as follows:

$$S_{lm} = \frac{1}{2}\int d^3\mathbf{r}(r_l j_m(\mathbf{r}) + r_m j_l(\mathbf{r})), \tag{A1}$$

where $\mathbf{j}(\mathbf{r})$ is the momentum–density operator. The operators S_{lm} are related to another set of Hermitian operators Q_{lm},

$$Q_{lm} \equiv \frac{m}{2}\int d^3r\, r_l r_m \rho(\mathbf{r}), \tag{A2}$$

by the equality

$$S_{lm} = \dot{Q}_{lm}. \tag{A3}$$

If we now define the unitary transformation

$$R_{lm} \equiv \exp(-i\epsilon S_{lm}),$$

then

$$R_{lm}\psi(\mathbf{r})R_{lm}^\dagger = \psi(\mathbf{r}) - i\epsilon[S_{lm}, \psi(\mathbf{r})] + (1/2!)(-i\epsilon)^2[S_{lm}, [S_{lm}, \psi(\mathbf{r})]] + \cdots.$$

[5] R. Kubo, J. Phys. Soc. Japan **12**, 1203 (1956); H. Mori, *ibid.* **11**, 1029 (1956); H. Mori, Phys. Rev. **112**, 1829 (1958).

Noting that

$$-i[S_{lm}, \psi(\mathbf{r})] = \tfrac{1}{2}(\delta_{lm} + r_l\nabla_m + r_m\nabla_l)\psi(\mathbf{r}),$$

we immediately obtain

$$R_{lm}\psi(\mathbf{r})R_{lm}^\dagger = \exp\left[\tfrac{1}{2}\epsilon\left(\delta_{lm} + r_l\frac{\partial}{\partial r_m} + r_m\frac{\partial}{\partial r_l}\right)\right]\psi(\mathbf{r}). \tag{A4}$$

We consider first the case $l = m$. From (A4) we have that

$$R_{ll}\psi(\mathbf{r})R_{ll}^\dagger = \exp\left[\tfrac{1}{2}\epsilon + \epsilon r_l\frac{\partial}{\partial r_l}\right]\psi(\mathbf{r}). \tag{A5}$$

Now, for an arbitrary function $f(x)$ we note that

$$\frac{\partial}{\partial\epsilon}f(xe^\epsilon) = x\frac{\partial}{\partial x}f(xe^\epsilon). \tag{A6}$$

Formal expansion of $f(xe^\epsilon)$ in a Taylor series in ϵ and use of (A6) results in

$$\begin{aligned} f(xe^\epsilon) &= f(x) + \epsilon\left[\frac{\partial}{\partial\epsilon}f(xe^\epsilon)\right]_0 \\ &\quad + \frac{1}{2!}\epsilon^2\left[\frac{\partial^2}{\partial\epsilon^2}f(xe^\epsilon)\right]_0 + \cdots \\ &= f(x) + \epsilon\left(x\frac{\partial}{\partial x}\right)f(x) + \frac{1}{2!}\epsilon^2\left(x\frac{\partial}{\partial x}\right)^2 f(x) + \cdots \\ &= \exp\left(\epsilon x\frac{d}{dx}\right)f(x). \end{aligned}$$

Thus, from (A5) we obtain

$$\begin{aligned} R_{11}\psi(\mathbf{r})R_{11}^\dagger &= \exp(\tfrac{1}{2}\epsilon)\psi(xe^\epsilon, y, z), \\ R_{22}\psi(\mathbf{r})R_{22}^\dagger &= \exp(\tfrac{1}{2}\epsilon)\psi(x, ye^\epsilon, z), \\ R_{33}\psi(\mathbf{r})R_{33}^\dagger &= \exp(\tfrac{1}{2}\epsilon)\psi(x, y, ze^\epsilon). \end{aligned} \tag{A7}$$

(A7) tells us that R_{11} has the effect of generating a scale transformation on the x coordinate of the field operator $\psi(\mathbf{r})$, and R_{22} and R_{33} have similar effects on the y and z coordinates.

We can now analyze (A2) for the case $l \neq m$ in a manner similar to the above. Indeed, we find that

$$\begin{aligned} R_{12}\psi(\mathbf{r})R_{12}^\dagger &= \psi(U_1(x, y, \epsilon), U_2(x, y, \epsilon), z), \\ R_{23}\psi(\mathbf{r})R_{23}^\dagger &= \psi(x, U_2(z, y, \epsilon), U_1(z, y, \epsilon)), \\ R_{31}\psi(\mathbf{r})R_{31}^\dagger &= \psi(U_1(x, z, \epsilon), y, U_2(x, z, \epsilon)), \end{aligned} \tag{A8}$$

where

$$\begin{aligned} U_1(x, y, \epsilon) &= x\cosh(\epsilon/2) + y\sinh(\epsilon/2), \\ U_2(x, y, \epsilon) &= y\cosh(\epsilon/2) + x\sinh(\epsilon/2). \end{aligned}$$

Thus, R_{lm} generates a two-dimensional "rotation" of the coordinates r_l and r_m in a hyperbolic space which preserves the quantity $r_l^2 - r_m^2$.

In concluding this section, we derive a class of sum rules, a special example of which was employed in the text in obtaining Eq. (33). We denote an equilibrium average using the Hamiltonian H_ϵ [see Eq. (17)] as $\langle \cdots \rangle_\epsilon$. Thus, the average of some operator $A(\mathbf{r}, t)$ in this ensemble is given by

$$\langle A(\mathbf{r}, 0)\rangle_\epsilon = \frac{\mathrm{Tr}\,\{\exp[\beta(\mu N - H_\epsilon)]A(\mathbf{r}, 0)\}}{\mathrm{Tr}\,\{\exp[\beta(\mu N - H_\epsilon)]\}} = \frac{\mathrm{Tr}\,\{\exp[\beta(\mu N - \bar{H}_\epsilon)]R^\dagger A(\mathbf{r}, 0)R\}}{\mathrm{Tr}\,\{\exp[\beta(\mu N - \bar{H}_\epsilon)]\}}, \quad \text{(A9)}$$

where

$$R \equiv \prod_l R_{ll},$$

$$\bar{H}_\epsilon = R^\dagger H_\epsilon R.$$

A direct differentiation of (A9) with respect to ϵ_l yields

$$\left[\frac{\partial}{\partial \epsilon_l}\langle A(\mathbf{r}, 0)\rangle_\epsilon\right]_{\epsilon=0} = -i\langle[A(\mathbf{r}t), S_{ll}(t)]\rangle_0 + \int_0^\beta d\lambda \int d^3\mathbf{r}'[\langle A(\mathbf{r}, -i\lambda)T_{ll}(\mathbf{r}')\rangle_0 - \langle A(\mathbf{r})\rangle_0\langle T_{ll}(\mathbf{r}')\rangle_0]. \quad \text{(A10)}$$

If we introduce the spectral function $\langle[A(\mathbf{r}), T_{ll}(\mathbf{r}')]\rangle_\omega$, then one easily shows [see (PG), Eq. 2.36] that the last term in (A10) may be written as

$$\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{\langle[A(\mathbf{r}), T_{ll}(\mathbf{r}')]\rangle_\omega}{\omega}.$$

Thus, (A10) becomes

$$\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{\langle[A(\mathbf{r}), T_{ll}(\mathbf{r}')]\rangle_\omega}{\omega} = \left[\frac{\partial}{\partial \epsilon}\langle A(\mathbf{r}, 0)\rangle_\epsilon\right]_0 + i\langle[A(\mathbf{r}t), S_{ll}(t)]\rangle_0, \quad \text{(A11)}$$

which is the desired result.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Relativistic Dynamics of a Point Charge in a Magnetic-Monopole Field

THOMAS P. MITCHELL
University of California, Santa Barbara, California
AND
J. A. BURNS*
Laboratory for Theoretical Studies, NASA-Goddard Space Flight Center, Greenbelt, Maryland

(Received 14 March 1968)

This paper derives and interprets the constants of the charge's motion. The physical meaning of these constants and their use in discussing the over-all motion of the charge are presented.

I. INTRODUCTION

Since the magnetic-monopole field is of use as a workable approximation to actual fields, e.g., exterior stellar magnetic fields, any direct physical insight which can be gained into the field's effect on the motion of a charged particle is of interest. The detailed geometry of the particle's trajectory was established by Poincaré[1] and Ferraro.[2] The trajectory is, in fact, a geodesic on a circular cone whose vertex coincides with the monopole position.

Van Allen[3] has utilized the differential-geometric properties of geodesics on circular cones to construct a constant of the motion and he has related this constant to an adiabatic invariant. The analysis to follow establishes the constants of the motion, including that of Van Allen, in a systematic way without recourse to differential geometry. It is shown that the rather complicated geometric interpretation of the particle's motion need not be involved in any way in establishing the existence and the physical meaning of the constants. Of course, if one wished, one could derive the complete trajectory as well as the constants by using the method presented here.

II. HAMILTONIAN FORMULATION

We use the standard spherical-polar coordinates (r, θ, φ) with the origin at the monopole. The orthogonal directions are denoted by the unit vectors $\mathbf{i}_r$, $\mathbf{i}_\theta$, and $\mathbf{i}_\varphi$. The magnetic field of the monopole $\mathbf{B} = \mathbf{i}_r B_0 r_0^2/r^2$ is derivable from the vector potential $\mathbf{A} = -(\mathbf{i}_\varphi B_0 r_0^2 \cot\theta)/r$, with B_0 (gauss) being a constant

* Research Associate, National Academy of Science–National Research Council; now at Cornell University.
[1] H. Poincaré, Compt. Rend. **123**, 530 (1896).
[2] V. C. A. Ferraro, *Electromagnetic Theory* (Athlone Press, University of London, 1956).
[3] J. A. Van Allen, J. Geophys. Res. **70**, 1240 (1965).

In concluding this section, we derive a class of sum rules, a special example of which was employed in the text in obtaining Eq. (33). We denote an equilibrium average using the Hamiltonian H_ϵ [see Eq. (17)] as $\langle \cdots \rangle_\epsilon$. Thus, the average of some operator $A(\mathbf{r}, t)$ in this ensemble is given by

$$\langle A(\mathbf{r}, 0)\rangle_\epsilon = \frac{\mathrm{Tr}\,\{\exp[\beta(\mu N - H_\epsilon)]A(\mathbf{r}, 0)\}}{\mathrm{Tr}\,\{\exp[\beta(\mu N - H_\epsilon)]\}} = \frac{\mathrm{Tr}\,\{\exp[\beta(\mu N - \bar{H}_\epsilon)]R^\dagger A(\mathbf{r}, 0)R\}}{\mathrm{Tr}\,\{\exp[\beta(\mu N - \bar{H}_\epsilon)]\}}, \quad \text{(A9)}$$

where

$$R \equiv \prod_l R_{ll},$$

$$\bar{H}_\epsilon = R^\dagger H_\epsilon R.$$

A direct differentiation of (A9) with respect to ϵ_l yields

$$\left[\frac{\partial}{\partial \epsilon_l}\langle A(\mathbf{r}, 0)\rangle_\epsilon\right]_{\epsilon=0} = -i\langle[A(\mathbf{r}t), S_{ll}(t)]\rangle_0 + \int_0^\beta d\lambda \int d^3\mathbf{r}'[\langle A(\mathbf{r}, -i\lambda)T_{ll}(\mathbf{r}')\rangle_0 - \langle A(\mathbf{r})\rangle_0\langle T_{ll}(\mathbf{r}')\rangle_0]. \quad \text{(A10)}$$

If we introduce the spectral function $\langle[A(\mathbf{r}), T_{ll}(\mathbf{r}')]\rangle_\omega$, then one easily shows [see (PG), Eq. 2.36] that the last term in (A10) may be written as

$$\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{\langle[A(\mathbf{r}), T_{ll}(\mathbf{r}')]\rangle_\omega}{\omega}.$$

Thus, (A10) becomes

$$\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{\langle[A(\mathbf{r}), T_{ll}(\mathbf{r}')]\rangle_\omega}{\omega} = \left[\frac{\partial}{\partial \epsilon}\langle A(\mathbf{r}, 0)\rangle_\epsilon\right]_0 + i\langle[A(\mathbf{r}t), S_{ll}(t)]\rangle_0, \quad \text{(A11)}$$

which is the desired result.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Relativistic Dynamics of a Point Charge in a Magnetic-Monopole Field

THOMAS P. MITCHELL
University of California, Santa Barbara, California
AND
J. A. BURNS*
Laboratory for Theoretical Studies, NASA-Goddard Space Flight Center, Greenbelt, Maryland

(Received 14 March 1968)

This paper derives and interprets the constants of the charge's motion. The physical meaning of these constants and their use in discussing the over-all motion of the charge are presented.

I. INTRODUCTION

Since the magnetic-monopole field is of use as a workable approximation to actual fields, e.g., exterior stellar magnetic fields, any direct physical insight which can be gained into the field's effect on the motion of a charged particle is of interest. The detailed geometry of the particle's trajectory was established by Poincaré[1] and Ferraro.[2] The trajectory is, in fact, a geodesic on a circular cone whose vertex coincides with the monopole position.

Van Allen[3] has utilized the differential-geometric properties of geodesics on circular cones to construct a constant of the motion and he has related this constant to an adiabatic invariant. The analysis to follow establishes the constants of the motion, including that of Van Allen, in a systematic way without recourse to differential geometry. It is shown that the rather complicated geometric interpretation of the particle's motion need not be involved in any way in establishing the existence and the physical meaning of the constants. Of course, if one wished, one could derive the complete trajectory as well as the constants by using the method presented here.

II. HAMILTONIAN FORMULATION

We use the standard spherical-polar coordinates (r, θ, φ) with the origin at the monopole. The orthogonal directions are denoted by the unit vectors $\mathbf{i}_r$, $\mathbf{i}_\theta$, and $\mathbf{i}_\varphi$. The magnetic field of the monopole $\mathbf{B} = \mathbf{i}_r B_0 r_0^2/r^2$ is derivable from the vector potential $\mathbf{A} = -(\mathbf{i}_\varphi B_0 r_0^2 \cot\theta)/r$, with B_0 (gauss) being a constant

* Research Associate, National Academy of Science–National Research Council; now at Cornell University.
[1] H. Poincaré, Compt. Rend. **123**, 530 (1896).
[2] V. C. A. Ferraro, *Electromagnetic Theory* (Athlone Press, University of London, 1956).
[3] J. A. Van Allen, J. Geophys. Res. **70**, 1240 (1965).

and r_0 (cm) an arbitrary radial distance from the origin. As the Lagrangian of a particle with rest mass m_0, charge e(esu) in this field, we adopt the form

$$L = \tfrac{1}{2}m_0 c g_{\alpha\beta}u^\alpha u^\beta + (e/c)g_{\alpha\beta}u^\alpha A^\beta,$$

in which u^α, the 4-velocity, is

$$u^\alpha = \begin{pmatrix} (\gamma/c)\dot{r} \\ (\gamma/c)\dot{\theta} \\ (\gamma/c)\dot{\varphi} \\ i\gamma \end{pmatrix},$$

the 4-potential A^α is

$$A^\alpha = \begin{pmatrix} 0 \\ 0 \\ (-B_0 r_0^2/r^2)\cot\theta\csc\theta \\ 0 \end{pmatrix},$$

the metric tensor for flat space–time is

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & r^2 & 0 & 0 \\ 0 & 0 & r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$, v being the 3-velocity magnitude. One finds in the standard way that the 4-momentum is

$$p_\alpha = \begin{pmatrix} m_0\gamma\dot{r} \\ m_0\gamma r^2\dot{\theta} \\ m_0\gamma r^2\sin^2\theta\dot{\varphi} - (e/c)B_0 r_0^2\cos\theta \\ im_0\gamma c \end{pmatrix},$$

and hence that the Hamiltonian $H = g_{\alpha\beta}p^\alpha u^\beta - L$ is

$$H = \frac{1}{2m_0 c}\left[p_1^2 + \frac{1}{r^2}p_2^2 + \frac{1}{r^2\sin^2\theta}\left(p_3 + \frac{eB_0}{c}r_0^2\cos\theta\right)^2 + p_4^2\right].$$

It is easily seen that this Hamiltonian has the constant value $-\frac{1}{2}m_0 c$. Thus the corresponding Hamilton–Jacobi equation is

$$\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial\theta}\right)^2 + \frac{1}{r^2\sin^2\theta}\left[\frac{\partial S}{\partial\varphi} + \frac{eB_0}{c}r_0^2\cos\theta\right]^2 - \frac{1}{c^2}\left(\frac{\partial S}{\partial t}\right)^2 = -m_0^2 c^2,$$

which separates in the form

$$S(r, \theta, \varphi, t) = S_1(r) + S_2(\theta) + p_3\varphi + Et,$$

where p_3 and E are constants. Accordingly, we can assert that the following constants of the motion exist:

$$p_4 = im_0\gamma c = iE/c = \text{const}, \quad (1)$$

$$p_3 = m_0\gamma r^2\sin^2\theta\dot{\varphi} - (e/c)B_0 r_0^2\cos\theta = \text{const}, \quad (2)$$

and

$$\frac{1}{\sin^2\theta}\left[p_3 + \frac{e}{c}B_0 r_0^2\cos^2\theta\right]^2 + \left(\frac{dS_2}{d\theta}\right)^2 = \text{const}. \quad (3)$$

These three constants correspond to conservation of the relative energy, conservation of the component of angular momentum in the direction $\theta = 0$, and conservation of the total orbital angular momentum about the monopole, respectively.

These constants contain much information about the properties of the system.

(*a*) *Particle motion:* The condition $\theta = \theta_0$, a constant, satisfies all the above equations for the constants; the charged particle moves on the surface of a cone whose vertex is at the origin.

(*b*) *An adiabatic invariant:* A moving charge distribution has a magnetic moment $\mu = (e/2\gamma m_0 c)\mathbf{M}$, where $\mathbf{M}$ is the orbital angular momentum of the charges. Hence (3) gives μ, normally only an adiabatic invariant, as a rigorous constant of the motion. This is related to Van Allen's constant,[3] which can be found explicitly by combining (1) and (3).

(*c*) *Reflection point:* The introduction of a radial force, due to an electric or a gravitational (for nonrelativistic motion) field, does not change the angular-momentum constants, since such a force has zero moment about the origin. However, the potential energy due to this added field should be included in (1). Doing this and using (3), one finds that a charged particle is reflected by the combined field at

$$r = [(g/2E_0)^2 + M_0^2/(2\gamma m_0 E_0)]^{\frac{1}{2}} - g/2E_0,$$

where E_0 and M_0 are the particle energy and orbital angular momentum initially and g is the numerator of the particular radial-force law.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Transmission and Reflection of Electromagnetic Waves at the Boundary of a Relativistic Collisionless Plasma

ERCÜMENT ÖZIZMIR*
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey

(Received 21 November 1967)

The exact solution of the transmission and reflection problem for transverse electromagnetic waves incident on a bounded plasma has been discussed to some extent by several authors. Shure considered the special cases of perpendicular incidence on nonrelativistic half-space and slab plasmas and made use of van Kampen–Case modes to construct the solution. For both half-space and slab plasmas, we generalize his results to (i) arbitrary temperatures (relativistic) and (ii) arbitrary angles of incidence. For simplicity, we consider the special case where the incident electric field is perpendicular to the plane of incidence and assume that particles are reflected specularly at the interface. We proceed somewhat differently from Shure, and use a Laplace transformation in obtaining our solution. We also show that present solutions can be expressed as a superposition of van Kampen–Case modes appropriate to a *relativistic plasma*.

I. INTRODUCTION

The exact solution of the transmission–reflection problem for transverse electromagnetic waves incident at the interface between a vacuum and a half-space plasma has been discussed to some extent by several authors.[1–3] Shure[1] and Felderhof[2] considered the special case of perpendicular incidence on a nonrelativistic plasma and made use of van Kampen[4]–Case[5] modes to construct the solution. Weston[3] recently extended their results to the case of oblique incidence, again for a nonrelativistic plasma which employs the Maxwell–Boltzmann distribution F_{MB} for the unperturbed state.

There are some undesirable features associated with the use of F_{MB} for the unperturbed state in this particular problem and they lead to nonphysical results. In particular, since F_{MB} is nonvanishing for all particle velocities, one finds that transverse waves propagating in the plasma are Landau damped.[1] The use of the appropriate relativistic Vlasov equation together with the Maxwell–Boltzmann (Jüttner[6]) distribution for the unperturbed distribution function, as was pointed out previously[7] in the context of the initial-value problem, shows that transverse waves propagating in the plasma are not Landau damped, since the phase velocity of these waves is always greater than the velocity of light. Felderhof[2] employed cutoff distributions to eliminate this nonphysical damping. It is also known that the analytical properties of the dispersion function are quite different when F_{MBJ} is used for the unperturbed state. Therefore, the point of a relativistic treatment is not only to extend the validity of the previous results to high-temperature plasmas, but also to eliminate, at the outset, the nonphysical aspects in the formulation of the problem.

In this paper we obtain a rigorous solution of the problem for a relativistic plasma when no external fields are present. We consider the cases of both half-space and slab plasmas with arbitrary angles of incidence, but consider, for simplicity, the special case where the incident electric field is perpendicular to the plane of incidence. We assume that particles are reflected specularly at the interface. In the case of a half-space plasma, we show that two basic requirements (corresponding to the boundedness of solutions at infinity and to the causality condition) on the analytical properties of the Laplace transformed fields suffice to determine *uniquely* (up to an arbitrary multiplicative constant which is determined by the intensity of the incident waves) the stationary solution of the coupled Maxwell–Vlasov equations. We wish to note that the electromagnetic properties of a relativistic plasma are also discussed in a series of papers by Silin,[8] who has given only a brief outline of his derivation. Our solution for a half-space plasma is in agreement with his result. In addition, we consider the case of a slab plasma. We have proceeded differently in obtaining our solutions using a Laplace transformation which appears to have certain advantages. For example, after a simple observation, we can express our solution as a superposition of van Kampen–Case stationary modes appropriate to a relativistic plasma

* Present address: Richmond College of The City University of New York, Staten Island, New York, N.Y.

[1] F. C. Shure, Ph.D. thesis, University of Michigan, 1962.

[2] B. U. Felderhof, Physica **29**, 662 (1963).

[3] V. H. Weston, Phys. Fluids **10**, 632 (1967).

[4] N. G. van Kampen, Physica **21**, 949 (1955).

[5] K. M. Case, Ann. Phys. (N.Y.) **7**, 349 (1959).

[6] F. Jüttner, Ann. Physik **34**, 856 (1911).

[7] K. Imre, Phys. Fluids **5**, 549 (1962). This paper also contains references to the related earlier work.

[8] V. P. Silin, Zh. Eksp. Teor. Fiz. **41**, 159 (1961) [Sov. Phys.—JETP **14**, 115 (1962)]. This paper also contains references to his earlier work.

(cf. Sec. V). (The use of these modes for a relativistic plasma does not seem to have been discussed in the literature.) Also, the case of a slab plasma can be treated by a straightforward extension of the present technique for a half-space plasma.

In Sec. II, we state the basic equations adopted for the description of the system. Solution for the fields is obtained in Sec. III under the above-mentioned assumption of certain analyticity requirements on the transformed field functions. These requirements impose certain integral conditions on the perturbed-particle distribution function at the interface. By the use of these conditions the fields can be calculated explicitly. Given these fields, we determine in Sec. IV the perturbed-particle distribution function uniquely and show that these conditions are actually satisfied, which proves the consistency of the solutions for the fields and the perturbed-particle distribution function. In Sec. V, we indicate the connection between the present result and the expression of solutions in terms of van Kampen–Case stationary modes. Explicit expressions are given in Sec. VI for the reflection, transmission, and absorption coefficients. We have also derived in this section an expression for the time-averaged heat flow as a function of position. Section VII is devoted to the case of a slab plasma. Finally, some relevant properties of the dispersion function are discussed in the Appendix.

II. BASIC EQUATIONS

The starting point of the present analysis will be the relativistic (linearized) Vlasov equation coupled with Maxwell's equations:

$$\gamma \frac{\partial f}{\partial t} + \mathbf{u} \cdot \frac{\partial f}{\partial \mathbf{x}} + \frac{ne}{m} \gamma \mathbf{E} \cdot \frac{\partial F_0}{\partial \mathbf{u}} = 0, \tag{1}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}; \quad \nabla \cdot \mathbf{B} = 0, \tag{2}$$

$$\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{j}; \quad \nabla \cdot \mathbf{E} = 4\pi\rho,$$

where n is the number density in the unperturbed state, $\mathbf{u} \equiv \gamma \mathbf{v}$ where $\mathbf{v}$ is the particle velocity,

$$\gamma \equiv \left(1 + \frac{u^2}{c^2}\right)^{\frac{1}{2}} = 1 \Big/ \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}},$$

and charge and current densities are given by

$$\mathbf{j} \equiv e \int \frac{d^3u}{\gamma} \mathbf{u} f(\mathbf{x}, \mathbf{u}, t),$$

$$\rho \equiv e \int d^3u f(\mathbf{x}, \mathbf{u}, t). \tag{3}$$

The equilibrium distribution function is taken as the relativistic Maxwell–Boltzmann (Jüttner) distribution, namely,

$$F_0 = \frac{1}{4\pi c^3} \frac{\beta}{K_2(\beta)} e^{-\beta\gamma}, \tag{4}$$

where $\beta = mc^2/kT$ and $K_2(\beta)$ is the modified Bessel function of the second kind.

We shall assume that the plasma fills the half-space $z > 0$. To obtain the stationary solutions of Eqs. (1) and (2) corresponding to a plane electromagnetic wave incident on the plasma, we look for solutions

$$\propto e^{-i\omega t} e^{i(k_x x + k_y y)} \tag{5}$$

and take the Laplace transformation with respect to z variable. Here, ω is a real, positive quantity. The resulting equations for transformed functions are

$$-i(\gamma\omega - \mathbf{k} \cdot \mathbf{u}) f(\mathbf{k}, \mathbf{u}, \omega) = u_z f(0) - (ne/m) \times \gamma \mathbf{E}(\mathbf{k}, \omega) \cdot (\partial F_0/\partial \mathbf{u}), \tag{6}$$

$$i\mathbf{k} \times \mathbf{E} = (i\omega/c)\mathbf{B} - \hat{\mathbf{x}} E_y(0) + \hat{\mathbf{y}} E_x(0); \quad i\mathbf{k} \cdot \mathbf{B} = B_z(0),$$

$$i\mathbf{k} \times \mathbf{B} = (4\pi/c)\mathbf{j} - (i\omega/c)\mathbf{E} - \hat{\mathbf{x}} B_y(0) + \hat{\mathbf{y}} B_x(0); \quad i\mathbf{k} \cdot \mathbf{E} = 4\pi\rho + E_z(0). \tag{7}$$

In these equations, $ik_z \equiv p$ is the Laplace transform variable. The continuity equation is obtained by dividing Eq. (6) by γ and integrating over $\mathbf{u}$ as

$$\omega\rho - \mathbf{k} \cdot \mathbf{j} = ij_z(0). \tag{8}$$

Without loss of generality, we can choose $k_y = 0$. For a given $\mathbf{k}$, one can construct the unit vectors

$$\hat{\mathbf{e}}_1(\mathbf{k}) \equiv \hat{\mathbf{y}} \times \hat{\mathbf{k}} = \hat{\mathbf{x}} \frac{k_z}{k} - \hat{\mathbf{z}} \frac{k_x}{k},$$

$$\hat{\mathbf{e}}_2(\mathbf{k}) \equiv \hat{\mathbf{y}}, \tag{9}$$

$$\hat{\mathbf{e}}_3(\mathbf{k}) \equiv \frac{\hat{\mathbf{k}}}{k} = \hat{\mathbf{x}} \frac{k_x}{k} + \hat{\mathbf{z}} \frac{k_z}{k},$$

where $k \equiv (k_x^2 + k_z^2)^{\frac{1}{2}}$. Then, any vector function $\mathbf{A}(\mathbf{k})$ has the unique decomposition

$$\mathbf{A}(\mathbf{k}) = \sum_{\alpha=1}^{3} A_\alpha(\mathbf{k}) \hat{\mathbf{e}}_\alpha(\mathbf{k}), \tag{10}$$

where

$$A_\alpha(\mathbf{k}) = \mathbf{A}(\mathbf{k}) \cdot \hat{\mathbf{e}}_\alpha(\mathbf{k}).$$

Using the decomposition of **E**, **B**, **j** in Eqs. (6) and (7), one obtains

$$\begin{aligned} -ikE_2 &= \frac{i\omega}{c} B_1 - \frac{k_z}{k} E_y(0), \\ ikE_1 &= \frac{i\omega}{c} B_2 + E_x(0), \\ ikB_3 &= B_z(0), \\ 0 &= \frac{i\omega}{c} B_3 - \frac{k_x}{k} E_y(0), \end{aligned} \tag{11}$$

$$\begin{aligned} -ikB_2 &= \frac{4\pi}{c} j_1 - \frac{i\omega}{c} E_1 - \frac{k_z}{k} B_y(0), \\ ikB_1 &= \frac{4\pi}{c} j_2 - \frac{i\omega}{c} E_2 - B_x(0), \\ ikE_3 &= 4\pi\rho + E_z(0), \\ 0 &= \frac{4\pi}{c} j_3 - \frac{i\omega}{c} E_3 - \frac{k_x}{k} B_y(0). \end{aligned} \tag{12}$$

The continuity equation may be written as

$$\omega\rho - kj_3 = ij_z(0). \tag{13}$$

In the following we restrict ourselves to the case in which the incident wave has an electric field polarized in the y direction (i.e., the electric field is perpendicular to the plane of incidence). In this case the solution leads to *purely* transverse waves, as will be seen in the following. The other case with incident electric field in the plane of incidence is somewhat more complicated and leads to both transverse and longitudinal waves within the plasma due to the existence of the boundary. However, it is felt that a full exposition of the simple case would be useful since it has not been discussed fully in the literature for a relativistic plasma. The latter case will be discussed in a subsequent report.

When the incident electric field is in the y direction we need only to consider the following equations in addition to the Vlasov equation:

$$\begin{aligned} -ikE_2 &= i\frac{\omega}{c} B_1 - \frac{k_z}{k} E_y(0), \\ ikB_1 &= \frac{4\pi}{c} j_2 - \frac{i\omega}{c} E_2 + B_x(0), \\ \omega B_z &= ck_x E_y. \end{aligned} \tag{14}$$

III. SOLUTION FOR THE ELECTRIC FIELD

From Eqs. (6) and (3) we obtain

$$j_2 \equiv j_y = ie\int \frac{d^3u}{\gamma} \frac{u_z u_y f(0)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}} - \frac{ine^2}{m} E_y \int d^3u \frac{u_y(\partial F_0/\partial u_y)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}}. \tag{15}$$

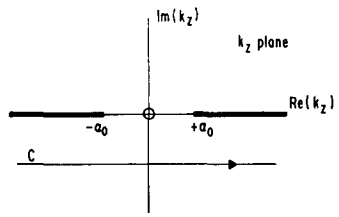


FIG. 1. The branch cuts of the dispersion function $\Lambda(k_z)$ in the complex k_z plane [$\alpha_0 \equiv (\omega/c)\cos\theta_0$].

From Eq. (15) and the first two of Eqs. (14), one obtains

$$\Lambda(k_z) iE_y(k_z) = R_2(k_z), \tag{16}$$

where

$$\begin{aligned} R_2(k_z) &\equiv -4\pi i\omega e \int \frac{d^3u}{\gamma} \frac{u_y u_z f(0)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}} \\ &\quad + [c^2 k_z E_y(0) - \omega c B_x(0)], \\ \Lambda(k_z) &\equiv c^2k^2 - \omega^2 + \omega_p^2 A(k_z), \\ A(k_z) &\equiv -\omega \int d^3u \frac{u_y(\partial F_0/\partial u_y)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}}, \\ \omega_p^2 &\equiv \frac{4\pi ne^2}{m}. \end{aligned}$$

The electric field as a function of position can then be written as

$$E_y(z) = \frac{1}{2\pi i} \int_C dk_z e^{ik_z z} iE_y(k_z), \tag{17}$$

where the contour C lies, in the complex k_z plane, parallel to the real axis and below all singularities of $E_y(k_z)$, as shown in Fig. 1.

It is shown in the Appendix that $\Lambda(k_z) = 0$ has either a pair of real or a pair of imaginary roots, depending on the values of β and ω/ω_p. Call them $\pm\kappa_0$. Also, in the complex k_z plane $\Lambda(k_z)$ is analytic everywhere, except along the cut which lies along the part of the real axis given by $|k_z| > \alpha_0$, where

$$\alpha_0 \equiv \frac{\omega}{c}(1 - \sin^2\theta_0)^{\frac{1}{2}} = \frac{\omega}{c}\cos\theta_0,$$

and $\sin\theta_0 = ck_x/\omega$, where θ_0 is the angle of incidence. Real roots, if any, are always in the open interval $(-\alpha_0, \alpha_0)$.

We now impose upon the solution the following physical requirements:

(1) $E_y(z)$ is bounded at $z = +\infty$;

(2) $E_y(z)$, $\mathbf{B}(z)$, and $f(z, \mathbf{u})$ for $z > 0$ consist only of waves travelling in the $+z$ direction.

In order to satisfy these requirements, the singularities of $E_y(k_z)$ must be confined to the region Re $(k_z) \geq 0$, Im $(k_z) \geq 0$ in the k_z plane. We must, therefore, have

$$R_2(-\kappa_0) = 0, \tag{18}$$

$$\frac{R_2^+(-k_z)}{\Lambda^+(-k_z)} = \frac{R_2^-(-k_z)}{\Lambda^-(-k_z)} \qquad (k_z > \alpha_0), \tag{19}$$

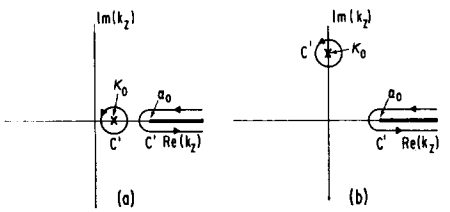


FIG. 2. The paths of integration C' in the complex k_z plane after deformation of the contour C.

where $\Lambda^{\pm}(-k_z) \equiv \lim_{\epsilon \to 0} \Lambda(-k_z \pm i\epsilon)$, and similarly for R_2. Note that (18) and (19) constitute certain integral conditions between the possible values of $f(0, u)$, $E_y(0)$, $B_x(0)$ which must be verified *a posteriori*.

Finally, we assume that particles are reflected specularly at the interface so that

$$f(0, u_x, u_y, u_z) = f(0, u_x, u_y, -u_z). \tag{20}$$

Recalling the definition of $R_2(k_z)$ and making use of Eqs. (18), (19), and (20), one readily obtains

$$R_2(+\kappa_0) = -2\omega c B_x(0), \tag{21}$$

$$iE_y^+(k_z) - iE_y^-(k_z) = -2\omega c B_x(0)\left(\frac{1}{\Lambda^+(k_z)} - \frac{1}{\Lambda^-(k_z)}\right). \tag{22}$$

In view of conditions (18) and (19) and assuming that $R_2(k_z)$ has no other singularities in the cut plane (which will be verified later) we may deform the contour C in Fig. 1 as shown in Fig. 2 to obtain, for $z > 0$,

$$E_y(z) = -2\omega c B_x(0)\left\{\frac{1}{\Lambda'(\kappa_0)} e^{i\kappa 0^z} + \frac{1}{2\pi i}\int_{a_0}^{\infty} dk_z e^{ik_z z}\left[\frac{1}{\Lambda^-(k_z)} - \frac{1}{\Lambda^+(k_z)}\right]\right\}, \tag{23}$$

where prime indicates the derivative of $\Lambda(k_z)$ with respect to k_z.

This completes the determination of the electric field as a function of z within the plasma. To be consistent, however, we must show that, for $E_y(z)$ given by (23), $f(0)$ does in fact satisfy conditions (18), (19), and (20).

Before proceeding to the determination of $f(z)$, let us note that Eq. (23) may be written more compactly as

$$E_y(z) = -2\omega c B_x(0)\frac{1}{2\pi i}\int_{C'} dk_z e^{ik_z z}\frac{1}{\Lambda(k_z)}, \tag{24}$$

where C' is the union of the contours shown in Fig. (2a) or (2b).

IV. DETERMINATION OF THE PERTURBED-PARTICLE DISTRIBUTION FUNCTION; PROOF OF CONSISTENCY

It is easily verified that, for $E_y(z)$ given by Eq. (24), the Vlasov equation

$$-i(\gamma\omega - k_x u_x) f(z) + u_z \frac{\partial f}{\partial z} = -\frac{ne}{m}\gamma\frac{\partial F_0}{\partial u_y} E_y(z) \tag{25}$$

is satisfied by

$$f(z) = \frac{nei}{m} 2\omega c B_x(0)\gamma \frac{\partial F_0}{\partial u_y}\frac{1}{2\pi i} \times \int_{C'} dk_z e^{ik_z z}\frac{1}{\Lambda(k_z)(\gamma\omega - \mathbf{k}\cdot\mathbf{u})}. \tag{26}$$

We now show that this solution is the unique solution of the Vlasov equation. For this purpose we consider the general solution f_H of the homogeneous Vlasov equation [$E_y(z) \equiv 0$] which may be added to (26), i.e.,

$$f_H = C(\mathbf{u}) \exp\left[\frac{i(\gamma\omega - k_x u_x)}{u_z} z\right] e^{-i\omega t} e^{-ik_x x}. \tag{27}$$

Since $f(z)$ given by (26) satisfies the condition of specular reflection, f_H also must satisfy the same condition. Therefore, we have

$$C(u_x, u_y, u_z) = C(u_x, u_y, -u_z).$$

Since

$$\gamma\omega - k_x u_x = \gamma\omega\left(1 - \frac{v_x}{c}\sin\theta_0\right)$$

is always positive, (27) corresponds to a wave coming in from $z = +\infty$ when $u_z < 0$. According to our requirement (cf. Sec. III), we must choose

$$C(u_x, u_y, u_z) = 0, \quad u_z < 0.$$

The condition of specular reflection then implies that $C \equiv 0$, i.e., $f_H \equiv 0$, which was to be shown.

Finally, we must show for consistency that $f(0)$ satisfies the conditions (18) and (19). To do this, let us first consider the quantity

$$Q(k_z) \equiv -4\pi i\omega e\int \frac{d^3u}{\gamma}\frac{u_y u_z f(0)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}} = \omega_p^2 \omega 2\omega c B_x(0)\int d^3u \frac{u_y u_z(\partial F_0/\partial u_y)}{\gamma\omega - k_x u_x - k_z u_z}\frac{1}{2\pi i} \times \int_{C'} dk_z' \frac{1}{\Lambda(k_z')(\gamma\omega - k_x u_x - k_z' u_z)}. \tag{28}$$

So long as k_z is not on C', we have

$$\frac{1}{(\gamma\omega - k_x u_x - k_z u_z)}\frac{1}{(\gamma\omega - k_x u_x - k_z' u_z)} = \frac{1}{u_z(k_z' - k_z)}\left\{\frac{1}{\gamma\omega - k_x u_x - k_z' u_z} - \frac{1}{\gamma\omega - k_x u_x - k_z u_z}\right\}. \tag{29}$$

Using (29) in (28), we obtain

$$\begin{aligned} Q(k_z) &= -\omega_p^2 2\omega c B_x(0) \\ &\quad \times \int_{C'} \frac{dk_z'}{2\pi i} \frac{1}{\Lambda(k_z')(k_z' - k_z)} \{A(k_z') - A(k_z)\} \\ &= -2\omega c B_x(0) \\ &\quad \times \int_{C'} \frac{dk_z'}{2\pi i} \frac{\Lambda(k_z') - c^2(k_z'^2 - k_z^2) - \Lambda(k_z)}{\Lambda(k_z')(k_z' - k_z)}. \end{aligned} \tag{30}$$

We note that

$$\begin{aligned} B_x(z) &= \frac{ic}{\omega} \frac{\partial E_y}{\partial z} \\ &= \frac{c}{\omega} 2\omega c B_x(0) \int_{C'} \frac{dk_z}{2\pi i} e^{ik_z z} \frac{k_z}{\Lambda(k_z)} \\ &= \frac{c}{\omega} 2\omega c B_x(0) \left\{ \frac{\kappa_0}{\Lambda'(\kappa_0)} e^{i\kappa_0 z} \right. \\ &\quad \left. + \int_{\alpha_0}^{\infty} \frac{dk_z}{2\pi i} k_z \left(\frac{1}{\Lambda^-(k_z)} - \frac{1}{\Lambda^+(k_z)} \right) \right\}, \end{aligned} \tag{31}$$

which, when evaluated at $z = 0$, yields the identity

$$1 = 2c^2 \int_{C'} \frac{dk_z}{2\pi i} \frac{k_z}{\Lambda(k_z)}. \tag{32}$$

Also, from Eq. (23) we have

$$E_y(o) = -2\omega c B_x(0) \int_{C'} \frac{dk_z}{2\pi i} \frac{1}{\Lambda(k_z)}. \tag{33}$$

Making use of (32) and (33) in (30), we obtain

$$\begin{aligned} R_2(k_z) &= Q(k_z) + c^2 k_z E_y(0) - \omega c B_x(0) \\ &= -2\omega c B_x(0) \left\{ \int_{C'} \frac{dk_z'}{2\pi i} \frac{1}{(k_z' - k_z)} \right. \\ &\quad \left. - \Lambda(k_z) \int_{C'} \frac{dk_z'}{2\pi i} \frac{1}{\Lambda(k_z')(k_z' - k_z)} \right\}, \end{aligned} \tag{34}$$

from which (18) and (19) follow immediately. Q.E.D. Also, we note that $R_2(k_z)$ has the analytical properties assumed in Sec. III.

V. VAN KAMPEN–CASE STATIONARY MODES

It may be of interest to indicate the van Kampen–Case stationary modes for the present problem. For this purpose, we define a column vector

$$\Psi_{k_x,\omega}(z, \mathbf{u}) \equiv \begin{pmatrix} f(z, \mathbf{u}) \\ E_y(z) \\ B_x(z) \\ B_z(z) \end{pmatrix}. \tag{35}$$

From Eqs. (24), (26), (34), and the last one of Eqs. (14), we may write

$$\begin{aligned} \Psi_{k_x,\omega}(z, \mathbf{u}) &= -2\omega c B_x(0) \frac{1}{2\pi i} \int_{C'} dk_z e^{ik_z z} \\ &\quad \times \frac{1}{\Lambda(k_z)} \begin{bmatrix} -\dfrac{nei}{m} \gamma \dfrac{\partial F_0}{\partial u_y} \dfrac{1}{\gamma\omega - \mathbf{k} \cdot \mathbf{u}} \\ 1 \\ -\dfrac{ck_z}{\omega} \\ \dfrac{ck_x}{\omega} \end{bmatrix} \\ &= -2\omega c B_x(0) \left\{ \frac{1}{\Lambda'(\kappa_0)} \Psi_{k_x,\omega}(\kappa_0, \mathbf{u}) e^{i\kappa_0 z} \right. \\ &\quad \left. + \int_{\alpha_0}^{\infty} dk_z e^{ik_z z} \left(\frac{1}{\Lambda^-} - \frac{1}{\Lambda^+} \right)_{(k_z)} \Psi_{k_x,\omega}(k_z, \mathbf{u}) \right\}, \end{aligned} \tag{36}$$

where

$$\Psi_{k_x,\omega}(\kappa_0, \mathbf{u}) \equiv \begin{bmatrix} -\dfrac{nei}{m} \gamma \dfrac{\partial F_0}{\partial u_y} \dfrac{1}{\gamma\omega - k_x u_x - \kappa_0 u_z} \\ 1 \\ -\dfrac{c\kappa_0}{\omega} \\ \dfrac{ck_x}{\omega} \end{bmatrix},$$

and for $k_z > \alpha_0$

$$\begin{aligned} &\Psi_{k_x,\omega}(k_z, \mathbf{u}) \\ &\equiv \begin{bmatrix} P \dfrac{1}{\gamma\omega - k_x u_x - k_z u_z} \\ \qquad - i\pi \left(\dfrac{\Lambda^+ + \Lambda^-}{\Lambda^+ - \Lambda^-} \right)_{(k_z)} \delta(\gamma\omega - \mathbf{k} \cdot \mathbf{u}) \\ 1 \\ -\dfrac{ck_z}{\omega} \\ \dfrac{ck_x}{\omega} \end{bmatrix}. \end{aligned}$$

Here $\Psi(\kappa_0)$ and $\Psi(k_z)$ are the van Kampen–Case stationary modes appropriate to the relativistic problem, and P denotes the principal value.

Equation (36) expresses the stationary solution (with an incident plane wave) of the half-space problem within the plasma in terms of the van Kampen–Case stationary modes appropriate to the relativistic plasma. Each of these modes represents a stationary solution of the infinite-medium problem and

is a distribution in the sense of Schwartz. The particular superposition of these modes given by (36) matches to a plane wave obliquely incident on the plasma to yield a stationary solution as we have shown.

VI. REFLECTION, TRANSMISSION, AND ABSORPTION COEFFICIENTS FOR HALF-SPACE PLASMA

Let us consider a plane electromagnetic wave in free space incident upon the plasma interface at an angle θ_0. Denoting by E_y^i and E_y^r the amplitudes of electric fields in the incident and reflected waves, respectively, for $z < 0$ we have

$$E_y(z) = \mathcal{E}_y^i(e^{iz(\omega/c)\cos\theta_0} + \rho e^{-iz(\omega/c)\cos\theta_0}), \quad (37)$$

where $\rho \equiv \mathcal{E}_y^i/\mathcal{E}_y^r$. For the magnetic field we have

$$B_x(z) = -\cos\theta_0 \mathcal{E}_y^i(e^{iz(\omega/c)\cos\theta_0} - \rho e^{-iz(\omega/c)\cos\theta_0}). \quad (38)$$

Using the continuity of the fields across the interface, we obtain the reflection coefficient R (i.e., the fraction of the incident energy which is reflected) as

$$R = |\rho|^2, \quad (39)$$

where

$$\rho \equiv \left[1 + \frac{B_x(0)}{E_y(0)\cos\theta_0}\right] \Big/ \left[1 - \frac{B_x(0)}{E_y(0)\cos\theta_0}\right], \quad (40)$$

and [cf. Eq. (24)]:

$$\frac{B_x(0)}{E_y(0)} = -\left[2\omega c \frac{1}{2\pi i}\int_{C'} dk_z \frac{1}{\Lambda(k_z)}\right]^{-1} \quad (41)$$

To obtain the transmission coefficient, we consider the electric field as $z \to \infty$ which is given by [cf. Eq. (23)]:

$$E_y^t = \mathcal{E}_y^t e^{i\kappa_0 z}, \quad (42)$$

where

$$\mathcal{E}_y^t \equiv -2\omega c B_x(0)\frac{1}{\Lambda'(\kappa_0)}. \quad (43)$$

We need only discuss the case when κ_0 is real; otherwise, there is no transmitted wave. We calculate the time-averaged Poynting vector

$$\langle \mathbf{S} \rangle = \frac{c}{8\pi}\mathbf{E}(z) \times \mathbf{B}^*(z) \quad (44)$$

for both the transmitted and incident waves; we obtain

$$\langle \mathbf{S}^t \rangle = \frac{c}{8\pi}|\mathcal{E}_y^t|^2 \frac{c}{\omega}(\kappa_0 \hat{\mathbf{z}} + k_x \hat{\mathbf{x}}), \quad (45)$$

$$\langle \mathbf{S}^i \rangle = \frac{c}{8\pi}|\mathcal{E}_y^i|^2 \frac{c}{\omega}\left(\frac{\omega}{c}\cos\theta_0 \hat{\mathbf{z}} + k_x \hat{\mathbf{x}}\right), \quad (46)$$

respectively. The transmission coefficient is then given by

$$T \equiv \frac{\hat{\mathbf{z}} \cdot \langle \mathbf{S}^t \rangle}{\hat{\mathbf{z}} \cdot \langle \mathbf{S}^i \rangle} = \frac{c\kappa_0}{\omega}\frac{1}{\cos\theta_0}\left|\frac{\mathcal{E}_y^t}{\mathcal{E}_y^i}\right|^2. \quad (47)$$

Finally, inserting (43) into (47) and using (37) with $z = 0$, we obtain

$$T = \frac{c\kappa_0}{\omega}\cos\theta_0 \frac{4\omega^2 c^2}{|\Lambda'(\kappa_0)|}|1 - \rho|^2, \quad (48)$$

where ρ is given by (40).

The fraction of the incident energy which is absorbed within the plasma is then given by

$$A = 1 - R - T. \quad (49)$$

It is interesting to note that the energy absorbed by the particles is removed from the system to infinity in the form of a steady heat flow. This is expected, since, otherwise, strictly stationary solutions (ω real) could not have existed. To make this point more explicit, we shall calculate the net flow of heat in the z direction, namely,

$$q_z^{(2)} = \int \frac{d^3u}{\gamma} mc^2 \gamma u_z f^{(2)}, \quad (50)$$

where $f^{(2)}$ denotes the second-order perturbed distribution function. We note that the first-order distribution function given by (26) does not yield any flow of heat, since it is an odd function of u_y.

The second-order distribution function satisfies the following equation:

$$\gamma\frac{\partial f^{(2)}}{\partial t} + u_x\frac{\partial f^{(2)}}{\partial x} + u_z\frac{\partial f^{(2)}}{\partial z} + \frac{ne}{m}E_y^{(2)}\frac{\partial F_0}{\partial u_y} = -\frac{e}{m}\{\mathrm{Re}\, E_y(\mathbf{x}, t)\}\gamma\frac{\partial}{\partial u_y}\{\mathrm{Re}\, f(\mathbf{x}, \mathbf{u}, t)\}, \quad (51)$$

where E_y and f are the first-order solutions obtained previously. Together with (51) we have also the second-order Maxwell equations.

To evaluate the time-averaged heat flow, we need only to consider the zero-frequency component $f_0^{(2)}$ of $f^{(2)}$, since $\pm 2\omega$ components yield zero through time averaging. Now, if one inserts the first-order solutions into the right-hand side of (51), one finds that the zero-frequency part of the right-hand side is also independent of x. This implies that $f_0^{(2)}$ is independent of x, since we are interested in solutions which are harmonic in x. Also, the static part of $E_0^{(2)}$ vanishes, since there are no gradients in the y direction. Accordingly, we find that $f_0^{(2)}$ satisfies the following equation:

$$u_z\frac{\partial f_0^{(2)}}{\partial z} = -\frac{e}{2m}\gamma\frac{\partial}{\partial u_y}\mathrm{Re}\{E^*(z)f(z)\}. \quad (52)$$

From (50) and (52) we obtain the following result for the time-averaged heat flow in the z direction:

$$\frac{\partial}{\partial z}\langle q_z^{(2)}(z)\rangle = -ec^2\,\mathrm{Re}\left\{E^*(z)\int d^3u\gamma\frac{\partial}{\partial u_y}f(z)\right\}. \quad (53)$$

The right-hand side of (53) may be simplified after inserting for $f(z)$ and by straightforward calculations. We obtain

$$\frac{\partial}{\partial z}\{\langle q_z^{(2)}(z)\rangle + \hat{\mathbf{z}}\cdot\langle \mathbf{S}(z)\rangle\} = 0, \tag{54}$$

where

$$\hat{\mathbf{z}}\cdot\langle \mathbf{S}(z)\rangle = \mathrm{Re}\left[-\frac{c}{8\pi}E_y^*(z)B_x(z)\right] \tag{55}$$

is the time-averaged net flow of electromagnetic energy in the z direction. Equation (54) expresses the conservation of the energy flow.

Since the heat flow is zero at $z = 0$ due to the specular reflection, from (54) we can write

$$q_z^{(2)}(z) = \hat{\mathbf{z}}\cdot\langle \mathbf{S}(0)\rangle - \hat{\mathbf{z}}\cdot\langle S(z)\rangle. \tag{56}$$

According to (56), heat flow builds up as z increases through the absorption of the energy by the particles from the electromagnetic wave and, as $z\to\infty$, reaches the asymptotic value

$$q_z^{(2)}(\infty) = \hat{\mathbf{z}}\cdot\langle \mathbf{S}^i(0) - \mathbf{S}^r(0) - \mathbf{S}^t\rangle, \tag{57}$$

where the superscripts refer to incident, reflected, and transmitted Poynting vectors. Also, the fact that heat flow attains a constant asymptotic value shows that the conversion of energy from electromagnetic energy to heat energy is a surface phenomenon (anomalous skin effect).

VII. SLAB PLASMA

We now consider a slab plasma whose faces are perpendicular to the z axis and situated at $z = 0$ and $z = a$. To avoid repetition, we begin here by indicating the modification necessary for adopting some of the results of the previous analysis to the present case.

We define the functions $\mathcal{E}_y(z)$, $\mathfrak{B}(z)$ which are identical to the fields $E_y(z)$, $\mathbf{B}(z)$ within the plasma and which vanish identically outside the plasma. It then follows that the Laplace transform of $\mathcal{E}_y(z)$ is given by the same expression as in Eq. (16) provided we replace in $R_2(k_z)$ the quantities $f(0)$, $E_y(0)$, $B_x(0)$ by $f(0) - e^{-ik_z a}f(a)$, $E_y(0) - e^{-ik_z a}E_y(a)$, $B_x(0) - e^{-ik_z a}B_x(a)$, respectively. Here again $ik_z \equiv p$ is the Laplace transform variable, and

$$\mathcal{E}_y(k_z) \equiv \int_0^\infty dz\, e^{-ik_z z}\mathcal{E}_y(z) = \int_0^a dz\, e^{-ik_z z}E_y(z). \tag{58}$$

Thus, in the case of a slab we have

$$\mathcal{E}_y(z) = \frac{1}{2\pi i}\int_C dk_z e^{ik_z z} i\mathcal{E}_y(k_z), \tag{59}$$

where

$$i\mathcal{E}_y(k_z) = \frac{R_2^{\mathrm{slab}}(k_z)}{\Lambda(k_z)}, \tag{60}$$

$$R_2^{\mathrm{slab}}(k_z) \equiv R^0(k_z) - e^{-ik_z a}R^a(k_z), \tag{61}$$

$$R^z(k_z) \equiv Q^z(k_z) + c^2 k_z E_y(z) - \omega c B_x(z), \tag{62}$$

$$Q^z(k_z) \equiv -4\pi i\omega e\int \frac{d^3u}{\gamma}\,\frac{u_y u_z\, f(z)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}}. \tag{63}$$

It is clear from (58) that if $E_y(z)$ is a bounded function of z for $0 < z < a$, then $E_y(k_z)$ is an entire function in the complex k_z plane. Therefore, we must have

$$R_2^{\mathrm{slab}}(\pm\kappa_0) = 0, \tag{64}$$

$$\frac{R_2^{\mathrm{slab}+}(k_z)}{\Lambda^+(k_z)} = \frac{R_2^{\mathrm{slab}-}(k_z)}{\Lambda^-(k_z)} \quad (|k_z| > \alpha_0). \tag{65}$$

It is seen that in the case of a slab one has twice as many conditions compared to the half-space case. In fact, this is necessary, since we have twice as many coefficients to determine, corresponding to the waves travelling in both directions within the slab.

The contour C in Eq. (59) can be shifted in the k_z plane, since the integrand is an entire function. We choose it to lie parallel to the real axis and *below* all singularities of the function $1/\Lambda(k_z)$.

From (62) one observes that, under certain restrictions on the velocity dependence of $f(0)$ and $f(a)$, $R^0(k_z)$ and $R^a(k_z)$ are analytic functions of k_z in the plane cut along $|k_z| > \alpha_0$. We assume this to be the case. It will be verified *a posteriori*.

Within the foregoing remarks, it is seen from Eq. (59) that $E_y(z) \equiv 0$ for $z < 0$ and $z > a$. Also it is seen that the second term in the expression for R_2^{slab} given by Eq. (61) does not contribute to the integral in Eq. (59) when $0 < z < a$, so that

$$\mathcal{E}_y(z) \equiv E_y(z) = \frac{1}{2\pi i}\int_C dk_z e^{ik_z z}\frac{R^0(k_z)}{\Lambda(k_z)} \quad (0 < z < a). \tag{66}$$

The contour C may now be deformed as in Fig. 3 to yield

$$\begin{aligned} E_y(z) = {} & e^{i\kappa_0 z}\frac{R^0(\kappa_0)}{\Lambda'(\kappa_0)} + e^{-i\kappa_0 z}\frac{R^0(-\kappa_0)}{\Lambda'(-\kappa_0)} \\ & + \frac{1}{2\pi i}\int_{\alpha_0}^{\infty} dk_z e^{ik_z z}\left[\frac{R^{0-}(k_z)}{\Lambda^-(k_z)} - \frac{R^{0+}(k_z)}{\Lambda^+(k_z)}\right] \\ & + \frac{1}{2\pi i}\int_{-\infty}^{-\alpha_0} dk_z e^{ik_z z}\left[\frac{R^{0-}(k_z)}{\Lambda^-(k_z)} - \frac{R^{0+}(k_z)}{\Lambda^+(k_z)}\right], \end{aligned} \tag{67}$$

since the contribution from the semicircle vanishes as the radius approaches infinity.

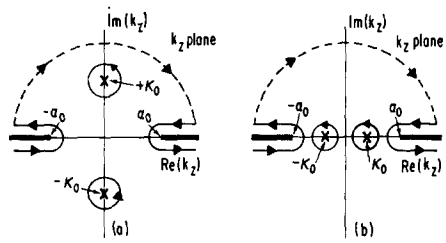


FIG. 3. The path of integration after deformation of the contour C.

Conditions (64) and (65) may now be utilized, together with the condition of specular reflection (at both faces of the slab), to determine the quantities

$$R^0(\pm\kappa_0), \tag{68}$$

$$\left(\frac{R^{0-}}{\Lambda^-} - \frac{R^{0+}}{\Lambda^+}\right)_{(k_z)} \quad (|k_z| > \alpha_0), \tag{69}$$

which are involved in the expression for $E_y(z)$ in (67). The result is

$$R^0(\pm\kappa_0)\sin\kappa_0 a = \mp i\omega c\{B_x(0)e^{\mp i\kappa_0 a} - B_x(a)\}, \tag{70}$$

$$\left(\frac{R^{0+}}{\Lambda^+} - \frac{R^{0-}}{\Lambda^-}\right)_{(k_z)}\sin k_z a = -i\omega c\left(\frac{1}{\Lambda^+} - \frac{1}{\Lambda^-}\right)_{(k_z)}\{B_x(0)e^{ik_z a} - B_x(a)\} \quad (|k_z| > \alpha_0). \tag{71}$$

There arises an ambiguity in obtaining the quantity given by (69) from Eq. (71), as one divides both sides by $\sin k_z a$, since this division is not permissible when

$$|k_z| = \left|\frac{n\pi}{a}\right| > \alpha_0 \quad (n = \text{integer}).$$

To overcome this difficulty, following Landau,[9] we consider ω as a complex number with a small, positive imaginary part which is allowed to go to zero after the solution is obtained completely. This amounts to the interpretation of the stationary solution in the limit (as $t \to \infty$) of an initial-value problem. The significance of this procedure in the present problem is that it completely removes the above-mentioned ambiguity. When ω has a small positive imaginary part, the singularities of $1/\Lambda(k_z)$ are shifted as in Fig. 4. Equation (71) now holds along the shifted branch cut where $\sin k_z a$ is non-vanishing. Thus, along the shifted branch cut we obtain

$$\left(\frac{R^{0+}}{\Lambda^+} - \frac{R^{0-}}{\Lambda^-}\right)_{(k_z)} = -\frac{i\omega c}{\sin k_z a}\left(\frac{1}{\Lambda^+} - \frac{1}{\Lambda^-}\right)_{(k_z)}\{B_x(0)e^{-ik_z a} - B_x(a)\}. \tag{72}$$

[9] L. Landau, J. Phys. (USSR) **10**, 25 (1946).

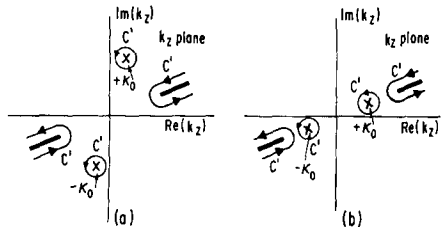


FIG. 4. The path of integration C' when ω has a slight positive imaginary part. [Note that the root $+\kappa_0$ moves into the quarter plane Re $(k_z) > 0$, Im $(k_z) > 0$.]

Making use of (70) and (72) in (67), for $0 < z < a$ we obtain

$$E_y(z) = -i\omega c\int_{C'}\frac{dk_z}{2\pi i}\,e^{ik_z z}\frac{1}{\Lambda(k_z)\sin k_z a} \times \{B_x(0)e^{-ik_z a} - B_x(a)\}. \tag{73}$$

A. Determination of the Perturbed-Particle Distribution Function

It is readily verified that a particular solution of the Vlasov equation (25) with $E_y(z)$ given by (73) is

$$f(z) = -\frac{ne\omega c}{m}\gamma\frac{\partial F_0}{\partial u_y} \times\int_{C'}\frac{dk_z}{2\pi i}\,e^{ik_z z}\frac{B_x(o)e^{-ik_z a} - B_x(a)}{(\gamma\omega - \mathbf{k}\cdot\mathbf{u})\Lambda(k_z)\sin k_z a}. \tag{74}$$

It is a straightforward matter to show that this solution satisfies the condition of specular reflection at $z = 0$ and $z = a$.

We now prove that (74) is the unique solution. To show this we consider the general solution

$$f_{\mathrm{H}}(z) = C(\mathbf{u})e^{i[(\gamma\omega - k_x u_x)/u_z]z} \quad (0 < z < a) \tag{75}$$

of the homogeneous $[E_y(z) \equiv 0]$ Vlasov equation which may be added to (74). Since (74) already satisfies the condition of specular reflection at both boundaries, $C(\mathbf{u})$ in (75) must be so chosen that $f_{\mathrm{H}}(z)$ also satisfies the same conditions. This implies (since ω has a slight imaginary part) that $C(\mathbf{u}) \equiv 0$. Q.E.D.

For consistency, we must show that, with $f(0)$ and $f(a)$ obtained from (74), the conditions (64) and (65) are satisfied. For this purpose we first calculate $Q^0(k_z)$ defined by (63):

$$Q^0(k_z) = i\omega c\Lambda(k_z)\int_{C'}\frac{dk_z'}{2\pi i}\frac{B_x(0)e^{-ik_z' a} - B_x(a)}{(k_z' - k_z)\Lambda(k_z')\sin k_z' a} - i\omega c\int_{C'}\frac{dk_z'}{2\pi i}\frac{B_x(0)e^{-ik_z' a} - B_x(a)}{(k_z' - k_z)\sin k_z' a} + i\omega c\int_{C'}\frac{dk_z'}{2\pi i}\frac{c^2(k_z' + k_z)}{\Lambda(k_z')\sin k_z' a} \times \{B_x(0)e^{-ik_z' a} - B_x(a)\}. \tag{76}$$

To put this result into a more convenient form, we note that, using (73), we may write

$$B_x(z) = \frac{ic}{\omega}\frac{\partial E_y}{\partial z} = ic^2\int_{C'}\frac{dk_z}{2\pi i}\frac{k_z}{\Lambda(k_z)\sin k_z a}[B_x(0)e^{-ik_z a} - B_x(a)]. \tag{77}$$

From (73) and (77), we have

$$E_y(0) = -i\omega c\int_{C'}\frac{dk_z}{2\pi i}\frac{[B_x(0)e^{-ik_z a} - B_x(a)]}{\Lambda(k_z)\sin k_z a}, \tag{78}$$

$$E_y(a) = -i\omega c\int_{C'}\frac{dk_z}{2\pi i}\frac{[B_x(0) - e^{ik_z a}B_x(a)]}{\Lambda(k_z)\sin k_z a}, \tag{79}$$

$$B_x(0) = ic^2\int_{C'}\frac{dk_z}{2\pi i}\frac{k_z}{\Lambda(k_z)\sin k_z a}[B_x(0)e^{-ik_z a} - B_x(a)], \tag{80}$$

$$B_x(a) = ic^2\int_{C'}\frac{dk_z}{2\pi i}\frac{k_z}{\Lambda(k_z)\sin k_z a}[B_x(0) - e^{ik_z a}B_x(a)]. \tag{81}$$

Using (78) and (80) in (76) and recalling (62), we obtain

$$R^0(k_z) = i\omega c\Lambda(k_z)\int_{C'}\frac{dk_z'}{2\pi i}\frac{[B_x(0)e^{-ik_z' a} - B_x(a)]}{(k_z' - k_z)\Lambda(k_z')\sin k_z' a} - i\omega c\int_{C'}\frac{dk_z'}{2\pi i}\frac{[B_x(0)e^{-ik_z' a} - B_x(a)]}{(k_z' - k_z)\sin k_z' a}. \tag{82}$$

Similarly, for $R^a(k_z)$ we obtain

$$R^a(k_z) = i\omega c\Lambda(k_z)\int_{C'}\frac{dk_z'}{2\pi i}\frac{[B_x(0) - e^{ik_z' a}B_x(a)]}{(k_z' - k_z)\Lambda(k_z')\sin k_z' a} - i\omega c\int_{C'}\frac{dk_z'}{2\pi i}\frac{[B_x(0) - e^{ik_z' a}B_x(a)]}{(k_z' - k_z)\sin k_z' a}. \tag{83}$$

One can show that $R^0(k_z)$, $R^a(k_z)$ are analytic functions of k_z in the cut plane and that conditions (64) and (65) are satisfied. We omit the details of this calculation here.

B. Reflection, Transmission, and Absorption Coefficients for the Slab Plasma

Rewrite Eqs. (78) and (79) as

$$E_y(0) = \mu B_x(0) + \nu B_x(a), \tag{84}$$

$$E_y(a) = -\nu B_x(0) - \mu B_x(a), \tag{85}$$

where

$$\mu \equiv -i\omega c\int_{C'}\frac{dk_z}{2\pi i}\frac{\cos k_z a}{\Lambda(k_z)\sin k_z a} = -i\omega c\int_{C'}\frac{dk_z}{2\pi i}\frac{e^{\pm ik_z a}}{\Lambda(k_z)\sin k_z a}, \tag{86}$$

$$\nu \equiv i\omega c\int_{C'}\frac{dk_z}{2\pi i}\frac{1}{\Lambda(k_z)\sin k_z a}. \tag{87}$$

From the continuity of the fields at $z = a$, we have

$$B_x(a) = -\cos\theta_0 E_y(a), \tag{88}$$

so that we obtain

$$\frac{E_y(0)}{B_x(0)} = \frac{\mu + (\nu^2 - \mu^2)\cos\theta_0}{1 - \mu\cos\theta_0}, \tag{89}$$

$$\frac{E_y(a)}{E_y(0)} = \frac{\nu}{\mu + (\nu^2 - \mu^2)\cos\theta_0}. \tag{90}$$

The reflection coefficient in the case of a slab is again given by Eq. (39), provided that in Eq. (40) one inserts for $E_y(0)/B_x(0)$ from Eq. (89).

The transmission coefficient is

$$T = \left|\frac{E_y^t}{E_y^i}\right|^2 = \left|\frac{E_y(a)}{E_y(0)}\right|^2\cdot|1 + \rho|^2, \tag{91}$$

where $E_y(a)/E_y(0)$ is given by (90).

C. Correspondence with the Fourier-Series Solution

To express our solution in the form of a Fourier series, we use the identity

$$-i\omega c\int_{C_1+C_2}\frac{dk_z}{2\pi i}\frac{e^{ik_z z}}{\Lambda(k_z)\sin k_z a}[B_x(0)e^{ik_z a} - B_x(a)] = 0 \quad (0 < z < a), \tag{92}$$

which is quite evident since, for $0 < z < a$, the contour C_1 (see Fig. 5) can be closed from above and C_2 from below.

By deforming contour C_1 continuously onto C_2, in view of (73) we obtain

$$E_y(z) = \frac{i\omega c}{a}\sum_{n=-\infty}^{+\infty}(-1)^n e^{in\pi(z/a)}\frac{\{B_x(0)e^{-in\pi} - B_x(a)\}}{\Lambda[(n\pi/a) - i\delta_n\epsilon]}, \tag{93}$$

where

$$\delta_n = 1, \quad n\pi/a > \alpha_0,$$
$$= -1, \quad n\pi/a < -\alpha_0,$$
$$= 0, \quad -\alpha_0 < n\pi/a < +\alpha_0.$$

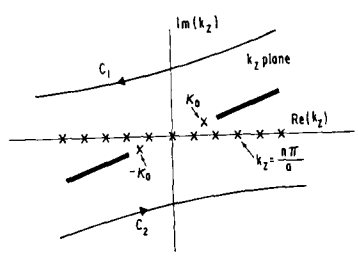


FIG. 5. The paths of integration C_1 and C_2 in the complex k_z plane.

Alternatively, we have

$$E_y(z) = \frac{i\omega c}{a}[B_x(0) - B_x(a)]\frac{1}{\Lambda(0)} + \frac{2i\omega c}{a}\sum_{0<n<\alpha_0 a/\pi}\frac{[B_x(0)-(-1)^n B_x(a)]}{\Lambda(n\pi/a)}\cos n\pi\frac{z}{a} + \frac{2i\omega c}{a}\sum_{n>\alpha_0 a/\pi}\frac{B_x(0)-(-1)^n B_x(a)}{\Lambda^-(n\pi/a)}\cos n\pi\frac{z}{a}. \quad (94)$$

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. J. M. Dawson for suggesting this problem and for his continuous interest. He is also grateful to Dr. M. Kruskal for his valuable criticism.

This work was performed under the auspices of the U.S. Atomic Energy Commission, Contract No. AT(30-1)-1238.

APPENDIX: THE DISPERSION FUNCTION

In this appendix we wish to discuss briefly some properties of the dispersion function in the complex k_z plane. We recall that

$$\Lambda(k_z) \equiv c^2k^2 - \omega^2 + \omega_p^2 A(k_z), \quad (A1)$$

where

$$A(k_z) \equiv -\omega\int d^3u\,\frac{u_y(\partial F_0/\partial u_y)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u}}. \quad (A2)$$

The denominator in (A2) is nonvanishing throughout the complex k_z plane, except along the parts of the real axis where $|k_z| > (\omega/c)(1-\sin^2\theta_0)^{\frac{1}{2}} \equiv \alpha_0$, and where $A(k_z)$ possesses, as will be seen below, two branch cuts.

We now proceed with the evaluation of (A2). First note that (for k_z real)

$$\frac{1}{\gamma\omega - k_x u_x - (k_z \pm i\epsilon)u_z} = \frac{1}{\gamma\omega - \mathbf{k}\cdot\mathbf{u} \mp i\epsilon}, \quad \text{if} \quad k_z > 0, \qquad = \frac{1}{\gamma\omega - \mathbf{k}\cdot\mathbf{u} \pm i\epsilon}, \quad \text{if} \quad k_z < 0. \quad (A3)$$

For $k_z > 0$ let us consider

$$\begin{aligned} A_{ij}(k_z + i\epsilon) &= -\omega\int d^3u\,\frac{u_i(\partial F_0/\partial u_j)}{\gamma\omega - \mathbf{k}\cdot\mathbf{u} - i\epsilon} \\ &= -i\omega\int_0^\infty dt\,e^{-\epsilon t}\int d^3u u_i\frac{\partial F_0}{\partial u_j}e^{-it(\gamma\omega-\mathbf{k}\cdot\mathbf{u})} \\ &= +i\frac{\omega\beta}{c^2}\int_0^\infty dt\,e^{-\epsilon t}\int\frac{d^3u}{\gamma}u_iu_jF_0e^{-it(\gamma\omega-\mathbf{k}\cdot\mathbf{u})} \\ &= -i\frac{\omega\beta}{c^2}\int_0^\infty dt\,e^{-\epsilon t}\frac{1}{t^2}\frac{\partial^2}{\partial k_i\partial k_j}\int\frac{d^3u}{\gamma}F_0e^{-it(\gamma\omega-\mathbf{k}\cdot\mathbf{u})}. \end{aligned} \quad (A4)$$

Carrying out the $\mathbf{u}$ integration using the polar coordinates with $\mathbf{k}$ as the polar axis by the use of the relation[10]

$$\int_0^\infty du\,\frac{u}{\gamma}e^{-ay}\sin yu = c^3y\frac{1}{(c^2y^2+a^2)}K_1[(c^2y^2+a^2)^{\frac{1}{2}}], \qquad \mathrm{Re}\,a > 0, \quad \mathrm{Re}\,c > 0, \quad (A5)$$

we obtain

$$\int\frac{d^3u}{\gamma}F_0e^{-it(\gamma\omega-\mathbf{k}\cdot\mathbf{u})} = \frac{\beta}{K_2(\beta)}\frac{1}{z}K_1(z), \quad (A6)$$

where $z \equiv [c^2k^2t^2 + (\beta + it\omega)^2]^{\frac{1}{2}}$. Inserting (A6) into (A4) and letting $\epsilon \to 0$, we obtain

$$A_{ij}^+(k_z) = +i\frac{\omega\beta^2}{K_2(\beta)}\int_0^\infty dt\left\{\frac{K_2(z)}{z^2}\delta_{ij} - c^2t^2\frac{K_3(z)}{z^3}k_ik_j\right\}. \quad (A7)$$

Since $k_y = 0$,

$$A_{yy}^+(k_z) = A^+(k_z) = +i\frac{\omega\beta^2}{K_2(\beta)}\int_0^\infty\frac{K_2(z)}{z^2} \quad (k_z > 0). \quad (A8)$$

When $c^2k^2 - \omega^2 > 0$, from (A8) we obtain

$$\mathrm{Re}\,A^+(k_z) = \frac{\beta}{K_2(\beta)}\sigma\int_0^\sigma dx\,\frac{K_2[\beta(1+\sigma^2-x^2)^{\frac{1}{2}}]}{1+\sigma^2-x^2}, \quad (A9)$$

$$\mathrm{Im}\,A^+(k_z) = \frac{\beta}{K_2(\beta)}\sigma\int_0^\infty dy\,\frac{K_2[\beta(y^2+1+\sigma^2)^{\frac{1}{2}}]}{y^2+1+\sigma^2}, \quad (A10)$$

where $\sigma \equiv \omega/(c^2k^2 - \omega^2)^{\frac{1}{2}}$.

The latter integral can be carried out explicitly to yield[11]

$$\mathrm{Im}\,A^+(k_z) = \frac{\beta}{K_2(\beta)}\sigma\frac{1}{2^{\frac{1}{2}}}\frac{\Gamma(\frac{1}{2})}{\beta^{\frac{1}{2}}(1+\sigma^2)^{\frac{3}{4}}}K_{\frac{3}{2}}[\beta(1+\sigma^2)^{\frac{1}{2}}]. \quad (A11)$$

Using the relation

$$A(k_z^*) = A^*(k_z), \quad (A12)$$

we find that

$$A^-(k_z) = [A^+(k_z)]^*. \quad (A13)$$

Since $\mathrm{Im}\,A^+ \neq 0$ for $k_z > \alpha_0$, we see that $A(k_z)$ has a branch cut along this part of the real axis. Finally, the relation

$$A(k_z) = A(-k_z) \quad (A14)$$

indicates that

$$A^\pm(-k_z) = A^\mp(k_z). \quad (A15)$$

Hence there is also a branch cut along $k_z < -\alpha_0$.

[10] A. Erdelyi, Ed., *Tables of Integral Transforms, Bateman Manuscript Project* (McGraw-Hill Book Company, Inc. 1954), Vol. 1, p. 75.

[11] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1962). Also note that we adopt the convention used in this book for the definition of $K_\nu(z)$.

When $c^2k^2 - \omega^2 < 0$, we find

$$A^+(k_z) = A^-(k_z) = \frac{\beta}{K_2(\beta)}\,\tilde{\sigma}\int_{\tilde{\sigma}}^{\infty} dy\,\frac{K_2[\beta(y^2+1-\tilde{\sigma}^2)^{\frac{1}{2}}]}{y^2+1-\tilde{\sigma}^2}, \quad \text{(A16)}$$

where

$$\tilde{\sigma} \equiv \frac{\omega}{(\omega^2 - c^2k^2)^{\frac{1}{2}}}.$$

It is noted that in this case $A(k_z)$ is real, as was expected.

The values of $A(k_z)$ elsewhere in the k_z plane may then be obtained by continuing $A^+(k_z)$ analytically in the cut plane.

Roots of the Dispersion Function $\Lambda(k_z)$

The number of zeros of the dispersion function throughout the cut plane can be determined by the argument principle. As k_z completes one tour along C_1, $\Lambda(k_z)$ traverses twice the contour C_2 sketched in Fig. 6. Since

$$\Lambda(\alpha_0) = \omega_p^2\,\frac{K_1(\beta)}{K_2(\beta)} \quad \text{(A17)}$$

is always positive, we find that there are always two roots of $\Lambda(k_z) = 0$ in the cut plane. According to the relations (A12) and (A14), if κ_0 is a root, then $-\kappa_0$, κ_0^*, and $-\kappa_0^*$ are also roots of the dispersion function. Therefore, roots must be either both real or both purely imaginary.

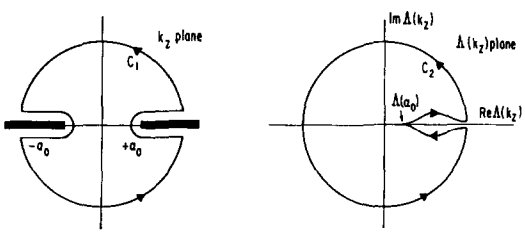


FIG. 6. The contours C_1 and C_2.

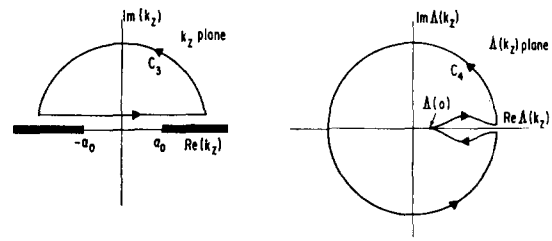


FIG. 7. The contours C_3 and C_4.

To obtain the condition which determines when the roots are purely imaginary, we let k_z run along the contour C_3 in Fig. 7. In this case, $\Lambda(k_z)$ traverses the full contour C_4 once. Since

$$\Lambda(0) = -\omega^2\,\frac{1}{\cos\theta_0}\left\{\cos^3\theta_0 - \frac{\omega_p^2}{\omega^2}\,\frac{\beta}{K_2(\beta)} \times \int_{1/\cos\theta_0}^{\infty} dy\,\frac{K_2[\beta(y^2+1-1/\cos^2\theta_0)^{\frac{1}{2}}]}{y^2+1-(1/\cos^2\theta_0)}\right\}, \quad \text{(A18)}$$

it follows that the roots are purely imaginary if

$$\cos^3\theta_0 < \frac{\omega_p^2}{\omega^2}\,\frac{\beta}{K_2(\beta)} \times \int_{1/\cos\theta_0}^{\infty} dy\,\frac{K_2[\beta(y^2+1-1/\cos^2\theta_0)^{\frac{1}{2}}]}{y^2+1-(1/\cos^2\theta_0)}. \quad \text{(A19)}$$

Otherwise, roots are real.

It is interesting to note that condition (A19), with an equal sign instead of inequality, determines the angle of "*total*" *reflection* from a half-space plasma as a function of ω/ω_p and β. [Of course, the reflection is never total (except in the zero-temperature limit) due to surface absorption (anomalous skin effect). Therefore, it is more appropriate to refer to this angle as the angle of "no transmission."]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Proof of the Fermion Superselection Rule without the Assumption of Time-Reversal Invariance

G. C. HEGERFELDT*† AND K. KRAUS
Institut für Theoretische Physik, Marburg, Germany
AND
E. P. WIGNER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

(Received 29 May 1968)

The superselection rule which separates states with integer angular momentum from those with half-integer angular momentum is proved using only rotational invariance.

The fermion or univalence superselection rule, which separates states with integer and half-integer angular momentum, was originally proved under the assumption of time-inversion invariance.[1] Recent experiments on *CP* violation, combined with the *CPT* theorem, now seem to question T as a rigorous symmetry. Another proof of the fermion superselection rule without the assumption of T invariance is thus desirable.

In Ref. 1, T invariance is not considered as a necessary assumption, but is merely used for technical reasons. Nevertheless, nobody as yet seems to have published a rigorous proof using only rotational invariance. Usually one gives the following plausibility argument: With two states ψ_+ and ψ_- of integer and half-integer angular momenta, respectively, a rotation by 2π around any axis transforms the state $\alpha\psi_+ + \beta\psi_-$ into $\alpha\psi_+ - \beta\psi_-$; since these states have to be physically indistinguishable, the superposition $\alpha\psi_+ + \beta\psi_-$ can not be a coherent one.

This argument, however, is valid only for a particular normalization of the rotation operators which ascribes to rotations by 2π the operators $+1$ and -1, respectively, in the subspaces of Hilbert space belonging to physical states with integer and half-integer angular momenta, respectively. Since this normalization, although very convenient, is only a mathematical convention and since the normalization might not be carried out in the subspaces independently, a rigorous proof must show in addition that the argument does not depend on the normalization of the rotation operators.

Such a proof, starting from ray representations of the rotation group, will be presented here. It turns out to be not much more complicated than the original proof of the fermion superselection rule in Ref. 1.

Pure states of a quantum system are described by unit rays $\Psi = \{\omega\psi\}$, $\|\psi\| = 1$, $\omega = e^{i\alpha}$, of a Hilbert space $\mathcal{H}$. The observables of the system are represented by Hermitian operators A on $\mathcal{H}$ with expectation values $(\psi, A\psi)$ in state Ψ. The value of this is, evidently, independent of the choice of ω in the correspondence $\Psi \to \omega\psi$. It is further assumed in quantum mechanics that the transition probability between any two states is an observable quantity. If the two states are characterized by the unit rays Ψ and Φ, the transition probability is given by $|(\psi, \varphi)|^2$, where ψ is any unit vector along the ray of Ψ and φ any unit vector along the ray of Φ. Again, one easily convinces oneself that the choice of these unit vectors from the rays Ψ and Φ, respectively, does not affect the value of the transition probability $|(\psi, \varphi)|^2$.

Before the establishment of the superselection rules, it was generally assumed that the vectors $\psi, \varphi, \cdots$, which correspond to physical states, form a Hilbert space; that is, that the linear combinations $\sum a_k\psi_k$, suitably normalized, also describe physical states for which the preceding postulates are valid. Naturally, $\sum a_k\psi_k$, and $\omega \sum a_k\psi_k$ describe the same state. The superselection rules limit the validity of this assumption and divide the set of all states into subsets which are called "coherent." The preceding statements apply within each coherent subset. In particular, if ψ and φ represent different states within the same coherent subset, there are states represented by the vectors $\cos\alpha\psi + e^{i\beta}\sin\alpha\psi$ for any α and β $(0 \leq \alpha, \beta < \pi)$, and they are all different. The transition probabilities between any pair of states within a coherent subset are physically meaningful.

It is well known,[2] furthermore, that if there are two different descriptions of the states *within a*

* Present address: The International Centre for Theoretical Physics, Trieste, Italy.

† Supported in part by the Bundesministerium für Wissenschaftliche Forschung, Bonn.

[1] G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. **88**, 101 (1952).

[2] This theorem is usually attributed to one of the present authors (E. P. W.), but it was probably known to others before. For a more recent proof, see V. Bargmann, J. Math. Phys. **5**, 862 (1964).

coherent subset of states, the second description attributing the rays $\Psi'', \Phi', X', \cdots$ to the states characterized by $\Psi, \Phi, X, \cdots$ in the first description, then one can choose unit vectors $\psi', \varphi', \chi', \cdots$ from the rays of $\Psi'', \Phi', X', \cdots$ to correspond to each unit vector $\psi, \varphi, \chi, \cdots$ of the rays $\Psi, \Phi, X, \cdots$ in such a way that the correspondence

$$\psi \leftrightarrow \psi', \quad \varphi \leftrightarrow \varphi', \quad \chi \leftrightarrow \chi', \quad \text{etc.}, \qquad (1)$$

either is unitary or it is antiunitary. It is true, also, that, if the choice of the $\psi', \varphi', \cdots$ is possible in such a way that the transformation from the $\psi, \varphi, \cdots$ is unitary, there is no choice $\omega\psi', \omega'\varphi', \cdots$ of these vectors which would render the transformation antiunitary[3] (and conversely). Finally, the choice of the $\psi', \varphi', \cdots$ from the rays $\Psi'', \Phi', \cdots$ is arbitrary only to within a *common* factor ω of modulus 1. Thus, once a vector ψ' is chosen from the ray Ψ'', the vectors $\varphi', \chi', \cdots$ are uniquely determined by the requirement that the correspondence (1) be unitary (if this can be accomplished at all) and also by the requirement that it be antiunitary (if this alternative applies).

If the underlying theory is rotationally invariant, there are many different descriptions of the total set of states. In particular, if there are two coordinate systems obtained from each other by a rotation R and if $\Psi, \Phi, \cdots$ are rays corresponding to certain states, one can introduce another description by attributing the rays $\Psi, \Phi, \cdots$ to states which have the same relation to the second coordinate system as the states originally described by $\Psi, \Phi, \cdots$ have with respect to the first coordinate system. Such states exist by rotational invariance. On the other hand, this second description will attribute rays $\Psi'', \Phi', \cdots$ to the states to which the first description attributed the rays $\Psi, \Phi, \cdots$. It then follows that there is either a unitary or an antiunitary operator O_R, which, if applied to any vector χ of a ray X characterizing a physical state in the first description, gives a vector $\chi' = O_R\chi$, which is in the ray X' which characterizes the same state in the second description. In particular, if $\psi, \varphi, \cdots$ are in the rays of $\Psi, \Phi, \cdots$, then $O_R\psi$, $O_R\varphi, \cdots$ will be in the rays of $\Psi'', \Phi', \cdots$; and if there are two operators O_R and O'_R with the described property, they can differ only in a constant factor of modulus 1. Finally (and this is an important point), if we consider only descriptions which attribute the rays $\Psi, \Phi, \cdots$ to the states which have the same relation to some coordinate system, which $\Psi, \Phi, \cdots$ have to the coordinate system used initially, i.e., if all coordinate systems use, from their own point of view, the same language to characterize states, then the operator O_R, which translates the description used originally to the description of the coordinate system rotated by R, will also translate the description of states used by *any* coordinate system into the description of the coordinate system obtained therefrom by the rotation R. This follows from the equivalence of all coordinate systems obtained from each other by rotations and from the postulate that each attributes the same ray Ψ to the state which, with respect to the coordinate system in question, has the same properties as the state described by the ray Ψ had in the original coordinate system.

It follows now that the product O_SO_R gives the same translation as O_{SR}, so that O_SO_R and O_{SR} can differ only in a constant factor. Since, for every rotation R, one can find another one (r) so that $R = r^2$, one has $O_R = \omega O_r^2$, and this is unitary no matter whether O_r is unitary or antiunitary. This applies to all rotations: all O_R are unitary. The translation operators which correspond to the unit element E of the rotation group are multiplications by the numbers of modulus 1. Furthermore, if O_R is a translation operator, so is O_R^{-1}; it corresponds to the rotation R^{-1}.

Let us now consider an operator $O_RO_SO_R^{-1}O_S^{-1}$. According to the preceding paragraph, this is a translation operator corresponding to the rotation $C = RSR^{-1}S^{-1}$. It is independent of the normalization of O_R and O_S, and this shows that one can define uniquely an operator for every commutator C in terms of the elements (in the preceding case R and S) of which it is a commutator.

In particular, if $C = E$ is the unit element, the corresponding operator is a number (of modulus 1), and this number, which may be a function $f(R, S)$ of R and S, is an invariant characteristic of the coherent subspace. The numerical function $f(R, S)$ has the following properties. Its value is 1 if O_R and O_S commute, and this will be always the case if they are members of a one-parametric subgroup. In this case, they can be represented as $\exp(iIt_1)$ and $\exp(iIt_2)$, respectively, where I is the infinitesimal operator of the subgroup. Furthermore, $f(R, S)$ is invariant under the substitution $R \to R' = TRT^{-1}$, $S \to S' = TST^{-1}$, where T is an arbitrary group element. This follows from the fact that one of the operators $O_{R'}$ is $O_TO_RO_T^{-1}$ and, similarly, one of the operators $O_{S'}$ is $O_TO_SO_T^{-1}$. Naturally, $f(R, S)$ is defined only if $RSR^{-1}S^{-1} = E$, i.e., for commuting R and S.

In the case of the rotation group, R and S commute only if (a) they are rotations about the same axis, or (b) if both R and S are rotations by π, about axes

[3] See E. P. Wigner, J. Math. Phys. **1**, 409, 414 (1960).

perpendicular to each other.[4] In the former case, R and S are members of the same one-parametric subgroup so that $f(R, S) = 1$ in this case. The condition that $f(R, S)$ is a single number is automatically satisfied in this case. The possible R, S of the second case can be obtained from each other by transformations $R \to TRT^{-1}$, $S \to TST^{-1}$, so that the corresponding $f(R, S)$ are equal for all R, S of the Case (b). The value of this quantity is, however, a characteristic of the coherent subspace in which the operators O_R are defined. Since it can assume only the values $+1$ and -1 if the O_R are normalized in the usual way,[5] it can assume only these values in any normalization. However, it must assume one of these values in any coherent subspace, and since it has the value 1 for states with integer spin, the value -1 for states with half-integer spin,[6] these cannot belong to the same coherent subspace.

The preceding argument made a minimum use of representation theory and made use of the properties of the rotation group only in the preceding paragraph. The rest of the considerations could be applied to any invariance group. The somewhat primitive argument involving a rotation by 2π was replaced by the consideration of the commutator of two rotations by π about axes perpendicular to each other, which, as are all operators, defined as commutators of other operators, is independent of the normalization of these.

The formal structure of quantum theories with superselection rules has been analyzed in detail by Jauch and Misra.[7] Under the assumption that there exists a complete set of commuting observables and that the "superobservables" have a discrete spectrum, it is shown that the state space $\mathcal{H}$ is the direct sum of subspaces $\mathcal{H}_i$ which are coherent in the sense defined above. For all observables A, $(\psi_i, A\psi_k) = 0$ if $\psi_i \in \mathcal{H}_i$, $\psi_k \in \mathcal{H}_k$, $i \neq k$. In such theories, pure states are in one-to-one correspondence with the unit rays of the coherent subspaces $\mathcal{H}_i$. For assume $\Psi = \{\omega\psi\}$ with $\psi = \sum_i \alpha_i \psi_i$, $\psi_i \in \mathcal{H}_i$, $\|\psi_i\| = 1$ to describe a pure state. Then,

$$(\psi, A\psi) = \sum_i |\alpha_i|^2 (\psi_i, A\psi_i) = \mathrm{Tr}\,(AW)$$

with $W = \sum_i |\alpha_i|^2 P_{\psi_i}$; i.e., the same state can as well be described by the density matrix W and is therefore pure if and only if all α_i except one are zero.[8] In other words, a "superposition" $\cos \alpha\psi_i + e^{i\beta} \sin \alpha\psi_k$ with $i \neq k$ does not produce a pure state, but a mixture[8] with a density matrix $W = \cos^2 \alpha P_{\psi_i} + \sin^2 \alpha P_{\psi_k}$, so that the relative phase β is unobservable. In this sense, different subspaces $\mathcal{H}_i$ and $\mathcal{H}_k$ are mutually incoherent.

Since, under rotations, pure states have to be transformed into pure states and transition probabilities must change continuously if one of the two states is rotated, a rotation R transforms all states of one coherent subspace into states belonging to this same subspace. This justifies our above assumption that the redescription of states induced by R occurs *within* a given coherent subset of states. The translation operator O_R of a theory with superselection rules is then the direct sum of operators $O_R^{(i)}$ for the coherent subspaces $\mathcal{H}_i$. This carries the theorem mentioned in Ref. 2 over to state spaces with superselection rules.

[4] If R and S commute, $SRS^{-1} = R$. However, if R is a rotation about v, so that $Rv = v$, then SRS^{-1} is a rotation about Sv. If the two are to be equal, Sv must be parallel to v. This is the case if $Sv = v$, i.e., if S is also a rotation about v, or if $Sv = -v$. The former case falls under (a); in the latter case, S must be a rotation by π about an axis perpendicular to v. Furthermore, if R is a rotation by φ about v, then SRS^{-1} is a rotation by φ about $-v$, i.e., a rotation by $-\varphi$ about v. If R and SRS^{-1} are to be identical, the rotations by φ and $-\varphi$ about v must be identical. This fixes $\varphi = \pi$ (unless $\varphi = 0$), i.e., R also is a rotation by π.

[5] The demonstration given by Wigner [Ann. Math. **40**, 149 (1939), p. 177] still appears to be the most simple.

[6] Denoting the rotations by π about the z and y axes by Z and Y, we note that the matrices which correspond to these rotations in $D^{(\frac{1}{2})}$, assuming the usual normalization, are is_z and is_y, respectively. Hence, in this case $O_Z O_Y O_Z^{-1} O_Y^{-1} = is_z is_y (is_z)^{-1} (is_y)^{-1} = s_z s_y s_z^{-1} s_y^{-1} = -1$. The same is then true of the Kronecker product $D^{\frac{1}{2}} x D^{\frac{1}{2}} x \cdots x D^{\frac{1}{2}}$ containing an odd number of factors. Since these Kronecker products contain all half-integer representations, $O_Z O_Y O_Z^{-1} O_Y^{-1} = -1$ for all of these. On the other hand, in the Kronecker products containing an even number of $D^{\frac{1}{2}}$, one has $O_Z O_Y O_Z^{-1} O_Y^{-1} = 1$, and this is true, therefore, for all integer j representations.

[7] J. M. Jauch, Helv. Phys. Acta **33**, 711 (1960); J. M. Jauch and B. Misra, *ibid.* **34**, 699 (1961).

[8] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Julius Springer-Verlag, Berlin, 1932) (English transl.: Princeton University Press, Princeton, 1955).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Nonlinear Boundary-Value Problems in One- and Two-Dimensional Composite Domains

JOSÉ CANOSA
IBM Scientific Center, Palo Alto, California

(Received 6 June 1968)

Some nonlinear boundary-value problems in one- and two-dimensional composite domains have been solved by a general eigenfunction-expansion method. The advantage of the method is that separable problems in more than one dimension can be solved almost as easily as one-dimensional problems. An optimum eigenfunction-expansion basis has been found that leads to accurate solutions with only a few terms in the expansion.

I. INTRODUCTION

This work is an extension of that reported in a previous paper,[1] referred to subsequently as I, on an expansion method for the solution of nonlinear boundary-value problems. In I, the method was applied only to one-dimensional problems in homogeneous domains; here we show that the method is much more general, and that it can be used for the solution of nonlinear boundary-value problems in composite multidimensional domains. Explicit solutions are given for one asymmetric problem in composite-slab geometry, and for a symmetric problem in two-dimensional r–z composite-cylindrical geometry. The importance of the method lies in the fact that its application to separable problems in more than one dimension does not involve a much greater effort than that required for one-dimensional problems.

II. STATEMENT OF THE PROBLEM AND FORMALISM

We consider homogeneous[2] boundary-value problems of the form

$$\nabla^2 y(\mathbf{x}) + p^2(\mathbf{x}) y = y^n, \quad y(\Gamma) = 0, \tag{1}$$

where Γ is the domain boundary, and $p^2(\mathbf{x})$ is a positive function which is piecewise continuous in the composite domain D. In the calculations presented, $p^2(\mathbf{x})$ assumes only constant values in the different subdomains of D. The exponent n is a positive number, and $n \geq 2$.

In one-dimensional problems it has been proved that the solution of (1) which is positive in all the domain is unique.[3] The uniqueness of the positive solution in multidimensional problems is now assumed without proof. It has some interest to mention that oscillating solutions of (1) have been found numerically. Our interest is in the positive distribution.

For clarity and completeness, we review briefly the method of solution which is given in detail in I. We assume that the solution of (1) is given in the form

$$y(\mathbf{x}) = \sum_{\nu=1}^{\infty} A_\nu \varphi_\nu(\mathbf{x}), \tag{2}$$

where φ_ν are the eigenfunctions of a certain eigenvalue problem associated with (1). The expansion (2) is now substituted into (1), truncated after j terms, and the orthogonality properties of the eigenfunctions φ_ν are used to eliminate the spatial dependence in the resulting expressions. In this way we obtain a coupled nonlinear algebraic system of j equations and j unknowns in the expansion coefficients A_ν of (2). In the problems in homogeneous domains considered in I, where $p^2(\mathbf{x})$ in (1) was constant, the choice of the expansion basis (2) was obvious: the eigenfunctions of the Helmholtz equation, i.e.,

$$\nabla^2 \varphi + \lambda \varphi = 0, \quad \varphi(\Gamma) = 0. \tag{3}$$

For the present problems in composite domains, this choice is not obvious. Although any complete set of eigenfunctions in D can be used in principle, one is interested in some optimum set that leads to the most accurate solution of (1) with the minimum of terms in (2). An expansion basis which seems to be optimum is defined by the eigenvalue problem associated with (1):

$$\nabla^2 \varphi(\mathbf{x}) + p^2(\mathbf{x}) \varphi - \lambda \varphi = 0, \quad \varphi(\Gamma) = 0. \tag{4}$$

The reasons are as follows: In one-dimensional problems, the maximum value of the positive solution of (1) is bounded[3]:

$$\lambda_1^{1/(n-1)} \leq M < [\text{maximum of } p^2(\mathbf{x})]^{1/(n-1)}, \tag{5}$$

where λ_1 is the fundamental eigenvalue of (4). The result (5) can be shown to be valid for two- and three-dimensional problems, in exactly the same way as was done in Ref. 3.[4] In all cases we have from (5):

$$M^{n-1} = O(\lambda_1). \tag{6}$$

[1] J. Canosa, J. Math. Phys. **8**, 2180 (1967).
[2] The extension of the method to inhomogeneous problems is immediate, as was shown in I.
[3] J. Canosa and J. Cole, J. Math. Phys. **9**, 1915 (1968).
[4] This follows from the necessary conditions that a function of two and three variables must satisfy to have a strict maximum.

In some asymptotic cases, i.e., when

$$p^2(\mathbf{x}) = \mu\rho(\mathbf{x}), \quad \mu \to \infty, \quad \rho(\mathbf{x}) = O(1), \tag{7}$$

we have the stronger result[3]

$$M^{n-1} \to \lambda_1. \tag{8}$$

In some vicinity of the maximum of the distribution, Eq. (1) can be approximated by

$$\nabla^2 y(\mathbf{x}) + p^2(\mathbf{x})y - M^{n-1}y = 0.$$

Because of (6) and (8), this equation is approximated by the equation satisfied by the fundamental eigenfunction of (4). Hence, in some vicinity of the maximum, the fundamental eigenfunction approaches the positive distribution given by (1) except for a constant factor. This fact is central in the choice of the eigenfunctions of (4) as our expansion basis. The results to be presented below justify this assumption, i.e., the fundamental-mode approximation obtained by keeping only the first term in (2) is a good approximation to the distribution given by (1) near its peak, even in strongly nonlinear and asymmetric cases.[5] In the two-dimensional problems considered, the one-mode approximation is not nearly so good as in one-dimensional problems, although its shape is quite close to that of the distribution near its peak.

The nonlinear algebraic system for the coefficients A_ν in (2) is

$$\begin{aligned}
&b_1^{11}A_1^2 + b_1^{22}A_2^2 + \cdots + b_1^{jj}A_j^2 \\
&\quad + 2b_2^{11}A_1A_2 + \cdots + 2b_j^{11}A_1A_j \\
&\quad + 2b^{1,2,3}A_2A_3 + \cdots + 2b^{1,2,j}A_2A_j \\
&\quad + \cdots + 2b^{1,j-1,j}A_{j-1}A_j = \lambda_1 b^{11}A_1, \\
&b_2^{11}A_1^2 + b_2^{22}A_2^2 + \cdots + b_2^{jj}A_j^2 + 2b_1^{22}A_1A_2 \\
&\quad + \cdots + 2b^{1,2,j}A_1A_j + 2b_3^{22}A_2A_3 \\
&\quad + \cdots + 2b_j^{22}A_2A_j + 2b^{2,j-1,j}A_{j-1}A_j = \lambda_2 b^{22}A_2, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (9) \\
&\qquad\qquad \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\
&\qquad\qquad \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\
&\qquad\qquad \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\
&b_j^{11}A_1^2 + b_j^{22}A_2^2 + \cdots + b_j^{jj}A_j^2 \\
&\quad + 2b^{1,2,j}A_1A_2 + \cdots + 2b_1^{jj}A_1A_j \\
&\quad + 2b^{2,3,j}A_2A_3 + \cdots + 2b_2^{jj}A_2A_j \\
&\quad + \cdots + 2b^{j,j-1,j}A_{j-1}A_j = \lambda_j b^{jj}A_j.
\end{aligned}$$

[5] It should be clearly understood that the choice of an expansion basis (2) for problem (1) is by no means unique, e.g., if we use that defined by

$$\nabla^2\varphi(\mathbf{x}) + \lambda p^2(\mathbf{x})\varphi = 0,$$

the approximations so obtained to the distribution given by (1) are found to be much poorer for the same number of terms in (2) than those obtained using (4).

It is worthwhile to note that this system has identical form as that obtained before for homogeneous domains [see Eq. (12) in I], and that it applies in principle to one-, two-, and three-dimensional problems (1). The notation in (9), which is valid for any geometry and number of dimensions, is as follows:

$$b_m^{ll} = \int_D \varphi_l^2\varphi_m\,dV, \quad b^{l,m,n} = \int_D \varphi_l\varphi_m\varphi_n\,dV,$$
$$b^{jj} = \int_D \varphi_j^2\,dV. \tag{10}$$

The eigenfunctions of (4) are normalized so that the maximum value of the fundamental mode is unity. In one-dimensional problems, the eigenfunctions of (4) can be shown to be orthogonal in the same way as in a standard Sturm–Liouville problem.[6] Although in separable two- and three-dimensional problems the eigenvalues and eigenfunctions require two and three indices, for convenience of notation they have been ordered in (9) and (10) with only one index according to the magnitude of the eigenvalues.

A. One-Dimensional Problem

The following asymmetric problem was chosen as an example because of its interest in nuclear-reactor theory, and also because it constitutes a severe test of the method:

$$\begin{gathered}
(d^2y/dx^2) + p^2(x)y = y^2, \\
y(0) = y(\pi) = 0, \\
p^2(x) = \begin{cases} p^2 > 1, & 0 \le x \le \pi/b, \quad b > 1, \\ 1, & \pi/b \le x \le \pi. \end{cases}
\end{gathered} \tag{11}$$

The necessary conditions for the existence of a positive solution of (11) have been given before.[3] It suffices to recall here that, when the average value of $p^2(x)$ over the domain

$$\overline{p^2(x)} \to 1, \tag{12a}$$

the solution of (11) approaches $\sin x$ except for a constant factor. But when

$$\overline{p^2(x)} \gg 1, \tag{12b}$$

the solution departs strongly from $\sin x$. With $p^2(x)$ given by (11), Eq. (4) is a Helmholtz-type equation in each subdomain, and its solution is elementary. We

[6] The Sturm–Liouville theory is discussed in standard mathematical physics textbooks with the requirement that $p^2(\mathbf{x})$ in (4) be continuous, and not piecewise continuous as in this paper. Eigenvalue problems such as (4) with piecewise continuous coefficients have been discussed frequently in technical articles [see, for example, S. S. Penner and S. Sherman, J. Chem. Phys. 15, 569 (1947)]. Also, these eigenvalue problems are common in quantum mechanics when the potential is a piecewise constant function of space.

get

$$\varphi = \sin (p^2 - \lambda)^{\frac{1}{2}}x, \qquad 0 \leq x \leq \pi/b,$$
$$\varphi = C \sin (1 - \lambda)^{\frac{1}{2}}(x - \pi), \quad \pi/b \leq x \leq \pi. \qquad (13)$$

Note that $p^2 > \lambda_1$ [Eq. (5)], and as the fundamental is the highest positive eigenvalue, the radicand $p^2 - \lambda$ in (13) is positive for all λ. However, the radicand $1 - \lambda$ might be positive or negative. The coefficients C assure the continuity of the eigenfunctions and their derivatives at the interface $x = \pi/b$. The matching conditions at $x = \pi/b$ give directly the transcendental equation for the eigenvalues

$$(p^2 - \lambda)^{\frac{1}{2}} \cos [(p^2 - \lambda)^{\frac{1}{2}}\pi/b] \sin \{[(1 - \lambda)^{\frac{1}{2}}(b - 1)/b]\pi\} + (1 - \lambda)^{\frac{1}{2}} \sin [(p^2 - \lambda)^{\frac{1}{2}}\pi/b] \times \cos \{[(1 - \lambda)^{\frac{1}{2}}(b - 1)/b]\pi\} = 0. \qquad (14)$$

Although $1 - \lambda$ might be positive or negative, Eq. (14) has only real roots.[7] A FORTRAN IV program was written to solve (14), which gives in one pass any desired number of eigenvalues,[8] and the corresponding eigenfunctions (13) and their zeros. The zeros are needed for an accurate numerical computation of the integrals (10), whose integrands are multiple products of the eigenfunctions. The integrands are continuous, but their derivatives have very strong discontinuities at the zeros of the eigenfunctions. The accurate computation of these integrals requires that the integration interval be divided at the zeros. The integrands are thus smooth and of the same sign in the resulting subintervals, where the integration can be performed numerically by any standard method (Simpson's or the trapezoidal rule).

It should be stressed that from a numerical point of view the only delicate part of the computation of the integrals is the above-mentioned one, and it is easily taken care of as stated. With the use of a digital computer, it is not significant whether the integrands are triple products of the eigenfunctions as in (10), higher-degree products, or nonsingular quotients involving the eigenfunctions. The method, the feasibility of which depends essentially on the easy and efficient computation of integrals of the type (10), can thus be equally applied to other types of algebraic nonlinearities, and not only to those given in (1).

[7] If the coefficient $p^2(x)$ in (4) is continuous, it is a general result of the Sturm–Liouville theory that all its eigenvalues are real [E. L. Ince, *Ordinary Differential Equations* (Dover Publications, Inc., New York, 1956), p. 238]. For $p^2(x)$ piecewise constant, we are unaware of an equally general proof of the result. However, the result can be shown to be true in each case [see H. S. Carslaw and J. G. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, London, 1959), 2nd ed. p. 324].

[8] The fundamental eigenvalue given by (14) is always positive, but for sufficiently large p^2 in (11), a finite number of positive eigenvalues also exist. The rest of the spectrum is given by an infinite number of negative eigenvalues. In most of the problems considered, only the fundamental eigenvalue is positive.

B. Numerical Results and Discussion

As an illustration of the power of the method, we have solved (11) for the following $p^2(x)$:

$$p^2(x) = \begin{cases} 25, & 0 \leq x \leq \pi/5, \\ 1, & \pi/5 \leq x \leq \pi. \end{cases} \qquad (15)$$

This is an example which is strongly nonlinear and asymmetric. Our main interest was to observe the convergence of the modal approximations on the "exact" numerical solution of the problem obtained by an independent numerical method. The results of the one-, two-, and three-mode approximations are shown in Fig. 1, together with the exact solution. One should note the striking accuracy of the one-mode approximation in the vicinity of the maximum of the distribution, as was suggested by the discussion following Eq. (8) above. It should also be noticed that the one-mode approximation requires hardly any computation, because after neglecting all terms but those with index 1, the modal coefficient A_1 in (2) is obtained explicitly from system (9),

$$y(x) = A_1\varphi_1(x) = (\lambda_1 b^{11}/b_1^{11})\varphi_1(x). \qquad (16)$$

It should be mentioned that the exact numerical solution was quite hard to obtain, because an oscillating solution with one zero inside the domain was unexpectedly found; this solution coincided practically with the positive solution in the range $0 \leq x \leq 0.2\pi$, e.g., its slope at $x = 0$ was 55.63, vs 55.91 for the positive solution. In short, the numerical separation of

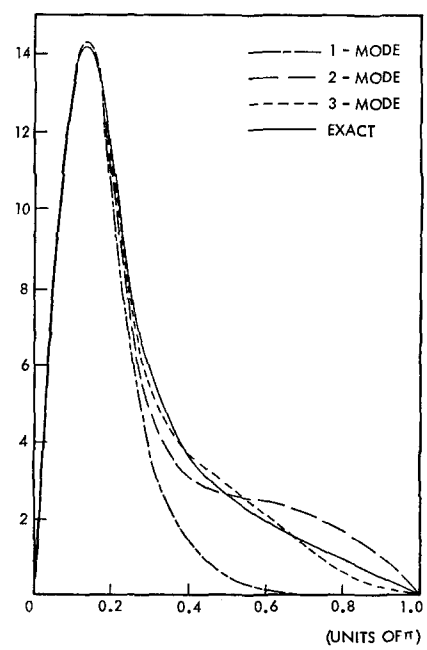


FIG. 1. First few mode approximations vs exact distribution.

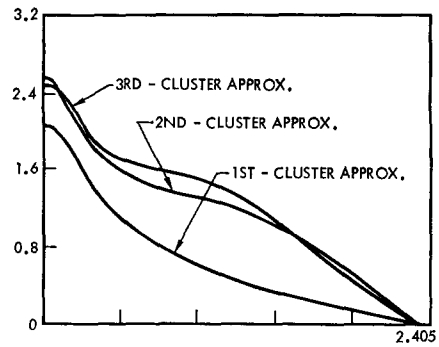


FIG. 2. First few mode approximations to the radial distribution in $z = 0$. [Case (21a).] The differences between the successive modal approximations within a cluster cannot be detected in the plots.

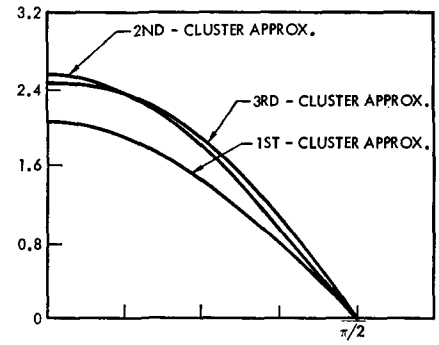


FIG. 3. First few mode approximations to the axial distribution in $r = 0$. Case (21a).

the two solutions required tedious numerical experimentation. To the contrary, system (9) for $j = 2$ and $j = 3$ was solved without any difficulty by the Newton–Raphson method[9] after only three iterations. Appropriate initial guesses for A_1, A_2, and A_3, giving the positive distribution were found to be, respectively, just as in I, A_1 [as given by (16)], $-A_1/10$, and $A_1/100$.

III. TWO-DIMENSIONAL PROBLEMS

Whenever the two- and three-dimensional eigenvalue problem (4) can be solved by separation of variables, the solution of the corresponding nonlinear problem (1) does not require a much greater effort than in one-dimensional problems. As an example, we have applied the method to the two-dimensional problem in the r–z composite cylindrical geometry given below. This problem is important in nuclear-reactor theory, because the discontinuous material properties of a reactor give rise naturally to composite problems of this type.

We have

$$\nabla^2 y(r, z) + 2p^2(r)y = y^2, \quad y(\Gamma) = 0,$$

$$p^2(r) = \begin{cases} p^2 > 1, & 0 \le r \le k_{01}/b, \quad b > 1, \\ 1, & k_{01}/b \le r \le k_{01}, \end{cases}$$

$$k_{01} = \text{first zero of } J_0(r). \tag{17}$$

$J_0(r)$ is the Bessel function of the first kind, of order zero. The cylinder dimensions are normalized such that its height is π and its radius is k_{01}. The coefficient 2 appearing in (17) is used for convenience of notation. The linear boundary-value problem

$$\nabla^2 y + 2p^2 y = 0, \quad y(\Gamma) = 0, \tag{18}$$

has a positive solution only for $p^2 = 1$ [$y = J_0(r) \cos z$]. It can be shown as in Ref. 3 that a necessary condition for the existence of a positive solution of (17) is that the average value

$$\overline{p^2(r)} > 1. \tag{19}$$

The greater $\overline{p^2(r)}$ is, the more the solution of (17) departs from that of (18). Analytically, the solution of the eigenvalue problem (4) corresponding to (17) is obtained by separation of variables, and is not more difficult in this two-dimensional case than in the one-dimensional problem treated before. Numerically, the eigenvalues are given by a somewhat complicated transcendental equation involving Bessel functions, and the eigenfunctions are given as products of a radial part involving Bessel functions and an axial part involving circular functions. The numerical calculations are quite straightforward with the use of a computer and are entirely analogous to those for one-dimensional problems; they will not be described here.[10] It is, however, of some interest to point out that the ordering by magnitude of the first few eigenvalues is as follows[11]:

$$\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{13}, \lambda_{22}, \lambda_{23}, \cdots \tag{20}$$

(the first index is axial and the second is radial). This order is the same as for the homogeneous eigenvalue problem (4) where $p^2(r) = \text{const} > 2$ (i.e., the Helmholtz equation).

The algebraic system (9) was solved by the Newton–Raphson method in the one-, two-, three-, four-, and five-mode approximations, for the two cases:

$$p^2(r) = \begin{cases} 8, & 0 \le r \le k_{01}/10, \\ 1, & k_{01}/10 \le r \le k_{01}, \end{cases} \tag{21a}$$

$$p^2(r) = \begin{cases} 12, & 0 \le r \le k_{01}/10, \\ 1, & k_{01}/10 \le r \le k_{01}. \end{cases} \tag{21b}$$

The results for case (21a) are given in Figs. 2 and 3, where we have plotted respectively the radial distributions on the radial plane $z = 0$, and the axial

[9] F. B. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill Book Co., Inc., New York, 1956), p. 451.

[10] J. Canosa, IBM Scientific Center (Palo Alto) Rep. No. 320-3240, 1968.

[11] The qualitative properties of the eigenvalue spectrum are the same as for one-dimensional problems, described in Ref. 8.

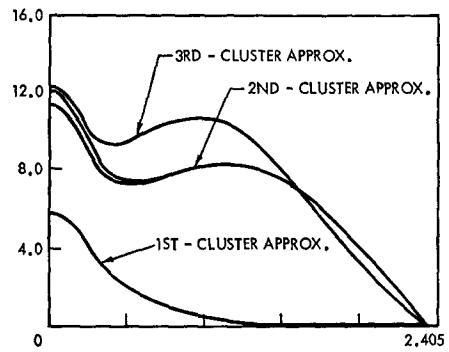


FIG. 4. First few mode approximations to the radial distribution in $z = 0$. Case (21b).

distributions on the axial plane $r = 0$. In Fig. 4, the radial distributions on the radial plane $z = 0$ are given for the case (21b). In both cases, the nonlinear effects are quite strong on the radial distributions and weak on the axial distributions. The nonlinear effects might be defined as the departure of the positive distribution from the fundamental eigenfunction of (4). The reason for the predominance of nonlinear effects in the radial direction is clear, because the problem is homogeneous axially, and composite radially. In one-dimensional problems if a j-mode approximation gives a distribution that is close to that given by the $(j-1)$-mode approximation, one is in general confident that the j-mode approximation is adequate. At least in the two-dimensional problems considered here, the different modal approximations cluster themselves according to the radial part of the eigenfunctions, that is, the accuracy of the successive approximations depends on how many of the following clusters of eigenfunctions are used:

$$\begin{aligned} &f_1(r)\cos z,\} \text{ first cluster,}\\ &\left.\begin{matrix} f_2(r)\cos z,\\ f_1(r)\cos 3z,\end{matrix}\right\} \text{ second cluster,}\\ &\left.\begin{matrix} f_3(r)\cos z,\\ f_2(r)\cos 3z,\\ f_3(r)\cos 3z,\end{matrix}\right\} \text{ third cluster,} \qquad (22)\\ &\quad\cdot\\ &\quad\cdot\\ &\quad\cdot \end{aligned}$$

When the order of a modal approximation is increased so as to include more than the first eigenfunction in the cluster, Figs. 2 and 4 show clearly that the accuracy of the solution does not increase appreciably. It is felt that the convergence of a certain modal approximation to the exact solution is assured if the distribution obtained does not vary significantly from a cluster to the following.

Figure 4 shows clearly that for case (21b) there is an appreciable difference from the second to the third cluster approximations. If a more accurate distribution were desired, it would be necessary to go to the next cluster of approximations involving the fourth radial eigenfunction $f_4(r)$. This behavior might be different in problems with strong nonlinear effects both in the radial and axial directions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Variational Solutions of Nonlinear Poisson-Boltzmann Boundary-Value Problems

N. ANDERSON AND A. M. ARTHURS
Department of Mathematics, University of York, Heslington, York, England

(Received 10 April 1968)

Variational solutions of one-dimensional nonlinear Poisson–Boltzmann boundary-value problems in the theory of colloids and plasmas are obtained. The accuracy of the solutions is measured in terms of upper and lower bounds for the field energy which result from complementary variational principles.

1. INTRODUCTION

The problem of solving the nonlinear Poisson–Boltzmann equation

$$d^2\varphi/dx^2 = f(\varphi), \quad 0 \leq x \leq L, \tag{1}$$

where

$$f(\varphi) = e^{\varphi} - e^{-\varphi}, \tag{2}$$

with boundary conditions

$$\varphi(0) = \varphi_1 \tag{3}$$

and

$$\varphi(L) = \varphi_2, \tag{4}$$

arises in the theory of colloids[1,2] ($L = \infty$, $\varphi_1 > 0$, $\varphi_2 = 0$) and in the theory of plasmas[3] (L finite, $\varphi_1 = 0$, $\varphi_2 > 0$). Here φ is the dimensionless quantity eV/kT, where V is the electric potential, e the proton charge, k Boltzmann's constant, T the particle temperature, and x is measured in units of the Debye length $\lambda_D = (kT/4\pi Ne^2)^{\frac{1}{2}}$, N being the particle density. When the function φ is much smaller than unity throughout the interval $(0, L)$, as is the case for sufficiently small V or high T, the problem is approximately linear with an elementary solution. However, in practice φ can be of the order of unity and the linear approach breaks down.

In this paper we present a variational approach to the problem. It is based on complementary variational principles[4,5] and leads to very accurate solutions for the potential.

2. COMPLEMENTARY VARIATIONAL PRINCIPLES

Using recent results[5] for boundary-value problems of the kind in Eqs. (1)–(4), we obtain the complementary variational principles

$$G(\Phi) \leq I(\varphi, u) \leq J(U), \tag{5}$$

where Φ and U are trial functions close to the exact solutions φ and u of (1) written as

$$d\varphi/dx = u, \quad -du/dx = -f(\varphi). \tag{6}$$

In (5),

$$G(\Phi) = \int_0^L \left\{ -\frac{1}{2}\left(\frac{d\Phi}{dx}\right)^2 - F(\Phi) \right\} dx,$$

$$\Phi(0) = \varphi_1, \quad \Phi(L) = \varphi_2, \tag{7}$$

$$I(\varphi, u) = \int_0^L \left\{ -\frac{1}{2}\left(\frac{d\varphi}{dx}\right)^2 - F(\varphi) \right\} dx, \tag{8}$$

$$J(U) = \int_0^L \left\{ \frac{1}{2}U^2 - F\left[f^{-1}\left(\frac{dU}{dx}\right)\right] + \left(\frac{dU}{dx}\right) f^{-1}\left(\frac{dU}{dx}\right) \right\} dx - [\varphi U]_{x=0}^{x=L}, \tag{9}$$

with

$$F(\varphi) = \int^{\varphi} f(\varphi')\, d\varphi'. \tag{10}$$

The principles (5) hold subject to the restriction

$$df/d\varphi \geq 0, \tag{11}$$

which is certainly satisfied by the function $f(\varphi)$ in (2).

3. COLLOID PROBLEM

Consider the boundary-value problem

$$d^2\varphi/dx^2 = f(\varphi) = e^{\varphi} - e^{-\varphi}, \quad x \geq 0, \tag{12}$$

where

$$\varphi(0) = A, \quad \lim_{L\to\infty} \varphi(L) = 0, \tag{13}$$

which arises in the Debye–Hückel theory of colloids.[1] When φ is much smaller than unity the solution is[2]

$$\varphi \sim A \exp(-x\sqrt{2}), \quad A \ll 1. \tag{14}$$

This suggests that for $A \sim 1$ we choose the trial functions

$$\Phi = Ae^{-\lambda x}, \tag{15}$$

[1] A. E. Alexander and P. Johnson, *Colloid Science* (Oxford University Press, London, 1949).
[2] R. P. Feynman, *The Feynman Lectures on Physics, Vol. II* (Addison-Wesley Publishing Co., Reading, Mass., 1964).
[3] F. Llewellyn-Jones, *The Glow-Discharge* (Methuen & Co., London, 1966).
[4] B. Noble, Univ. Wisconsin Math. Res. Center Report 473, 1964.
[5] A. M. Arthurs and P. D. Robinson, Proc. Cambridge Phil. Soc. (to be published 1969).

and

$$U = d(Ae^{-\mu x})/dx, \tag{16}$$

where the parameters λ and μ are found from the stationary conditions

$$\partial G/\partial\lambda = 0, \quad \partial J/\partial\mu = 0. \tag{17}$$

Taking $L = 10$, which is sufficiently large, and $A = 1$, we obtain

$$\lambda = 1.45 \quad \text{and} \quad \mu = 1.48, \tag{18}$$

and the corresponding functionals

$$G = -20.7222 \quad \text{and} \quad J = -20.7213. \tag{19}$$

The results in (19) provide upper and lower bounds for $I(\varphi, u)$ in (8), which for these one-dimensional problems is a measure of the field energy. Thus in terms of the field energy, the trial function

$$\Phi = e^{-1.45x} \tag{20}$$

is very good and may be taken to represent the potential in this problem.

As A decreases from unity it is clear that the parameter λ in (15) decreases from 1.45 and tends to $\sqrt{2}$.

4. PLASMA PROBLEM

Next we consider the problem of solving

$$d^2\varphi/dx^2 = f(\varphi) = e^{\varphi} - e^{-\varphi}, \quad 0 \leq x \leq L, \tag{21}$$

subject to

$$\varphi(0) = 0, \quad \varphi(L) = B, \tag{22}$$

which arises in plasma theory.[3] The form of the exact solution near $x = L$ suggests that we take the trial function

$$\Phi = \frac{Be^{-\alpha(L-x)}(1 - e^{-\alpha x})}{(1 - e^{-\alpha L})} \tag{23}$$

in (7), and the trial function

$$U = \frac{d}{dx}\left\{\frac{Be^{-\beta(L-x)}(1 - e^{-\beta x})}{(1 - e^{-\beta L})}\right\} \tag{24}$$

in (9), where the parameters α and β are found from the stationary conditions

$$\partial G/\partial\alpha = 0, \quad \partial J/\partial\beta = 0. \tag{25}$$

Calculations have been performed for $B = 1$ and $B = 2$ and $L = 1$ (1.0) 5, and the results are given in Tables I and II. Since the accuracy of the trial functions is judged by the closeness of the bounds G and J, we see that the function (23) improves as L increases and B decreases, and provides an accurate representation of the potential in this problem.

TABLE I. Parameters for $B = 1$.

L	α	$G(\alpha)$	β	$J(\beta)$	$J - G$
1	0.93	−2.8097	1.06	−2.7785	0.0312
2	1.31	−4.7278	1.42	−4.7226	0.0052
3	1.39	−6.7227	1.46	−6.7212	0.0015
4	1.43	−8.7222	1.48	−8.7211	0.0011
5	1.44	−10.7222	1.48	−10.7213	0.0009

TABLE II. Parameters for $B = 2$.

L	α	$G(\alpha)$	β	$J(\beta)$	$J - G$
1	1.05	−5.4003	1.23	−5.2148	0.1855
2	1.43	−7.0993	1.61	−7.0467	0.0526
3	1.51	−9.0792	1.65	−9.0470	0.0322
4	1.53	−11.0771	1.66	−11.0476	0.0295
5	1.54	−13.0767	1.67	−13.0479	0.0288

5. CONCLUDING REMARKS

Solutions of two kinds of one-dimensional nonlinear Poisson–Boltzmann problems have been found using complementary variational principles. These principles provided upper and lower bounds for a certain functional, and the accuracy of the variational solutions was measured by the closeness of the bounds.

For similar problems in higher dimensions the method[5] adopted here still applies. The main difficulty of the method in the nonlinear case is to find trial functions which are not too far removed from the exact ones.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Reduction of Reducible Representations of the Poincaré Group to Standard Helicity Representations

HARRY E. MOSES
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.*

(Received 28 October 1967; Revised Manuscript Received 26 January 1968)

In this paper we introduce realizations of the generators of the Poincaré group for real and imaginary masses which are close in form to the Lomont–Moses realizations for zero mass. These realizations (which we call "standard helicity realizations or representations") are characterized by the way that the infinitesimal generators are given in terms of the helicity operator. We also give the global form of the realizations and discuss in detail the realizations for the case that they are unitary and irreducible. We then show how any reducible representation of the Poincaré group for which the infinitesimal generators of the translation and rotation subgroups are Hermitian and integrable and for which the space–time generators are integrable (but not necessarily Hermitian) can be reduced to the standard helicity realizations. In the case that the reducible representation is unitary, this process enables one to reduce the reducible representation to irreducible unitary standard helicity representations. Finally, we show how the Foldy–Shirokov realizations for real mass are related to the standard helicity representation.

1. INTRODUCTION

Lomont and this author[1] showed how reducible unitary representations of the Poincaré group could be reduced to irreducible unitary representations when it was assumed that only real nonzero-mass and zero-mass representations were contained in the reducible representations. The irreducible representations were in the form of the Foldy–Shirokov realization for nonzero mass[2,3] and in the Lomont–Moses realization for zero mass.[4] In extending the reduction procedure to cases in which the reducible representations contained imaginary-mass components, the present writer found it convenient to introduce realizations for imaginary mass which are analogous to those of Ref. 4 for zero mass. In addition, for certain applications which we shall give in later papers, it was found useful to round out these realizations by giving the analogous realization for real nonzero mass.

These realizations for the real and imaginary masses formally are very similar. In fact the infinitesimal generators corresponding to translation and rotation are identical, while the space–time infinitesimal generators are very close in form. These realizations are characterized by the prominence enjoyed by the helicity operator, especially in the infinitesimal generators for rotation. The infinitesimal generators will be given explicitly. We shall call these realizations "standard helicity realizations or representations." We shall also give the global form for the irreducible unitary representations of the Poincaré group in the helicity realization.

We then go back to our original objective and show how reducible representations of the Poincaré group can be reduced to the standard helicity realizations. Any mass, including the imaginary mass, can be assumed to be contained in the reducible representation. The only representations which are assumed absent from the reducible representation are those of the homogeneous Lorentz group, for which the infinitesimal generators corresponding to energy and linear-momentum components are identically zero. Our assumptions about the reducible representation are that the infinitesimal generators corresponding to the energy and components of the linear momentum and those corresponding to the components of the angular momentum are Hermitian and integrable, while the infinitesimal generators corresponding to space–time transformations are integrable but not necessarily Hermitian. Of course, reducible unitary representations are included among these.

In later papers we shall use the technique given herein to obtain the Clebsch–Gordan coefficients of the direct products of the representations. We shall also use this technique to expand relativistic wavefunctions into modes which transform like the standard helicity representations for all masses, including imaginary masses. Indeed, the author has already given such an expansion when mass-zero modes only are assumed to be contained in the expansion of the wavefunction.[5]

In the present paper we show how the Foldy–Shirokov realizations for real nonzero mass are related to the standard helicity realization of the present paper.

* Operated with support from the U.S. Advanced Research Projects Agency.

[1] J. S. Lomont and H. E. Moses, J. Math. Phys. **8**, 837 (1967).

[2] L. L. Foldy, Phys. Rev. **102**, 568 (1956).

[3] Yu. M. Shirkov, Zh. Eksp. Teor. Fiz. **33**, 1196 (1957) [Sov. Phys.—JETP **13**, 240 (1961)].

[4] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962).

[5] H. E. Moses, J. Math. Phys. **9**, 16 (1968).

For the sake of brevity we do not prove the reduction algorithm in the present paper. However, the proof is close to that used in Ref. 1 and we believe that the reader can reconstruct the proof needed in the present paper from that reference.

2. STANDARD HELICITY REALIZATIONS FOR ANY MASS

A. Commutation Rules for the Infinitesimal Generators: The Casimir Operator Associated with Mass

The infinitesimal generators of the Poincaré group are the energy operator $H = P^0 = -P_0$, the operators corresponding to the components of the linear momentum $P_i = P^i$ $(i = 1, 2, 3)$, the operators corresponding to the components of the angular momentum J_i, and the space–time infinitesimal generators $\mathfrak{J}_i$. These operators satisfy the commutation rules:

$$\begin{aligned}
[H, P_i] &= 0, \\
[H, J_i] &= 0, \\
[P_i, P_j] &= 0, \\
[J_i, J_j] &= i \sum_k \epsilon_{ijk} J_k, \\
[J_i, P_j] &= i \sum_k \epsilon_{ijk} P_k, \\
[J_i, \mathfrak{J}_j] &= i \sum_k \epsilon_{ijk} \mathfrak{J}_k, \\
[\mathfrak{J}_i, \mathfrak{J}_j] &= -i \sum_k \epsilon_{ijk} J_k, \\
[\mathfrak{J}_i, H] &= i P_i, \\
[\mathfrak{J}_i, P_j] &= i \delta_{ij} H.
\end{aligned} \tag{2.1}$$

In (2.1), ϵ_{ijk} is the usual antisymmetric three-index symbol. As is well known, the Casimir operator C defined by

$$C = H^2 - \sum_i P_i^2 \tag{2.2}$$

commutes with all of the infinitesimal generators. Our realizations are confined to operators C which are real scalars c. Later, when we consider reducible representations in general, C will be a Hermitian operator with eigenvalues c. The realizations which we give in the present section will then be in a subspace corresponding to the eigenvalue c.

We write

$$\begin{aligned}
c &= m^2, \quad \text{for} \quad c \geq 0, \\
&= -\kappa^2, \quad \text{for} \quad c < 0.
\end{aligned} \tag{2.3}$$

We take both m and κ to be nonnegative in (2.3). When $c \geq 0$, the realizations correspond to the realizations for which the mass is m. When $c < 0$, the realizations correspond to the imaginary mass $i\kappa$.

B. Commutation Rules for the Infinitesimal Generators of the Little Groups

It is now necessary to introduce the infinitesimal generators of three groups which are associated with the three cases that the mass be nonzero and real, that the mass be imaginary, and that the mass be zero, i.e., that $c > 0$, $c < 0$, and $c = 0$, respectively. These three groups correspond to the little groups used by Wigner[6] in his original work in obtaining and classifying the irreducible unitary ray representations of the Poincaré group. For each of these three groups, there are three infinitesimal generators which we denote by T_1, T_2, and M, respectively. They satisfy the commutation rules

$$\begin{aligned}
[T_1, M] &= -iT_2, \\
[T_2, M] &= iT_1, \\
[T_1, T_2] &= iA(c)M,
\end{aligned} \tag{2.4}$$

where $A(c)$ is the following function of c:

$$\begin{aligned}
A(0) &= 0, \\
A(c) &= c/|c| \quad \text{for} \quad c \neq 0.
\end{aligned} \tag{2.4'}$$

It is seen that, for nonzero real-mass realizations, the generators of the little group satisfy the commutation rules of the generators of the rotation group. That is, letting $S_1 = T_1$, $S_2 = T_2$, $S_3 = M$, the operators S_i satisfy $[S_1, S_2] = iS_3$ (cyc). For zero-mass representations, the generators satisfy the commutation rules for the Euclidean group in the plane, while for imaginary-mass representations the generators satisfy the commutation rules for the rotation group in pseudo-Euclidean space. The first two groups are, of course, well known. The third group has been extensively studied by Bargmann.[7] In Sec. 2D we give the irreducible Hermitian representations of the generators of the three groups and also give the corresponding irreducible unitary ray representations of the groups, obtained from the integration of the infinitesimal generators in a suitable parameter space.

Realizations of the operators T_i and M will be described in terms of a real variable λ. The variable λ may be discrete or continuous or both, depending on the particular realization which is used. The range of λ also depends on the realization. In fact, λ may consist of a set of real variables. For a given realization, we introduce for the operators T_i and M, respectively, the kernels of which we denote by $T_i(\lambda \mid \lambda')$ and $M(\lambda \mid \lambda')$. For values of λ which are discrete, these kernels are matrices. For continuous ranges of the variable they may be symbolic functions, i.e.,

[6] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[7] V. Bargmann, Ann. Math. **48**, 568 (1947).

distributions. These kernels or matrices are required to satisfy the commutation rules

$$\begin{aligned}
\sum_{\lambda''}[T_1(\lambda\,|\,\lambda'')M(\lambda''\,|\,\lambda') - M(\lambda\,|\,\lambda'')T_1(\lambda''\,|\,\lambda')] &= -iT_2(\lambda\,|\,\lambda'),\\
\sum_{\lambda''}[T_2(\lambda\,|\,\lambda'')M(\lambda''\,|\,\lambda') - M(\lambda\,|\,\lambda'')T_2(\lambda''\,|\,\lambda')] &= iT_1(\lambda\,|\,\lambda'),\\
\sum_{\lambda''}[T_1(\lambda\,|\,\lambda'')T_2(\lambda''\,|\,\lambda') - T_2(\lambda\,|\,\lambda'')T_1(\lambda''\,|\,\lambda')] &= iA(c)M(\lambda\,|\,\lambda').
\end{aligned} \tag{2.4''}$$

In (2.4'') we have assumed that λ is a discrete variable. In the regions for which λ'' is continuous, the summation is to be replaced by integration.

C. Standard Helicity Representations

Each representation is characterized by a value of c and a realization of the infinitesimal generators of the little group $T_i(\lambda\,|\,\lambda')$, $M(\lambda\,|\,\lambda')$ which satisfy (2.4''). We introduce a space of suitably differentiable complex functions $\{f(c, \epsilon, \mathbf{p}, \lambda)\}$, where ϵ is the sign of the energy and can take on either of the values $+1$ or -1, $\mathbf{p}$ denotes collectively three real variables p_1, p_2, p_3, each of which can take on all real values, the variable λ has the same range and character as the variable λ associated with the realization of the little group. We also require that, for the case $c < 0$, $f(c, \epsilon, \mathbf{p}, \lambda)$ vanishes when $p = |\mathbf{p}| < \kappa$.

In writing down the standard helicity realizations, it is convenient for brevity to suppress the appearance of the variable λ. Hence we write $f(c, \epsilon, \mathbf{p})$ for $f(c, \epsilon, \mathbf{p}, \lambda)$. Furthermore, $T_if(c, \epsilon, \mathbf{p})$ and $Mf(c, \epsilon, \mathbf{p})$ are used to denote

$$\sum_{\lambda'} T_i(\lambda\,|\,\lambda')f(c, \epsilon, \mathbf{p}, \lambda') \quad \text{and} \quad \sum_{\lambda'} M(\lambda\,|\,\lambda')f(c, \epsilon, \mathbf{p}, \lambda'),$$

respectively.

Let us also use the following definitions:

$$\begin{aligned}
p &= |\mathbf{p}|,\\
\omega(c, p) &= [p^2 + c]^{\frac{1}{2}},\\
\nabla_i &= \partial/\partial p_i,\\
\mathbf{p}\cdot\mathbf{T} &= p_1T_1 + P_2T_2,\\
B(c) &= [|c|]^{\frac{1}{2}} \quad \text{for} \quad c \neq 0,\\
B(0) &= 1.
\end{aligned} \tag{2.5}$$

We can now write the standard helicity realizations in terms of the infinitesimal generators acting on the functions $f(c, \epsilon, \mathbf{p})$:

$$\begin{aligned}
P_if(c, \epsilon, \mathbf{p}) &= p_if(c, \epsilon, \mathbf{p}),\\
Hf(c, \epsilon, \mathbf{p}) &= \epsilon\omega(c, p)f(c, \epsilon, \mathbf{p}),\\
J_1f(c, \epsilon, \mathbf{p}) &= -i(\mathbf{p}\times\nabla)_1f(c, \epsilon, \mathbf{p}) + \frac{p_1}{p + p_3}Mf(c, \epsilon, \mathbf{p}),\\
J_2f(c, \epsilon, \mathbf{p}) &= -i(\mathbf{p}\times\nabla)_2f(c, \epsilon, \mathbf{p}) + \frac{p_2}{p + p_3}Mf(c, \epsilon, \mathbf{p}),\\
J_3f(c, \epsilon, \mathbf{p}) &= -i(\mathbf{p}\times\nabla)_3f(c, \epsilon, \mathbf{p}) + Mf(c, \epsilon, \mathbf{p}),\\
\mathfrak{J}_1f(c, \epsilon, \mathbf{p}) &= \epsilon\left\{i\omega(c, p)\nabla_1 + \frac{p_2}{p(p + p_3)}\omega(c, p)M + \frac{B(c)}{p^2}\left[\frac{p_1(\mathbf{p}\cdot\mathbf{T})}{p + p_3} - pT_1\right]\right\}f(c, \epsilon, \mathbf{p}),\\
\mathfrak{J}_2f(c, \epsilon, p) &= \epsilon\left\{i\omega(c, p)\nabla_2 - \frac{p_1}{p(p + p_3)}\omega(c, p)M + \frac{B(c)}{p^2}\left[\frac{p_2(\mathbf{p}\cdot\mathbf{T})}{p + p_3} - pT_2\right]\right\}f(c, \epsilon, \mathbf{p}),\\
\mathfrak{J}_3f(c, \epsilon, \mathbf{p}) &= \epsilon\left\{i\omega(c, p)\nabla_3 + \frac{B(c)}{p^2}(\mathbf{p}\cdot\mathbf{T})\right\}f(c, \epsilon, \mathbf{p}).
\end{aligned} \tag{2.6}$$

That the generators given by (2.6) satisfy the commutation rules (2.1) can be verified by direct computation using the commutation rules (2.4'') for the generators of the little group.

The mass-zero case has already been given in Ref. 4. It should be mentioned that in Ref. 8 realizations are given such that P_i, H, and J_i have the same form as (2.6). However, the expressions for $\mathfrak{J}_i$ of Ref. 8 differ from ours, which we believe to be simpler. Our realizations for the nonzero-mass cases differ considerably in appearance from the helicity realizations of Refs. 9 and 10.

The operator M is essentially the helicity operator. Let us define w^0 by

$$w^0 = \mathbf{P}\cdot\mathbf{J} = \sum_i P_iJ_i. \tag{2.7}$$

Then from (2.6) it is clear that

$$w^0f(c, \epsilon, \mathbf{p}) = pMf(c, \epsilon, \mathbf{p}). \tag{2.8}$$

Equation (2.8) is the principal reason for calling the representations (2.6) helicity representations.

Our space of functions $\{f(c, \epsilon, \mathbf{p})\}$ is not yet a Hilbert space, for we have not introduced an inner product. We pick out one of several possible inner products as being particularly useful in describing irreducible Hermitian representations of the infinitesimal generators. Then the "standard" inner product of two functions $f^{(1)}(c, \epsilon, \mathbf{p})$ and $f(c, \epsilon, \mathbf{p})$ is defined to be

$$(f^{(1)}, f) = \sum_\lambda \int \frac{d\mathbf{p}}{\omega(c, p)} f^{(1)*}(c, \epsilon, \mathbf{p}, \lambda)f(c, \epsilon, \mathbf{p}, \lambda). \tag{2.9}$$

[8] D. Korff, J. Math. Phys. **5**, 869 (1964).
[9] A. Chakrabarti, J. Math. Phys. **7**, 949 (1966).
[10] V. I. Ritus, Zh. Eksp. Teor. Fiz. **40**, 352 (1961) [Sov. Phys.—JETP **13**, 240 (1961)].

The inner product (2.9) assures us that the part of the infinitesimal generators which are independent of T_i and M are Hermitian. We shall later see that, from a global point of view, we require a summation over ϵ also for the case $c < 0$.

We call the matrices of the realizations of the infinitesimal generators of the little groups Hermitian if

$$M(\lambda \mid \lambda') = M^*(\lambda' \mid \lambda),$$
$$T_i(\lambda \mid \lambda') = T_i^*(\lambda' \mid \lambda). \tag{2.10}$$

The following theorem is then easy to prove:

Theorem: A necessary and sufficient condition that the realizations (2.6) with the inner product (2.9) be irreducible and Hermitian is that the kernels $M(\lambda \mid \lambda')$, $T_i(\lambda \mid \lambda')$ constitute an irreducible Hermitian representation of operators which satisfy (2.4).

The irreducible representations of the Poincaré group obtained from the above theorem are labeled by the Casimir operator c, the sign of energy ϵ, and the labels necessary to distinguish the irreducible representations of T_i and M, the infinitesimal generators of the little group.

When we integrate the irreducible infinitesimal generators, as we shall do shortly, we shall obtain the irreducible unitary ray representations of the Poincaré group in the "standard helicity realization." While the labels that are used to distinguish the irreducible representations of the infinitesimal generators can also be used for the integrated representations in the case of zero mass and real nonzero mass, the quantum number ϵ cannot be used as a label in the imaginary-mass case. That is, for the imaginary-mass case, ϵ is an invariant under infinitesimal transformations, but not under finite transformations.

D. Irreducible Representations of the Infinitesimal Generators of the Little Group: Integration of the Infinitesimal Generators of the Little Group

For use in describing the irreducible Hermitian representations (2.6) we give the irreducible Hermitian representations of the generators of the little groups for all c. We also give certain integrated forms of the generators which are used in the integrated (i.e., global) form of the helicity representation of the Poincaré group. In the realizations which we give, the kernel $M(\lambda \mid \lambda')$ is diagonal:

$$M(\lambda \mid \lambda') = \lambda \delta_{\lambda\lambda'}. \tag{2.11}$$

The requirement that the irreducible Hermitian infinitesimal generators (2.6) be integrable leads to the requirement that λ have either only integer values or only half-odd integer values (see the discussion in Sec. 4 of Ref. 1).

The integrated form of the generators which we require are $e^{i\alpha M}$ and $e^{i\boldsymbol{\theta}\cdot\mathbf{T}}$, where $\boldsymbol{\theta}\cdot\mathbf{T} = \theta_1 T_1 + \theta_2 T_2$. The quantities α, θ_1, θ_2 are any real numbers. It is convenient to define φ and θ by

$$\theta = [\theta_1^2 + \theta_2^2]^{\frac{1}{2}},$$
$$\theta_1 = \theta \cos\varphi, \tag{2.12}$$
$$\theta_2 = \theta \sin\varphi.$$

The matrix elements of $\exp i\alpha M$ are denoted by $\exp(i\alpha M)(\lambda \mid \lambda')$ and the matrix elements of

$$\exp(i\boldsymbol{\theta}\cdot\mathbf{T})$$

by $\exp(i\boldsymbol{\theta}\cdot\mathbf{T})(\lambda \mid \lambda')$. From (2.11) we have for all representations

$$e^{i\alpha M}(\lambda \mid \lambda') = e^{i\alpha\lambda}\delta_{\lambda\lambda'}. \tag{2.13}$$

We now give the irreducible representations of the infinitesimal generators explicitly.

1. Case $c > 0$

The irreducible representations correspond to particles of real nonzero mass $m = c^{\frac{1}{2}}$. The generators T_1, T_2, M are just the infinitesimal generators of the rotation group whose representations are discussed in many places (see, e.g., Ref. 11). Each irreducible representation is characterized by a nonnegative integer or half-odd integer which corresponds to the spin of the particle. The Casimir operator of the little group is $T_1^2 + T_2^2 + M^2 = s(s+1)I$, where I is the identity operator. The variable λ takes on the values $-s, -s+1, \cdots, s-1, s$. Let us define T^+ and T^- in the usual way:

$$T^+ = T_1 + iT_2,$$
$$T^- = T_1 - iT_2. \tag{2.14}$$

Denoting the corresponding matrices by $T^+(\lambda \mid \lambda')$ and $T^-(\lambda \mid \lambda')$ we have

$$T^+(\lambda \mid \lambda') = [(s-\lambda')(s+\lambda'+1)]^{\frac{1}{2}}\delta_{\lambda,\lambda'+1},$$
$$T^-(\lambda \mid \lambda') = [(s+\lambda')(s-\lambda'+1)]^{\frac{1}{2}}\delta_{\lambda,\lambda'-1}. \tag{2.14'}$$

Now let $P_n^{(\alpha,\beta)}(x)$ be the Jacobi polynomial in the notation of Ref. 12. Let us define

$$S(s,\lambda,\lambda',x) = P_{s-\lambda}^{(\lambda-\lambda',\lambda+\lambda')}(x),$$
$$Y_s^{\lambda,\lambda'}(\theta,\varphi) = (-1)^{\lambda-\lambda'}(\tfrac{1}{2})^{\lambda+1}[(2s+1)/\pi]^{\frac{1}{2}}\left[\frac{(s-\lambda)!\,(s+\lambda)!}{(s-\lambda')!\,(s+\lambda')!}\right]^{\frac{1}{2}}$$
$$\times e^{i(\lambda-\lambda')\varphi}[\sin\theta]^{\lambda-\lambda'}[1+\cos\theta]^{\lambda'}S(s,\lambda,\lambda',\cos\theta). \tag{2.15}$$

[11] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957).

[12] G. Szego, "Orthogonal Polynomials" in *American Mathematical Society Colloquium Publication, Vol. 23* (American Mathematical Society, Providence, Rhode Island, 1959), revised ed.

The functions $Y_s^{\lambda,\lambda'}(\theta, \varphi)$ have been called by the author "generalized surface harmonics" and their properties are discussed in Ref. 13. Then

$$e^{i\theta\cdot\mathbf{T}}(\lambda \mid \lambda') = [4\pi/(2s+1)]^{\frac{1}{2}}(i)^{\lambda'-\lambda}Y_s^{\lambda,\lambda'*}(\theta, \varphi). \quad (2.16)$$

2. Case $c = 0$

For $c = 0$ the infinitesimal generators T_1, T_2, M satisfy the commutation rules for the infinitesimal generators for the Euclidean group in the plane. This group is discussed in detail in Ref. 14. The Casimir operator which characterizes the irreducible representations is $T_1^2 + T_2^2 = rI$, where r is any nonnegative real number. The irreducible representations of the little group fall into two types.

(a) $r = 0$: In this case M is a scalar, which, in accordance with our earlier discussion, must be either an integer or half-odd integer. The matrices corresponding to T_1, T_2 are identically zero. This representation of the little group yields zero-mass representations of the Poincaré group which are identified with particles of finite spin which is equal to the absolute value of M.

(b) $r > 0$: For such representations, the variable λ takes on either all integer values or all half-odd integer values. Defining $T^{\pm}$ as in (2.14) we have

$$\begin{aligned} T^{+}(\lambda \mid \lambda') &= r\delta_{\lambda,\lambda'+1}, \\ T^{-}(\lambda \mid \lambda') &= r\delta_{\lambda,\lambda'-1}, \\ e^{i\theta\cdot\mathbf{T}}(\lambda \mid \lambda') &= (i)^{\lambda-\lambda'}e^{-i(\lambda-\lambda')\varphi}J_{\lambda-\lambda'}(\theta r). \end{aligned} \quad (2.17)$$

In (2.17), $J_n(x)$ is the notation for the Bessel function of nth order as customarily used.

3. Case $c < 0$

The little group is that for rotations in pseudo-Euclidean space. Its representations have been completely investigated in Ref. 7. The notation of Ref. 7 is modified somewhat to conform to the notation of the present paper. The Casimir operator is $T_1^2 + T_2^2 - M^2 = qI$. The trivial representation $T_i = M = 0$ is one of the acceptable representations. The nontrivial irreducible representations fall into two classes: (a) the continuous class, and (b) the discrete class. Each class is further subdivided in two cases.

(*a*) *The Continuous Class:* (1) Integral case. In this case $q > 0$, λ takes on all positive and negative integer values including zero. This case is labeled C_q^0 by Bargmann:

$$\begin{aligned} T^{+}(\lambda \mid \lambda') &= [q + \lambda'(\lambda'+1)]^{\frac{1}{2}}\delta_{\lambda,\lambda'+1}, \\ T^{-}(\lambda \mid \lambda') &= [q + \lambda'(\lambda'-1)]^{\frac{1}{2}}\delta_{\lambda,\lambda'-1}, \\ e^{i\theta\cdot\mathbf{T}}(\lambda \mid \lambda') &= \frac{1}{(\lambda-\lambda')!}\prod_{p=1}^{\lambda-\lambda'}[q + (\lambda'+p)(\lambda'+p-1)]^{\frac{1}{2}}(\tfrac{1}{2})^{\lambda} \\ &\quad \times [ie^{-i\varphi}\sinh\theta]^{\lambda-\lambda'}[1+\cosh\theta]^{\lambda'} \\ &\quad \times F(\tfrac{1}{2}+\lambda+\sigma, \tfrac{1}{2}+\lambda-\sigma, 1+\lambda-\lambda', -\sinh^2\tfrac{1}{2}\theta), \\ &\qquad\qquad \text{for} \quad \lambda \geq \lambda', \\ &= \frac{1}{(\lambda'-\lambda)!}\prod_{p=1}^{\lambda'-\lambda}[q + (\lambda+p)(\lambda+p-1)]^{\frac{1}{2}}(\tfrac{1}{2})^{\lambda'} \\ &\quad \times [ie^{i\varphi}\sinh\theta]^{\lambda'-\lambda}[1+\cosh\theta]^{\lambda} \\ &\quad \times F(\tfrac{1}{2}+\lambda'+\sigma, \tfrac{1}{2}+\lambda'-\sigma, 1+\lambda'-\lambda, -\sinh^2\tfrac{1}{2}\theta), \\ &\qquad\qquad \text{for} \quad \lambda \leq \lambda'. \end{aligned} \quad (2.18)$$

In Eq. (2.18), $\sigma = [q - \frac{1}{4}]^{\frac{1}{2}}$. Furthermore, $F(a, b, c, x)$ is the hypergeometric function. Because of the symmetry of the hypergeometric function in a and b, one can take either sign of the square root in the definition of σ.

(2) Half-odd integral case. In this case $q > \frac{1}{4}$, and λ takes on all positive and negative half-odd integer values. Equations (2.18) continue to hold. This case is labeled $C_q^{\frac{1}{2}}$ by Bargmann.

(*b*) *Discrete Class:* (1) Maximal-λ case. Let k be any nonnegative integer or half-odd integer. Then $\lambda = -k, -(k+1), -(k+2), \cdots$. For this case $q = k(1-k)$. This case is labeled D_k^{-}:

$$\begin{aligned} T^{+}(\lambda \mid \lambda') &= [(\lambda'+k)(\lambda'-k+1)]^{\frac{1}{2}}\delta_{\lambda,\lambda'+1}, \\ T^{-}(\lambda \mid \lambda') &= [(\lambda'-k)(\lambda'+k-1)]^{\frac{1}{2}}\delta_{\lambda,\lambda'-1}, \\ e^{i\theta\cdot\mathbf{T}}(\lambda \mid \lambda') &= \left[\frac{(k-\lambda'-1)!\,(-k-\lambda)!}{(k-\lambda-1)!\,(-k-\lambda')!}\right]^{\frac{1}{2}}\left[ie^{-i\varphi}\sinh\frac{\theta}{2}\right]^{\lambda-\lambda'} \\ &\quad \times \left[\cosh\frac{\theta}{2}\right]^{\lambda+\lambda'}S(-k, \lambda, \lambda', \cosh\theta), \quad \text{for} \quad \lambda \geq \lambda', \\ &= \left[\frac{(k-\lambda-1)!\,(-k-\lambda')!}{(k-\lambda'-1)!\,(-k-\lambda)!}\right]^{\frac{1}{2}}\left[ie^{i\varphi}\sinh\frac{\theta}{2}\right]^{\lambda'-\lambda} \\ &\quad \times \left[\cosh\frac{\theta}{2}\right]^{\lambda+\lambda'}S(-k, \lambda', \lambda, \cosh\theta), \quad \text{for} \quad \lambda \leq \lambda'. \end{aligned} \quad (2.19)$$

In (2.19) the function S is given in terms of the Jacobi polynomial by the first of Eqs. (2.15).

(2) Minimal-λ case: Let k be any nonnegative integer or half-odd integer. Then $\lambda = k, k+1, k+2, \cdots$. Also $q = k(1-k)$. This case is labeled D_k^{+}. The matrices $T^{\pm}(\lambda \mid \lambda')$ have the same form as in (2.19), though it should be remembered that the

[13] H. E. Moses, Ann. Phys. (N.Y.) **41**, 166 (1967).

[14] E. P. Wigner, *The Application of Group Theory to the Special Functions of Mathematical Physics*, Lecture Notes by J. D. Talman, Spring Term of 1955, Princeton University.

range of λ differs from that of (2.19). Also,

$$e^{i\theta\cdot T}(\lambda \mid \lambda') = (-1)^{\lambda'-k}\left[\frac{(k+\lambda-1)!\,(\lambda'-k)!}{(k+\lambda'-1)!\,(\lambda-k)!}\right]^{\frac{1}{2}}\left[ie^{-i\varphi}\sinh\frac{\theta}{2}\right]^{\lambda-\lambda'} \times\left[\cosh\frac{\theta}{2}\right]^{-(\lambda+\lambda')} S(-k,-\lambda',\lambda,-\cosh\theta), \quad \text{for } \lambda\geq\lambda',$$

$$= (-1)^{\lambda-k}\left[\frac{(k+\lambda'-1)!\,(\lambda-k)!}{(k+\lambda-1)!\,(\lambda'-k)!}\right]^{\frac{1}{2}}\left[ie^{i\varphi}\sinh\frac{\theta}{2}\right]^{\lambda'-\lambda} \times\left[\cosh\frac{\theta}{2}\right]^{-(\lambda+\lambda')} S(-k,-\lambda,\lambda',-\cosh\theta), \quad \text{for } \lambda\leq\lambda'. \tag{2.20}$$

E. Global Representations of the Poincaré Group in the Standard Helicity Realization

We are now in a position to integrate the infinitesimal generators (2.6) when they correspond to an irreducible representation of the Poincaré group. First, however, it will be useful to parametrize the Poincaré group. The transformations of the Poincaré group are obtained by considering the transformations of a four-vector x^μ with $x^0 = -x_0$, $x^i = x_i$ $(i = 1, 2, 3)$. It is convenient to consider the general transformation as being the product of three particular transformations which we now give.

The transformation

$$x^{\mu'} = x^\mu - a^\mu, \tag{2.21}$$

where a^μ are real numbers, is designated by $T(a^\mu)$. This transformation is called a translation given by a^μ.

The second transformation corresponds to a pure rotation which is described by a vector $\boldsymbol{\theta}$ where the axis of rotation is given by the direction of $\boldsymbol{\theta}$ and the angle of rotation is given by $\theta = |\boldsymbol{\theta}|$. Under this transformation,

$$x^{o'} = x^o,$$
$$\mathbf{x}' = \mathbf{x}\cos\theta + \frac{1-\cos\theta}{\theta^2}(\boldsymbol{\theta}\cdot\mathbf{x})\boldsymbol{\theta} - \frac{\sin\theta}{\theta}(\boldsymbol{\theta}\times\mathbf{x}), \tag{2.22}$$

where $\mathbf{x}$ is the vector constructed from x_i $(i = 1, 2, 3)$. This transformation is denoted by $R(\boldsymbol{\theta})$ and will be called a rotation given by the vector $\boldsymbol{\theta}$. (The vector $\boldsymbol{\theta}$ as used here is not to be confused with $\boldsymbol{\theta}$ in $\boldsymbol{\theta}\cdot\mathbf{T}$.)

The third transformation is a pure Lorentz transformation and is characterized by the vector $\boldsymbol{\beta}$. The direction of $\boldsymbol{\beta}$ is opposite to the direction of the moving frame as observed in the original frame and $\beta = |\boldsymbol{\beta}|$ is related to the relative velocity of the moving and fixed frames of reference by

$$\cosh\beta = [1 - v^2]^{-\frac{1}{2}}. \tag{2.23}$$

Under this transformation

$$x^{o'} = x^o\cosh\beta + \boldsymbol{\beta}\cdot\mathbf{x}(\sinh\beta)/\beta,$$
$$\mathbf{x}' = \mathbf{x} + \boldsymbol{\beta}(\boldsymbol{\beta}\cdot\mathbf{x})\frac{\cosh\beta - 1}{\beta^2} + \boldsymbol{\beta}x^o\frac{\sinh\beta}{\beta}. \tag{2.24}$$

This transformation is denoted by $L(\boldsymbol{\beta})$ and will be called the pure Lorentz transformation associated with $\boldsymbol{\beta}$. The transformations $T(a^\mu)$, $R(\boldsymbol{\theta})$, and $L(\boldsymbol{\beta})$ are discussed in somewhat more detail in Ref. 15.

In the global representation of the Poincaré group, the operators which correspond to the group elements $T(a^\mu)$, $R(\boldsymbol{\theta})$, and $L(\boldsymbol{\beta})$ are $\exp[ia^\mu P_\mu]$, $\exp[i\boldsymbol{\theta}\cdot\mathbf{J}]$, $\exp[i\boldsymbol{\beta}\cdot\mathfrak{J}]$, respectively. In terms of the space of functions $\{f(c, \epsilon, \mathbf{p})\}$, the first two of these operators are given as follows:

$$\exp[ia^\mu P_\mu]f(c,\epsilon,\mathbf{p}) = \exp\{i[\mathbf{a}\cdot\mathbf{p} - \epsilon\omega(c,p)a^0]\}f(c,\epsilon,\mathbf{p}),$$
$$\exp[i\boldsymbol{\theta}\cdot\mathbf{J}]f(c,\epsilon,\mathbf{p}) = \exp[2i\Phi(\boldsymbol{\theta},\mathbf{p})M]f(c,\epsilon,\mathbf{p}'), \tag{2.25}$$

where

$$\mathbf{p}' = \mathbf{p}\cos\theta + [(1-\cos\theta)/\theta^2](\boldsymbol{\theta}\cdot\mathbf{p})\boldsymbol{\theta} + [(\sin\theta)/\theta](\boldsymbol{\theta}\times\mathbf{p}), \tag{2.25'}$$

and $\Phi(\boldsymbol{\theta}, \mathbf{p})$ is given by

$$\tan\Phi(\boldsymbol{\theta},\mathbf{p}) = \frac{[\boldsymbol{\theta}\cdot\mathbf{p} + \theta_3 p]\tan(\theta/2)}{\theta(p+p_3) + (\boldsymbol{\theta}\times\mathbf{p})_3\tan(\theta/2)}. \tag{2.25''}$$

The form for $\exp[i\boldsymbol{\theta}\cdot\mathbf{J}]$ is essentially the same as for the massless case of finite spin given earlier in Ref. 15.

To find $\exp[i\boldsymbol{\beta}\cdot\mathfrak{J}]$ is rather more difficult. We first give $\exp[i\beta\mathfrak{J}_3]$.

We need the following well-known algorithm: Let A and B be two operators. We have

$$e^{-A}Be^{A} = \sum_{n=0}\frac{\{B,A\}^{(n)}}{n!}, \tag{2.26}$$

where $\{B, A\}^{(n)}$ is defined by induction by means of commutators

$$\{B,A\}^{(n)} = [\{B,A\}^{(n-1)}, A], \quad \{B,A\}^{(0)} = B. \tag{2.26'}$$

Then let us define the vector $\boldsymbol{\xi}$ by

$$\begin{aligned}\xi &= |\boldsymbol{\xi}|,\\ \beta_3/\beta &= \cos\xi,\\ \beta_1/\beta &= [(\sin\xi)/\xi]\xi_2,\\ \beta_2/\beta &= -[(\sin\xi)/\xi]\xi_1,\\ \xi_3 &= 0.\end{aligned} \tag{2.27}$$

[15] H. E. Moses, Ann. Phys. (N.Y.) **41**, 158 (1967).

We have from (2.26) and the commutation rules (2.1) that

$$e^{-i\xi\cdot J}\mathcal{J}_3 e^{i\xi\cdot J} = \boldsymbol{\beta}\cdot\mathcal{J}/\beta, \tag{2.28}$$

so that

$$\exp[i\boldsymbol{\beta}\cdot\mathcal{J}] = e^{-i\xi\cdot J}e^{i\beta\mathcal{J}_3}e^{i\xi\cdot J}. \tag{2.29}$$

Unlike the results (2.25), it is convenient to treat nonzero real mass, zero mass, and imaginary mass separately:

$$\exp[i\beta\mathcal{J}_3]f(0,\epsilon,\mathbf{p}) = \exp[i\epsilon\sigma(p,p_3,\beta,\epsilon)\mathbf{p}\cdot\mathbf{T}]f(0,\epsilon,\mathbf{p}'), \tag{2.30}$$

where

$$p_1' = p_1, \quad p_2' = p_2,$$
$$p_3' = p_3\cosh\beta - \epsilon p\sinh\beta, \tag{2.30'}$$

and

$$\sigma(p,p_3,\beta,\epsilon) = \tanh\beta/(p^2 - \epsilon p p_3\tanh\beta). \tag{2.30''}$$

In Eq. (2.30) and later, $\mathbf{p}\cdot\mathbf{T} = p_1T_1 + p_2T_2$. For $c > 0$,

$$\exp[i\beta\mathcal{J}_3]f(c,\epsilon,\mathbf{p}) = \exp[i\epsilon(p^2 - p_3^2)^{-\frac{1}{2}}\sigma(p,p_3,\beta,\epsilon)\mathbf{p}\cdot\mathbf{T}]f(c,\epsilon,\mathbf{p}'), \tag{2.31}$$

where

$$p_1' = p_1, \quad p_2' = p_2,$$
$$p_3' = p_3\cosh\beta - \epsilon\omega(c,p)\sinh\beta, \tag{2.31'}$$

and

$$\tan\sigma(p,p_3,\beta,\epsilon) = \frac{[c(p^2 - p_3^2)]^{\frac{1}{2}}\tanh\beta}{p^2 - \epsilon p_3\omega(c,p)\tanh\beta}. \tag{2.31''}$$

For $c < 0$,

$$\exp[i\beta\mathcal{J}_3]f(c,\epsilon,\mathbf{p}) = \exp[\tfrac{1}{2}i(p^2 - p_3^2)^{-\frac{1}{2}}\sigma(p,p_3,\beta,\epsilon)\mathbf{p}\cdot\mathbf{T}]f(c,\epsilon',\mathbf{p}'), \tag{2.32}$$

where $\mathbf{p}'$ is given by (2.31') and ϵ' is given as a function of $\mathbf{p}$ and β through

$$\epsilon'\omega(c,p') = \epsilon\omega(c,p)\cosh\beta - p_3\sinh\beta, \tag{2.32'}$$

where $p' = |\mathbf{p}'|$. It is easily seen from (2.31') that ϵ' takes on the value $+1$ or -1. Finally,

$$\sigma(p,p_3,\beta,\epsilon) = \varphi(p,p_3) - \varphi(p',p_3'), \tag{2.32''}$$

where

$$\varphi(p,p_3) = \log\frac{|(p^2 - p_3^2)^{\frac{1}{2}}\omega(c,p) + (-c)^{\frac{1}{2}}p_3|}{|(p^2 - p_3^2)^{\frac{1}{2}}\omega(c,p) - (-c)^{\frac{1}{2}}p_3|} \tag{2.32'''}$$

and where p_3' is given by (2.31') and $p' = [p_1^2 + p_2^2 + (p_3')^2]^{\frac{1}{2}}$.

We can now find $\exp[i\boldsymbol{\beta}\cdot\mathcal{J}]f(c,\epsilon,\mathbf{p})$ through the use of (2.29) and the second of Eqs. (2.25). Though the process of obtaining the result is straightforward and leads to explicit expressions, they are so cumbersome in form that we do not give the results in this paper.

3. ALGORITHM FOR REDUCING REDUCIBLE REPRESENTATIONS OF THE POINCARÉ GROUP

We now give the algorithm for reducing reducible representations of the Poincaré group to standard helicity representations. A reducible representation is given in terms of a Hilbert space of functions of a set of variables which we denote collectively by ζ. The variables included in ζ may be discrete and/or continuous. An element of the Hilbert space is denoted by $g(\zeta)$. Any operator acting on a function in this space has a superscript ζ to indicate that it is acting on the ζ variable in this representation. Thus, for example, $J_3^\zeta g(\zeta)$ gives the function obtained by applying the operator J_3 on the function $g(\zeta)$ in the reducible representation. We call this representation the ζ representation.

Elements of the Hilbert space upon which the operators act in the standard helicity representations are denoted, as in Sec. 2, by $f(c,\epsilon,\mathbf{p},\lambda)$. Operators acting in the standard helicity representation have the superscript $\mathbf{p}$ to indicate this fact. Thus $J_3^{\mathbf{p}}f(c,\epsilon,\mathbf{p},\lambda)$ is the element of the Hilbert space obtained by operating with J_3 upon the function $f(c,\epsilon,\mathbf{p},\lambda)$ in the standard helicity representation. The way that the infinitesimal generators of the Poincaré group act in the standard helicity representation is given by Eqs. (2.6) where we would now adjoin the superscript $\mathbf{p}$ and indicate the variable λ explicitly. It is our objective to express every function $g(\zeta)$ as a linear combination of functions $f(c,\epsilon,\mathbf{p},\lambda)$ such that $A^\zeta g(\zeta)$ is the same linear combination of functions $A^{\mathbf{p}}f(c,\epsilon,\mathbf{p},\lambda)$, where A is any of the infinitesimal generators of the Poincaré group.

We now state our assumptions.

A. Assumptions

(1) The infinitesimal generators are given explicitly in the ζ representation.

(2) In this representation, the generators H, P_i $(i = 1, 2, 3)$ are Hermitian and integrable.

(3) The generators J_i are Hermitian and integrable, i.e., $[\exp i(\boldsymbol{\theta}\cdot\mathbf{J})]^\zeta$ exists and is unitary for every real vector $\boldsymbol{\theta}$.

(4) The generator $\mathcal{J}_3$ is integrable, i.e., the operator $[\exp(i\beta\mathcal{J}_3)]^\zeta$ exists for all β.

It is to be noted that we do not require $\mathcal{J}_3$ to be Hermitian. Indeed for some applications, such as the

reduction of relativistic wavefunctions for zero mass, the requirement that $\mathfrak{J}_3$ be Hermitian cannot be satisfied, though the integrability condition is fulfilled (see Ref. 5).

Step 1: Determine the spectrum of the Casimir operator C given by

$$C = H^2 - \sum_i P_i^2. \tag{3.1}$$

We see that the spectrum of C can be found, since the Hermitian operator is given in the ζ representation. We are thus led to a conventional eigenvalue problem. As before, the eigenvalues of C are denoted by c. The spectrum of C may be discrete and/or continuous.

Step 2a: For each value of $c > 0$ and for $\epsilon = 1$ and -1, find all linearly independent functions $g(\zeta; c, \epsilon, \lambda)$ which satisfy

$$\begin{aligned} P_i^\zeta g(\zeta; c, \epsilon, \lambda) &= 0, \quad i = 1, 2, 3, \\ H^\zeta g(\zeta; c, \epsilon, \lambda) &= \epsilon m g(\zeta; c, \epsilon, \lambda), \end{aligned} \tag{3.2a}$$

where

$$m = (c)^{\frac{1}{2}}.$$

The variable λ is used to label the linearly independent solutions of (3.2a) and similar equations to follow. The variable λ may actually represent a set of variables.

Step 2b: For $c = 0$ and both values of ϵ, find the linearly independent solutions of

$$\begin{aligned} P_i^\zeta g(\zeta; 0, \epsilon, \lambda) &= 0 \quad (i = 1, 2), \\ P_3^\zeta g(\zeta; 0, \epsilon, \lambda) &= g(\zeta; 0, \epsilon, \lambda), \\ H^\zeta g(\zeta; 0, \epsilon, \lambda) &= \epsilon g(\zeta; 0, \epsilon, \lambda). \end{aligned} \tag{3.2b}$$

Step 2c: For each value of $c < 0$, find the linearly independent solutions of

$$\begin{aligned} P_i^\zeta g(\zeta; c, \lambda) &= 0 \quad (i = 1, 2), \\ P_3^\zeta g(\zeta; c, \lambda) &= \kappa g(\zeta; c, \lambda) \quad (\kappa = |c|^{\frac{1}{2}}), \\ H^\zeta g(\zeta; c, \lambda) &= 0. \end{aligned} \tag{3.2c}$$

It can be shown that Eqs. (3.2) have nontrivial solutions for some values of c and ϵ. It is also clear that the solutions are not unique, since new solutions can be made by taking linear combinations in the variable λ. We discuss later the effect of this lack of uniqueness, which is equivalent to the lack of uniqueness in the choice of the variable λ.

Step 3a: For each value of $c > 0$ and both values of ϵ, find the kernels $M(\lambda \mid \lambda')$, $R_1(\lambda \mid \lambda')$, and $R_2(\lambda \mid \lambda')$ from

$$\begin{aligned} J_3^\zeta g(\zeta; c, \epsilon, \lambda) &= \sum_{\lambda'} M(\lambda' \mid \lambda) g(\zeta; c, \epsilon, \lambda'), \\ J_i^\zeta g(\zeta; c, \epsilon, \lambda) &= \sum_{\lambda'} R_i(\lambda' \mid \lambda) g(\zeta; c, \epsilon, \lambda') \quad (i = 1, 2). \end{aligned} \tag{3.3a}$$

Step 3b: For $c = 0$ and for both values of ϵ, these kernels are to be obtained from

$$\begin{aligned} J_3^\zeta g(\zeta; 0, \epsilon, \lambda) &= \sum_{\lambda'} M(\lambda' \mid \lambda) g(\zeta; 0, \epsilon, \lambda'), \\ [-\epsilon \mathfrak{J}_1 - J_2]^\zeta g(\zeta; 0, \epsilon, \lambda) &= \sum_{\lambda'} R_1(\lambda' \mid \lambda) g(\zeta; 0, \epsilon, \lambda'), \\ [-\epsilon \mathfrak{J}_2 + J_1]^\zeta g(\zeta; 0, \epsilon, \lambda) &= \sum_{\lambda'} R_2(\lambda' \mid \lambda) g(\zeta; 0, \epsilon, \lambda'). \end{aligned} \tag{3.3b}$$

Step 3c: For each value of $c < 0$, these kernels are to be obtained from

$$\begin{aligned} J_3^\zeta g(\zeta; c, \lambda) &= \sum_{\lambda'} M(\lambda' \mid \lambda) g(\zeta; c, \lambda'), \\ \mathfrak{J}_i^\zeta g(\zeta; c, \lambda) &= \sum_{\lambda'} R_i(\lambda' \mid \lambda) g(\zeta; c, \lambda'), \quad (i = 1, 2). \end{aligned} \tag{3.3c}$$

For continuous values of λ the sums are to be replaced by integrals. Furthermore, for such continuous values the kernels may be symbolic functions. The kernels in general depend on c and ϵ. The existence of these kernels and their properties as a representation of the infinitesimal generators of the little groups is given by the following theorem:

Theorem 1: The kernels of Eqs. (3.3) always exist. The kernel $M(\lambda \mid \lambda')$ is a representation of an operator which is equivalent to a Hermitian operator whose eigenvalues are integers and/or half-odd integers. Furthermore, the kernels $M(\lambda \mid \lambda')$, $R_i(\lambda \mid \lambda')$ satisfy the commutation rules (2.4″) in which the kernels $T_i(\lambda \mid \lambda')$ are replaced by $R_i(\lambda \mid \lambda')$.

We now continue the procedure for the reduction.

Step 4: Let us define $\langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle$ as the following function of its arguments:

$$\begin{aligned} \langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle &= \{[\exp i(\boldsymbol{\omega} \cdot \mathbf{J})][\exp iv\mathfrak{J}_3]\}^\zeta g(\zeta; c, \epsilon, \lambda), \\ &\qquad \text{for} \quad c \geq 0, \\ &= \{[\exp i(\boldsymbol{\omega} \cdot \mathbf{J})][\exp iv\mathfrak{J}_3]\}^\zeta g(\zeta; c, \lambda), \\ &\qquad \text{for} \quad c < 0, \end{aligned} \tag{3.4}$$

where the real vector $\boldsymbol{\omega}$ and the real number v are related in a one-to-one way to the vector $\mathbf{p}$ whose components take on all real values as follows.

For $c > 0$,

$$\begin{aligned} p = |\mathbf{p}| &= m \sinh |v| = \epsilon m \sinh v, \\ \epsilon &= \operatorname{sgn} v. \end{aligned} \tag{3.4′}$$

For $c = 0$,

$$p = e^{\epsilon v}. \tag{3.4″}$$

For $c < 0$,

$$\begin{aligned} p &= \kappa \cosh v, \\ \epsilon &= \operatorname{sgn} v. \end{aligned} \tag{3.4‴}$$

For all values of c,

$$\begin{aligned} \omega_3 &= 0, \\ p_1 &= -p(\omega_2/\omega)\sin\omega, \\ p_2 &= p(\omega_1/\omega)\sin\omega, \\ p_3 &= p\cos\omega, \\ \omega &= |\boldsymbol{\omega}|. \end{aligned} \tag{3.4''''}$$

We now present the principal result of the present paper as a theorem.

Theorem 2: Every function $g(\zeta)$ upon which the infinitesimal generators act in the ζ representation can be expanded in the following form:

$$g(\zeta) = \int dc \sum_{\epsilon} \sum_{\lambda} \int \frac{d\mathbf{p}}{\omega(c,p)} \langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle f(c, \epsilon, \mathbf{p}, \lambda), \tag{3.5}$$

such that

$$A^{\zeta} g(\zeta) = \int dc \sum_{\epsilon} \sum_{\lambda} \int \frac{d\mathbf{p}}{\omega(c,p)} \langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle A^{\mathbf{p}} f(c, \epsilon, \mathbf{p}, \lambda), \tag{3.5'}$$

for all A^{ζ} and $A^{\mathbf{p}}$, where, in (3.5'), A^{ζ} and $A^{\mathbf{p}}$ are the infinitesimal generators in the ζ and standard helicity representations, respectively. The kernel $M(\lambda \mid \lambda')$ which appears in the standard helicity representations (2.6) is the kernel $M(\lambda \mid \lambda')$ which appears in Eqs. (3.3). The kernels $T_i(\lambda \mid \lambda')$ of (2.6) are obtained from the kernels $R_i(\lambda \mid \lambda')$ of Eqs. (3.3) in the following way.

For $c > 0$,

$$\begin{aligned} T_1(\lambda \mid \lambda') &= -R_2(\lambda \mid \lambda'), \\ T_2(\lambda \mid \lambda') &= R_1(\lambda \mid \lambda'). \end{aligned} \tag{3.5''}$$

For $c = 0$,

$$T_i(\lambda \mid \lambda') = R_i(\lambda \mid \lambda'). \tag{3.5'''}$$

For $c < 0$,

$$\dot{T}_i(\lambda \mid \lambda') = -\epsilon R_i(\lambda \mid \lambda'). \tag{3.5''''}$$

In the expressions (3.5) and (3.5'), the integral over c is to be replaced by summation over the discrete values of this variable. Also, the summation over λ is to be replaced by integrations for continuous values of λ.

As we have said, the way of choosing λ is not unique. Let us assume that we had picked another variable, which we shall call β, to label the linearly independent solutions of (3.2). Let us define $\langle \zeta \mid c, \epsilon, \mathbf{p}, \beta \rangle$ and $f(c, \epsilon, \mathbf{p}, \beta)$ in a manner similar to that used for the definition of these quantities in terms of λ. The variable β may have quite a different character from the label λ. For example, if λ with an infinite number of values is a discrete variable, β could be a continuous variable.

The following theorem is easily proved using the notion of linear independence of the solutions of (3.2).

Theorem 3: There exist two kernels $C(\lambda \mid \beta)$ and $D(\beta \mid \lambda)$ such that

$$\begin{aligned} \sum_{\beta} C(\lambda \mid \beta) D(\beta \mid \lambda') &= \delta_{\lambda,\lambda'}, \\ \sum_{\lambda} D(\beta \mid \lambda) C(\lambda \mid \beta') &= \delta_{\beta,\beta'}, \\ \langle \zeta \mid c, \epsilon, \mathbf{p}, \beta \rangle &= \sum_{\lambda} D(\lambda \mid \beta) \langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle, \\ \langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle &= \sum_{\beta} C(\beta \mid \lambda) \langle \zeta \mid c, \epsilon, \mathbf{p}, \beta \rangle, \\ f(c, \epsilon, \mathbf{p}, \beta) &= \sum_{\lambda} C(\beta \mid \lambda) f(c, \epsilon, \mathbf{p}, \lambda), \\ f(c, \epsilon, \mathbf{p}, \lambda) &= \sum_{\beta} D(\lambda \mid \beta) f(c, \epsilon, \mathbf{p}, \beta). \end{aligned} \tag{3.6}$$

In (3.6), summations and Kronecker-delta functions are to be replaced by integrations and Dirac-delta functions for continuous ranges of the variables λ and β.

It is now convenient to indicate the inner product in the ζ representation explicitly. The inner product of two functions $g^{(1)}(\zeta)$ and $g(\zeta)$ will be written

$$\int g^{(1)*}(\zeta) g(\zeta)\, dm(\zeta),$$

where the asterisk means, as usual, the complex conjugate and $m(\zeta)$ is the measure function associated with the representation. It is useful to regard any function $g(\zeta)$ as the inner product of an abstract ket $|\Phi\rangle$ and a bra $\langle\zeta|$. Thus,

$$\begin{aligned} g(\zeta) &= \langle \zeta \mid \Phi \rangle, \\ g^{(1)}(\zeta) &= \langle \zeta \mid \Phi^{(1)} \rangle, \end{aligned} \tag{3.7}$$

and the resolution of the identity is

$$\int |\zeta\rangle\, dm(\zeta)\, \langle\zeta| = I, \tag{3.8}$$

where I is the identity operator. Thus the inner product of two functions of the Hilbert space in the ζ representation can be written

$$\begin{aligned} \int g^{(1)*}(\zeta) g(\zeta)\, dm(\zeta) &= \int \langle \Phi^{(1)} \mid \zeta \rangle\, dm(\zeta) \langle \zeta \mid \Phi \rangle \\ &= \langle \Phi^{(1)} \mid \Phi \rangle. \end{aligned} \tag{3.9}$$

We regard the transformation functions $\langle \zeta \mid c, \epsilon, \mathbf{p}, \lambda \rangle$ as being the product of the bra $\langle\zeta|$ and an improper ket $|c, \epsilon, \mathbf{p}, \lambda\rangle$. We can show that the set of kets $|c, \epsilon, \mathbf{p}, \lambda\rangle$ spans the Hilbert space, each ket being a simultaneous eigenket of C, sgn H, and P_i.

Let us now add another assumption to our previous one, namely that the operators $\mathfrak{J}_i$ are Hermitian so that the entire set of infinitesimal generators is now Hermitian and integrable. Then we have the following theorem.

Theorem 4: Under the above assumptions, the variable λ can always be so chosen that the kets $|c, \epsilon, \mathbf{p}, \lambda\rangle$ satisfy the completeness and orthogonality relations

$$\langle c, \epsilon, \mathbf{p}, \lambda \mid c', \epsilon', \mathbf{p}', \lambda'\rangle = \omega(c, p)\delta(c - c')\delta(\mathbf{p} - \mathbf{p}')\delta_{\epsilon,\epsilon'}\delta_{\lambda,\lambda'},$$

$$\sum_{\epsilon}\sum_{\lambda}\int |c, \epsilon, \mathbf{p}, \lambda\rangle \frac{dc\, d\mathbf{p}}{\omega(c, p)} \langle c, \epsilon, \mathbf{p}, \lambda| = I. \quad (3.10)$$

Furthermore, expanding $g(\zeta)$ as in (3.5) and likewise $g^{(1)}(\zeta)$ in terms of $f^{(1)}(c, \epsilon, \mathbf{p}, \lambda)$, we have

$$\sum_{\epsilon}\sum_{\lambda}\int f^{(1)*}(c, \epsilon, \mathbf{p}, \lambda) f(c, \epsilon, \mathbf{p}, \lambda) \frac{dc\, d\mathbf{p}}{\omega(c, p)} = \int g^{(1)*}(\zeta) g(\zeta)\, dm(\zeta) = \langle \Phi^{(1)} \mid \Phi\rangle. \quad (3.11)$$

Finally, with this choice of λ, the kernels $M(\lambda \mid \lambda')$ and $T_i(\lambda \mid \lambda')$ are Hermitian, i.e.,

$$M(\lambda \mid \lambda') = M^*(\lambda' \mid \lambda), \quad T_i(\lambda \mid \lambda') = T_i^*(\lambda' \mid \lambda). \quad (3.12)$$

Because of (3.10) and (3.11) we have

$$f(c, \epsilon, \mathbf{p}, \lambda) = \langle c, \epsilon, p, \lambda \mid \Phi\rangle. \quad (3.13)$$

The way that the variable λ is to be chosen to make (3.10)–(3.13) valid is given in Ref. 1. The choice is not unique. In practice, we have found it possible to guess a suitable variable by using the requirement (3.12). However, such guessing requires some detailed knowledge of the properties of reducible representations of the little group. We shall go into this matter further when we discuss applications in later papers.

We can now use Theorem 4 for the last step in reducing completely reducible sets of infinitesimal generators of the Poincaré group which are Hermitian and integrable.

Step 5: Having introduced the variables λ for which Eqs. (3.10)–(3.13) hold, reduce the Hermitian kernels of (3.12). In this reduction, a new basis is introduced in the λ space such that, in this basis, the infinitesimal generators of the little groups M, T_i are completely reduced. One then uses this basis instead of the original λ basis. This new basis corresponds to the choice of a new set of linearly independent kets in accordance with Theorem 3. We refrain from details, since they are now obvious.

APPENDIX: RELATION BETWEEN THE FOLDY–SHIROKOV REALIZATIONS AND THE STANDARD HELICITY REALIZATIONS FOR PARTICLES OF REAL MASS

The Hermitian Foldy–Shirokov realizations of the infinitesimal generators of the Poincaré group for a single real mass m and sign of energy ϵ are characterized by a set of Hermitian spin matrices or kernels $\{S_i(\lambda \mid \lambda')\}$, where λ is a real variable which can take on any values compatible with the set of matrices being an integrable representation of the infinitesimal generators of the rotation group. For the moment, we do not require that the representation be irreducible. Thus, assuming that λ is a discrete variable, the matrices must satisfy the commutation relations

$$\sum_{\lambda'} \{S_1(\lambda \mid \lambda'')S_2(\lambda'' \mid \lambda') - S_2(\lambda \mid \lambda'')S_1(\lambda'' \mid \lambda')\} = iS_3(\lambda \mid \lambda') \quad (A1)$$

cyclically.

The functions upon which the infinitesimal generators act are denoted by $g(\mathbf{p}, \lambda)$, where $\mathbf{p}$ extends over the entire three-dimensional space and λ has the same range and character as in the kernels S_i.

The inner product of two functions $g^{(1)}$ and g are given by

$$\sum_{\lambda}\int g^{(1)*}(p, \lambda) g(p, \lambda) \frac{d\mathbf{p}}{\omega(c, p)}, \quad c = m^2. \quad (A2)$$

In order to write the expressions for the Foldy–Shirokov realizations compactly, it is convenient to suppress the variable λ and use $S_i g(\mathbf{p})$ to mean

$$\sum_{\lambda'} S_i(\lambda \mid \lambda') g(\mathbf{p}, \lambda').$$

Then the Foldy–Shirokov realization is given by

$$\begin{aligned} P_i g(\mathbf{p}) &= p_i g(\mathbf{p}), \\ Hg(\mathbf{p}) &= \epsilon\omega(c, p) g(\mathbf{p}), \\ J_i g(\mathbf{p}) &= -i(\mathbf{p} \times \nabla)_i g(\mathbf{p}) + S_i g(\mathbf{p}), \\ \mathfrak{J}_i g(\mathbf{p}) &= \epsilon\left[i\omega(c, p)\nabla_i g(\mathbf{p}) + \sum_{j,k} \frac{\epsilon_{ijk} p_j}{\omega(c, p) + m} S_k g(\mathbf{p})\right]. \end{aligned} \quad (A3)$$

In (A3), ϵ_{ijk} is the usual antisymmetric three-index symbol.

Let us denote the functions upon which the infinitesimal generators act in the standard helicity representation by $f(\mathbf{p}, \lambda)$. These are the same as the functions previously denoted by $f(c, \epsilon, \mathbf{p}, \lambda)$ in Sec. 2. On suppressing the λ variable, the standard helicity realization is given by (2.6). The inner product is (A2) as in the Foldy–Shirokov realization. The

standard helicity realization is unitarily equivalent to the Foldy–Shirokov realization, provided the masses m are the same, the signs of energy ϵ are the same, and provided a unitary transformation in the λ variables exists that transforms the kernels T_1, T_2, M (which we take to be Hermitian) of the helicity representation into the kernels S_i of the Foldy–Shirokov realization.

We now give the relationship between the functions $g(\mathbf{p}, \lambda)$ and $f(\mathbf{p}, \lambda)$ and thus have the unitary transformation which relates the two realizations under the condition that

$$T_i(\lambda \mid \lambda') = S_i(\lambda \mid \lambda'), \quad i = 1, 2,$$
$$M(\lambda \mid \lambda') = S_3(\lambda \mid \lambda'). \tag{A4}$$

Of course, if the kernels M, T_i are unitarily equivalent to the kernels S_i, one can always make a transformation in the λ variable of the Foldy–Shirokov realization or of the standard helicity realization so that (A4) is valid.

Theorem:

$$f(\mathbf{p}) = [\exp -i\tfrac{1}{2}\pi S_3][\exp -i(\boldsymbol{\omega} \cdot \mathbf{S})]g(\mathbf{p}), \tag{A5}$$

where $\boldsymbol{\omega}$ is given in terms of $\mathbf{p}$ by (3.4‴).

Also,

$$g(\mathbf{p}) = [\exp i(\boldsymbol{\omega} \cdot \mathbf{S})][\exp i\tfrac{1}{2}\pi S_3]f(\mathbf{p}). \tag{A5'}$$

In (A5) and (A5') we have used the convention that if A is a matrix with the element $A(\lambda \mid \lambda')$, then $Af(\mathbf{p})$ means $\sum_{\lambda'} A(\lambda \mid \lambda')f(\mathbf{p}, \lambda')$. A similar statement holds for $Ag(\mathbf{p})$.

Now let us consider the case in which the kernels S_i constitute an *irreducible* representation of the rotation group such that $\mathbf{S}^2 = s(s+1)I$, where I is the identity operator in the λ space and s is a nonnegative integer or half-odd integer. Then the representations of the Poincaré group that we are dealing with are irreducible and correspond to representations for a particle of mass m, sign of energy ϵ, and spin s. The variable λ takes on the $2s+1$ values $s, s-1, \cdots, -s+1, -s$. Let us assume that the kernels $S_i(\lambda \mid \lambda')$ have the standard form given by $M(\lambda \mid \lambda')$ and $T_i(\lambda \mid \lambda')$ of (2.11) and (2.14). Let us further define the polar angle of $\mathbf{p}$ by

$$p_1 = p \sin\theta \cos\varphi,$$
$$p_2 = p \sin\theta \sin\varphi, \tag{A6}$$
$$p_3 = p \cos\theta.$$

It is now no longer convenient to suppress the variable λ. On using (2.16), and the following two relationships discussed in Ref. 13:

$$(-1)^{\lambda'-\lambda} Y_s^{\lambda,\lambda'*}(\theta, \varphi) = Y_s^{\lambda',\lambda}(\theta, \varphi), \tag{A7}$$

$$\sum_{\lambda''=-s}^{s} Y_s^{\lambda,\lambda''}(\theta, \varphi) Y_s^{\lambda',\lambda''*}(\theta, \varphi) = \frac{2s+1}{4\pi}\,\delta_{\lambda,\lambda'}, \tag{A8}$$

(A5) and (A5') become, respectively,

$$f(\mathbf{p}, \lambda) = \left[\frac{4\pi}{2s+1}\right]^{\frac{1}{2}} \left[\exp -i\frac{\pi}{2}\lambda\right] \times \sum_{\lambda'} Y_s^{\lambda',\lambda}(\theta, \varphi) g(\mathbf{p}, \lambda'), \tag{A9}$$

$$g(\mathbf{p}, \lambda) = \left[\frac{4\pi}{2s+1}\right]^{\frac{1}{2}} \sum_{\lambda'} Y_s^{\lambda,\lambda'*}(\theta, \varphi) \times \left[\exp i\frac{\pi}{2}\lambda'\right] f(\mathbf{p}, \lambda'). \tag{A9'}$$

Equations (A9) and (A9') are similar to results in Refs. 16 and 9 where transformations are introduced from spin variables to helicity variables. However, our transformations have the additional restriction on them that they induce a transformation between two specified realizations of the Poincaré group. This restriction is reflected in the choice of phases.

[16] M. Jacob and G. C. Wick, Ann. Phys. (N.Y.) 7, 404 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Determination of the Amplitude from the Differential Cross Section by Unitarity*

ROGER G. NEWTON
Indiana University, Bloomington, Indiana

(Received 7 July 1968)

Banach-space fixed-point theorems are used to prove two results: (a) If the differential scattering cross section is smooth and small enough, relative to the wavelength of the relative motion of the colliding particles, there always exists an amplitude function which satisfies elastic unitarity and whose squared modulus equals a given differential cross section. (b) Under somewhat stronger conditions this amplitude is uniquely determined (except for the sign of its real part) by the generalized optical theorem (unitarity) and it can be constructed by iterating the latter. The condition for (a) ensures *a priori* that the real part of the amplitude cannot vanish at any angle, and that for (b) implies that its real part cannot be smaller than twice its imaginary part. These results are then generalized to inelastic and production processes.

1. INTRODUCTION

The question we wish to examine here is whether the requirement of unitarity of the S matrix, i.e., the generalized optical theorem, or conservation of flux *without any additional restrictions*, determines the phase of the quantum-mechanical scattering amplitude once the differential cross section $d\sigma/d\Omega$ is experimentally given for all angles (at one energy). Surprisingly enough, very little seems to be known about this question, even though it has important practical applications as well as theoretical implications. In the context of the so-called inverse scattering problem one always starts with the assumed knowledge of the complex amplitude, even though experimentally it is always only its absolute magnitude that is given. From a practical point of view, the phase shifts have usually been determined in the low-energy region by starting at energies where only s waves exist, and then continuing upwards in energy where the higher partial waves enter one after another.[1]

2. ELASTIC UNITARITY AND SPHERICAL SYMMETRY

Let us restrict ourselves first to energy at which only elastic scattering is possible. Then the conservation of flux (based on Hermiticity of the Hamiltonian) leads to unitarity of the scattering matrix. The scattering amplitude $A(\mathbf{n}', \mathbf{n})$, the square of whose absolute magnitude is the differential cross section for scattering from the direction $\mathbf{n}$ of the relative momentum to direction $\mathbf{n}'$, then satisfies the *generalized optical theorem*[2]

$$4\pi k^{-1} \operatorname{Im} A(\mathbf{n}', \mathbf{n}) = \int d\mathbf{n}'' A^*(\mathbf{n}'', \mathbf{n}') A(\mathbf{n}'', \mathbf{n}), \quad (1)$$

where k is the wavenumber of the relative motion of the particles, and the integral on the right-hand side extends over all directions of the unit vector $\mathbf{n}''$. For $\mathbf{n}' = \mathbf{n}$ we obtain the *optical theorem* for forward scattering

$$4\pi k^{-1} \operatorname{Im} A(\mathbf{n}, \mathbf{n}) = \int d\mathbf{n}' \, |A(\mathbf{n}', \mathbf{n})|^2, \quad (2)$$

the right-hand side being the observable total scattering cross section for incidence from the direction $\mathbf{n}$.

If the interparticle force is spherically symmetric, the amplitude $A(\mathbf{n}', \mathbf{n})$ can depend only on the angle θ between $\mathbf{n}$ and $\mathbf{n}'$; we set $\cos\theta = \mathbf{n} \cdot \mathbf{n}'$. Let us write

$$A(\mathbf{n}', \mathbf{n}) = k^{-1} F(\cos\theta) \quad (3)$$

so that F is dimensionless. Then (2) reads

$$\operatorname{Im} F(1) = \tfrac{1}{2} \int_{-1}^{1} dx \, |F(x)|^2, \quad (4)$$

where $x = \cos\theta$. A change of variables of integration allows (1) to be written in the form

$$\operatorname{Im} F(x) = \frac{1}{2\pi} \iint dy \, dz \, \frac{F(y) F^*(z)}{(1 - x^2 - y^2 - z^2 + 2xyz)^{\frac{1}{2}}}. \quad (5)$$

The region of integration is the interior of the ellipse, inscribed in the square $-1 \leq x, y \leq 1$, in which the radicand in the integral is positive.

* This work was supported in part by the National Science Foundation and the U.S. Army Research Office, Durham.

[1] The ambiguities that occur in the case of nucleon–nucleon scattering, because of the spin if not all polarizations are measured, are of no concern to us here.

[2] See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Publ. Co., New York, 1966). The history of the optical theorem is briefly traced there on p. 28. If the interparticle force is not spherically symmetric, we need time-reversal invariance as well.

Let us write

$$F(x) \equiv G(x)e^{i\varphi(x)},$$

with $G(x) = |F(x)| = k[d\sigma(x)/d\Omega]^{\frac{1}{2}}$. Then (5) becomes

$$\sin \varphi(x) = \iint dy\, dz H(x, y, z) \cos [\varphi(y) - \varphi(z)], \quad (6)$$

where we have defined the integral kernel

$$H(x, y, z) = \frac{G(y)G(z)}{2\pi G(x)(1 - x^2 - y^2 - z^2 + 2xyz)^{\frac{1}{2}}}, \quad (7)$$

when $1 - x^2 - y^2 - z^2 + 2xyz \geq 0$, and zero otherwise.

For $x = 1$ this reduces to (4) or

$$\sin \varphi(1) = \tfrac{1}{2}\int_{-1}^{1} \frac{dx G^2(x)}{G(1)}. \quad (6')$$

We now consider the function $G(x) \geq 0$, $-1 \leq x \leq 1$, given. Its square is the experimentally observed differential scattering cross section $d\sigma(x)/d\Omega$ in units of the reduced wavelength $\lambda\!\!\!^{-}$. Then (6) constitutes a nonlinear integral equation for the phase function $\varphi(x)$. The question we want to study is: *Under what conditions does (6) have a unique solution?* It should be stressed that this question has two distinct parts, about neither of which much is known: the *existence* of a solution, and its *uniqueness*. The first part can be rephrased in this form: Does unitarity impose restrictions on the possible functional dependence of the differential cross section on the scattering angle? The second part is the question: Does unitarity alone allow us to determine the amplitude if the differential cross section is given?

Partial answers to both questions can be given immediately on quite trivial grounds. Equation (1) shows directly that if A satisfies it, then so does $-A^*$; the over-all sign of the real part of A cannot be determined. In other words, if $\varphi(x)$, $-1 \leq x \leq 1$, is a solution of (6), then, clearly, so is $\pi - \varphi(x)$. So if there is one solution of (6), then there are two.[3] In order to be able to talk freely about further possible ambiguities in the solution, we shall call it *essentially* unique if it is unique except for this, or the solution is essentially unique if $\sin \varphi(x)$ is uniquely determined.[4]

As for the *existence* of a solution, the optical theorem (4), or (6'), to which (6) reduces for $x = 1$, shows an immediate restriction. We must have

$$G(1) \geq \tfrac{1}{2}\int_{-1}^{1} dx\, G^2(x). \quad (8)$$

Otherwise there can be no (real) solution $\varphi(x)$ of (6). In other words, the optical theorem trivially implies that the square root of the forward scattering cross section, multiplied by twice the wavelength, must be larger than the total cross section.

Are there any other restrictions on the differential cross section implied by unitarity? That there must be such restrictions can be made plausible as follows.

Suppose that for a given $G(x)$ the function $\varphi(x)$ solves (6). Let us now decrease $G(x)$ only in a small neighborhood Δ of width ϵ around $x = x_0$. Clearly we can make ϵ so small that, no matter how small G is made in Δ, for all values of x not in Δ, the right-hand side of (6) changes by as little as we like. Hence the original solution $\varphi(x)$ need be changed only "infinitesimally" in order to solve the altered equation (6) everywhere except inside Δ. But inside Δ we need only make $G(x)$ small enough in order to prevent any $|\sin \varphi(x)| \leq 1$ from being able to solve (6).

The foregoing argument shows that at every point x, $-1 \leq x \leq 1$, (6) implies a lower bound on G relative to its values everywhere else in the interval. In the case of $x = 1$ this restriction is given by (8), but for other values of x it cannot be explicitly stated. The class of functions $G(x)$ for which (6) possesses a solution is considerably smaller than the set that satisfies (8).

It must not be supposed, on the other hand, that (6) implies that the function

$$Q(x) \equiv \iint dy\, dz H(x, y, z), \quad (9)$$

which is an obvious *a priori* upper bound for $\sin \varphi(x)$, must not be too large. Consider the special case in which F consists of a single partial wave,

$$F(x) = F_l(x) \equiv (2l + 1)e^{i\delta_l} \sin \delta_l P_l(x), \quad \text{Im } \delta_l = 0.$$

We know that this is a solution of (5) if $P_l(x)$ is the Legendre polynomial of order l. In other words, if

$$G(x) = (2l + 1)\,|\sin \delta_l P_l(x)|,$$

then a solution of (6) is

$$e^{i\varphi(x)} = e^{i\delta_l} \operatorname{sgn} [\sin \delta_l P_l(x)].$$

But since, for $l > 0$, the function $P_l(x)$ has zeros in $-1 \leq x \leq 1$, $Q(x)$ must have infinities there. A situation in which G, and hence the cross section, has zeros is admittedly very special. But we can approximate this state of affairs by adding to $F_l(x)$ small

[3] There is of course always the even more trivial ambiguity that φ is determined at every point only *modulo* 2π. But this introduces no ambiguity in the amplitude.

[4] The problem of removing the two-fold ambiguity inherent in the unitarity equations by the use of other requirements has been studied by D. Bessis and A. Martin, Nuovo Cimento **52**, 719 (1967). Explicit examples of essential nonuniqueness were given by J. H. Crichton, Nuovo Cimento **45**, 256 (1966).

admixtures of other partial waves. The function G will then generally not vanish anywhere in $-1 \leq x \leq 1$, and $\varphi(x)$ will be continuous.

We now come to our first result, which is a sufficient criterion for the *existence* of a solution to (6):

Theorem 1: If the differential cross section is a continuous function of x and if the function $Q(x)$ of (9) is such that, for all x in $-1 \leq x \leq 1$,

$$Q(x) \leq M < 1, \tag{10}$$

then a solution to (6) exists. Furthermore, if $\varphi(x)$ is a solution of (6) and (10) holds, then

$$(1 - M^2)^{\frac{1}{2}} Q(x) \leq \sin \varphi(x) \leq Q(x). \tag{11}$$

Proof: The proof goes as follows. The equation

$$\psi(x) = \sin^{-1} \iint dy\, dz H(x, y, z) \cos [\varphi(y) - \varphi(z)] \tag{12}$$

defines a nonlinear mapping $\varphi \to \psi$ which we may symbolize by

$$\psi = \mathcal{M}(\varphi). \tag{12'}$$

Let us restrict the domain of the map to the region

$$0 \leq \varphi(x) \leq \tfrac{1}{2}\pi \tag{12''}$$

and make $\psi(x)$ unique by the requirement

$$0 \leq \psi(x) < \tfrac{1}{2}\pi, \tag{13}$$

which is possible because of (10). Now let us introduce the norm

$$\|\psi\| \equiv \sup_{-1 \leq x \leq 1} |\psi(x)| \tag{14}$$

which makes the set of bounded functions on the interval $-1 \leq x \leq 1$ into a Banach space. With the restriction (10), $\mathcal{M}$ maps the sphere $\|\varphi\| \leq \frac{1}{2}\pi$ into $\|\psi\| \leq \sin^{-1} M < \frac{1}{2}\pi$ and hence into itself. Now if the differential cross section is a continuous function of x in $0 \leq x \leq 1$ and (10) holds, then it is easy to see that the range of the map (12) is an equicontinuous set. By Ascoli's theorem the image of the sphere $\|\varphi\| \leq \frac{1}{2}\pi$ is therefore compact. Schauder's fixed-point theorem[5] then implies that $\mathcal{M}$ has at least one fixed point, and hence (6) has at least one solution.

Now it is obvious that $Q(x)$ is an upper bound on $\sin \varphi(x)$ if $\varphi(x)$ solves (6). But it then follows that

$$\cos [\varphi(y) - \varphi(z)] \geq \cos \varphi_{\max} \geq (1 - M^2)^{\frac{1}{2}}.$$

The left-hand inequality in (11) then follows from (6). This completes the proof.

It should be noticed that if the differential cross section is small compared to the wavelength and not strongly angle-dependent, then the two inequalities (11) strongly restrict the phase of the amplitude.

Our result should be compared with what a naive partial wave analysis would lead one to believe. Suppose that a given differential cross section is expressible as a linear combination of Legendre polynomials of order up to $2L$. It is then natural to assume that the amplitude contains no partial waves of order higher than L. But the cross section has $(2L + 1)$ real coefficients, and such an amplitude, because of unitarity, has only $(L + 1)$ real parameters. Therefore the expressibility in terms of unitary amplitudes appears to imply severe restrictions on the class of cross sections made up of $(2L + 1)$ Legendre polynomials. In view of Theorem 1 it is clear that these restrictions originate not in the unitarity condition, but in the assumption that the amplitude contains no partial waves higher than L. In order to form the most general cross section made up of $2L + 1$ Legendre polynomials, amplitudes containing infinitely many partial waves are required. (It is of course impossible that the amplitude contains a finite number of Legendre polynomials larger than $L + 1$.)

Theorem 1 gives a partial answer to the question of restrictions on the differential cross section implied by unitarity. If the differential cross section is small and smooth enough so that (10) holds, then no restrictions are implied. But it tells us nothing about the essential uniqueness of the solution for the amplitude, nor does it tell us how to construct the latter. These questions are answered in our main result, whose hypotheses are, however, more restrictive than (10).

Theorem 2: If $M < 5^{-\frac{1}{2}}$, then (6) has a unique solution that satisfies (12), and this solution can be obtained by iterating (6).

Proof: The proof is based directly on the principle of contraction mapping[6] in a Banach space as follows: Let $0 \leq \alpha < \beta < 1$. Then $\sin^{-1} \alpha < \sin^{-1} \beta$ and

$$\frac{\sin^{-1} \beta - \sin^{-1} \alpha}{\beta - \alpha} < \frac{d}{d\beta} \sin^{-1} \beta = \frac{1}{(1 - \beta^2)^{\frac{1}{2}}}$$

because $\sin^{-1} \alpha$ is concave upwards. Therefore, for all $0 \leq \alpha \leq A$, $0 \leq \beta \leq A$, where $A < 1$,

$$\left| \frac{\sin^{-1} \beta - \sin^{-1} \alpha}{\beta - \alpha} \right| \leq (1 - A^2)^{-\frac{1}{2}}. \tag{15}$$

Using the fact that the identity

$$\cos \alpha - \cos \beta = 2 \sin \tfrac{1}{2}(\alpha - \beta) \sin \tfrac{1}{2}(\alpha + \beta)$$

implies

$$|\cos \alpha - \cos \beta| \leq |\alpha - \beta|,$$

[5] See, for example, *Nonlinear Integral Equations*, P. M. Anselone, Ed. (The University of Wisconsin Press, Madison, 1964), p. 10.

[6] See, for example, Ref. 5, p. 13.

let us compare the equations

$$\psi_1 = \mathcal{M}(\varphi_1) \quad \text{and} \quad \psi_2 = \mathcal{M}(\varphi_2)$$

by means of (10) and (15):

$$\begin{aligned}
&|\psi_1(x) - \psi_2(x)| \\
&\quad \leq (1 - M^2)^{-\frac{1}{2}} \iint dy\, dz H(x, y, z) \\
&\qquad \times |\cos[\varphi_1(y) - \varphi_1(z)] - \cos[\varphi_2(y) - \varphi_2(z)]| \\
&\quad \leq (1 - M^2)^{-\frac{1}{2}} \iint dy\, dz H(x, y, z) \\
&\qquad \times |[\varphi_1(y) - \varphi_2(y)] - [\varphi_1(z) - \varphi_2(z)]| \\
&\quad \leq 2(1 - M^2)^{-\frac{1}{2}} \iint dy\, dz H(x, y, z)\, |\varphi_1(y) - \varphi_2(y)|.
\end{aligned}$$

The last step uses the symmetry of $H(x, y, z)$ under interchange of y and z. This inequality can be written in terms of the norm defined in (14),

$$\|\psi_1 - \psi_2\| \leq \gamma \|\varphi_1 - \varphi_2\|, \tag{16}$$

where, because of (10),

$$\gamma = 2M(1 - M^2)^{-\frac{1}{2}}.$$

Therefore, if $\gamma < 1$, i.e., if $M < 5^{-\frac{1}{2}}$, then (12′) is a *contraction mapping* in a Banach space. The theorem then follows. However, since the principle of contraction mapping is not very generally familiar, and the proof is simple, we prove the theorem directly.

Consider the sequence of iterations of (6), defined by

$$\sin \varphi_{n+1}(x) = \iint dy\, dz H(x, y, z) \cos[\varphi_n(y) - \varphi_n(z)] \tag{17}$$

or, in terms of (12′),

$$\varphi_{n+1} = \mathcal{M}(\varphi_n).$$

Then, by (16),

$$\|\varphi_{n+1} - \varphi_n\| \leq \gamma \|\varphi_n - \varphi_{n-1}\|$$

with $\gamma < 1$. Repetition leads to

$$\|\varphi_{n+1} - \varphi_n\| \leq M\gamma^n$$

which implies that $\{\varphi_n\}$ is a Cauchy sequence. Hence it converges pointwise [because of the choice of norm (14)] to a limit function $\varphi(x)$: For a given ϵ there exists an N so that, for all $n > N$,

$$\|\varphi - \varphi_n\| < \epsilon.$$

Let ψ be the image of φ:

$$\psi = \mathcal{M}(\varphi).$$

Then, according to (16),

$$\|\psi - \varphi_{n+1}\| \leq \gamma \|\varphi - \varphi_n\| < \gamma\epsilon$$

and, consequently, $\|\varphi_n - \psi\| \to 0$ as $n \to \infty$. Therefore, $\varphi = \psi$ for each x, and the limit function φ satisfies (6).

Finally, let φ_1 and φ_2 be two solutions of (6). Then (16) tells us that

$$\|\varphi_1 - \varphi_2\| \leq \gamma \|\varphi_1 - \varphi_2\|,$$

which implies that $\|\varphi_1 - \varphi_2\| = 0$ because $\gamma < 1$. Hence the solution obtained by iteration is the *only* solution of (6). Q.E.D.

To state the result of our theorem in more directly physical terms is to say that if, for all $x = \cos\theta$,

$$\left[\frac{d\sigma(x)}{d\Omega}\right]^{\frac{1}{2}} \geq \frac{5^{\frac{1}{2}}}{\lambda} \iint dy\, dz \frac{\{[d\sigma(y)/d\Omega][d\sigma(z)/d\Omega]\}^{\frac{1}{2}}}{(1 - x^2 - y^2 - z^2 + 2xyz)^{\frac{1}{2}}}, \tag{18}$$

then the amplitude is essentially uniquely determined by unitarity, and it can be obtained by iteration of (6). If (18) holds then it follows from (11) that $\tan\varphi(x) < \frac{1}{2}$ for all $-1 \leq x \leq 1$. Hence the real part of the amplitude must certainly be larger than twice its imaginary part. If the number $5^{\frac{1}{2}}$ in (18) is replaced by 1, we are sure that an amplitude that obeys unitarity exists, but we are not sure of its uniqueness. That condition entails that the real part of the amplitude can certainly not vanish at any angle.

One may wonder at this point if the difference between the hypotheses of the two theorems is merely a technical difficulty or if the essential uniqueness of the solution may get lost when $5^{\frac{1}{2}} < M \leq 1$. That it may indeed get lost may be argued as follows.

Let us write

$$G(x) \equiv \xi g(x)$$

and normalize $g(x)$ so that

$$\sup_{-1 \leq x \leq 1} \frac{1}{2\pi} \iint dy\, dz \frac{g(y)g(z)/g(x)}{(1 - x^2 - y^2 - z^2 + 2xyz)^{\frac{1}{2}}} = 1, \tag{19}$$

assuming that $Q(x)$ is bounded. Then Theorem 2 tells us that (6) has a well-defined solution for $\xi < 5^{-\frac{1}{2}}$. We may consider this solution $\varphi(x)$ as a function of ξ, even for complex values of ξ. Clearly then (6) will define φ also in a neighborhood of the real positive ξ axis so long as $|\xi| < 5^{-\frac{1}{2}}$. Let us differentiate (6) with respect to ξ, and indicate the derivative by a subscript:

$$\begin{aligned}
\varphi_\xi(x) \cos\varphi(x) = \xi^{-1}\varphi(x) - 2\xi \iint dy\, dz\, h(x, y, z) \\
\times \varphi_\xi(y) \sin[\varphi(y) - \varphi(z)],
\end{aligned} \tag{20}$$

where h is defined in terms of g just as H is in (7) defined in terms of G, and we have used the symmetry of $h(x, y, z)$ under interchange of y and z. Now (20) constitutes a linear integral equation for $\varphi_\xi(x)$, assuming that $\varphi(x)$ is given. It will have a unique solution unless the homogeneous equation has a solution, i.e., unless unity is an eigenvalue of the linear operator whose integral kernel is

$$\xi f(x, y) \equiv -2\xi \sec \varphi(x) \int dz\, h(x, y, z) \times \sin [\varphi(y) - \varphi(z)], \quad (21)$$

the limits of the integral being

$$xy \pm (1 - x^2)^{\frac{1}{2}}(1 - y^2)^{\frac{1}{2}}.$$

So long as ξf does not have unity as an eigenvalue, the derivative φ_ξ is well defined and thus $\varphi(x)$ is a regular *analytic* function of ξ. Now if for all ψ and, for all $-1 \leq x \leq 1$,

$$\left| \xi \int_{-1}^{1} dy\, f(x, y)\psi(y) \right| < \|\psi\|, \quad (22)$$

the eigenvalues of ξf are all strictly less than unity and hence the solution $\varphi(x)$ of (6) is a regular analytic function of ξ. But because of the second inequality in (11) we have

$$\left| \xi \int_{-1}^{1} dy\, f(x, y)\psi(y) \right| \leq 2Q(x)[1 - Q^2(x)]^{-\frac{1}{2}} \|\psi\|,$$

which is less than $\|\psi\|$ on the condition that $Q(x) < 5^{-\frac{1}{2}}$ for all x. So we meet again the hypothesis of Theorem 2.

We interpret this result as follows. Where the derivative of $\varphi(x)$ with respect to ξ is ill defined we have a branch point above which $\varphi(x)$ is no longer single-valued. There is then more than one solution of (6). When $0 \leq \xi < 5^{-\frac{1}{2}}$, this cannot happen. In the region $5^{-\frac{1}{2}} \leq \xi \leq 1$ there still always exists at least one solution of (6), but it may branch into several solutions at points, where ξf has the eigenvalue 1. When $\xi > 1$ there may not exist any solutions of (6), depending on the shape of $G(x)$. That is also the region where the phase $\varphi(x)$ may pass through $\frac{1}{2}\pi$, or the real part of the amplitude may vanish. For certain shapes of $G(x)$ an increase of ξ beyond a value at which $\varphi(x) = \frac{1}{2}\pi$ for some x, makes a (real) solution of (6) impossible.[7]

[7] That the dividing points for the three regions are exactly $\xi = 5^{-\frac{1}{2}}$ and $\xi = 1$ may have something to do with our choice of metric (14). For other choices the division may look somewhat different. It should also be noted that the possible existence of more than one solution when $\xi > 5^{-\frac{1}{2}}$ is conjectural. We have not produced any actual examples.

3. GENERALIZATION

It is easily seen that both theorems can be immediately generalized to cases in which there is no spherical symmetry. We then write

$$F(\mathbf{n}', \mathbf{n}) = G(\mathbf{n}', \mathbf{n}) \exp i\varphi(\mathbf{n}', \mathbf{n})$$

and the generalized optical theorem reads

$$4\pi G(\mathbf{n}', \mathbf{n}) \sin \varphi(\mathbf{n}', \mathbf{n}) = \int d\mathbf{n}'' G(\mathbf{n}'', \mathbf{n}')G(\mathbf{n}'', \mathbf{n}) \cos [\varphi(\mathbf{n}'', \mathbf{n}') - \varphi(\mathbf{n}'', \mathbf{n})]. \quad (23)$$

We set

$$Q(\mathbf{n}', \mathbf{n}) = \frac{1}{4\pi} \int d\mathbf{n}'' \frac{G(\mathbf{n}'', \mathbf{n}')G(\mathbf{n}'', \mathbf{n})}{G(\mathbf{n}', \mathbf{n})}$$

and find that if $Q < 1$ for all $\mathbf{n}$ and $\mathbf{n}'$ then (23) has a solution, and if $Q < 5^{-\frac{1}{2}}$ then it has a unique solution in $0 \leq \varphi < \frac{1}{2}\pi$ obtainable by iteration. In other words, for $Q < 1$, elastic unitarity imposes no restriction on the angle dependence of the differential cross section. If $Q < 5^{-\frac{1}{2}}$, then the generalized optical theorem determines the phase of the amplitude essentially uniquely.

What can we say if the particles have spin or inelastic processes are possible? Under the assumption of time-reversal invariance the generalized optical theorem can in both cases be written in the form

$$4\pi k_\alpha^{-1} \operatorname{Im} A_{\beta\alpha}(\mathbf{n}', \mathbf{n}) = \sum_\gamma \int d\mathbf{n}'' A^*_{\gamma\beta}(\mathbf{n}'', \mathbf{n}')A_{\gamma\alpha}(\mathbf{n}'', \mathbf{n}), \quad (24)$$

where the indices α, β, and γ refer either to the channels or to the helicities, or both, and the differential cross section for a reaction from α to β is given by

$$(d\sigma/d\Omega)\,(\mathbf{n}'\beta, \mathbf{n}\alpha) = |A_{\beta\alpha}(\mathbf{n}', \mathbf{n})|^2.$$

Defining

$$k_\alpha A_{\beta\alpha}(\mathbf{n}', \mathbf{n}) \equiv F_{\beta\alpha}(\mathbf{n}', \mathbf{n}) = G_{\beta\alpha}(\mathbf{n}', \mathbf{n}) \exp i\varphi_{\beta\alpha}(\mathbf{n}', \mathbf{n}),$$

with $G_{\beta\alpha}(\mathbf{n}', \mathbf{n}) \geq 0$, we get

$$4\pi G_{\beta\alpha}(\mathbf{n}', \mathbf{n}) \sin \varphi_{\beta\alpha}(\mathbf{n}', \mathbf{n}) = \sum_\gamma \int d\mathbf{n}'' G_{\gamma\beta}(\mathbf{n}'', \mathbf{n}')G_{\gamma\alpha}(\mathbf{n}'', \mathbf{n}) \times \cos [\varphi_{\gamma\beta}(\mathbf{n}'', \mathbf{n}') - \varphi_{\gamma\alpha}(\mathbf{n}'', \mathbf{n})]. \quad (25)$$

The previous arguments can now be carried through in the same form provided that we use the norm (14) with the sup taken over all directions of $\mathbf{n}$ and $\mathbf{n}'$ and over all values of the subscripts. If the set of functions

$$Q_{\beta\alpha}(\mathbf{n}', \mathbf{n}) \equiv \frac{1}{4\pi} \sum_\gamma \int d\mathbf{n}'' \frac{G_{\gamma\beta}(\mathbf{n}'', \mathbf{n}')G_{\gamma\alpha}(\mathbf{n}'', \mathbf{n})}{G_{\beta\alpha}(\mathbf{n}', \mathbf{n})} \quad (26)$$

is bounded by 1,

$$Q_{\beta\alpha}(\mathbf{n}', \mathbf{n}) < 1,$$

for all $\mathbf{n}$ and $\mathbf{n}'$ and all α and β, then unitarity imposes no restrictions on the cross sections. If for all $\mathbf{n}$, $\mathbf{n}'$, α, and β:

$$Q_{\beta\alpha}(\mathbf{n}', \mathbf{n}) < 5^{-\frac{1}{2}},$$

then all phases $\varphi_{\beta\alpha}(\mathbf{n}', \mathbf{n})$ are essentially uniquely determined by (25). It is important to note, however, that the cross sections of *all* coupled energetically possible processes must be known for this procedure to be feasible. This means if production is possible, that is, if more than two particles may appear in final states, then even the cross sections for more than two particles in both initial and final states must be known. Since these are never experimentally accessible, unitarity is useless as a tool to obtain the phase of amplitudes above production thresholds.

In the case of elastic scattering of particles with spin the determination of all elements of the scattering amplitude may proceed somewhat differently. In principle (if not usually in practice) we may assume that the spin–density matrix ρ_{inc} of the incoming beam can be controlled, and the spin–density matrix of the scattered wave can be measured.[8] Since

$$(\rho_{\text{scatt}})_{\alpha\beta} = \sum_{\gamma\delta} A_{\alpha\gamma} A^*_{\beta\delta} (\rho_{\text{inc}})_{\gamma\delta},$$

all products $A_{\alpha\lambda} A^*_{\beta\delta}$ are (in principle) subject to measurement by scattering experiments. There is then only a single phase (as a function of the scattering angles) that need be determined by unitarity. The fixed-point theorems will then serve again. Similarly, if less than complete information is available for the incident or scattered beams. These cases are too complicated for profitable general discussion and should be considered in detail for each individual case of practical interest.

Note added in proof. After this paper was finished I received a preprint by Martin[9] with similar and, in part, somewhat stronger results. The steps contained in his Eqs. (17) to (20) are easily transferred to our context and, together with (11), lead to

$$\|\sin \psi_1 - \sin \psi_2\| \leq \tfrac{1}{2}\gamma \, \|\sin \varphi_1 - \sin \varphi_2\|$$

in place of (16). Therefore Theorem 2 holds whenever $M < 2^{-\frac{1}{2}}$. Martin also proved uniqueness for $M < 0.79$, but whether the solution can be obtained by iteration when $M > 2^{-\frac{1}{2}}$ is not known. I am indebted to Dr. Martin for sending me a preprint of his work. The work by Crichton[4] was not known to me until I read his paper.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge that my interest in this problem was stimulated some time ago by a conversation with André Martin. I also found a discussion with C. van Winter very useful.

[8] For the measurability of the spin density matrix, see R. G. Newton and B. Young, Ann. Phys. (N.Y.) **49**, (1968) (to be published).

[9] André Martin, Nuovo Cimento (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Interpretation of $\Gamma^i_{[jk]}$ in a Huygens Model*

MURRAY MURASKIN
University of Nebraska, Lincoln, Nebraska, and University of North Dakota, Grand Forks, North Dakota

(Received 12 June 1967)

We describe a classical field theory based on Huygens' principle which is characterized by an additional degree of freedom which, to our knowledge, has not been discussed previously. This additional degree of freedom asserts that the forward propagation cone is different from the backward cone. The purpose of this paper is to find a function which describes this degree of freedom, and next, to understand this effect in reference to other theories. We find that this additional degree of freedom can be described by means of $\Gamma^i_{[jk]}$. The object $\Gamma^i_{[jk]}$ is then related to the torsion of a preferred-frame geometric theory. The additional degree of freedom is of interest since it enables one to introduce $\Gamma^i_{[jk]}$ in a framework involving characteristic equations, described by g_{ij}, and bicharacteristics described by $\Gamma^i_{(jk)}$, such that the role of $\Gamma^i_{[jk]}$ can be understood. Also, the theory furnishes a generalized framework for gravitational theory. Paths with noncontinuous slopes appear also in Feynman's path-integral approach. Thus, this type of discontinuity has physical interest here also, although we do not pursue this point in this paper.

I. INTRODUCTION

There are several types of discontinuities that one can study in field theory. In this paper, we investigate a particular kind of discontinuity. We shall consider a theory with noncontinuous propagation vectors.

We propose a model where such discontinuities appear. The model is that of a classical field theory based on Huygens' principle. By Huygens' principle, we mean that a point is a source of disturbance. Although we do not assume the most general Huygens model, our model is general enough to describe noncontinuous propagation vectors. We assume that at a point we have a forward propagation cone and a backward cone which is different from the forward cone. Then we cannot, in general, define trajectories with continuous propagation vectors along them.

We find a function such that when this function is nonzero, the backward cone is different from the forward cone. Thus, if this function is nonzero, propagation vectors are, in general, noncontinuous. The function that does this, $\Gamma^i_{[jk]}$, is antisymmetric in two indices. Although the Huygens propagation is described with respect to a Minkowski coordinate system, we show that $\Gamma^i_{[jk]}$ is related to the torsion of a preferred-frame geometric theory.

In this paper, we do not propose any field equations for $\Gamma^i_{[jk]}$. What we do propose is a framework in which to work. The gravitational equations

$$G_{\alpha\beta}(\{^{\sigma}_{\tau\lambda}\}) \equiv kT_{\alpha\beta}$$

are consistent with this framework. The torsion was introduced by Einstein and Schrödinger[1] to generalize the gravitational theory. Thus, it is of interest to understand what features this new object can give rise to within a framework consistent with the gravitational equations. The framework is based on general principles like Huygens' principle. Huygens' principle is already common to Minkowski field theories.

In the Huygens propagation model, g_{ij} represents the forward cone. $\Gamma^i_{(jk)}$ describes the propagation between points. $\Gamma^i_{[jk]}$ gives the change of backward cone into forward cone. We make the assumption that, if a set of variables completely describes the way propagation vectors behave, then they may be taken as field variables.

It is instructive to consider mathematical models consistent with the framework of Sec. II. We find a smaller set of variables that satisfy the propagation formulas. These are the sixteen variables e^{α}_{i}. These variables can be thought of as defining a coordinate transformation that takes us to a system where the propagation is of a trivial form. All the dynamical effects are, then, in the coordinate system (geometric theory). The geometric counterpart of the propagation variable g_{ij} is $\eta_{\alpha\beta} = e^i_{\alpha}e^j_{\beta}g^{(0)}_{ij}$, which is the transform of the Minkowski metric. The connection is $\Gamma^{\alpha}_{\beta\gamma} = e^{\alpha}_{m}(\partial e^m_{\beta}/\partial x^{\gamma})$. Thus, the sixteen-variable Huygens theory corresponds to a nonsymmetric preferred-frame geometric theory.

II. GENERALIZED PROPAGATION

In this section we seek a description for the backward cone such that it is different from the forward cone.

Let k^i be a propagation vector in Minkowski space. The forward propagation cone at a point is composed of an infinite number of propagation

* Supported in part by the Nebraska Research Council and in part by the National Science Foundation Grant GP-5373.

[1] E. Schrödinger, *Space Time Structure* (Cambridge University Press, New York, 1950); A. Einstein, *Meaning of Relativity* (Princeton University Press, Princeton, N.J., 1955), 5th ed.

vectors, one for each spatial direction. We label the propagation vectors $k^i_A, k^i_{A'}, \cdots$. The simplest forward cone is given by

$$g^{(0)}_{ij}k^ik^j = 0, \tag{1}$$

where $g^{(0)}_{ij} = \text{diag}\,(-1, -1, -1, +1)$. We shall consider the more general propagation cone given by

$$g_{ij}k^ik^j = 0. \tag{2}$$

A propagation vector at the point P will change, in general, during the propagation. We define δk^i by $\delta k^i = k^i(Q^-) - k^i(P)$. $k^i(Q^-)$ designates the value of the propagation vector just before the neighboring point Q is reached. We assume

$$\frac{\delta k^i}{d\lambda} + \Gamma^i_{(jk)}k^jk^k = 0. \tag{3}$$

λ is a parameter along the propagation path and $\Gamma^i_{(jk)} = \frac{1}{2}(\Gamma^i_{jk} + \Gamma^i_{kj})$. dx^k refers to the displacement between P and Q. We have in lowest order

$$\delta k^i + \Gamma^i_{(jk)}k^j\,dx^k = 0. \tag{4}$$

Thus it is assumed that $\Gamma^i_{(jk)}$ completely determines the change of a propagation vector between points.

The Ath propagation vector of the forward cone, k^i_A, at the point P obeys

$$g_{ij}(P)k^i_A(P)k^j_A(P) = 0. \tag{5}$$

We follow this propagation vector as it changes on propagation to the neighboring point Q according to (4). At Q, this propagation vector furnishes us with one of the generators of the backward cone at Q. The backward cone at Q is written as

$$g'_{ij}(Q; k^m)k^i(Q^-)k^j(Q^-) = 0.$$

We have allowed g'_{ij} to depend in such a way on direction since the g'_{ij} are not independent variables and thus, we must be prepared for the more general situation. The Ath generator of the backward cone at Q is given by

$$g'_{ij}(Q; P)k^i_A(Q^-)k^j_A(Q^-) = 0. \tag{6}$$

From (5) and (6), we get for the Ath propagation vector

$$g'_{ij}(Q; P)k^i_A(Q^-)k^j_A(Q^-) - g_{ij}(P)k^i_A(P)k^j_A(P) = 0. \tag{7}$$

$g'_{ij}(Q; P)$ constitutes ten variables and can be thought of as defining a fictitious cone at Q, by means of $g'_{ij}(Q; P)k^ik^j = 0$ for all k^i at this point. The Ath generator of this fictitious cone is a generator of the backward cone at Q. Using $k^i_A(P) = k^i_A(Q^{-i}_A) - \delta k$ with δk^i_A given by (4), we get

$$(\delta g_{ij} - \Gamma^t_{(ik)}g_{tj}\,dx^k - \Gamma^t_{(jk)}g_{it}\,dx^k)k^i_Ak^j_A = 0, \tag{8}$$

where $\delta g_{ij} \equiv g'_{ij}(Q; P) - g_{ij}(P)$. Equation (8) is valid to lowest order in dx^k. To lowest order, we evaluate g_{ij}, Γ^i_{jk} at point Q. There is at this point some arbitrariness in the solution of (8). We shall discuss the situation where we have

$$\delta g_{ij} - \Gamma^t_{(ik)}g_{tj}\,dx^k - \Gamma^t_{(jk)}g_{it}\,dx^k = 0. \tag{9}$$

Remember that we are not considering the most general Huygens model, but a model that is general enough to describe the forward propagation cone different from the backward propagation cone.

The difference between the forward cone at Q and the forward cone at P is written as $dg_{ij} = g_{ij}(Q) - g_{ij}(P)$. We make the continuity assumption that $dg_{ij} = (\partial g_{ij}/\partial x^k)\,dx^k$ for any neighboring propagation points. The necessary condition that the Ath propagation vector be discontinuous at Q is

$$\delta g_{ij} \neq dg_{ij}, \tag{10}$$

where

$$dg_{ij} - \delta g_{ij} = g_{ij}(Q) - g'_{ij}(Q; P). \tag{11}$$

That is, only if the cones $g_{ij}(Q)$ and $g'_{ij}(Q; P)$ are different can the propagation vector in the Ath spatial direction be discontinuous. We define Dg_{ij} as

$$Dg_{ij} \equiv dg_{ij} - \delta g_{ij}. \tag{12}$$

Thus, to have a theory with discontinuous propagation vectors, we must have from (10) that $g_{ij;k}\,dx^k \neq 0$, where

$$g_{ij;k} \equiv \frac{\partial g_{ij}}{\partial x^k} - \Gamma^t_{(ik)}g_{tj} - \Gamma^t_{(jk)}g_{it}. \tag{13}$$

Since $g_{ij;k}$ is nonzero, we write

$$g_{ij;k} = \Gamma_{j[ik]} + \Gamma_{i[jk]} \equiv \Gamma^t_{[ik]}g_{tj} + \Gamma^t_{[jk]}g_{it} \tag{14}$$

where $\Gamma^i_{[jk]} \equiv -\Gamma^i_{[kj]}$. We do not consider terms of the type $\Lambda^t_{(ik)}g_{tj} + \Lambda^t_{(jk)}g_{it}$ on the right-hand side. They are of the same form as the terms appearing in (9), and we consider all of the backward-cone type contributions as arising from δk^i. If $\Lambda^t_{(ik)}$ were not zero, then at a point we would have both $\Gamma^t_{(ik)}$ and $\Lambda^t_{(ik)}$ appearing. Now, under a coordinate transformation, the two fields get mixed up. It is natural to require that the operations δ and D have an invariant character under coordinate transformations. Thus, we take $\Lambda^t_{(ik)} = 0$. Using (9), (12), (13), and (14),

we get

$$Dg_{ij} = \Gamma^t_{[ik]}g_{tj}\,dx^k + \Gamma^t_{[jk]}g_{it}\,dx^k. \tag{15}$$

$\Gamma^i_{[jk]}$ like the other basic fields $\Gamma^i_{(jk)}$ and g_{ij}, is independent of direction. In any particular spatial direction, we get that the propagation vector is continuous if $\Gamma^i_{[jk]} = 0$, since we have $dg_{ij} = \delta g_{ij}$ for all source points P, and thus the backward cone at Q is the same as the forward cone there. Suppose we went through the analysis assuming that the backward cone is not direction-dependent; since the right-hand side of the resulting Eq. (15) is direction-dependent when $\Gamma^i_{[jk]} \neq 0$, the left-hand side must be also (barring accidents). Thus, we have a contradiction. Therefore we conclude that the backward cone is direction-dependent and the forward cone is not. Hence, the two cones are different. Thus, the function $\Gamma^i_{[jk]}$ is what is responsible for the backward cone being different from the forward cone.

The propagation is our probe of the field theory. Hence, we shall make the hypothesis that if a set of variables describes completely the way the propagation vectors behave, then these variables constitute a representation of the fields. Our field variables are then g_{ij}, $\Gamma^i_{(jk)}$, and $\Gamma^i_{[jk]}$.

We have made the assumption that we have continuous propagation between points (3) and discontinuities arising at a point (14). We may get an intuitive feeling for discontinuities at a point by thinking of a point as a depot where contributions come in from all spatial directions. At the point, they interact with each other. Thus, in general, we can't expect continuous propagation vectors at a point.

From generality we cannot expect all Minkowski points to be propagation points. A propagation point is a point where g_{ij} and Γ^i_{jk} are nonzero. The set of propagation points determines a space. In the sixteen-variable formulation of Sec. III, the propagation-space displacements $dx^i(x)$ satisfy $dx^i(x) = e^i_\alpha(x)\,dx^\alpha$. [See discussion under Eq. (31).]

A general coordinate transformation would introduce, in addition to the propagation variables, coordinate fields. Minkowski coordinate systems are, thus, preferred on grounds of simplicity. The notion of a preferred set of frames will carry over in the geometric formulation in Sec. IV.[2]

III. SIXTEEN-VARIABLE FORMULATION

We reconsider Einstein's[3] sixteen-variable basic-vector approach as we can reproduce the formulas, so far, in terms of this lesser number of variables. One of the basis vectors is of a timelike character. We construct the basis-vector system such that the propagation vectors as looked at from these basis vectors describe the simplest-type propagation. That is, since the number of variables is sixteen, we can think of these variables as defining a coordinate transformation that takes us from the system of general propagation to a system where the propagation is of the simplest type, while the dynamical effects are in the coordinate system (geometric theory). Then we have

$$g_{ij} = g_{\alpha\beta}e^\alpha_i e^\beta_j \tag{16}$$

and $g_{\alpha\beta}$ has a diagonal form $(-1, -1, -1, 1)$.

We define the dual basis vectors by

$$e^\alpha_i e^i_\beta = \delta^\alpha_\beta. \tag{17}$$

We also have

$$e^\alpha_i e^i_\beta = \delta^\alpha_\beta. \tag{18}$$

The e^α_i vary from point to point and thus can be considered dynamical variables.

We define Γ^j_{ik} by

$$\frac{\partial e^\alpha_i}{\partial x^k} = \Gamma^j_{ik}e^\alpha_j, \tag{19}$$

where $de^\alpha_i = (\partial e^\alpha_i/\partial x^k)\,dx^k$ and where dx^k is a displacement in propagation space. Thus, we have

$$de^\alpha_i = \Gamma^j_{ik}e^\alpha_j\,dx^k. \tag{20}$$

From (19) we have

$$\Gamma^t_{ik} = e^t_\alpha\frac{\partial e^\alpha_i}{\partial x^k} = -e^\alpha_i\frac{\partial e^t_\alpha}{\partial x^k}, \tag{21}$$

and thus we have

$$\Gamma^t_{(ik)} = \tfrac{1}{2}e^t_\alpha\left(\frac{\partial e^\alpha_i}{\partial x^k} + \frac{\partial e^\alpha_k}{\partial x^i}\right) \tag{22}$$

and

$$\Gamma^t_{[ik]} = \tfrac{1}{2}e^t_\alpha\left(\frac{\partial e^\alpha_i}{\partial x^k} - \frac{\partial e^\alpha_k}{\partial x^i}\right). \tag{23}$$

We decompose de^α_i into two parts in the case where dx^k is the displacement between a point and one of its source points,

$$de^\alpha_i = \delta e^\alpha_i + De^\alpha_i. \tag{24}$$

Since the propagation is to be of the simplest kind in the α, β system, we have

$$\delta k^\alpha = \delta(k^i e^\alpha_i) = 0 \tag{25}$$

for all propagation vectors. We then satisfy (25), using (4) with

$$\delta e^\alpha_i = \Gamma^j_{(ik)}e^\alpha_j\,dx^k. \tag{26}$$

We use $g_{ik} = e^\alpha_i e^\beta_k g_{\alpha\beta}$ and since $\delta g_{\alpha\beta} = 0$, we get

$$\delta g_{ij} = \Gamma^t_{(ik)}g_{tj}\,dx^k + \Gamma^t_{(jk)}g_{it}\,dx^k \tag{27}$$

[2] V. Fock, *Theory of Space Time and Gravitation* (Pergamon Press, Inc., New York, 1959).

[3] A. Einstein, Ann. Inst. Henri Poincaré **1**, 1 (1931).

which is Eq. (9). We write

$$dg_{ij} = d(e^\alpha_i e^\beta_j g_{\alpha\beta}) = \delta g_{ij} + Dg_{ij}. \tag{28}$$

From (20), (24), and (26), we get

$$De^\alpha_i = \Gamma^j_{[ik]} e^\alpha_j \, dx^k. \tag{29}$$

From (16) and (21), we get that

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma^t_{ik} g_{tj} + \Gamma^t_{jk} g_{it} \tag{30}$$

is identically satisfied. Also, from (27), (28), and (30), we get (15).

We have, here, expressed the variables g_{ij}, $\Gamma^i_{(jk)}$, $\Gamma^i_{[jk]}$ in terms of a lesser number, the sixteen basis-vector variables such that our previous formulas hold.

We can make arbitrary constant basis-vector Lorentz transformations in the α, β system since the α, β system is defined by $g_{\alpha\beta} = \text{diag}(-1, -1, -1, +1)$ and $\delta k^\alpha = 0$.

IV. GEOMETRIC VARIABLES

We expect from generality that not all Minkowski points will be propagation points. Thus, the distance ds^2 between propagation points will depend upon position. When we transform to a system where the propagation is trivial, the effect is reflected in the geometry. Thus, we see why a nontrivial line element comes out in the study of Huygens model. The argument is made explicit as follows. In Minkowski space the distance between two neighboring points in the space of propagation points is

$$ds^2(x) = g^{(0)}_{ij} \, dx^i(x) \, dx^j(x), \tag{31}$$

where $g^{(0)}_{ij}$ is the Minkowski metric and dx^i is the displacement between the two propagation points. We may transform to the system of trivial propagation using $dx^i(x) = e^i_\alpha(x) \, dx^\alpha$. The displacements between neighboring propagation points, dx^α, are x independent since they are associated with a trivial propagation. Thus, we have

$$ds^2(x) = \eta_{\alpha\beta}(x) \, dx^\alpha \, dx^\beta, \tag{32}$$

with the metric of the geometric theory given by

$$\eta_{\alpha\beta}(x) = e^i_\alpha(x) e^j_\beta(x) g^{(0)}_{ij}. \tag{33}$$

We define $\Gamma^\rho_{\beta\alpha}$ by

$$\Gamma^\rho_{\beta\alpha} \equiv e^\rho_m \frac{\partial e^m_\beta}{\partial x^\alpha}, \tag{34}$$

where $\partial/\partial x^\alpha = e^m_\alpha(\partial/\partial x^m)$. $\Gamma^\rho_{\beta\alpha}$ is then related to Γ^j_{ik} by

$$\Gamma^j_{ik} = -\Gamma^\rho_{\beta\alpha} e^\beta_i e^\alpha_k e^j_\rho. \tag{35}$$

From (34) we get

$$\frac{\partial e^\sigma_i}{\partial x^\rho} = -\Gamma^\sigma_{\alpha\rho} e^\alpha_i \tag{36}$$

and

$$de^\alpha_i = -\Gamma^\alpha_{\rho\beta} e^\rho_i \, dx^\beta. \tag{37}$$

We note, as a result of (33) and (34), that $\eta_{\alpha\beta}$ and $\Gamma^\rho_{\beta\alpha}$ satisfy the equation

$$\frac{\partial \eta_{\alpha\beta}}{\partial x^\gamma} - \Gamma^\sigma_{\alpha\gamma}\eta_{\sigma\beta} - \Gamma^\sigma_{\beta\gamma}\eta_{\sigma\alpha} = 0. \tag{38}$$

This equation is of the type satisfied by g_{ij} and Γ^i_{jk} [Eq. (14)]. When $\Gamma^\alpha_{[\beta\gamma]} = 0$, the geometric connection is determined by the Christoffel relations. Thus, as a result of (38) and from (37), we identify $\Gamma^\alpha_{\beta\gamma}$ with the connection of the geometric theory. The change of e^α_i [Eq. (37)] is then interpreted as equal to the change of e^α_i under parallel transport.

Assuming

$$\frac{\partial^2 e^i_\gamma}{\partial x^\tau \partial x^\beta} = \frac{\partial^2 e^i_\gamma}{\partial x^\beta \partial x^\alpha}, \tag{39}$$

we get from (36) that

$$R^\sigma{}_{\chi\beta\alpha} = 0, \tag{40}$$

where

$$R^\sigma{}_{\chi\beta\alpha} = \frac{\partial \Gamma^\sigma_{\chi\alpha}}{\partial x^\beta} - \frac{\partial \Gamma^\sigma_{\chi\beta}}{\partial x^\alpha} + \Gamma^\sigma_{\delta\beta}\Gamma^\delta_{\chi\alpha} - \Gamma^\sigma_{\delta\alpha}\Gamma^\delta_{\chi\beta}. \tag{41}$$

From (39) and (40), we get

$$R^\sigma{}_{\chi\beta\alpha}(\Gamma_{(\ )}) + R^\sigma{}_{\chi\beta\alpha}(\Gamma_{[\ ]}) + \Gamma^\lambda_{[\chi\alpha]}\Gamma^\sigma_{(\lambda\beta)} - \Gamma^\lambda_{(\chi\beta)}\Gamma^\sigma_{[\lambda\alpha]} - \Gamma^\lambda_{[\chi\beta]}\Gamma^\sigma_{(\lambda\alpha)} + \Gamma^\lambda_{(\chi\alpha)}\Gamma^\sigma_{[\lambda\beta]} = 0. \tag{42}$$

Thus, $R^\sigma_{\chi\beta\gamma}(\Gamma_{(\ )})$ is not zero if $\Gamma^\alpha_{[\beta\gamma]} \neq 0$.

We have discussed the geometric theory that corresponds to the generalized propagation theory in the sixteen-variable formulation. The torsion is related to $\Gamma^j_{[ik]}$ via (35). We thus see that the torsion is nonzero when the forward propagation cone is different from the backward propagation cone.

A generalized gravitation theory of the type discussed above was considered by Einstein.[3] (We have introduced in addition the operations δ and D.) A theory based not only on e^α_i but also on $\Gamma_\alpha{}^\beta{}_k$ and e^α_i is briefly discussed in the Appendix. These generalized preferred-frame gravitational frameworks are related to propagation theories having in common that the backward propagation cone is different from the forward cone.

V. DISCUSSION

We have described a classical field-theory model characterized by an additional degree of freedom which, to our knowledge, has not been discussed

previously. This additional degree of freedom is that the forward-propagation cone is different from the backward cone. We are able to describe this effect by means of $\Gamma^i_{[jk]}$, which we show is related to the torsion of a preferred-frame geometric theory.

In the Huygens model, we can think of the field as being made up of "particles" (in the sense of a well-defined trajectory between points) coming in to a point from the different spatial directions. There are an infinite number of propagation vectors coming out from a point. Since a propagation vector is not continuous, we cannot say to which of these outgoing vectors a particular ingoing vector should be joined in making a Huygens path. That is, we can follow a particle from a point to a neighboring point but not, in general, beyond. The interpretation we give from this picture is that an infinite number of particles come into a point and are destroyed, and a new set of infinite particles are subsequently created and leave the point.

As a result of (2) we have a different velocity of propagation in each spatial direction. In a theory with an index of refraction, the velocity of propagation is not that of light in empty space. Thus, we can say that (2) describes propagation in a medium. The idea in this paper is that the medium is not external to the field theory but is given by the fields g_{ij}, Γ^i_{jk} themselves.

Let us examine the situation where we do not make any reduction of field variables. (We further discuss this case in the Appendix.) If we take as a special case $\Gamma^i_{[jk]} = 0$ and $R_{ik}(\{^i_{jk}\}) = 0$,then we have equations having the form of the free-field gravitational equations. $R^i_{jkl}(\{^i_{jk}\})$ is not necessarily zero, since this does not follow from $R_{ik}(\{^j_{jk}\}) = 0$. In this situation (14) is the usual Christoffel relations. Equation (3) gives the bicharacteristics of these field equations. Equation (2) is also satisfied as a consequence of the characteristic equations. Thus, the equations of the paper are satisfied in the special case.

Although we do not propose field equations, the interpretation of $\Gamma^i_{[jk]}$ is not dependent on field equations.

The additional degree of freedom of different backward and forward cones is of interest since it enables one to introduce $\Gamma^i_{[jk]}$ in a framework involving characteristic equations, described by g_{ij}, and bicharacteristics described by $\Gamma^i_{(jk)}$, such that the role of $\Gamma^i_{[jk]}$ can be understood. Also, the theory furnishes a generalized framework for gravitational theory. Paths with noncontinuous slopes appear also in Feynman's path-integral approach.[4] Thus, this type of discontinuity has physical interest here also, although we do not pursue this point in this paper.

ACKNOWLEDGMENTS

The author wishes to sincerely thank Professor John M. McKinley, Professor Saul T. Epstein, Professor Thomas A. Morgan, and Professor J. Plebánski for discussing this paper.

APPENDIX: VIERBEIN FORMALISM

The Vierbein formalism[5] has, instead of (19), the equation

$$\frac{\partial e^\alpha_i}{\partial x^k} - \Gamma^t_{ik} e^\alpha_t - \bar{\Gamma}_\lambda{}^\alpha{}_k e^\lambda_i = 0. \tag{A1}$$

The spin connection is then given up to a constant vector by[6]

$$\Gamma_k = \tfrac{1}{4}\bar{\Gamma}_{\alpha\beta k}\tfrac{1}{2}(\gamma^\beta\gamma^\alpha - \gamma^\alpha\gamma^\beta). \tag{A2}$$

In the text $\bar{\Gamma}_\alpha{}^\beta{}_k = 0$, so that the spin connection is zero.

The Vierbein formalism gives an alternative formulation of the system g_{ij}, Γ^i_{jk} with no reduction of variables. We again express g_{ij} in terms of e^α_i by

$$g_{ij} = e^\alpha_i e^\beta_j g_{\alpha\beta}. \tag{A3}$$

We do not wish to bring in six additional field variables. Thus, we allow for arbitrary x dependent α, β Lorentz transformations at each point.

From (17) we get

$$\frac{\partial e^i_\alpha}{\partial x^k} + \Gamma^i_{jk} e^j_\alpha + \bar{\Gamma}_\alpha{}^\beta{}_k e^i_\beta = 0. \tag{A4}$$

We can express Γ^i_{jk} in terms of e^i_α and $\bar{\Gamma}_\alpha{}^\beta{}_k$ by multiplying (A4) by e^α_m. This gives

$$\Gamma^i_{mk} = -e^\alpha_m \frac{\partial e^i_\alpha}{\partial x^k} - \bar{\Gamma}_\alpha{}^\beta{}_k e^i_\beta e^\alpha_m. \tag{A5}$$

Using (A3) and (A5), we see that (14) is identically satisfied provided

$$\bar{\Gamma}_{\alpha\beta k} = -\bar{\Gamma}_{\beta\alpha k} \tag{A6}$$

where $\bar{\Gamma}_{\alpha\beta k} = g_{\beta\sigma}\bar{\Gamma}_\alpha{}^\sigma{}_k$.

Equation (14) is forty equations that express $\Gamma^i_{(jk)}$ in terms of the independent variables g_{ij} and $\Gamma^i_{[jk]}$ (see Ref. 1, p. 66) via

$$\Gamma^m_{ik} = \begin{Bmatrix} m \\ ik \end{Bmatrix} + g^{ml}g_{it}\Gamma^t_{[kl]} + g^{ml}g_{kt}\Gamma^t_{[il]} + \Gamma^m_{[ik]}, \tag{A7}$$

where $\{^m_{ik}\}$ are the Christoffels. The formulation above replaces g_{ij} and $\Gamma^i_{[jk]}$ by e^i_α and $\bar{\Gamma}_\alpha{}^\beta{}_k$ as independent variables by means of (A3) and (A5). The six extra variables in the latter set means that we also have six arbitrary functions appearing as mentioned before.

[4] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Co., New York, 1965). Paths with noncontinuous slopes appear also in E. Nelson, Phys. Rev. **150**, 1079 (1966).

[5] H. Weyl, Phys. Rev. **77**, 699 (1950); Z. Physik **56**, 330 (1929); see also D. Sciama, J. Math. Phys. **2**, 472 (1961); T. Kibble, J. Math. Phys. **2**, 212 (1961).

[6] V. Bargmann, Sitzber. Preuss. Akad. Wiss. Physik Math. Kl. **25**, 356 (1932).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

On the Clebsch-Gordan Series of a Semisimple Lie Algebra

D. RADHAKRISHNAN
Department of Physics and Astrophysics, University of Delhi, Delhi, India

(Received 29 May 1968)

The problem of determining the multiplicity of an irreducible representation of a semisimple Lie algebra, in the decomposition of the product of two such representations, is reduced to one of solving a system of linear equations. This is achieved by using some properties of partition functions which occur in the formulas for the multiplicity of a weight in a representation. It is shown that one need not know the partition function explicitly.

I. INTRODUCTION

Since the advent of unitary symmetries in elementary particle physics, it has become necessary to know the ("external") multiplicity of an irreducible representation (IR) in the decomposition of the product of two IR's of a semisimple Lie algebra. For low-rank algebras and small dimensions of the IR's, the Young tableaux method has been useful. Nevertheless, many authors[1,2] have been interested in arriving at algebraic expressions for the multiplicity.

This problem has been shown[2] to be connected to finding the ("internal") multiplicity of a weight in a given IR. Kostant[3] has given a formula for this in terms of partition functions. An expression is derived by Steinberg,[1] which gives the external multiplicity in terms of the same functions, and which involves a double sum over the Weyl group. However, these functions are very complicated[4] except for specific low-rank algebras.

It is the aim of this paper to calculate the external multiplicity without an explicit knowledge of the internal multiplicity and, in turn, of the partition function. In Sec. II, a lemma of Tarski[5,6] is used in the formulation of the problem. In Sec. III, we show how the problem reduces to one of solving a system of linear equations, the unknowns being the external multiplicities. The coefficient matrix is shown to be triangular and nonsingular, thus ensuring uniqueness of the solutions.

Some examples are presented in the Appendix, which illustrate the usefulness of this method. Applications to large-rank algebras and higher-dimensional IR's are possible, in principle, but tedious to perform manually.

[1] N. Jacobson, *Lie Algebras* (Interscience Publ. Inc., New York, 1962), p. 262.
[2] B. Gruber, Ann. Inst. Henri Poincaré **8**, 43 (1968).
[3] B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959).
[4] D. Radhakrishnan and T. S. Santhanam, J. Math. Phys. **8**, 2206 (1967); D. Radhakrishnan, Delhi University preprint, 1968 (unpublished).
[5] J. Tarski, J. Math. Phys. **4**, 569 (1963).
[6] V. B. Mandel'tsveig, Sov. Math. Dokl. **6**, 851 (1965).

II. FORMULATION OF THE PROBLEM

Let G be a complex semisimple Lie algebra, and let its system of positive roots (relative to a given Cartan subalgebra and lexicographic ordering) be

$$\Delta \equiv \{\alpha_1, \cdots, \alpha_s\}.$$

Let $D(\lambda)$ be a finite-dimensional IR of G with highest weight λ. The Clebsch–Gordan series is

$$D(\lambda) \otimes D(\lambda') = \oplus \sum_{\lambda''} \gamma_{\lambda''} D(\lambda''), \tag{1}$$

where $\gamma_{\lambda''}$ is the multiplicity of $D(\lambda'')$ in the decomposition of the product on the left.

In terms of characters we can write (1) as

$$\chi_\lambda(\varphi)\chi_{\lambda'}(\varphi) = \sum_{i=1}^{n} \gamma_{\lambda_i} \chi_{\lambda_i}(\varphi), \tag{2}$$

where the terms on the right are ordered for convenience such that $i < j$ implies $\lambda_i > \lambda_j$. In general, $\lambda_1 \geq \lambda + \lambda'$, while λ_n corresponds to the one-dimensional IR. We will show later that $\lambda_1 = \lambda + \lambda'$.

Weyl's character formula is

$$\chi_\lambda(\varphi) = \frac{\sum_{S \in W} \delta_S \exp i[S(\beta + \lambda), \varphi]}{\sum_{S \in W} \delta_S \exp i[S(\beta), \varphi]}. \tag{3a}$$

Also,

$$\chi_\lambda(\varphi) = \sum_{\nu \in D(\lambda)} m_\lambda(\nu) \exp i[\nu, \varphi], \tag{3b}$$

where $\varphi \equiv (\varphi_1, \cdots \varphi_l)$ are l real parameters, l being the rank of the algebra. W is the Weyl group and $\delta_S = \pm 1$ according to whether S is an even or odd reflection, respectively. β is half the sum of positive roots, and $m_\lambda(\nu)$ is the multiplicity of the weight ν in $D(\lambda)$. The bracket is the Cartan–Killing form on G, by which one can identify the Cartan subalgebra with its dual space.[3]

Substituting for χ_λ in (2) from (3a), and for $\chi_{\lambda'}$ and χ_{λ_i} from (3b), we obtain

$$\sum_{S \in W} \delta_S \sum_{\nu' \in D(\lambda')} m_{\lambda'}(\nu') \exp i[S(\beta + \lambda) + \nu', \varphi] = \sum_{i=1}^{n} \gamma_{\lambda_i} \sum_{S \in W} \delta_S \sum_{\nu_i \in D(\lambda_i)} m_{\lambda_i}(\nu_i) \exp i[S(\beta) + \nu_i, \varphi].$$

Multiplying by $\exp -i[\beta + \nu, \varphi]$, where ν is some dominant weight, and using orthogonality properties[2,7] on integration over all parameters, we get

$$\sum_{S\in W}\delta_S \sum_{\nu'\in D(\lambda')} m_{\lambda'}(\nu')\delta_{S(\beta+\lambda)+\nu',\beta+\nu} = \sum_{i=1}^{n}\gamma_{\lambda_i}\sum_{S\in W}\delta_S \sum_{\nu_i\in D(\lambda_i)} m_{\lambda_i}(\nu_i)\delta_{S(\beta)+\nu_i,\beta+\nu},$$

where $\delta_{x,y}$ is the Kronecker symbol.

Thus,

$$\sum_{S\in W}\delta_S m_{\lambda'}\{(\beta+\nu)-S(\beta+\lambda)\} = \sum_{i=1}^{n}\gamma_{\lambda_i}\sum_{S\in W}\delta_S m_{\lambda_i}\{(\beta+\nu)-S(\beta)\}. \quad (4)$$

Kostant[3] has given an expression for internal multiplicity, as follows:

$$m_\lambda(\nu) = \sum_{S\in W}\delta_S P[S(\beta+\lambda)-(\beta+\nu)], \quad (5)$$

where $P(\mu)$ is the partition function, i.e., the number of ways of writing μ as a linear combination of all the positive roots with nonnegative integral coefficients.

According to a lemma of Tarski,[5,6]

$$P^{(\alpha_k)}(\mu) \equiv P(\mu) - P(\mu-\alpha_k)$$

is a partition function over Δ with some α_k removed. Thus,

$$P^{(\alpha_1,\cdots,\alpha_m)}(\mu) \equiv \sum_{j_1=0}^{1}\cdots\sum_{j_m=0}^{1}(-1)^{\Sigma_1^m j_i}P\left(\mu-\sum_{i=1}^{m}j_i\alpha_i\right) \quad (6)$$

is the partition function[6] over the system $\Delta - \{\alpha_1,\cdots,\alpha_m\}$. Obviously, when $m = s$,

$$P^{(\Delta)}(\mu) = \delta_{\mu,0}, \quad (7)$$

since, by the lemma, $P^{(\Delta)}(\mu)$ is a partition function over zero roots.[6]

Consider the quantity

$$m_\lambda^{(\alpha_1,\cdots,\alpha_m)}(\nu) \equiv \sum_{j_1=0}^{1}\cdots\sum_{j_m=0}^{1}(-1)^{\Sigma_1^m j_i}m_\lambda\left(\nu+\sum_{i=1}^{m}j_i\alpha_i\right).$$

It is understood that $m_\lambda(\nu+\sum j_i\alpha_i) = 0$ when $\nu + \sum j_i\alpha_i \notin D(\lambda)$.

Using (5) and (6), we obtain

$$m_\lambda^{(\alpha_1,\cdots,\alpha_m)}(\nu) = \sum_{S\in W}\delta_S\sum_{j_1=0}^{1}\cdots\sum_{j_m=0}^{1}(-1)^{\Sigma_1^m j_i} \times P\left[S(\beta+\lambda)-(\beta+\nu)-\sum_{i=1}^{m}j_i\alpha_i\right] = \sum_{S\in W}\delta_S P^{(\alpha_1,\cdots,\alpha_m)}[S(\beta+\lambda)-(\beta+\nu)],$$

using the lemma.

[7] H. Weyl, *Classical Groups* (Princeton University Press, Princeton, N.J., 1946), p. 200.

Clearly,

$$m_\lambda^{(\Delta)}(\nu) = \sum_{S\in W}\delta_S\delta_{S(\beta+\lambda),\beta+\nu} \quad (8)$$

from (7).

Going back to (4), we see that

$$\sum_{i=1}^{n}\gamma_{\lambda_i}\sum_{S\in W}\delta_S m_{\lambda_i}^{(\Delta)}\{(\beta+\nu)-S(\beta)\} = \sum_{i=1}^{n}\gamma_{\lambda_i}\sum_{S\in W}\delta_S\sum_{j_1=0}^{1}\cdots\sum_{j_s=0}^{1}(-1)^{\Sigma^s j_k} \times m_{\lambda_i}\left\{(\beta+\nu)-S(\beta)+\sum_{k=1}^{s}j_k\alpha_k\right\} = \sum_{j_1=0}^{1}\cdots\sum_{j_s=0}^{1}(-1)^{\Sigma_1^s j_k}\sum_{S\in W}\delta_S \times m_{\lambda'}\left\{(\beta+\nu)-S(\beta+\lambda)+\sum_{k=1}^{s}j_k\alpha_k\right\} = \sum_{S\in W}\delta_S m_{\lambda'}^{(\Delta)}\{(\beta+\nu)-S(\beta+\lambda)\}.$$

Hence, using (8), we obtain

$$\sum_{i=1}^{n}\gamma_{\lambda_i}\sum_{S,S'\in W}\delta_{SS'}\delta_{S'(\beta+\lambda_i)+S(\beta),2\beta+\nu} = \sum_{S,S'\in W}\delta_{SS'}\delta_{S'(\beta+\lambda')+S(\beta+\lambda),2\beta+\nu}. \quad (9)$$

Here, $\delta_{SS'} \equiv \delta_S\delta_{S'}$.

III. LINEAR EQUATIONS

Since ν can be any one of the λ_i in (9), there are as many equations as there are unknowns, which are the multiplicities γ_{λ_i}.

When $\nu = \lambda_1 > \lambda + \lambda'$, the right side of (9) is always zero, since $2\beta + \nu$ will be greater than the highest value of $S'(\beta+\lambda') + S(\beta+\lambda)$, which is obtained when $S = S' = I$. On the left of (9), only $i = 1$ contributes, the Kronecker symbol being unity here. All the other terms are zero, due to the same reason cited above. Thus, $\gamma_{\lambda_1} = 0$. We thus see that $\lambda_1 = \lambda + \lambda'$, and $\gamma_{\lambda+\lambda'} = 1$ always. These facts are already known.[8]

We can rewrite (9) as

$$\sum_{i=1}^{n}A_{ki}\gamma_{\lambda_i} = B_k, \quad (k = 1,\cdots,n), \quad (10)$$

where

$$A_{ki} = \sum_{S,S'\in W}\delta_{SS'}\delta_{S'(\beta+\lambda_i)+S(\beta),2\beta+\lambda_k} \quad (11)$$

and

$$B_k = \sum_{S,S'\in W}\delta_{SS'}\delta_{S'(\beta+\lambda')+S(\beta+\lambda),2\beta+\lambda_k}.$$

Now, A_{kk} = unity, since only $S = S' = I$ contributes here, while $A_{ki} = 0$ when $k < i$, i.e., when $\lambda_k > \lambda_i$. We thus find that the coefficient matrix is

[8] N. Straumann, Helv. Phys. Acta. **38**, 56 (1965).

triangular and nonsingular, ensuring the uniqueness of the solutions.

From (10) and (11), we can write, for $k \neq 1$,

$$\gamma_{\lambda_k} = \sum_{S,S' \in W} \delta_{SS'} \delta_{S'(\beta+\lambda')+S(\beta+\lambda),\,2\beta+\lambda_k} - \sum_{i=1}^{k-1} \gamma_{\lambda_i} \sum_{S,S' \in W} \delta_{SS'} \delta_{S'(\beta+\lambda_i)+S(\beta),\,2\beta+\lambda_k}. \quad (12)$$

We thus obtain a kind of "*recurrence*" relation, with which, starting with the IR with highest weight $\lambda + \lambda'$, we can determine the multiplicities of the various IR's in steps. Some examples will be given in the Appendix to illustrate the usefulness of (12).

Note added in proof; It has been recently realized by the author that the intermediate expression (9) can also be got by substituting Eq. (3a) in (2), multiplying by $\exp -[2\beta + \nu, \varphi]$, and integrating. The interest, if any, in this procedure is in the contents of Sec. III. Note that the reduction of the problem to one of solving a system of linear equations depends crucially on the properties of strictly dominant weights.

ACKNOWLEDGMENTS

The author is grateful to Professor R. C. Majumdar and Professor F. C. Auluck for their interest in this work. He also thanks Dr. Bhama Srinivasan and Sri T. S. Santhanam for useful discussions.

APPENDIX

We consider the algebra A_2 whose root system consists of all vectors of the form $e_i - e_j$ $(i \neq j = 1, 2, 3)$ which lie in a plane $x_1 + x_2 + x_3 = 0$ in a three-dimensional space. The e_i are unit vectors. Here, $\beta = (1, 0, -1)$.

Consider for example, $\lambda = \lambda' = (1, 0, -1)$. Then, $\lambda + \lambda' = (2, 0, -2)$. Thus only the following IR's can occur in the Clebsch–Gordan series:

$$\begin{array}{ll} D^{27}(2,0,-2); & D^{21}(\tfrac{5}{3},\tfrac{5}{3},-\tfrac{10}{3}); \\ D^{24}(\tfrac{5}{3},\tfrac{2}{3},-\tfrac{7}{3}); & D^{15}(\tfrac{5}{3},-\tfrac{1}{3},-\tfrac{4}{3}); \\ D^{\bar{15}}(\tfrac{4}{3},\tfrac{4}{3},-\tfrac{8}{3}); & D^{15*}(\tfrac{4}{3},\tfrac{1}{3},-\tfrac{5}{3}); \\ D^{10}(2,-1,-1); & D^{\bar{10}}(1,1,-2); \\ D^{8}(1,0,-1); & D^{6}(\tfrac{4}{3},-\tfrac{2}{3},-\tfrac{2}{3}); \\ D^{\bar{6}}(\tfrac{2}{3},\tfrac{2}{3},-\tfrac{4}{3}); & D^{3}(\tfrac{2}{3},-\tfrac{1}{3},-\tfrac{1}{3}); \\ D^{\bar{3}}(\tfrac{1}{3},\tfrac{1}{3},-\tfrac{2}{3}); & D^{1}(0,0,0). \end{array}$$

The superscripts refer to the dimensionalities of the IR's.

Applying the procedure given in Sec. III, one finds that the multiplicities of those IR's with nonintegral components for the highest weight are zero. The nontrivial set of linear equations is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 \\ -1 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{27} \\ \gamma_{10} \\ \gamma_{\bar{10}} \\ \gamma_8 \\ \gamma_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 2 \end{bmatrix},$$

where we have labeled the γ's by the corresponding dimensionalities. Thus

$$\gamma_1 = \gamma_{\bar{10}} = \gamma_{10} = \gamma_{27} = 1, \quad \gamma_8 = 2,$$

which is a well-known result.

Consider next the algebra G_2, which has a root system similar to A_2, with the additional six roots $e_i - 2e_j + e_k$ $(i \neq j \neq k = 1, 2, 3)$. The components of a weight $m = (m_1, m_2, m_3)$ are integers, with $m_1 \geq 0$, $m_2 \leq 0$, $m_3 \leq 0$. The Weyl group, of order twelve, consists of all permutations, along with permutations with total change of sign of the components. Here, $\beta = (3, -1, -2)$.

Let $\lambda = (1, 0, -1)$ and $\lambda' = (2, -1, -1)$. The only IR's which can occur in the decomposition are

$$\begin{array}{ll} D^{64}(3,-1,-2); & D^{27}(2,0,-2); \\ D^{14}(2,-1,-1); & D^{7}(1,0,-1); \\ D^{1}(0,0,0). & \end{array}$$

The linear equations are

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 \\ 0 & -2 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{64} \\ \gamma_{27} \\ \gamma_{14} \\ \gamma_7 \\ \gamma_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}.$$

Hence,

$$\gamma_7 = \gamma_{64} = \gamma_{27} = 1, \quad \gamma_1 = \gamma_{14} = 0.$$

Symbolically,

$$7 \otimes 14 = 64 \oplus 27 \oplus 7.$$

Let $\lambda = (2, -1, -1)$ and $\lambda' = (2, 0, -2)$. The IR's which can occur are

$$\begin{array}{ll} D^{189}(4,-1,-3); & D^{77'}(4,-2,-2); \\ D^{77}(3,0,-3); & D^{64}(3,-1,-2); \\ D^{27}(2,0,-2); & D^{14}(2,-1,-1); \\ D^{7}(1,0,-1); & D^{1}(0,0,0). \end{array}$$

Here, we find that

$$\gamma_{189} = \gamma_{77} = \gamma_{64} = \gamma_{27} = \gamma_{14} = \gamma_7 = 1,$$
$$\gamma_{77'} = \gamma_1 = 0.$$

Symbolically,

$$14 \otimes 27 = 189 \oplus 77 \oplus 64 \oplus 27 \oplus 14 \oplus 7.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Theorems on the Ising Model with General Spin and Phase Transition

MASUO SUZUKI
The Institute for Solid State Physics, University of Tokyo, Roppongi, Tokyo

(Received 5 July 1968)

The theorem of Lee and Yang has been extended to the ferromagnetic Ising model with arbitrarily mixed spin values of $S_j = \frac{1}{2}$, 1, and $\frac{3}{2}$, including the case of equal spin values as a special one. Namely, it has been proved that the zeros of the partition function for the above Ising model with higher spin values lie on the unit circle in the fugacity plane (or complex magnetic-field plane). Expressions for general correlation functions in Ising ferromagnets with higher spin values have been derived in terms of the above generalized theorem. By the use of these expressions, the relations among the critical indices are discussed and the same results are obtained as those predicted by the scaling-law approach.

1. INTRODUCTION

Recently, the critical behavior of the second-order phase transition was investigated by the use of the distribution of the zeros of the partition function in the fugacity plane[1–4] and also in the complex temperature plane.[5–11] The known distributions of zeros in the Ising model are one-dimensional in both complex planes. But it is noted that in other systems, such as the modified Slater KDP model,[12] there exists a two-dimensional distribution of zeros, in which case the critical behaviour shows a quite different character.[9] In this paper, it is shown that the zeros of the ferromagnetic Ising model with arbitrarily mixed spin values of $S_j = \frac{1}{2}$, 1, and $\frac{3}{2}$ lie on a unit circle in the fugacity plane, which has been expected to hold from a computer experiment for finite lattices.[13] Exact solutions for $S = \frac{3}{2}$ and $S = 1$ have been known only in the one-dimensional Ising model.[14]

2. PROPERTIES OF THE PARTITION FUNCTION OF THE ISING MODEL

Consider a crystal lattice with an Ising spin of spin S_j at the jth lattice site. The total energy of this system is given by

$$\mathcal{H} = -\sum_{i>j} J_{ij} s_i s_j - mH \sum_j s_j, \tag{2.1}$$

where the symbol $i > j$ means that each pair (i,j) should be counted only once, and s_j is the z component of the spin operator ($s_j = S_j, S_j - 1, \cdots, -S_j$). The partition function Ξ_N is

$$\Xi_N = \mathrm{Tr} \exp\left(\sum_{i>j} K_{ij} s_i s_j + h \sum_j s_j\right), \tag{2.2}$$

where

$$K_{ij} = J_{ij}/kT \quad \text{and} \quad h = mH/kT. \tag{2.3}$$

It is convenient to introduce the following function f_n of the variables $z_1, \cdots, z_n$:

$$f_n(z_1, \cdots, z_n; S_1, \cdots, S_n) = \sum_{s_1=-S_1}^{S_1} \cdots \sum_{s_n=-S_n}^{S_n} \prod_{i>j} x_{ij}^{-s_i s_j} \prod_j z_j^{s_j}, \tag{2.4}$$

which corresponds to the partition function in an inhomogeneous field, putting

$$x_{ij} = e^{-K_{ij}} \quad \text{and} \quad z_j = e^{-h_j}. \tag{2.5}$$

Expanding the above function with respect to the variable z_1, the following recurrence formula is easily obtained:

$$f_n(\{z_j\}; \{S_j\}) = \sum_{k=-S_1}^{S_1} z_1^k A_k, \tag{2.6}$$

where

$$A_k = f_{n-1}(\{z_j x_{1j}^{-k}\}; S_2, \cdots, S_n). \tag{2.7}$$

It is of use to notice that the following simple relations hold in general for real parameters $\{x_{ij}\}$:

$$\text{(i)} \qquad f_n^*(\{z_j\}; \{S_j\}) = f_n(\{z_j^*\}; \{S_j\}), \tag{2.8}$$

$$\text{(ii)} \qquad f_n(\{1/z_j\}; \{S_j\}) = f_n(\{z_j\}; \{S_j\}), \tag{2.9}$$

and

(iii) if $|z_2| = \cdots = |z_n| = 1$, then the following symmetry relations hold:

$$A_{-k} = A_k^*; \quad k = S_1, S_1 - 1, \cdots, -S_1. \tag{2.10}$$

In this paper, we prove that the zeros of the partition function of the ferromagnetic Ising model

[1] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[2] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **38**, 289, 744, 1225 (1967).
[3] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **39**, 349 (1968).
[4] R. Abe, Progr. Theoret. Phys. (Kyoto) **38**, 72, 568 (1967).
[5] M. E. Fisher, *Lectures in Theoretical Physics VIIC* (University of Colorado Press, Boulder, Colorado, 1964), p. 1.
[6] G. L. Jones, J. Math. Phys. **7**, 2000 (1966).
[7] R. Abe, Progr. Theoret. Phys. (Kyoto) **37**, 1070 (1966); **38**, 322 (1967).
[8] S. Grossmann and W. Rosenhauer, Z. Physik **207**, 138 (1967).
[9] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **38**, 1243 (1967).
[10] S. Ono, Y. Karaki, M. Suzuki, and C. Kawabata, Phys. Letters **24A**, 703 (1967); J. Phys. Soc. Japan **25**, 54 (1968).
[11] S. Katsura, Progr. Theoret. Phys. (Kyoto) **38**, 1415 (1967).
[12] F. Y. Wu, Phys. Rev. Letters **18**, 605 (1967).
[13] C. Kawabata, M. Suzuki, S. Ono, and Y. Karaki, Phys. Letters (to be published).
[14] M. Suzuki, B. Tsujiyama, and S. Katsura, J. Math. Phys. **8**, 124 (1967).

with arbitrarily mixed spin values of $S_j = \frac{1}{2}$, 1, and $\frac{3}{2}$ lie on a unit circle in the fugacity plane, including the case of all $S_j = S$ ($S = \frac{1}{2}$, 1, or $\frac{3}{2}$) as a special one. Namely, we shall prove the following theorem:

Theorem 1: In the ferromagnetic system ($J_{ij} \geq 0$), all the roots of the equation

$$\Xi_N(\{x_{ij}\}; z; S_1, \cdots, S_N) = 0$$

are on the unit circle for $S_j = \frac{1}{2}$, 1, or $\frac{3}{2}$.

It is clear that

$$f_N(z, z, \cdots z; \{S_j\}) = \Xi_N(z; \{S_j\}). \tag{2.11}$$

Therefore, Theorem 1 is an immediate consequence of the following theorem:

Theorem 2: If $f_n(z_1, \cdots, z_n; \{S_j\}) = 0$ and $|z_1| \geq 1, \cdots, |z_n| \geq 1$, then

$$|z_1| = |z_2| = \cdots = |z_n| = 1$$

for $0 < x_{ij} \leq 1$ and $S_j = \frac{1}{2}$, 1, or $\frac{3}{2}$.

In the following sections, we assume that all the x's are different from 1. The proof can then be easily generalized to include the case when one or more of the x's are equal to 1. We prove Theorem 2 by mathematical induction as did Lee and Yang.

3. LEMMAS

Consider the following conjugate reciprocal equation:

$$a_0 z^m + a_1 z^{m-1} + \cdots + a_m = 0,$$
$$a_{m-k} = a_k^* \quad \text{and} \quad a_0 \neq 0. \tag{3.1}$$

Lemma 1: If $z_1 = re^{i\phi}$ ($r \neq 1$) is a solution of the equation, then $z_2 = (1/r)e^{i\phi}$ is another solution of Eq. (3.1).

This is evident.

Lemma 2: If m is odd, one of the roots of Eq. (3.1), at least, must have the absolute magnitude equal to 1.

Proof: If we assume that for all the roots $\{z_j\}$, $|z_j| \neq 1$, then, in terms of Lemma 1, the following $m + 1$ roots should exist:

$$z_1 = r_1 e^{i\phi_1}, \quad z_2 = (1/r_1)e^{i\phi_1}, \cdots, z_m = r_p e^{i\phi_p},$$
$$z_{m+1} = (1/r_p)e^{i\phi_p}; \quad p = \tfrac{1}{2}(m + 1).$$

This is contradictory to the Gauss theorem.

Lemma 3: If $|a_0| \geq |a_1|$ in Eq. (3.1) for $m = 1$, 2, or 3, then all the roots of Eq. (3.1) must have the absolute value equal to 1.

Proof: The case of $m = 1$ is evident from $a_1 = a_0^*$. Next, if we assume that one of the roots of Eq. (3.1) is expressed as $z_1 = re^{i\phi}$ ($r \neq 1$) in the case of $m = 2$, then the other root in terms of Lemma 1 is given by $z_2 = (1/r)e^{i\phi}$. Now the coefficients of Eq. (3.1) must satisfy the relation

$$a_1/a_0 = -(r + 1/r)e^{i\phi}. \tag{3.2}$$

Consequently, the following inequality should hold for one of the roots, at least, to have an absolute magnitude unequal to 1:

$$|a_1/a_0| = r + 1/r > 2. \tag{3.3}$$

Therefore, if $|a_0| \geq |a_1|$, at least, then all the roots must have an absolute magnitude equal to 1.

In the case of $m = 3$, in terms of Lemma 2, one of the roots can be written as $z_1 = e^{i\theta}$. If we assume that one of the other roots is expressed as $z_2 = re^{i\phi}$ ($r \neq 1$), then the third root is given by $z_3 = (1/r)e^{i\phi}$. Now, we obtain the following relation between the roots and the coefficients:

$$a_1/a_0 = -e^{i\theta} - (r + 1/r)e^{i\phi}. \tag{3.4}$$

Consequently, the following inequality should hold for one of the roots, at least, to have an absolute value unequal to 1:

$$\left|\frac{a_1}{a_0}\right| \geq \left|\frac{a_1}{a_0} + e^{i\theta}\right| - 1 = r + \frac{1}{r} - 1 > 1. \tag{3.5}$$

Therefore, if $|a_0| \geq |a_1|$, then all roots must have an absolute magnitude equal to 1.

The fact that Lemma 3 is proven only for $m = 1$, 2, and 3 is what limits the proof to $S_j \leq \frac{3}{2}$. Lemma 3 can be extended to higher m as the following theorem:

If

$$|a_0| \geq |a_1| + \cdots |a_{q-1}| + \left(\frac{m+1}{2} - q\right)|a_q|,$$
$$q = [m/2], \tag{3.6}$$

then all the roots of Eq. (3.1) have the absolute value equal to 1.

The proof and applications of this theorem will be reported in the forthcoming paper. If the sufficient condition (3.6) could be replaced by the inequality

$$|a_0| \geq |a_1| \geq |a_2| \geq \cdots \geq |a_q|, \tag{3.7}$$

then the proof in the subsequent sections would be valid to the case of $S_j \geq 2$. However, we can easily

find a counter example to the condition (3.7) for $m \geq 4$.

4. PROOF IN THE CASE OF $n = 1, 2$

In this section, Theorem 2 is proved for $n = 1$ and $n = 2$. In the case of $n = 1$, the proof is evident from the fact that the roots of the equation $f_1(z, S) = 0$ are given by

$$\omega_k = \exp\{2\pi ik/(2S + 1)\}, \quad k = 1, 2, \cdots, 2S. \tag{4.1}$$

In the case of $n = 2$, the relevant polynomial is expressed as

$$f_2(z_1, z_2; S_1, S_2) = \sum_{k=-S_1}^{S_1} z_1^k f_1(z_2 x_{12}^{-k}; S_2), \tag{4.2}$$

where $S_j = \frac{1}{2}$, 1, or $\frac{3}{2}$, $j = 1, 2$.

Assume that there exists a set of z's equal to z_1 and z_2 such that

$$f_2(z_1, z_2; S_1, S_2) = 0, \tag{4.3}$$

and

$$|z_1| > 1 \quad \text{and} \quad |z_2| \geq 1. \tag{4.4}$$

Regarding z_2 as a function of z_1 defined by (4.3), one obtains a limit $\mathscr{y}_2$ for z_2 as $z_1 \to \infty$, which satisfies the equation

$$f_1(\mathscr{y}_2 x^{-S_1}; S_2) = 0, \quad x = x_{12}. \tag{4.5}$$

From the discussion for $n = 1$, one obtains

$$|\mathscr{y}_2 x^{-S_1}| = 1.$$

Now by condition $0 < x < 1$ we have

$$|\mathscr{y}_2| < 1.$$

Therefore, if we assume (4.3) and (4.4), there should exist a set of values z_1', z_2' such that

$$f_2(z_1', z_2'; S_1, S_2) = 0$$

and

$$|z_1'| > 1, \quad |z_2'| = 1. \tag{4.6}$$

On the other hand, let us consider the equation

$$f_2(z_1, z_2; S_1, S_2) = 0, \tag{4.7}$$

for $|z_2| = 1$. As the function $z^{S_2} f_1(z, S_2)$ is a polynomial of the order $2S_2$, it can be decomposed into the form

$$z^{S_2} f_1(z; S_2) = \prod_{k=1}^{2S_2} (z - \omega_k), \tag{4.8}$$

where

$$\omega_k = \exp\{2\pi ik/(2S_2 + 1)\}$$

and $|\omega_k| = 1$. If we define a function $\phi(t)$ by

$$\phi(t) = |f_1(tz_2; S_2)|^2, \quad \text{for } |z_2| = 1,$$

then we obtain easily

$$\phi'(t) = \phi(t) \sum_{k=1}^{2S_2} (1 - t^{-2})/|t^{\frac{1}{2}} - \omega_k/t^{\frac{1}{2}} z_2|^2, \tag{4.9}$$

and consequently

$$\phi'(t) > 0, \quad \text{for } t > 1. \tag{4.10}$$

Namely, $\phi(t)$ is a monotonically increasing function for $t > 1$. Therefore, the following inequality is obtained:

$$|f_1(z_2 x^{-S_1}; S_2)| \geq |f_1(z_2 x^{-S_1+1}; S_2)| \tag{4.11}$$

for $0 < x < 1$ and $|z_2| = 1$. The sign of equality in Eq. (4.11) is valid for $S_1 = \frac{1}{2}$. By the use of Lemma 3 and the formula (4.2), the root z_1 of Eq. (4.7) must satisfy

$$|z_1| = 1,$$

which contradicts (4.6). This means that the assumption (4.4) is not true. Thus, Theorem 2 has been proved for $n = 1$ and $n = 2$.

5. MAIN PARTS OF INDUCTION

Assume that Theorem 2 is true for $n = m - 2$ and $n = m - 1$, but not true for $n = m$. We shall show that the above assumption leads to the inconsistent situation that the results of the following two parts are contradictory to each other.

Part 1: Under the above assumption that Theorem 2 is not true for $n = m$, there exists a set of z's equal to $z_1, \cdots, z_m$ such that

$$f_m(z_1, \cdots, z_m, S_1, \cdots, S_m) = 0 \tag{5.1}$$

and

$$|z_1| > 1, \quad \text{and } |z_2|, \ |z_3|, \cdots, |z_m|, \ \forall \geq 1. \tag{5.2}$$

We can repeat Lee–Yang's procedure in a generalized form as follows. Keeping $z_3, \cdots, z_m$ fixed and regarding z_2 as a function of z_1 defined by (5.1), one obtains a limit $\mathscr{y}_2$ for z_2 as $z_1 \to \infty$, which is given by the equation of the order $2S_2$,

$$f_{m-1}(\mathscr{y}_2 x_{12}^{-S_1}, z_3 x_{13}^{-S_1}, \cdots, z_m x_{1m}^{-S_1}; \{S_j\}) = 0, \tag{5.3}$$

unless the coefficient of the highest order in Eq. (5.3),

$$f_{m-2}(\{z_j x_{1j}^{-S_1} x_{2j}^{-S_2}\}; \{S_j\}; j \geq 3), \tag{5.4}$$

vanishes. Now write

$$\mathscr{y}_2 x_{12}^{-S_1} = \zeta_2, \quad z_3 x_{13}^{-S_1} = \zeta_3, \cdots, z_m x_{1m}^{-S_1} = \zeta_m. \tag{5.5}$$

Equation (5.3) reduces to

$$f_{m-1}(\zeta_2, \cdots, \zeta_m; \{S_j\}) = 0. \tag{5.6}$$

Now, by condition (5.2),

$$|\zeta_j| > |z_j| \geq 1, \quad j = 3, 4, \cdots, m. \tag{5.7}$$

Under the assumption that Theorem 2 is true for $n = m - 1$, one obtains

$$|\mathcal{Z}_2| < |\zeta_2| < 1,$$

unless the polynomial (5.4) vanishes. In fact, the polynomial (5.4) does not vanish from the inequality (5.7) and from the assumption that Theorem 2 is true for $n = m - 2$. Therefore, keeping $z_3, \cdots, z_m$ fixed one can increase $|z_1|$ and define z_2 as a continuous function of z_1. Since by (5.2), $|z_2|$ starts to be > 1 in absolute magnitude and tends to a limit < 1 in absolute magnitude as $z_1 \to \infty$, there must be a value equal to z_1' so that z_2 assumes a value z_2' equal to 1 in absolute magnitude, i.e.,

$$f_m(z_1', z_2', z_3, \cdots, z_m; \{S_j\}) = 0 \tag{5.8}$$

and

$$|z_1'| > 1, \quad |z_2'| = 1, \quad |z_3|, \cdots, |z_m| \geq 1.$$

We can fix $z_2', z_4, \cdots, z_m$ and regard z_3 as a function of z_1' and follow the same procedure as mentioned above. Continuing this way we finally get a set of values $z_1'', \cdots, z_m''$ such that

$$f_m(z_1'', \cdots, z_m''; \{S_j\}) = 0$$

and

$$|z_1''| > 1, \quad |z_2''| = \cdots = |z_m''| = 1. \tag{5.9}$$

Part 2: In this paragraph, we shall prove that if $|z_2| = \cdots = |z_m| = 1$ and

$$f_m(\{z_j\}; \{S_j\}) = 0, \tag{5.10}$$

then

$$|z_1| = 1. \tag{5.11}$$

Proof: As the function $z_k^{S_k} f_{m-1}(\{z_j\}; \{S_j\})$ is a polynomial of the order $2S_k$ with respect to the variable z_k, it can be decomposed into the form

$$z_k^{S_k} f_{m-1}(\{z_j\}; \{S_j\}) = B_k \prod_{j=1}^{2S_k} (z_k - \omega_{j,k}(\{z\}_k)), \tag{5.12}$$

where $\{z\}_k$ indicates a set of variables $z_2, \cdots, z_{k-1}, z_{k+1}, \cdots, z_m$. Theorem 2 for $n = m - 1$ asserts that if $|z_2| \geq 1, \cdots, |z_{k-1}| \geq 1, |z_{k+1}| \geq 1, \cdots, |z_m| \geq 1$, then

$$|\omega_{j,k}(\{z\}_k)| \leq 1. \tag{5.13}$$

Now define a function $\phi_{m-1}(t_2, \cdots, t_m)$ by

$$\phi_{m-1}(t_2, \cdots, t_m) = |f_{m-1}(t_2 z_2, \cdots, t_m z_m; \{S_j\})|^2 \quad \text{for} \quad |z_2| = \cdots = |z_m| = 1. \tag{5.14}$$

We obtain easily

$$\frac{\partial \phi_{m-1}}{\partial t_k} = \phi_{m-1} \sum_{j=1}^{2S_k} (1 - |\omega_{j,k}(\{tz\}_k)|^2/t_k^2) \Big/ \left| t_k^{\frac{1}{2}} - \frac{\omega_{j,k}(\{tz\}_k)}{t_k^{\frac{1}{2}} z_k} \right|^2 > 0, \quad \text{for all} \quad t_k > 1 \tag{5.15}$$

Namely, $\phi_{m-1}(t_2, \cdots, t_m)$ is a monotonically increasing function of all t_k for $t_k > 1$. Therefore, the following inequality is obtained:

$$|f_{m-1}(\{x_{1j}^{-k} z_j\}; \{S_j\})| \geq |f_{m-1}(\{x_{1j}^{-k'} z_j\}; \{S_j\})|, \quad \text{for} \quad |z_j| = 1 \quad \text{and} \quad k > k' \geq 0. \tag{5.16}$$

In particular, from Eq. (2.7), we obtain

$$|A_{S_1}| \geq |A_{S_1 - 1}| \quad \text{for} \quad |z_2| = \cdots = |z_m| = 1. \tag{5.17}$$

The sign of equality in Eq. (5.17) is valid for $S_1 = \frac{1}{2}$. Therefore, in terms of Lemma 3 and the recurrence formula (2.6), one obtains that if $|z_2| = \cdots = |z_m| = 1$ and

$$f_m(\{z_j\}; \{S_j\}) = 0,$$

then

$$|z_1| = 1.$$

Thus, the results of Part 1 and Part 2 contradict each other, which means that Theorem 2 must hold for $n = m$.

This completes the proof of Theorem 2 by induction.

6. EXPRESSIONS FOR THERMODYNAMIC QUANTITIES AND CORRELATION FUNCTIONS

In terms of Theorem 1, the partition function of the ferromagnetic Ising model with arbitrarily mixed spin values of $S_j = \frac{1}{2}$, 1, and $\frac{3}{2}$ is expressed as

$$\Xi_N = \left(\prod_{i>j} x_{ij}^{-S_i S_j} \right) z^{-(S_1 + \cdots + S_N)} \prod_{j=1}^{N} \prod_{k=1}^{2S_k} (z - \omega_{j,k}), \tag{6.1}$$

where

$$\omega_{j,k} = \exp(i\theta_{j,k}).$$

It is evident that if $\omega = e^{i\theta}$ is a zero of Ξ_N, then $\omega' = e^{-i\theta}$ is also another zero. Neglecting a constant term, the free energy of the above system for infinite N is represented as

$$-\frac{F}{kT} = \int_0^\pi g_{\{S\}}(\theta, t) \log 2(\cosh h - \cos\theta)\, d\theta, \tag{6.2}$$

where $t = (T - T_c)/T_c$, $h = mH/kT$, and $g_{\{S\}}(\theta, t)$ is the distribution function of the zeros of the partition function in the fugacity plane under the configuration of spin $\{S\}$. The magnetization is obtained from Eq. (6.2):

$$M = m \sinh h \int_0^\pi \frac{g_{\{S\}}(\theta, t)}{\cosh h - \cos\theta}\, d\theta, \tag{6.3}$$

and the spontaneous magnetization is expressed in term of $g_{\{S\}}(\theta, t)$ as follows:

$$M_s = \pi m g_{\{s\}}(0, t). \tag{6.4}$$

The susceptibility is derived from Eqs. (6.3) and (6.4) as follows:

$$\chi_0 = \frac{m^2}{kT} \int_0^\pi \frac{g_{\{S\}}(\theta, t) - g_{\{S\}}(0, t)}{1 - \cos\theta} d\theta. \tag{6.5}$$

Correlation functions of *even number* of spins can be expressed as[3]

$$\langle \hat{s}_i \cdots \hat{s}_j \rangle = 1 + \int_0^\pi \frac{\rho_e^{\{S\}}(\theta, i \cdots j, t)}{2(\cosh h - \cos\theta)} g_{\{S\}}(\theta, t)\, d\theta, \tag{6.6}$$

where $\hat{s}_j = s_j/S_j$ and $\rho_e^{\{S\}}$ is the spectral intensity of the correlation function $\{\hat{s}_i \cdots \hat{s}_j\}$. Correlation functions of *odd number* of spins can be also expressed as[3]

$$\langle \hat{s}_i \hat{s}_j \cdots \hat{s}_m \rangle = \int_0^\pi \frac{(\sinh h)\rho_o^{\{S\}}(\theta, ij \cdots m, t)}{\cosh h - \cos\theta} g_{\{S\}}(\theta, t)\, d\theta. \tag{6.7}$$

The distribution function $g_{\{S\}}(\theta, t)$ and the spectral intensity $\rho_o^{\{S\}}(\theta, ij \cdots m, t)$ satisfy the following normalization conditions

$$\int_0^\pi g_{\{S\}}(\theta, t)\, d\theta = \frac{1}{2}, \tag{6.8}$$

and

$$\int_0^\pi g_{\{S\}}(\theta, t)\rho_o^{\{S\}}(\theta, ij \cdots m, t)\, d\theta = 1. \tag{6.9}$$

These expressions can be fruitfully used in order to investigate the singularities of the thermodynamic quantities and correlation functions near the transition point.[2-4]

7. DISCUSSION

As was discussed in previous papers,[2-4] the distribution function of zeros for small θ and t, in terms of Eqs. (6.4) and (6.5), has been derived in the form

$$g(\theta, t) = t^\beta f(\theta t^{-\beta-\gamma}). \tag{7.1}$$

As usual, α, β, and γ denote the indices of specific heat, spontaneous magnetization, and susceptibility, respectively. The magnetic equation of state is given by

$$M = t^\beta \phi_m(ht^{-\beta-\gamma}), \tag{7.2}$$

in terms of Eqs. (6.3) and (7.1), which agrees with the results predicted by the scaling law.[15-17] Namely, we obtain the relations[15-18]

$$\alpha + 2\beta + \gamma = 2 \tag{7.3}$$

and

$$\alpha + \beta(1 + \delta) = 2, \tag{7.4}$$

where δ is the index of the critical magnetization. The index of the correlation length ν is expressed as[3]

$$\nu = (2\beta + \gamma)/d, \tag{7.5}$$

in terms of Eqs. (6.6) and (6.7), where d is the dimensionality of the system.

As a special case, Theorem 1 holds for all $S_j = S$ ($S = \frac{1}{2}$, 1, or $\frac{3}{2}$). Quite recently, Asano[19] proved the case of $S = 1$. As another special case, Theorem 1 can be applied to the ferromagnetic Ising model consisting of two sublattices with different spin values.

ACKNOWLEDGMENTS

We wish to thank T. Asano for sending us his paper prior to publication, though the work was carried out independently.

This work has partially been financed by the Sakkokai foundation.

[15] B. Widom, J. Chem. Phys. **43**, 3898 (1965).
[16] L. P. Kadanoff, Physics **2**, 263 (1966).
[17] L. P. Kadanoff *et al.*, Rev. Mod. Phys. **39**, 395 (1967).
[18] J. W. Essam and M. E. Fisher, J. Chem. Phys. **38**, 802 (1963).
[19] T. Asano, Progr. Theoret. Phys. (Kyoto) **40** (1968)(to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Eigenvalue Problem for Lagrangian Systems. III

E. M. BARSTON
Courant Institute of Mathematical Sciences, New York University, New York, New York

(Received 13 September 1967)

The quadratic Lagrangian eigenvalue problem $[\lambda^2 P + \lambda Q - (L + B)]\zeta = 0$ and the associated time-dependent problem $P\ddot{\zeta} + iQ\dot{\zeta} + (L + B)\zeta(t) = 0$ are investigated for the case where P, Q, and B are bounded linear Hermitian operators in Hilbert space, P is positive and invertible, L possesses a positive completely continuous Hermitian inverse, and $L + B > 0$. Existence and completeness theorems for the eigenvectors as well as variational characterizations of the eigenvalues are given, and the general solution of the time-dependent problem is obtained in terms of an eigenvector expansion. Finally, these results are applied to the problem of small oscillations of a rotating elastic string.

I. INTRODUCTION

Existence and completeness theorems for the eigenvectors of quadratic Lagrangian eigenvalue problems involving completely continuous Hermitian operators have been given in two earlier papers.[1,2] Apart from finite-dimensional examples, the great majority of physical problems involve differential operators. We deal herein with one such type of problem, illustrated by the vibrations of a rotating elastic string, where the differential operator possesses a completely continuous positive inverse. This problem can be reduced to an equivalent problem already treated in Ref. 2, and existence and completeness theorems as well as variational characterizations of the eigenvalues are thereby obtained. The time-dependent problem is also discussed, and the solution is obtained as an eigenvector expansion.

A note as to notation: Theorem 2.III denotes Theorem III of Ref. 2, while Theorem 1.V denotes Theorem V of Ref. 1.

II. EIGENVALUES, EIGENVECTORS, ETC.

Theorem I: Let P, Q, B, and K be bounded linear Hermitian operators on and into the Hilbert space E, and suppose that K is positive semidefinite and completely continuous. Let L be a linear operator with domain $D_L \subset E$ and range $R_L \subset E$ with the property that $KL = I$ on D_L, and suppose that $PK(E) \subset R_L$, $QK(E) \subset R_L$, and $BK(E) \subset R_L$. Then

$$\{\lambda^2 P + \lambda Q - (L + B)\}\zeta = 0, \quad \zeta \in D_L, \quad \lambda \neq 0, \quad (1)$$

holds if and only if

$$\{\lambda^{-2}(I + kBk) - \lambda^{-1}kQk - kPk\}\eta = 0, \quad (2)$$

with $\zeta = k\eta$, $\eta = kL\zeta$, where $k \equiv K^{\frac{1}{2}}$ and the operators kBk, kQk, and kPk are completely continuous and Hermitian.

Proof: Suppose Eq. (1) holds. Then $\zeta \in D_L$ implies $k^2 L\zeta = \zeta$. Define $\eta \equiv kL\zeta$, so that $k\eta = \zeta$. Equation (2) follows at once by multiplying Eq. (1) on the left with $\lambda^{-2}k$. Now suppose Eq. (2) holds. Multiplying Eq. (2) by λ^2 and setting $\zeta \equiv k\eta$, we obtain

$$\eta = kH_\lambda\zeta, \quad H_\lambda \equiv \lambda^2 P + \lambda Q - B, \quad (3)$$

$$\zeta = k\eta = k^2 H_\lambda\zeta = KH_\lambda KH_\lambda\zeta = Kf, \quad (4)$$

where $f \equiv H_\lambda KH_\lambda\zeta \in R_L$. Thus for some $x \in D_L$ we have $f = Lx$, which leads to $\zeta = Kf = KLx = x$. Therefore $\zeta \in D_L$ and $L\zeta = f = H_\lambda KH_\lambda\zeta = H_\lambda\zeta$, which gives Eq. (1), and $\eta = kH_\lambda\zeta$ then implies $\eta = kL\zeta$. Since K is a completely continuous positive Hermitian operator, so is $k = K^{\frac{1}{2}}$, and the complete continuity of kBk, kQk, and kPk follows from the boundedness of B, Q, and P.

Theorem II: Let the hypothesis of Theorem I hold, $L + B > 0$ on D_L ($D_L \neq 0$), $P \geq 0$, and $\|kPk\| + \|kQk\| > 0$. Then Eq. (1) possesses a nonempty sequence [finite or infinite, with at most 2(dim R_K) elements] of eigenvectors $\{\zeta_n\}$ ($n = 1, 2, \cdots$) with real nonzero eigenvalues λ_n having the following properties:

(A) $\zeta_n \in D_L$, $\{\lambda_n^2 P + \lambda_n Q - (L + B)\}\zeta_n = 0$, $n = 1, 2, \cdots$. (5)

(B) $\{|\lambda_n|\}$ is a monotonic, nondecreasing sequence and $|\lambda_n| \to \infty$ as $n \to \infty$. If $P > 0$ on R_k and dim $R_K = \infty$, there are an infinite number of $\lambda_n > 0$ as well as an infinite number of $\lambda_n < 0$.

(C) Dim $M(\lambda_n) < \infty$, where $M(\lambda_n)$ is the linear manifold of eigenvectors with eigenvalue λ_n.

(D) $\lambda_m\lambda_n(\zeta_m, P\zeta_n) + (\zeta_m, [L + B]\zeta_n) = \delta_{mn}$, $m, n = 1, 2, \cdots$. (6)

[1] E. M. Barston, J. Math. Phys. **8**, 523 (1967).
[2] E. M. Barston, J. Math. Phys. **8**, 1886 (1967).

(E) Let either of the following statements hold:

(E1) $\inf_E \frac{(\zeta, P\zeta)}{(\zeta, \zeta)} > 0, \quad \bar{D}_L = E, \quad \dot{y} \in R_k, \quad y \in R_k,$

(E2) $P = I, \quad B = 0, \quad \dot{y} \in R_k, \quad y \in R_K.$

Then

$$y = \sum_m \beta_m \zeta_m, \tag{7}$$

$$\dot{y} = \sum_m \beta_m \lambda_m \zeta_m, \tag{8}$$

where

$$\beta_m = ([L + B]\zeta_m, y) + \lambda_m (P\zeta_m, \dot{y}). \tag{9}$$

(F) Let

$$\inf_E \frac{(\zeta, P\zeta)}{(\zeta, \zeta)} > 0,$$

and $\bar{D}_L = E$ or $P = I$ and $B = 0$. Then, for $y \in D_L$, $\dot{y} \in R_k$, and γ_m $(m = 1, 2, \cdots, n)$ any n complex numbers, we have

$$\begin{aligned}
\left\| y - \sum_1^n \gamma_m \zeta_m \right\|_T^2 + \left\| \dot{y} - \sum_1^n \gamma_m \lambda_m \zeta_m \right\|_P^2 &= \left\| y - \sum_1^n \beta_m \zeta_m \right\|_T^2 \\
&\quad + \left\| \dot{y} - \sum_1^n \beta_m \lambda_m \zeta_m \right\|_P^2 + \sum_1^n |\beta_m - \gamma_m|^2 \\
&\geq \left\| y - \sum_1^n \beta_m \zeta_m \right\|_T^2 + \left\| \dot{y} - \sum_1^n \beta_m \lambda_m \zeta_m \right\|_P^2 \\
&= \|y\|_T^2 + \|\dot{y}\|_P^2 - \sum_1^n |\beta_m|^2, \qquad (10)
\end{aligned}$$

$$\|y\|_T^2 + \|\dot{y}\|_P^2 = \sum_m |\beta_m|^2, \tag{11}$$

where the β_m are given by Eq. (9), $T \equiv L + B$, and $\|f\|_P^2 \equiv (f, Pf)$.

Proof: Theorem I implies that, for $\lambda_n \neq 0$, Eq. (1) [and (5)] is equivalent to

$$\{\lambda_n^{-2}(I + kBk) - \lambda_n^{-1} kQk - kPk\}\eta_n = 0, \tag{12}$$

with

$$\zeta_n = k\eta_n, \quad \eta_n = kL\zeta_n. \tag{13}$$

$L + B > 0$ on D_L implies that

$$\Delta \equiv \inf_E \frac{(\xi, [I + kBk]\xi)}{(\xi, \xi)} > 0,$$

so that there exists a positive bounded Hermitian operator r with a positive bounded Hermitian inverse r^{-1}, such that $r^2 = I + kBk$ $(r = [I + kBk]^{\frac{1}{2}})$. Indeed, suppose $\Delta \leq 0$. We have $\Delta = 1 + \lambda$, where

$$\lambda \equiv \inf_E \frac{(\xi, kBk\xi)}{(\xi, \xi)},$$

so that $\lambda \leq -1$. Then there exists $\zeta \in E$, $\zeta \neq 0$, such that $kBk\zeta = \lambda\zeta$, $k\zeta \neq 0$, and we have $k\zeta = \lambda^{-1}k^2Bk\zeta = \lambda^{-2}k^2 f$, where $f \equiv Bk^2Bk\zeta \in R_L$. Therefore $k\zeta \in D_L$, $Bk\zeta = Bk^2Lk\zeta \in R_L$, so that $\lambda Lk\zeta = Bk\zeta$. Thus $(k\zeta, [L + B]k\zeta) = (1 + \lambda)(k\zeta, Lk\zeta)$ and $(k\zeta, Lk\zeta) = (KLk\zeta, Lk\zeta) = (Lk\zeta, KLk\zeta) \geq 0$, so that $\lambda \leq -1$ implies $(k\zeta, [L + B]k\zeta) \leq 0$ for $k\zeta \neq 0$, which contradicts $L + B > 0$ on D_L. Hence $\Delta > 0$, and $r = [I + kBk]^{\frac{1}{2}}$ has the properties stated above. Equation (12), and therefore Eq. (1) (for $\lambda_n \neq 0$), is then equivalent to

$$\{\omega_n^2 I - \omega_n iA - H\}\xi_n = 0, \tag{14}$$

where

$$\omega_n \equiv \lambda_n^{-1}, \quad iA \equiv r^{-1}kQkr^{-1}, \quad H \equiv r^{-1}kPkr^{-1} \tag{15}$$

and

$$\begin{aligned}
\xi_n &\equiv r\eta_n = rkL\zeta_n, \\
\zeta_n &= k\eta_n = kr^{-1}\xi_n. \qquad (16)
\end{aligned}$$

The problem is now reduced to the case of Theorem 2.II, since $P \geq 0$ and $\|kPk\| + \|kQk\| > 0$ imply $H \geq 0$ and $\|H\| + \|A\| > 0$. If $P > 0$ on R_k and $\dim R_K = \infty$, then H has infinitely many positive eigenvalues. Theorem 1.V then implies the existence of an infinite number of positive (as well as negative) λ_n. Statements (A), (B), and (C) follow from Theorem 2.II (A), (B), and (C). Note that $R_{iA} \subset R_{r^{-1}k}$ and $R_H \subset R_{r^{-1}k}$, so that we could restrict Eq. (14) to the Hilbert space $\bar{R}_{r^{-1}k}$ with no loss of eigenvectors. Since $\dim \bar{R}_{r^{-1}k} = \dim R_K$, Theorem 2.II yields $2(\dim R_K)$ as the maximum number of eigenvectors ζ_n. Equations (15) and (16) and Theorem 2.II (D) give

$$(rkL\zeta_m, rkL\zeta_n) + \lambda_m\lambda_n(\xi_m, r^{-1}kPkr^{-1}\xi_n) = \delta_{mn}. \tag{17}$$

Now,

$$\begin{aligned}
(\xi_m, r^{-1}kPkr^{-1}\xi_n) &= (kr^{-1}\xi_m, Pkr^{-1}\xi_n) \\
&= (\zeta_m, P\zeta_n)
\end{aligned}$$

and

$$\begin{aligned}
(rkL\zeta_m, rkL\zeta_n) &= (kr^2kL\zeta_m, L\zeta_n) \\
&= (k[I + kBk]kL\zeta_m, L\zeta_n) \\
&= ([k^2 + k^2Bk^2]L\zeta_m, L\zeta_n) \\
&= (\zeta_m, [L + B]\zeta_n),
\end{aligned}$$

so that (D) holds. In the circumstance that $\bar{D}_L = E$, we have $\bar{R}_K = E(D_L \subset R_K)$, $K > 0$, $k > 0$, and $H > 0$ if $P > 0$. Therefore the set $S(x)$ [cf. Theorem 2.II (E) and (F)] satisfies $\bar{S}(x) = E$. If $P = I$ and

$B = 0$, then $r^{-1} = I$, $H^{\frac{1}{2}} = k$, and $S(x) = R_k$. Let (E1) or (E2) hold. Then there exist $x \in E$, $\dot{x} \in S(x)$, such that $y = kr^{-1}\dot{x}$, $\dot{y} = kr^{-1}x$, and Theorem 2.II (F) yields

$$H^{\frac{1}{2}}x = \sum_m \beta_m \lambda_m H^{\frac{1}{2}}\xi_m \tag{18}$$

and

$$\dot{x} = \sum_m \beta_m \xi_m, \tag{19}$$

where

$$\beta_m \equiv \lambda_m^{-1}\alpha_m = \lambda_m(\xi_m, Hx) + (\xi_m, \dot{x}). \tag{20}$$

We have

$$\begin{aligned}(\xi_m, Hx) &= (\xi_m, r^{-1}kPkr^{-1}x)\\ &= (Pkr^{-1}\xi_m, kr^{-1}x) = (P\zeta_m, \dot{y}),\\ (\xi_m, \dot{x}) &= (rkL\zeta_m, \dot{x}) = (L\zeta_m, kr^2r^{-1}\dot{x})\\ &= (L\zeta_m, k[I + kBk]r^{-1}\dot{x})\\ &= (L\zeta_m, y) + (L\zeta_m, KBy)\\ &= ([L + B]\zeta_m, y),\end{aligned}$$

so that Eq. (9) holds. Equations (16), (19), and the fact that kr^{-1} is bounded yield Eq. (7). Equation (8) follows from

$$\inf_E \frac{(\zeta, P\zeta)}{(\zeta, \zeta)} > 0$$

and Eq. (18), since

$$\begin{aligned}&\left\| H^{\frac{1}{2}}\left\{x - \sum_1^n \beta_m\lambda_m\xi_m\right\}\right\|^2\\ &= \left(x - \sum_1^n \beta_m\lambda_m\xi_m, r^{-1}kPkr^{-1}\left[x - \sum_1^n \beta_m\lambda_m\xi_m\right]\right)\\ &= \left(\dot{y} - \sum_1^n \beta_m\lambda_m\zeta_m, P\left[\dot{y} - \sum_1^n \beta_m\lambda_m\zeta_m\right]\right).\end{aligned}$$

Let $y \in D_L$, $\dot{y} \in R_k$. Then there exists $x \in E$ such that $\dot{y} = kr^{-1}x$, and $y = kr^{-1}\dot{x}$ for $\dot{x} = rkLy$. Statement (F) is now an immediate consequence of Theorem 2.II (G) and (H), Eqs. (15), (16), and (20), and the fact that, for $\zeta \in D_L$, $\eta \in E$, we have

$$\begin{aligned}\|rkL\zeta\|^2 &= (rkL\zeta, rkL\zeta)\\ &= (kr^2kL\zeta, L\zeta) = (k[I + kBk]kL\zeta, L\zeta)\\ &= (\zeta, [L + B]\zeta) = \|\zeta\|_T^2,\end{aligned} \tag{21}$$

$$\|\eta\|_H^2 = (\eta, H\eta) = (kr^{-1}\eta, Pkr^{-1}\eta) = \|kr^{-1}\eta\|_P^2. \tag{22}$$

This completes the proof of Theorem II.

Theorem III: Let the hypothesis of Theorem I hold, $P \geq 0$, $L + B > 0$ on D_L, and suppose that the positive and negative eigenvalues of Eq. (1) are arranged in the nondecreasing and nonincreasing sequences $\lambda_1^+ \leq \lambda_2^+ \leq \cdots$, $\lambda_1^- \geq \lambda_2^- \geq \cdots$, respectively. Denote the corresponding eigenvectors which satisfy Eq. (6) by $\zeta_n^\pm$. Then

$$(\lambda_n^+)^{-1} = \max_{\eta \in s_n^+} F(\eta), \tag{23}$$

$$(\lambda_n^-)^{-1} = \min_{\eta \in s_n^-} F(\eta), \tag{24}$$

$$(\lambda_n^+)^{-1} = \min_{\substack{\phi_m \in E\times E\\ m=1,2,\cdots,n-1}} \max_{\substack{(\phi_m,\eta)_2=0\\ m=1,2,\cdots,n-1\\ \eta \in D_L\times D_L}} F(\eta), \tag{25}$$

$$(\lambda_n^-)^{-1} = \max_{\substack{\phi_m \in E\times E\\ m=1,2,\cdots,n-1}} \min_{\substack{(\phi_m,\eta)_2=0\\ m=1,2,\cdots,n-1\\ \eta \in D_L\times D_L}} F(\eta), \tag{26}$$

where

$$(\phi, \eta)_2 \equiv (\phi_1, \eta_1) + (\phi_2, \eta_2), \tag{27}$$

$$\begin{aligned}s_n^\pm \equiv \{\eta \mid \eta \in D_L \times D_L, &([L + B]\zeta_m^\pm, \eta_2)\\ &+ \lambda_m^\pm(P\zeta_m^\pm, \eta_1) = 0, \quad m = 1, 2, \cdots, n-1\},\end{aligned} \tag{28}$$

and

$$F(\eta) \equiv \frac{2\,\mathrm{Re}\,(\eta_1, P\eta_2) + (\eta_2, Q\eta_2)}{(\eta_1, P\eta_1) + (\eta_2, [L + B]\eta_2)}. \tag{29}$$

Proof: Let

$$\eta = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \in D_L \times D_L.$$

Set $u_1 \equiv rkL\eta_1$, $u_2 \equiv rkL\eta_2$, so that $\eta_1 = kr^{-1}u_1$, $\eta_2 = kr^{-1}u_2$ (cf. Theorems I and II). Then we have

$$\begin{aligned}&(\eta_1, P\eta_2) = (u_1, Hu_2),\ (\eta_2, Q\eta_2) = (u_2, iAu_2),\\ &(\eta_1, P\eta_1) = (u_1, Hu_1),\ (\eta_2, [L + B]\eta_2) = (u_2, u_2),\end{aligned} \tag{30}$$

where H and iA are given by Eq. (15). Referring to Eqs. (14)–(16) and Theorem 2.III, we have the correspondence $\omega_n^\pm = (\lambda_n^\pm)^{-1}$, $\xi_m^\pm = rkL\zeta_m^\pm$, $\zeta_m^\pm = kr^{-1}\xi_m^\pm$, so that $\eta \in s_n^\pm$ implies

$$u = \begin{pmatrix}u_1\\ u_2\end{pmatrix} \in V_n^\pm,$$

and therefore Theorem 2.III gives

$$(\lambda_n^+)^{-1} \geq \max_{\eta \in s_n^+} F(\eta), \tag{31}$$

$$(\lambda_n^-)^{-1} \leq \min_{\eta \in s_n^-} F(\eta). \tag{32}$$

Equations (23) and (24) now follow immediately from the fact that

$$F(\eta) = (\lambda_n^\pm)^{-1} \quad \text{for} \quad \eta = \begin{pmatrix}\lambda_n^\pm\zeta_n^\pm\\ \zeta_n^\pm\end{pmatrix} \in s_n^\pm.$$

Suppose $\eta \in D_L \times D_L$ and $(\phi_m, \eta)_2 = 0$, $m = 1, 2, \cdots, n-1$, where

$$\phi_m = \begin{pmatrix}\lambda_m^+P\zeta_m^+\\ [L + B]\zeta_m^+\end{pmatrix}.$$

This is equivalent to the statement that $\eta \in s_n^+$, so that Eq. (23) implies

$$(\lambda_n^+)^{-1} \geq \min_{\substack{\phi_m \in E \times E \\ m=1,2,\cdots,n-1}} \max_{\substack{(\phi_m,\eta)_2=0 \\ m=1,2,\cdots,n-1 \\ \eta \in D_L \times D_L}} F(\eta). \quad (33)$$

Let $\{\phi_m\}_{m=1}^{n-1}$ be any $n-1$ vectors in $E \times E$, and define

$$e_m \equiv \begin{pmatrix} \lambda_m^+ \zeta_m^+ \\ \zeta_m^+ \end{pmatrix}$$

for $m = 1, 2, \cdots, n$. Equation (6) implies that the $\{e_m\}_{m=1}^n$ is linearly independent and thus spans an n-dimensional subspace $M_n \subset D_L \times D_L \subset E \times E$. Therefore, there exists a nonzero vector

$$\eta = \sum_1^n \alpha_m e_m \in M_n,$$

such that $(\phi_m, \eta)_2 = 0$ for $m = 1, 2, \cdots, n-1$. Since

$$F\left(\sum_1^n \alpha_m e_m\right) = \frac{\sum_1^n |\alpha_m|^2 (\lambda_m^+)^{-1}}{\sum_1^n |\alpha_m|^2} \geq (\lambda_n^+)^{-1}, \quad (34)$$

we have

$$\max_{\substack{(\phi_m,\eta)_2=0 \\ m=1,2,\cdots,n-1 \\ \eta \in D_L \times D_L}} F(\eta) \geq (\lambda_n^+)^{-1}, \quad (35)$$

and Eq. (25) follows from Eqs. (33) and (35). The proof of Eq. (26) is similar.

Lemma I: Let L and P be positive linear operators with domain and range in the Hilbert space E,

$$D_L \subset D_P,$$

$$\delta_1 \equiv \inf_{D_L} \frac{(\zeta, L\zeta)}{(\zeta, \zeta)} > 0, \quad \delta_2 \equiv \inf_{D_L} \frac{(\zeta, P\zeta)}{(\zeta, \zeta)} > 0,$$

and let the infinite sequences of vectors $\{\zeta_m\} \in D_L$ and real numbers $\{\lambda_m\}$ satisfy

$$\lambda_m \lambda_n (\zeta_m, P\zeta_n) + (\zeta_m, L\zeta_n) = \delta_{mn},$$

$m = 1, 2, \cdots$. Then, for any sequence of numbers $\{C_m\}$, $\sum_1^\infty |C_m|^2 < \infty$ implies that

$$\sum_1^\infty C_m \zeta_m \quad \text{and} \quad \sum_1^\infty C_m \lambda_m \zeta_m$$

both converge in E.

Proof: Let $y_n \equiv \sum_1^n C_m \zeta_m$ and $z_n \equiv \sum_1^n C_m \lambda_m \zeta_m$. Then,

$$\begin{aligned}(y_n - y_m, L[y_n - y_m]) + (z_n - z_m, P[z_n - z_m]) \\ = \left(\sum_{m+1}^n C_k \zeta_k, L \sum_{m+1}^n C_l \zeta_l\right) + \left(\sum_{m+1}^n C_k \lambda_k \zeta_k, P \sum_{m+1}^n C_l \lambda_l \zeta_l\right) \\ = \sum_{k,l=m+1}^n \bar{C}_k C_l \{(\zeta_k, L\zeta_l) + \lambda_k \lambda_l (\zeta_k, P\zeta_l)\} = \sum_{m+1}^n |C_k|^2. \end{aligned} \quad (36)$$

Therefore

$$\begin{aligned}\|y_n - y_m\|^2 &\leq \delta_1^{-1}(y_n - y_m, L[y_n - y_m]) \\ &\leq \delta_1^{-1} \sum_{m+1}^n |C_k|^2 \end{aligned} \quad (37)$$

and

$$\begin{aligned}\|z_n - z_m\|^2 &\leq \delta_2^{-1}(z_n - z_m, P[z_n - z_m]) \\ &\leq \delta_2^{-1} \sum_{m+1}^n |C_k|^2, \end{aligned} \quad (38)$$

so that $\sum_1^\infty |C_m|^2 < \infty$ implies that $\{y_m\}$ and $\{z_n\}$ are Cauchy sequences and hence converge in E.

Lemma II: Let the hypothesis of Lemma I hold, $|\lambda_m| \to \infty$ as $m \to \infty$, and $\sum_1^\infty |C_m|^2 \lambda_m^2 < \infty$. Then the series

$$\sum_1^\infty C_m (i\lambda_m)^l e^{i\lambda_m t} \zeta_m$$

converges uniformly in $t (-\infty < t < \infty)$ to a limit $\eta_l(t) \in E$ for $l = 0, 1, 2$, and $(d/dt)\eta_l(t) = \eta_{l+1}(t)$ for $t \in (-\infty, \infty)$ and $l = 0, 1$.

Proof: $\sum_1^\infty |C_m|^2 \lambda_m^2 < \infty$ and $|\lambda_m| \to \infty$ as $m \to \infty$ implies $\sum_1^\infty |C_m|^2 < \infty$. Let

$$\xi_n^{(l)}(t) \equiv \sum_1^n C_m (i\lambda_m)^l e^{i\lambda_m t} \zeta_m$$

for $l = 0, 1, 2$. It follows from Eqs. (37) and (38) of Lemma I that

$$\|\xi_n^{(l)}(t) - \xi_m^{(l)}(t)\|^2 \leq \Delta_l \sum_{m+1}^n |C_k|^2 \lambda_k^{P(l)}, \quad (39)$$

where

$$\Delta_l \equiv \begin{cases} \delta_1^{-1}, & l = 0, \\ \delta_2^{-1}, & l = 1, 2, \end{cases} \qquad P(l) \equiv \begin{cases} 0, & l = 0, 1, \\ 2, & l = 2. \end{cases}$$

Thus $\{\xi_n^{(l)}(t)\}$ is a uniform Cauchy sequence and converges uniformly to a limit element $\eta_l(t) \in E$ for $t \in (-\infty, \infty)$ and $l = 0, 1, 2$. Let $t_0 \in (-\infty, \infty)$. We define

$$\phi_n^{(l)}(t) \equiv \begin{cases} \dfrac{\xi_n^{(l)}(t) - \xi_n^{(l)}(t_0)}{t - t_0}, & t \neq t_0, \\ \xi_n^{(l+1)}(t_0), & t = t_0, \end{cases} \quad \begin{matrix} l = 0, 1, \\ n = 1, 2, \cdots, \end{matrix}$$

$$\psi_l(t) \equiv \begin{cases} \dfrac{\eta_l(t) - \eta_l(t_0)}{t - t_0}, & t \neq t_0, \\ \eta_{l+1}(t_0), & t = t_0, \end{cases} \quad l = 0, 1,$$

and we have

$$\lim_{n \to \infty} \phi_n^{(l)}(t) = \psi_l(t) \quad \text{for } t \in (-\infty, \infty) \quad \text{and} \quad l = 0, 1. \quad (40)$$

We show that the convergence is uniform. Suppose

$t \neq t_0$. Then,

$$\phi_n^{(l)}(t) - \phi_m^{(l)}(t) = (t - t_0)^{-1} \sum_{m+1}^{n} \{C_k(i\lambda_k)^l e^{i\lambda_k t_0}[e^{i\lambda_k(t-t_0)} - 1]\zeta_k\} \quad (41)$$

so that Eqs. (37) and (38) give

$$\|\phi_n^{(l)}(t) - \phi_m^{(l)}(t)\|^2 \leq \Delta_l \sum_{m+1}^{n} |C_k[e^{i\lambda_k(t-t_0)} - 1](t - t_0)^{-1}|^2 \leq 2\Delta_l \sum_{m+1}^{n} |C_k|^2 \lambda_k^2, \quad l = 0, 1, \quad (42)$$

where we have used the fact that for real λ and θ, $\theta \neq 0$,

$$|(e^{i\lambda\theta} - 1)\theta^{-1}| \leq 2^{\frac{1}{2}} |\lambda|. \quad (43)$$

Since $\sum_1^\infty |C_k|^2 \lambda_k^2 < \infty$, Eq. (42) implies that $\{\phi_n^{(l)}(t)\}$ is a uniform Cauchy sequence for $t \neq t_0$ and $l = 0, 1$, so we conclude from Eq. (40) that $\phi_n^{(l)}(t)$ converges to $\psi_l(t)$ uniformly for all $t \in (-\infty, \infty)$ and $l = 0, 1$. For $l = 0, 1$ and each n, $\phi_n^{(l)}(t)$ is continuous at t_0, so the uniform convergence of $\phi_n^{(l)}(t)$ to $\psi_l(t)$ implies that $\psi_l(t)$ is continuous at t_0, i.e.,

$$\lim_{t \to t_0} \left\| \frac{\eta_l(t) - \eta_l(t_0)}{t - t_0} - \eta_{l+1}(t_0) \right\| = 0, \quad l = 0, 1. \quad (44)$$

Therefore $(d\eta_l/dt)(t_0) = \eta_{l+1}(t_0)$ for $l = 0, 1$. Since t_0 was an arbitrary point of $(-\infty, \infty)$, the proof is complete.

Theorem IV: Let the hypothesis of Theorem II as well as (E1) or (E2) of Theorem II be satisfied, and suppose that

$$\sum_1^\infty |C_m|^2 \lambda_m^6 < \infty,$$

where $\{C_m\}$ is a sequence of complex numbers and $\{\zeta_m\}$ and $\{\lambda_m\}$ are the eigenvectors and eigenvalues obtained in Theorem II. Then the series

$$\sum_1^\infty C_m(i\lambda_m)^n e^{i\lambda_m t}\zeta_m$$

converges uniformly in $t(-\infty < t < \infty)$ for $n = 0, 1, 2, 3, 4$. Let

$$\zeta(t) \equiv \sum_1^\infty C_m e^{i\lambda_m t}\zeta_m, \quad -\infty < t < \infty. \quad (45)$$

Then

$$\frac{d^n\zeta(t)}{dt^n} = \sum_1^\infty C_m(i\lambda_m)^n e^{i\lambda_m t}\zeta_m, \quad -\infty < t < \infty, \quad (46)$$

for $n = 1, 2, 3, 4$, $\zeta(t) \in D_L$ for $t \in (-\infty, \infty)$, and we have

$$P\ddot{\zeta} + iQ\dot{\zeta} + [L + B]\zeta(t) = 0, \quad -\infty < t < \infty, \quad (47)$$

and

$$C_m = ([L + B]\zeta_m, \zeta(0)) + \lambda_m(P\zeta_m, -i\dot{\zeta}(0)). \quad (48)$$

Proof: We first show that

$$\inf_{D_L} \frac{(\zeta, [L + B]\zeta)}{(\zeta, \zeta)} > 0.$$

Let $L' \equiv L + B$, $K' \equiv kr^{-2}k$ (see Theorems I and II), and $k' \equiv (K')^{\frac{1}{2}}$. Then K' and k' are both ≥ 0, Hermitian, and completely continuous, and $K'L' = I$ on D_L, since, for $\zeta \in D_L$,

$$kr^{-2}k(L + B)\zeta = kr^{-2}(I + kBk)kL\zeta = k^2L\zeta = \zeta.$$

Now $\zeta = k'u$ for $u = k'L'\zeta$, so that $(\zeta, L'\zeta) = (k'u, L'\zeta) = (u, k'L'\zeta) = (u, u)$ and

$$(\zeta, \zeta) = (k'u, k'u) = (u, K'u).$$

Therefore

$$\inf_{D_L} \frac{(\zeta, L'\zeta)}{(\zeta, \zeta)} \geq \inf_E \left[\frac{(u, K'u)}{(u, u)}\right]^{-1} = \left[\sup_E \frac{(u, K'u)}{(u, u)}\right]^{-1} = \|K'\|^{-1} > 0.$$

$$\sum_1^\infty |C_m|^2 \lambda_m^6 < \infty$$

implies

$$\sum_1^\infty |C_m\lambda_m^2|^2 \lambda_m^2 < \infty$$

and

$$\sum_1^\infty |C_m|^2 \lambda_m^2 < \infty \quad (|\lambda_m| \to \infty \text{ for } m \to \infty).$$

Therefore, by Lemma II, the series

$$\eta_l(t) \equiv \sum_1^\infty C_m(i\lambda_m)^l e^{i\lambda_m t}\zeta_m, \qquad l = 0, 1, 2,$$

$$\eta_{l+2}(t) \equiv \sum_1^\infty [C_m(i\lambda_m)^2](i\lambda_m)^l e^{i\lambda_m t}\zeta_m, \quad l = 0, 1, 2,$$

converge uniformly for $t \in (-\infty, \infty)$ and $l = 0, 1, 2$, and we have $d\eta_l/dt = \eta_{l+1}$ for $l = 0, 1, 2, 3$. Thus Eq. (46) holds. Now

$$-K\{P\ddot{\zeta} + iQ\dot{\zeta} + B\zeta\} = K\sum_1^\infty C_m e^{i\lambda_m t}\{\lambda_m^2 P\zeta_m + \lambda_m Q\zeta_m - B\zeta_m\} = K\sum_1^\infty C_m e^{i\lambda_m t}L\zeta_m = \sum_1^\infty C_m e^{i\lambda_m t}KL\zeta_m = \sum_1^\infty C_m e^{i\lambda_m t}\zeta_m;$$

i.e.,

$$\zeta(t) = -Kf, \quad f \equiv P\ddot{\zeta} + iQ\dot{\zeta} + B\zeta. \quad (49)$$

Therefore $\dot{\zeta} = -Kf^{\cdot}$, $\ddot{\zeta} = -Kf^{\cdot\cdot}$, so that $-f = PKf^{\cdot\cdot} + iQKf^{\cdot} + BKf \in R_L$, since $PK(E) \subset R_L$, $QK(E) \subset R_L$, and $BK(E) \subset R_L$. Then $-f = Lu$ for some $u \in D_L$, and Eq. (49) gives $\zeta = -Kf = KLu = u$; i.e., $\zeta \in D_L$ and $L\zeta + f = 0$, which is Eq. (47). We have $\dot{\zeta} = -Kf^{\cdot}$ and $\zeta \in D_L$ for all $t \in (-\infty, \infty)$, so that, in particular, $\dot{\zeta}(0) \in R_K$ and $\zeta(0) \in D_L \subset R_K$. Thus Eqs. (7) and (8) of Theorem II give

$$\sum_1^\infty \beta_m \zeta_m = \zeta(0) = \sum_1^\infty C_m \zeta_m, \tag{50}$$

$$i\sum_1^\infty \beta_m \lambda_m \zeta_m = \dot{\zeta}(0) = \sum_1^\infty C_m i\lambda_m \zeta_m, \tag{51}$$

where $\beta_m = ([L + B]\zeta_m, \zeta(0)) + \lambda_m(P\zeta_m, -i\dot{\zeta}(0))$; i.e.,

$$\sum_1^\infty \gamma_m \zeta_m = 0 = \sum_1^\infty \gamma_m \lambda_m \zeta_m$$

for $\gamma_m = C_m - \beta_m$. Therefore

$$\begin{aligned} 0 &= \left([L + B]\zeta_n, \sum_1^\infty \gamma_m \zeta_m\right) + \lambda_n\left(P\zeta_n, \sum_1^\infty \gamma_m \lambda_m \zeta_m\right) \\ &= \sum_1^\infty \gamma_m\{([L + B]\zeta_n, \zeta_m) + \lambda_n\lambda_m(P\zeta_n, \zeta_m)\} = \gamma_n, \end{aligned} \tag{52}$$

which establishes Eq. (48).

III. THE ROTATING ELASTIC STRING

Consider a straight free elastic string of length R, lying within a fixed plane P, in which the string will later be assumed to rotate. Let r be the coordinate and $\mathbf{e}_r$ the unit vector along the length of the string ($0 \le r \le R$), $\mathbf{e}_\theta$ the unit vector in P orthogonal to $\mathbf{e}_r$, $K(r)$ the product of the cross-sectional area of the string and its modulus of elasticity at the point r, $\mu(r)$ its mass per unit length, and let $\boldsymbol{\xi}$ (in P) be the displacement vector associated with a deformation of the string. We write

$$\boldsymbol{\xi} = \xi_r \mathbf{e}_r + \xi_\theta \mathbf{e}_\theta, \tag{53}$$

and for the strain ϵ we obtain

$$\epsilon = \left[\left(1 + \frac{\partial \xi_r}{\partial r}\right)^2 + \left(\frac{\partial \xi_\theta}{\partial r}\right)^2\right]^{\frac{1}{2}} - 1. \tag{54}$$

Assuming the (deformed) string to rotate about $r = 0$ with the angular velocity Ω, we have

$$\dot{\mathbf{e}}_r = \Omega \mathbf{e}_\theta, \quad \dot{\mathbf{e}}_\theta = -\Omega \mathbf{e}_r, \tag{55}$$

so that the kinetic energy T of the deformed string can be written as

$$T = \tfrac{1}{2}\int_0^R \mu(r)\{(\dot{\xi}_r - \Omega\xi_\theta)^2 + [\dot{\xi}_\theta + \Omega(\xi_r + r)]^2\}\, dr. \tag{56}$$

For the potential energy V we have

$$\begin{aligned} V &= \int_0^R \frac{AE}{2}\epsilon^2\, dr \\ &= \tfrac{1}{2}\int_0^R K(r)\left\{\left[\left(1 + \frac{\partial \xi_r}{\partial r}\right)^2 + \left(\frac{\partial \xi_\theta}{\partial r}\right)^2\right]^{\frac{1}{2}} - 1\right\}^2 dr. \end{aligned} \tag{57}$$

Let the string be in a state of steady rotation, with angular velocity Ω about $r = 0$, with the deformation given by $\xi_r = \xi_0(r)$ and $\xi_\theta \equiv 0$. We take the ends of the deformed string to be fixed in the frame of rotation and consider small oscillations about this state of steady motion. To this end, we expand $\boldsymbol{\xi}$ in a power-series expansion in the perturbation parameter δ, viz.,

$$\begin{aligned} \xi_r &= \xi_0(r) + \delta\xi_{r1}(r, t) + \delta^2\xi_{r2} + \cdots, \\ \xi_\theta &= 0 + \delta\xi_{\theta 1}(r, t) + \delta^2\xi_{\theta 2} + \cdots. \end{aligned} \tag{58}$$

Lagrange's equations for ξ_r and ξ_θ then yield, to first order in δ and for $0 \le r \le R$,

$$\frac{d}{dr}\left[K(r)\frac{d\xi_0}{dr}\right] + \Omega^2\mu(r)\xi_0(r) = -\Omega^2\mu r, \tag{59}$$

$$\mu[\ddot{\xi}_{r1} - 2\Omega\dot{\xi}_{\theta 1} - \Omega^2\xi_{r1}] - \frac{\partial}{\partial r}\left\{K(r)\frac{\partial \xi_{r1}}{\partial r}\right\} = 0, \tag{60}$$

$$\mu[\ddot{\xi}_{\theta 1} + 2\Omega\dot{\xi}_{r1} - \Omega^2\xi_{\theta 1}] - \frac{\partial}{\partial r}\left\{\gamma(r)\frac{\partial \xi_{\theta 1}}{\partial r}\right\} = 0, \tag{61}$$

where

$$\gamma(r) \equiv K(r)\frac{d\xi_0}{dr}\left(1 + \frac{d\xi_0}{dr}\right)^{-1}$$

and our boundary conditions are

$$\xi_{r1} = \xi_{\theta 1} = 0 \quad \text{at} \quad r = 0 \quad \text{and} \quad r = R. \tag{62}$$

The rotating string is assumed to be under tension so that $d\xi_0/dr > 0$, $0 \le r \le R$, and therefore $K(r) > 0$ implies $\gamma(r) > 0$ for $0 \le r \le R$. Equations (60) and (61) take the form of Eq. (47) in the Hilbert space $L_2[0, R] \times L_2[0, R]$, provided that we set

$$\begin{aligned} &P = \begin{pmatrix} \mu & 0 \\ 0 & \mu \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & i2\Omega\mu \\ -i2\Omega\mu & 0 \end{pmatrix}, \quad B = 0, \\ &L = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}, \quad L_1 = -\frac{d}{dr}\left[K\frac{d}{dr}\right] - \mu\Omega^2, \\ &L_2 = -\frac{d}{dr}\left[\gamma\frac{d}{dr}\right] - \mu\Omega^2, \quad \zeta = \begin{pmatrix} \xi_{r1} \\ \xi_{\theta 1} \end{pmatrix}. \end{aligned} \tag{63}$$

We take

$$\begin{aligned} D_{L_1} &= D_{L_2} \\ &= D \equiv \{\xi \mid \xi \in C^2[0, R],\ \xi(0) = \xi(R) = 0\}, \end{aligned}$$

$D_L = D \times D$ (note that $\bar{D}_L = E$), and assume that $\mu > 0$, $\mu \in C[0, R]$, $K > 0$, $K \in C^1[0, R]$, $\gamma > 0$, $\gamma \in C^1[0, R]$, and that Ω is sufficiently small (or K and γ are sufficiently large) so that L_1 and L_2 are both positive-definite on D. Then L_1 and L_2 have completely continuous positive Hermitian inverses K_1 and K_2 (integral operators with Green's functions as kernels) on $L_2[0, R]$ with the following properties: $K_1L_1 = I = K_2L_2$ on D, $R_{K_1} \subset C[0, R]$, and $R_{K_2} \subset C[0, R]$. We also have $R_{L_1} \supset C[0, R]$ and $R_{L_2} \supset C[0, R]$. Therefore,

$$K = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix}$$

is a completely continuous positive Hermitian operator on E and satisfies $KL = I$ on D_L, $PK(E) \subset R_L$, and $QK(E) \subset R_L$, so that all the previous theorems are applicable. In particular, the system possesses infinitely many positive as well as negative eigenfrequencies λ_m, and the associated eigenmodes ζ_m are complete [in the sense of Eqs. (7) and (8)] for y and $\dot{y}$ in $D \times D$.

ACKNOWLEDGMENTS

The author would like to thank Dr. R. W. Dickey and Professor Harold Weitzner for many helpful suggestions and for reading the manuscript. The work presented here was supported by the Magneto-Fluid Dynamics Division, Courant Institute of Mathematical Sciences, New York University, under contract AT(30-1)1480 with the U.S. Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Asymptotic Fields in Some Models of Quantum Field Theory. I

RAPHAEL HØEGH-KROHN
University of Oslo

(Received 19 September 1967)

A quantum field with nonlocal interaction is considered. We prove under a proper smoothness condition on the interaction that the asymptotic limits of the annihilation–creation operators exist. The asymptotic limits are then used to prove that the state space decomposes as a tensor product of an incoming (outgoing) Fock space and a zero-particle space.

1. INTRODUCTION

In quantum theories the scattering operator is of great significance in relating theory with observation. The scattering operator is given relative to a decomposition of the total-energy operator H into two parts. The free energy H_0 and the interaction energy V:

$$H = H_0 + V.$$

H and H_0 are self-adjoint operators on a Hilbert space $\mathcal{H}$. The pure states of the system are represented by elements in $\mathcal{H}$. If at the time zero the state is represented by ψ_0, then at the time t it is represented by $e^{itH}\psi_0$. Asymptotically, for very large t, we expect the system to behave as if there were no interaction, that is, as if the energy operator were H_0:

$$e^{itH}\psi_0 \sim e^{itH_0}\psi_\pm, \quad \text{for } t \to \pm\infty.$$

The scattering operator S is now defined by

$$S\psi_- = \psi_+$$

and describes the asymptotic time transition of the system. We see that existence of S corresponds to the existence of the asymptotic limits

$$\lim_{t\to\pm\infty} e^{-itH}e^{itH_0} = W_\pm,$$

since

$$\psi_0 = W_\pm\psi_\pm.$$

As is well known, there is also another way of describing scattering, or asymptotic transition from very early to very late times, than by the scattering operator using asymptotic states as described above, that is, by using asymptotic operators. The advantages of using asymptotic operators in connection with quantum field theories were pointed out by Friedrichs[1] as well as by Kato and Mugibayashi.[2]

In a quantum system, the observable quantities are represented by operators on the Hilbert space $\mathcal{H}$ and, if the quantity at the time zero is represented by A_0,

[1] K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (American Mathematical Society, Providence, R.I., 1965).
[2] Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **30**, 103 (1963).

$D_L = D \times D$ (note that $\bar{D}_L = E$), and assume that $\mu > 0$, $\mu \in C[0, R]$, $K > 0$, $K \in C^1[0, R]$, $\gamma > 0$, $\gamma \in C^1[0, R]$, and that Ω is sufficiently small (or K and γ are sufficiently large) so that L_1 and L_2 are both positive-definite on D. Then L_1 and L_2 have completely continuous positive Hermitian inverses K_1 and K_2 (integral operators with Green's functions as kernels) on $L_2[0, R]$ with the following properties: $K_1L_1 = I = K_2L_2$ on D, $R_{K_1} \subset C[0, R]$, and $R_{K_2} \subset C[0, R]$. We also have $R_{L_1} \supset C[0, R]$ and $R_{L_2} \supset C[0, R]$. Therefore,

$$K = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix}$$

is a completely continuous positive Hermitian operator on E and satisfies $KL = I$ on D_L, $PK(E) \subset R_L$, and $QK(E) \subset R_L$, so that all the previous theorems are applicable. In particular, the system possesses infinitely many positive as well as negative eigenfrequencies λ_m, and the associated eigenmodes ζ_m are complete [in the sense of Eqs. (7) and (8)] for y and $\dot{y}$ in $D \times D$.

ACKNOWLEDGMENTS

The author would like to thank Dr. R. W. Dickey and Professor Harold Weitzner for many helpful suggestions and for reading the manuscript. The work presented here was supported by the Magneto-Fluid Dynamics Division, Courant Institute of Mathematical Sciences, New York University, under contract AT(30-1)1480 with the U.S. Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Asymptotic Fields in Some Models of Quantum Field Theory. I

RAPHAEL HØEGH-KROHN
University of Oslo

(Received 19 September 1967)

A quantum field with nonlocal interaction is considered. We prove under a proper smoothness condition on the interaction that the asymptotic limits of the annihilation–creation operators exist. The asymptotic limits are then used to prove that the state space decomposes as a tensor product of an incoming (outgoing) Fock space and a zero-particle space.

1. INTRODUCTION

In quantum theories the scattering operator is of great significance in relating theory with observation. The scattering operator is given relative to a decomposition of the total-energy operator H into two parts. The free energy H_0 and the interaction energy V:

$$H = H_0 + V.$$

H and H_0 are self-adjoint operators on a Hilbert space $\mathcal{H}$. The pure states of the system are represented by elements in $\mathcal{H}$. If at the time zero the state is represented by ψ_0, then at the time t it is represented by $e^{itH}\psi_0$. Asymptotically, for very large t, we expect the system to behave as if there were no interaction, that is, as if the energy operator were H_0:

$$e^{itH}\psi_0 \sim e^{itH_0}\psi_\pm, \quad \text{for } t \to \pm\infty.$$

The scattering operator S is now defined by

$$S\psi_- = \psi_+$$

and describes the asymptotic time transition of the system. We see that existence of S corresponds to the existence of the asymptotic limits

$$\lim_{t\to\pm\infty} e^{-itH}e^{itH_0} = W_\pm,$$

since

$$\psi_0 = W_\pm\psi_\pm.$$

As is well known, there is also another way of describing scattering, or asymptotic transition from very early to very late times, than by the scattering operator using asymptotic states as described above, that is, by using asymptotic operators. The advantages of using asymptotic operators in connection with quantum field theories were pointed out by Friedrichs[1] as well as by Kato and Mugibayashi.[2]

In a quantum system, the observable quantities are represented by operators on the Hilbert space $\mathcal{H}$ and, if the quantity at the time zero is represented by A_0,

[1] K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (American Mathematical Society, Providence, R.I., 1965).
[2] Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **30**, 103 (1963).

then at the time t it is represented by $e^{-itH}A_0e^{itH}$. Asymptotically, we expect the system to behave as if H_0 were the energy operator, or

$$e^{-itH}A_0e^{itH} \sim e^{-itH_0}A_\pm e^{itH_0}, \quad \text{for } t \to \pm\infty.$$

The scattering is now given by the transition

$$A_- \to A_+,$$

describing the asymptotic transition of the system. We see that even if there exist no asymptotic states there still may exist asymptotic operators and we can, therefore, still study scattering of the system. We expect the mappings $A_0 \to A_\pm$ to preserve the algebraic relations, so if we substitute $e^{itH}A_0e^{-itH_0}$ for A_0 in the asymptotic relation above we have

$$e^{-itH}e^{itH_0}A_0e^{-itH_0}e^{itH} \sim A_\pm, \quad \text{for } t \to \pm\infty,$$

or the following relations:

$$\lim_{t\to\pm\infty} e^{-itH}e^{itH_0}A_0e^{-itH_0}e^{itH} = A_\pm.$$

Together with the specification of the sense in which the limit is to be taken, we use this as a definition of the asymptotic operators $A_\pm$.

In two earlier papers[3] the author studied perturbation by annihilation–creation operators, using a technique based on "gentleness." This technique, however, was not able to deal with the case of V containing pure annihilation and pure creation terms. In that case it is well known that there is a vacuum renormalization, i.e., a general shift of the whole spectrum of the energy operator (see, for instance, Ref. 1) and this causes considerable difficulties for the above mentioned technique.

On the other hand, Kato and Mugibayashi[2] studied perturbations where V contained pure annihilation and pure creation terms, by an adaption of Cook's method[4] to the study of asymptotic limits of annihilation–creation operators, but with a very strong restriction on V, namely that the kernels of V be finite-dimensional.

2. QUANTUM FIELD WITH NONLOCAL INTERACTION OR PERTURBATION BY ANNIHILATION–CREATION OPERATORS

The free-energy operator H is given with respect to a specific representation of the Hilbert space $\mathcal{H}$, the so-called Fock representation. An element f in $\mathcal{H}$ is given by a sequence f_n of complex-valued functions, where f_0 is just a complex constant and f_n is a function $f_n(x_1, \cdots, x_n)$ of n variables $x_1, \cdots, x_n$, and each x_i is a variable in the Euclidian 3-space E_3. We consider only the case of one fermion field interacting with itself. The reason for this is partly one of notational convenience and partly the fact that, if the interaction were more than quadratic in the boson field, some of the proofs would become more complicated.

That we consider only one fermion field means that the functions $f_n(x_1, \cdots, x_n)$ are all antisymmetric, i.e.,

$$f_n(x_1, \cdots, x_n) = \frac{1}{n!}\sum_\sigma (-1)^\sigma f_n(x_{\sigma(1)}, \cdots, x_{\sigma(n)}),$$

where the summation runs over all permutations of the indices $1, \cdots, n$. The inner product in $\mathcal{H}$ is given by

$$(f, g) = \sum_{n=0}^{\infty} n! \int \cdots \int f_n(x_1, \cdots, x_n) \times g_n(x_1, \cdots, x_n)\, dx_1 \cdots dx_n.$$

Let Ω be the self-adjoint operator in $L_2(E_3)$:

$$\Omega = (-\Delta + m^2)^{\frac{1}{2}}$$

on its natural domain of definition, where Δ is the Laplacian

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

and m a nonnegative constant called the "mass of the free fermion." H_0 is then given by

$$(H_0f)_n(x_1, \cdots, x_n) = \sum_{i=1}^{n} \Omega_i f_n(x_1, \cdots, x_n),$$

where Ω_i is the operator Ω operating on the variable x_i, and $(H_0f)_0 = 0$. H_0 is obviously self-adjoint on its natural domain of definition. The interaction operator V is now given in terms of the annihilation–creation operators. The annihilation operator $a(x)$ is defined for $x \in E_3$ by

$$(a(x)f)_n(x_1, \cdots, x_n) = (n+1)f_{n+1}(x, x_1, \cdots, x_n)$$

and the creation operator $a^*(x)$ as the adjoint of $a(x)$. $a(x)$ and $a^*(x)$ are both improper operations but their definition is easily made precise in the following way. Let $h \in L_2(E_3)$; define

$$a(h) = \int a(x)h(x)\, dx$$

and $a^*(h)$ as the adjoint of $a(h)$. The definition of the integral above is, of course,

$$(a(h)f)_n = (n+1)\int dx\, h(x)f_{n+1}(x, x_1, \cdots, x_n).$$

It is easily verified that, due to the fact that

[3] R. Høegh-Krohn, Proc. Natl. Acad. Sci. **58**, 2187 (1967); Commun. Pure Appl. Math. **21**, 313, 343 (1968).
[4] J. M. Cook, J. Math. & Phys. **36**, 82 (1957).

$f_n(x_1, \cdots, x_n)$ is antisymmetric, we have the following anticommutation relations characteristic of a fermion field:

$$\begin{aligned} a(h)a(g) + a(g)a(h) &= 0, \\ a^*(h)a^*(g) + a^*(g)a^*(h) &= 0, \\ a^*(h)a(g) + a(g)a^*(h) &= \int g(x)h(x)\,dx. \end{aligned} \tag{1}$$

The last equality implies that

$$\|a(h)\| \leq \|h\|_2,$$

so that, for $h \in L_2$, $a(h)$ is a bounded operator. The interaction V is now given in the following way:

$$V = \sum_{0 \leq k,l \leq N} V_{kl},$$

where V_{kl} is given in terms of the kernel

$$V_{kl}(x_1, \cdots, x_k \,|\, y_1, \cdots, y_l),$$

antisymmetric in $x_1, \cdots, x_k$ and $y_1, \cdots, y_l$, in the following way:

$$\begin{aligned} V_{kl} = \int \cdots \int & V_{kl}(x_1, \cdots, x_k \,|\, y_1, \cdots, y_l) \\ & \times a^*(x_1) \cdots a^*(x_k) a(y_1) \cdots a(y_l) \\ & \times dx_1 \cdots dx_k\, dy_1 \cdots dy_l. \end{aligned}$$

V_{kl} may also be defined explicitly by

$$\begin{aligned} (V_{kl}f)_n(x_1, \cdots, x_n) &= \text{asym} \binom{n}{l} l! \int \cdots \int dy_1 \cdots dy_l \\ &\quad \times V_{kl}(x_1, \cdots, x_k \,|\, y_1, \cdots, y_l) \\ &\quad \times f_m(y_1, \cdots, y_l, x_{k+1}, \cdots, x_n), \end{aligned}$$

where $n = k - l + m$ and asym is short for the antisymmetrization with respect to the variables $x_1, \cdots, x_n$. Since $a(h)$ is bounded for $h \in L_2$, we see that, for $V_{kl}(x_1, \cdots, x_k \,|\, y_1, \cdots, y_l)$ smooth enough, the operator V is bounded. We do assume that V is symmetric, i.e.,

$$\begin{aligned} V_{kl}(x_1, \cdots, x_k \,|\, y_1, \cdots, y_l) \\ = V_{lk}(y_1, \cdots, y_l \,|\, x_1, \cdots, x_k). \end{aligned}$$

Since V is bounded, $H = H_0 + V$ is also a self-adjoint operator with the same domain as H_0.

We now prove the existence of the asymptotic annihilation and creation operators. Since V contains pure annihilation and pure creation terms, i.e., terms of the form V_{0l}, and V_{k0}, we know that the scattering operator or asymptotic states do not exist (see, for instance, Refs. 1, 3, or 5), but Friedrichs[1] argued that even with the pure annihilation and creation terms present the asymptotic annihilation and creation operators would still exist. The argument proceeded by indicating how the "clouding terms" would cancel each other asymptotically. Independently, Kato and Mugibayashi[2] proved, in a very special case with finite-dimensional kernels $V_{kl}(x_1, \cdots, x_k \,|\, y_1, \cdots, y_l)$, that the asymptotic operators did exist. We introduce the anticommutator of two operators A and B, by $\{A, B\} = AB + BA$, and the commutator by $[A, B] = AB - BA$, and observe that the definition of H_0 together with (1) gives us the following relations:

$$\begin{aligned} [H_0, a(h)] &= a(-\Omega h), \\ [H_0, a^*(h)] &= a^*(\Omega h), \\ \{a(h), a(g)\} &= \{a^*(h), a^*(g)\} = 0, \\ \{a(h), a^*(g)\} &= \int h(x)g(x)\,dx. \end{aligned} \tag{2}$$

A set of operators H_0, $a(h)$, and $a^*(h)$—such that $(a(h))^* = a^*(h)$, $a(h)$ is linear in h, and satisfies the relations in (2)—is called a "free-fermion field with mass m." If in addition there is in the Hilbert space $\mathcal{H}$ an element ϕ such that

$$H_0\phi = \omega\phi, \quad a(h)\phi = 0, \quad \text{for all} \quad h \in L_2(E_3),$$

and the smallest closed subspace containing ϕ and invariant under $a^*(h)$, for all $h \in L_2(E_3)$, is the Hilbert space $\mathcal{H}$ itself, we say that we have a "Fock representation of the free-fermion field with vacuum energy ω," and ϕ is called the "vacuum" state.

We observe that our annihilation and creation operators together with the free-energy operator form a free-fermion field with mass m and that $\mathcal{H}$ is the Fock representation with vacuum-energy zero.

We define for $h \in L_2(E_3)$:

$$\begin{aligned} a_t(h) &= e^{-itH}e^{itH_0}a(h)e^{-itH_0}e^{itH}, \\ a_t^*(h) &= e^{-itH}e^{itH_0}a^*(h)e^{-itH_0}e^{itH}, \end{aligned} \tag{3}$$

and we observe that

$$\begin{aligned} e^{itH_0}a(h)e^{-itH_0} &= a(e^{-it\Omega}h), \\ e^{itH_0}a^*(h)e^{-itH_0} &= a^*(e^{it\Omega}h). \end{aligned} \tag{4}$$

Lemma 1: Let $D_\Omega \subset L_2(E_3)$ be the domain of Ω and D_0 the domain of H_0. For $h \in D_\Omega$, $a(h)$, and $a^*(h)$ leave D_0 invariant, i.e.,

$$a(h)D_0 \subset D_0, \quad a^*(h)D_0 \subset D_0, \quad \text{for} \quad h \in D_\Omega.$$

Proof: Let $f \in D_0$. From (2) we get the result that $H_0a(h)f = a(h)H_0f - a(\Omega h)f$, and this proves that $a(h)f \in D_0$. The argument is identical for $a^*(h)$.

[5] A. L. Chestyakov, Dokl. Akad. Nauk SSSR **158**, 66 (1964).

Since D_0 is also the domain of H, we find that e^{iHt}, as well as e^{itH_0}, leaves D_0 invariant. Hence we get from (3) that, for $f \in D_0$, $a_t(h)f$ is strongly differentiable with respect to t, for $h \in D_\Omega$, and

$$\frac{d}{dt} a_t(h)f = -ie^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f,$$

$$\frac{d}{dt} a_t^*(h)f = -ie^{-itH}[V, e^{itH_0}a^*(h)e^{-itH_0}]e^{itH}f. \quad (5)$$

Integrating with respect to t, we get

$$(a_s(h) - a(h))f = -i\int_0^s e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f\,dt,$$

$$(a^*(h) - a^*(h)) = -i\int_0^s e^{-itH}[V, e^{itH_0}a^*(h)e^{-itH_0}]e^{itH}f\,dt. \quad (6)$$

Observing that the operators on both sides are bounded operators, we conclude, since D_0 is dense in $\mathcal{K}$ and since D_Ω is dense in $L_2(E_3)$, that

$$a_s(h) - a(h) = -i\int_0^s e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}\,dt,$$

$$a_s^*(h) - a^*(h) = -i\int_0^s e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}\,dt, \quad (7)$$

for all h in $L_2(E_3)$.

Lemma 2: Set $h_t(x) = (e^{it\Omega}h)(x)$. For $h \in C_0^\infty$, i.e., infinitely differentiable with compact support, we have that

$$\sup_x |h_t(x)| \leq C\,|t|^{-\frac{3}{2}}.$$

Proof: $h_t(x)$ is in fact the positive-frequency solution of the hyperbolic equation

$$\left(\frac{\partial^2}{\partial t^2} - \Delta + m^2\right)h_t(x) = 0,$$

with the initial condition $h(x)$. It is well known that this tends to zero in L_∞ faster than $C\,|t|^{-\frac{3}{2}}$ when initial condition $h(x)$ is in C_0^∞.

From (7) we see that the question of the existence of the asymptotic annihilation and creation operators is equivalent to the question of convergence of the integrals on the right-hand side of (7).

Assumption 1: The kernels $V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l)$ of V have a representation in the following form:

$$V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l) = \sum a_{i_1,\cdots,i_k,\cdots,j_l} \operatorname*{asym}_x \operatorname*{asym}_y u_{i_1}^1(x_1)\cdots u_{i_k}^k(x_k) \times v_{j_1}^1(y_1)\cdots v_{j_l}^l(y_l),$$

where

$$\|u_j^p\|_1 = \|v_j^q\|_1 = \|u_i^p\|_\infty = \|v_j^q\|_\infty = 1$$

and

$$\sum |a_{i_1,\cdots,i_k,j_1,\cdots,j_l}| < \infty.$$

We see that V is a bounded operator from this assumption, since $\|u\|_2^2 \leq \|u\|_\infty \cdot \|u\|_1$.

We now state a theorem.

Theorem 1: $a_t(h)$ and $a_t^*(h)$ converge in the operator norm as $t \to \pm\infty$, for all $h \in L_2(E_3)$, and the limits

$$a_\pm(h) = \operatorname*{norm\,lim}_{t\to\pm\infty} a_t(h), \quad a_\pm^*(h) = \operatorname*{norm\,lim}_{t\to\pm\infty} a_t^*(h)$$

satisfy the same relations as do $a(h)$ and $a^*(h)$, namely,

$$\{a_\pm(h), a_\pm(g)\} = \{a_\pm^*(h), a_\pm^*(g)\} = 0,$$

$$\{a_\pm(h), a_\pm^*(g)\} = \int h(x)g(x)\,dx.$$

Proof: First let $h \in C_0^\infty$. We prove that the integrals on the right-hand side of (7) converge in the norm. Consider, therefore, the norm of the integrand

$$\|e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}\| = \|[V, a(h_{-t})]\| \leq \sum_{k,l\leq N} \|[V_{kl}, a(h_{-t})]\|.$$

By Assumption 1 and Lemma 2, this is bounded by $C\,|t|^{-\frac{3}{2}}$ which is integrable at infinity. Hence the integrals on the right-hand side of (7) converge both at plus and minus infinity. This proves that $a_t(h)$ and $a_t^*(h)$ converge in the norm for all $h \in C_0^\infty$. Observing that $\|a_t(h)\| \leq \|h\|_2$ and $\|a_t^*(h)\| \leq \|h\|_2$, so that $a_t(h)$ and $a_t^*(h)$ map $L_2(E_3)$ into the operator algebra uniformly bounded in t, we get, from the fact that C_0^∞ is dense in $L_2(E_3)$, that $a_t(h)$ converge in the norm for all $h \in L_2(E_3)$. This proves the first part of the theorem. We get the second part as an immediate consequence of the first, observing that $a_t(h)$ and $a_t^*(h)$ satisfy the same anticommutation relations for all t.

Lemma 3: $e^{itH_0}a(h)e^{-itH_0}$ tends strongly to zero as $t \to \pm\infty$ for all $h \in L_2(E_3)$.

Proof: Since $e^{itH_0}a(h)e^{-itH_0}$ is uniformly bounded for $h \in L_2(E_3)$ and $f \in \mathcal{K}$, it is enough to prove that the norm tends to zero for f in a dense set of $\mathcal{K}$ and h in C_0^∞. Therefore let $f = \{f_0, f_1, \cdots, f_m, 0, 0, \cdots\}$, where each $f_n \in C_0$ for $0 < n \leq m$. From Lemma 2 we now get $\|e^{itH_0}a(h)e^{-itH_0}f\| \leq C\,|t|^{-\frac{3}{2}}$ and this proves the lemma.

Theorem 2: If ϕ is an eigenstate of H, i.e.,

$$H\phi = \lambda\phi,$$

then $a_\pm(h)\phi = 0$ for all $h \in L_2(E_3)$.

Proof:

$$\begin{aligned}\|a_t(h)\phi\| &= \|e^{-itH}e^{itH_0}a(h)e^{-itH_0}e^{itH}\phi\| \\ &= \|e^{-it(H-\lambda)}e^{itH_0}a(h)e^{-itH_0}\phi\| \\ &= \|e^{itH_0}a(h)e^{-itH_0}\phi\|,\end{aligned}$$

which tends to zero as $t \to \pm\infty$ by Lemma 3.

Theorem 3: The operators H, $a_+(h)$, and $a_+^*(h)$, and the operators H, $a_-(h)$, and $a_-^*(h)$ constitute free-fermion fields with mass m; i.e., the following commutation relations are satisfied for $h \in D\Omega$:

$$[H, a_\pm(h)] = a_\pm(-\Omega h),$$
$$[H, a_\pm^*(h)] = a_\pm^*(\Omega h).$$

Proof: From Lemma 1 it follows that, for $h \in D_\Omega$, $a_t(h)$ and $a_t^*(h)$ both leave D_0 invariant. Hence for $f \in D_0$ we have the following identity:

$$\begin{aligned}Ha_t(h)f &= a_t(h)Hf + [H, a_t(h)]f \\ &= a_t(h)Hf + e^{-itH}[H, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f \\ &= a_t(h)Hf + e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f \\ &\quad + e^{-itH}[H_0, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f \\ &= a_t(h)Hf + e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f \\ &\quad + e^{-itH}e^{itH_0}a(-\Omega h)e^{-itH_0}e^{itH}f.\end{aligned}$$

Since we have already proved that the second term on the right-hand side tends to zero, we have that the right-hand side tends strongly to

$$a_\pm(h)Hf + a_\pm(-\Omega h)f.$$

Making use of the fact that H is a closed operator we get that $a_\pm(h)f \in D_0$ and that

$$Ha_\pm(h)f = a_\pm(h)Hf + a_\pm(-\Omega h)f.$$

This proves the first commutation relation; the second is proved in the same way.

Corollary: Let $h \in D_\Omega$; then the operators $a_\pm(h)$ and $a_\pm^*(h)$ leave D_0 or the domain of H invariant.

3. THE TOTAL ENERGY OPERATOR H

It is an interesting fact that Theorem 3 may be used to get some information on the spectrum of H.

Since H is bounded below, i.e., there exists a constant w_0 such that

$$(f, Hf) \geq w_0(f,f), \quad \text{for all} \quad f \in D_0,$$

we have that the spectral decomposition of H takes the form

$$H = \int_{w_0}^{\infty} \lambda \, dE_\lambda.$$

As $-\Delta$ is a positive operator, we see that $\Omega = (-\Delta + m^2)^{\frac{1}{2}}$ is bounded below by m, i.e.,

$$(h, \Omega h) \geq m(h, h), \quad \text{for all} \quad h \in D_\Omega.$$

Lemma 4: Let $f \in \mathcal{H}$, then $f = E_\lambda f$ if and only if there exists a constant c such that

$$\|e^{tH}f\| \leq ce^{t\lambda}, \quad \text{for all} \quad t > 0.$$

Proof:

$$\|e^{tH}f\|^2 = \int e^{2t\tilde{\lambda}} \, d(f, E_{\tilde{\lambda}}f).$$

From the theory of the Laplace transform we know that the existence of a constant c_1 such that

$$\int e^{2t\tilde{\lambda}} \, d(f, E_{\tilde{\lambda}}f) \leq c_1 e^{2t\lambda}, \quad \text{for all} \quad t > 0,$$

is equivalent with the measure $d(f, E_{\tilde{\lambda}}f)$ having support bounded above by λ. But this is the same as $f = E_\lambda f$.

Lemma 5: For any $h \in L_2(E_3)$ we have

$$a_\pm(h)E_\lambda\mathcal{H} \subset E_{\lambda-m}\mathcal{H}.$$

Proof: From Theorem 3, we have that, for $f \in D_0$ and $h \in D_\Omega$,

$$[H, a_\pm(h)]f = -a_\pm(\Omega h)f.$$

This gives us that

$$e^{itH}a_\pm(h)e^{-itH}f = a_\pm(e^{-it\Omega}h)f.$$

Since the operators in this equation are uniformly bounded in f as well as in h, we get that this equation is valid for all $f \in \mathcal{H}$ and all $h \in L_2(E_3)$. Hence,

$$e^{itH}a_\pm(h)f = a_\pm(e^{-it\Omega}h)e^{itH}f.$$

For $f = E_\lambda f$, the right-hand side of this equation is analytic in t for $\operatorname{Im} t > 0$, since Ω is bounded below. So by analytic continuation for $h \in L_2(E_3)$ and $f \in E_\lambda\mathcal{H}$, we have that

$$e^{tH}a_\pm(h)f = a_\pm(e^{-t\Omega}h)e^{tH}f,$$

for $t > 0$. Hence,

$$\begin{aligned}\|e^{tH}a_\pm(h)f\| &= \|a_\pm(e^{-t\Omega}h)e^{tH}f\| \\ &\leq \|a_\pm(e^{-t\Omega}h)\| \, \|e^{tH}f\| \\ &\leq \|e^{-t\Omega}h\|_2 \, \|e^{tH}E_\lambda f\| \\ &\leq ce^{-tm}e^{t\lambda} = ce^{t(\lambda-m)}.\end{aligned}$$

By Lemma 4, $a_\pm(h)f \in E_{\lambda-m}\mathcal{H}$, and this proves Lemma 5.

Since H is bounded below by w_0, we have that

$E_{w_0-\epsilon} = 0$ for all $\epsilon > 0$. By Lemma 5 we get that $a_+(h)$ and $a_-(h)$ annihilate $E_{w_0-\epsilon+m}$ for all $h \in L_2(E_3)$.

Let V_+^n and V_-^n be the maximal closed subspaces in $\mathcal{K}$ annihilated by all operators of the type $a_+(h_1)\cdots a_+(h_{n+1})$ and $a_-(h_1)\cdots a_-(h_{n+1})$, respectively, where $h_1 \cdots h_{n+1}$ is in $L_2(E_3)$.

Lemma 6: H is reduced by V_+^n and by V_-^n, i.e.,

$$HV_+^n \subset V_+^n, \quad HV_-^n \subset V_-^n.$$

Moreover, $V_\pm^n \subset V_\pm^{n+1}$ and

$$\mathcal{K} = \overline{\bigcup_n V_+^n} = \overline{\bigcup_n V_-^n}.$$

Proof: Since

$$a_\pm(h_1)\cdots a_\pm(h_{n+1})e^{itH} = e^{itH}a_\pm(e^{it\Omega}h_1)\cdots a_\pm(e^{it\Omega}h_{n+1}),$$

we see that $V_\pm^n$ is invariant under e^{itH}, hence H is reduced by $V_\pm^n$.

That $V_\pm^n \subset V_\pm^{n+1}$ follows from the anticommutation relations of the $a_\pm(h)$. From Lemma 5 we get that $E_{w_0-\epsilon+nm}\mathcal{K} \subset V_\pm^n$, and this proves that $\bigcup_n V_\pm^n$ is dense in $\mathcal{K}$.

Let us now define the incoming and outgoing n-particle spaces $\mathcal{K}_+^n$ and $\mathcal{K}_-^n$ by

$$\mathcal{K}_\pm^0 = V_\pm^0, \quad \mathcal{K}_\pm^n = V_\pm^n - V_\pm^{n-1}, \quad \text{for } n \geq 1.$$

Lemma 7: $\mathcal{K}_\pm^n$ is the smallest closed subspace containing all vectors of the form

$$a_\pm^*(h_1)\cdots a_\pm^*(h_n)v,$$

where $v \in \mathcal{K}_\pm^0$, and $h_1, \cdots, h_n$ is in $L_2(E_3)$. H is reduced by $\mathcal{K}_\pm^*$ and $\mathcal{K} = \sum_{n=0}^\infty \mathcal{K}_\pm^n$ where the sum is a direct sum of Hilbert spaces.

Proof: The last part of the lemma follows immediately from Lemma 6. To prove the first part we see that the anticommutation relations for $a_\pm^*(h)$ and $a_\pm(g)$ imply that $a_\pm^*(h_1)\cdots a_\pm^*(h_n)v$ is in $V_\pm^n$ but not in $V_\pm^{n-1}$, hence in $\mathcal{K}_\pm^n$.

Let f be in $\mathcal{K}_\pm^n$ and suppose that f is orthogonal to all vectors of the form $a_\pm^*(h_1)\cdots a_\pm^*(h_n)v$ with $v \in \mathcal{K}_\pm^0$. Then

$$0 = (f, a_\pm^*(h_1), \cdots, a_\pm^*(h_n)v) = (a_\pm(\bar{h}_n), \cdots, a_\pm(\bar{h}_1)f, v).$$

Since $f \in \mathcal{K}_\pm^n$, $a_\pm(\bar{h}_n)\cdots a_\pm(\bar{h}_1)f$ is in $\mathcal{K}_\pm^0$ because it is annihilated by all $a_\pm(h)$. By the identity above it is orthogonal to all elements in $\mathcal{K}_\pm^0$, hence

$$a_\pm(\bar{h}_n)\cdots a_\pm(\bar{h}_1)f = 0.$$

This gives us that $f \in V_\pm^{n-1}$; but, as $f \in \mathcal{K}_\pm^n$, f must be zero and the lemma is proved.

Theorem 4: The Hilbert space $\mathcal{K}$ decomposes in two ways as a tensor product of two Hilbert spaces

$$\mathcal{K} = \mathcal{F} \otimes \mathcal{K}_+^0, \quad \mathcal{K} = \mathcal{F} \otimes \mathcal{K}_-^0,$$

where $\mathcal{F}$ is the Fokk representation of a free-fermion field with mass m, and $\mathcal{K}_\pm^0$ is the outgoing (incoming) zero-particle spaces.

According to these decompositions the total-energy operator H decomposes as a sum

$$H = H_0 \otimes 1 + 1 \otimes H_+^0,$$
$$H = H_0 \otimes 1 + 1 \otimes H_-^0,$$

where H_0 is the free-energy operator with mass m in $\mathcal{F}$, and $H_\pm^0$ is the restriction of H to $\mathcal{K}_\pm^0$.

Proof: The identification of $\mathcal{K}$ with $\mathcal{F} \otimes \mathcal{K}_\pm^0$ is given by

$$a_\pm^*(h_1)\cdots a_\pm^*(h_n)v \leftrightarrow a^*(h_1)\cdots a^*(h_n)\phi_0 \otimes v,$$

where $a^*(h)$ is the creation operator in $\mathcal{F}$, ϕ_0 is the vacuum state in $\mathcal{F}$, and $v \in \mathcal{K}_\pm^0$.

The anticommutation relations, together with Lemma 7, gives us that this is a norm-preserving identification of all $\mathcal{K}$ with $\mathcal{F} \otimes \mathcal{K}_\pm^0$. The rest of the theorem now follows from the commutation relations between H and $a_\pm^*(h)$, i.e., from Theorem 3.

ACKNOWLEDGMENTS

Some of the ideas in this paper were inspired by the work of Y. Kato and N. Mugibayashi.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Green's Functions for Electromagnetic Waves in Moving Lossy Media

K. C. CHEN AND J. L. YEN
University of Toronto, Toronto, Ontario, Canada

(Received 19 April 1968)

The equations governing the potentials for electromagnetic waves in moving lossy media are obtained by using a covariant formulation. A versatile form of the time-dependent Green's function is derived by first transforming the wave equation to the normal form and then applying the Fourier-integral technique. The time-harmonic Green's function is also obtained.

INTRODUCTION

The problem of electrodynamics of moving lossy media was recently treated by Besieris and Compton,[1] based on the indefinite form of Maxwell–Minkowski equations. In this paper we utilize a covariant formulation previously used by Lee and Papas[2] in treating the lossless case to derive the equations governing the potentials of moving lossy media. Besieris and Compton applied the Riemann matrix to solve for the time-dependent Green's function. We show that the simplest way of solving such a wave equation is to use an affine transformation. This transformation enables us to bring the wave equation to its normal form, and thus simplifies the integration considerably. In addition, since the transformation includes an arbitrary parameter, we obtain a versatile representation for the time-dependent Green's function. As a result, we are able to identify our form of the time-dependent Green's function with that obtained by applying the Lorentz transformation to the time-dependent Green's function in a stationary lossy medium. Finally, we obtain the time-harmonic Green's function by applying the Fourier integral to the time-dependent Green's function.

FORMULATION OF THE PROBLEM

In the rest frame Σ' of a moving medium, the scalar and vector potentials satisfy the inhomogeneous equations

$$\left(\nabla'^2 - \frac{n'^2}{c^2}\frac{\partial^2}{\partial t'^2} - \frac{\sigma' n'^2}{\epsilon' c^2}\frac{\partial}{\partial t'}\right)\mathbf{A}' = -\mu'\mathbf{J}', \quad (1)$$

$$\left(\nabla'^2 - \frac{n'^2}{c^2}\frac{\partial^2}{\partial t'^2} - \frac{\sigma' n'^2}{\epsilon' c^2}\frac{\partial}{\partial t'}\right)\phi = -\frac{\rho'}{\epsilon'}, \quad (2)$$

where μ', ϵ', and σ' are the permeability, permittivity, and conductivity of the medium, and $n' = c(\mu'\epsilon')^{\frac{1}{2}}$ is its index of refraction.

Transforming (1) and (2), with the help of a 4-vector,[2] to the Σ frame, we obtain

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{(n'^2-1)}{c^2}\gamma^2\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)^2 - \sigma'\mu'\gamma\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\right]\mathbf{A}$$
$$= -\mu'\mathbf{J} - \frac{\mu'\gamma^2}{c^2 n'^2}(n'^2-1)\mathbf{v}\mathbf{v}\cdot\mathbf{J} + \frac{\mu'(n'^2-1)}{n'^2}\rho\mathbf{v}, \quad (3)$$

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{(n'^2-1)}{c^2}\gamma^2\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)^2 - \sigma'\mu'\gamma\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\right]\phi$$
$$= -c^2\mu'\left[1 - \frac{(n'^2-1)}{n'^2}\gamma^2\right]\rho - \frac{\mu'(n'^2-1)}{n'^2}\gamma^2\mathbf{v}\cdot\mathbf{J}. \quad (4)$$

Equations (3) and (4) are identical to the equations given by Lee and Papas,[2] except that we have inserted the operator

$$-\sigma'\mu'\gamma\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)$$

to account for the conductivity of the medium. We then rewrite (3) and (4) in the following form[3]:

$$\left(L - p\frac{\partial}{\partial z} - q\frac{\partial}{\partial t}\right)\mathbf{A}$$
$$= -\mu'\mathbf{J} - \frac{\mu'\gamma^2}{c^2 n'^2}(n'^2-1)\mathbf{v}\mathbf{v}\cdot\mathbf{J} + \frac{\mu'(n'^2-1)}{n'^2}\rho\mathbf{v}, \quad (5)$$

$$\left(L - p\frac{\partial}{\partial z} - q\frac{\partial}{\partial t}\right)\phi = -\frac{1}{\epsilon'}\left[\frac{\rho}{a} - \frac{(n'^2-1)}{(1-\beta^2)c^2}\mathbf{v}\cdot\mathbf{J}\right], \quad (6)$$

[1] I. M. Besieris and R. T. Compton, J. Math. Phys. **8**, 2445 (1967).
[2] K. S. H. Lee and C. H. Papas, J. Math. Phys. **5**, 1668 (1964).
[3] There is an algebraic error in Ref. 1. It can be easily checked that (18a) in Ref. 1 does not follow from (17). The correct equations are (5) and (6) of the present work. Because of this error in Ref. 1, no attempt is made to compare our results with theirs.

where

$$L = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{1}{a}\frac{\partial^2}{\partial t^2} - 2\frac{\Omega}{a}\frac{\partial^2}{\partial t \partial z} + \left(\frac{\Omega^2}{a} - \mu'\epsilon' a\right)\frac{\partial^2}{\partial t^2}, \quad (7)$$

$$p = \sigma'\mu'\gamma v, \quad q = \sigma'\mu'\gamma,$$

and

$$a = \frac{1-\beta^2}{1-n'^2\beta^2}, \quad \beta = \frac{v}{c}, \quad \Omega = \frac{(n'^2-1)v}{(1-n'^2\beta^2)c^2},$$

$$\gamma = \frac{1}{(1-\beta^2)^{\frac{1}{2}}}.$$

We define the time-dependent Green's function $G(\mathbf{R}, \tau)$ as the solution of

$$\left(L - p\frac{\partial}{\partial z} - q\frac{\partial}{\partial t}\right)G(\mathbf{R}, \tau) = \delta(\mathbf{R})\delta(\tau), \quad (8)$$

where

$$\mathbf{R} = \mathbf{r} - \mathbf{r}', \quad \tau = t - t'.$$

GREEN'S FUNCTIONS

Any second-order hyperbolic partial differential equation with constant coefficients can be brought into the normal form by an affine transformation.[4] In the general case, one can include an arbitrary parameter in the transformation. This allows the derivation of different forms of the time-dependent Green's function. The affine transformation with three particular values of parameter is discussed in the Appendix.

Let us introduce the transformation

$$t_1 = \pm \frac{1}{2|a\mu'\epsilon'\lambda|^{\frac{1}{2}}}[(\lambda \pm 1)t + (\lambda\alpha \pm \delta)z], \quad (9)$$

$$z_1 = \pm \frac{1}{2|a\mu'\epsilon'\lambda|^{\frac{1}{2}}}[(\lambda \mp 1)t + (\lambda\alpha \mp \delta)z], \quad (10)$$

where the signs should be chosen so that t_1 increases with t, and α and δ are given by

$$\alpha = \Omega + a(\mu'\epsilon')^{\frac{1}{2}} \quad (11)$$

and

$$\delta = \Omega - a(\mu'\epsilon')^{\frac{1}{2}}. \quad (12)$$

It is shown in the Appendix that (8) transforms into

$$-\frac{\partial^2 G}{\partial t_1^2} + \frac{\partial^2 G}{\partial z_1^2} + \epsilon_1\left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} - p_1\frac{\partial G}{\partial z_1} - q_1\frac{\partial G}{\partial t_1}\right) = (1/\mu'\epsilon')^{\frac{1}{2}}\delta(t_1 - t_1')\delta(z_1 - z_1')\delta(\boldsymbol{\rho} - \boldsymbol{\rho}'), \quad (13)$$

where

$$\epsilon_1 = \begin{cases} 1, & \text{for } a > 0 \quad (\text{or } n'\beta < 1), \\ -1, & \text{for } a < 0 \quad (\text{or } n'\beta > 1), \end{cases} \quad (14)$$

and

$$p_1 = p(\partial z_1/\partial z) + q(\partial z_1/\partial t), \quad (15)$$

$$q_1 = p(\partial t_1/\partial z) + q(\partial t_1/\partial t). \quad (16)$$

The solution of (13) for $n'\beta < 1$ is

$$G(\mathbf{R}, \tau) = \frac{1}{(2\pi)^4(\mu'\epsilon')^{\frac{1}{2}}}\int \frac{\exp(j\mathbf{k}_1 \cdot \mathbf{R}_1 + j\omega_1\tau_1)\, d\mathbf{k}_1\, d\omega_1}{(\omega_1 - jq/2)^2 - (k_{1z} + jp_1/2)^2 - k_x^2 - k_y^2 - (q_1^2 - p_1^2)/4}, \quad (17)$$

where the integrations are along the real axes, and $\mathbf{R}_1 = \mathbf{r}_1 - \mathbf{r}_1'$, $\tau_1 = t_1 - t_1'$. The integration can be simplified by introducing the transformations $\omega_1' = \omega_1 - jq/2$, $k_{1z}' = k_{1z} + jp/2$ and then deforming the contours to the real ω_1', k_{1z}' axes. Since no singularity is enclosed, we obtain

$$G(\mathbf{R}, \tau) = \frac{1}{(2\pi)^4(\mu'\epsilon')^{\frac{1}{2}}} \times \int \frac{\exp(j\mathbf{k}_1' \cdot \mathbf{R}_1 + j\omega_1'\tau_1)\, d\mathbf{k}_1'\, d\omega_1'}{\omega_1'^2 - k_{1z}'^2 - k_x^2 - k_y^2 - (q_1^2 - p_1^2)/4} \times \exp(-q_1\tau_1/2 + p_1 Z_1/2), \quad (18)$$

where

$$Z_1 = z_1 - z_1'.$$

In terms of the spherical polar coordinate $k_1' = (k_x^2 + k_y^2 + k_{1z}'^2)^{\frac{1}{2}}$,

$$\theta_k = \tan^{-1}\frac{k_{1z}'}{(k_x^2 + k_y^2)^{\frac{1}{2}}}$$

and $\phi_k = \tan^{-1}(k_y/k_x)$ in the transformed space, the integrations are carried out with respect to θ_k and ϕ_k. Then, (18) reduces to

$$G(\mathbf{R}, \tau) = \frac{1}{(2\pi)^3(\mu'\epsilon')^{\frac{1}{2}}} \times \exp(-q_1\tau_1/2 + p_1 Z_1/2)\frac{1}{R_1}\frac{\partial}{\partial R_1} \times \int_{-\infty}^{\infty}\int_0^{\infty} \frac{\exp(jk_1'R_1 + j\omega_1'\tau_1)}{\omega_1'^2 - k_1'^2 - (q_1^2 - p_1^2)/4}\, dk_1'\, d\omega_1'. \quad (19)$$

where

$$R_1 = [(x - x')^2 + (y - y')^2 + (z_1 - z_1')^2]^{\frac{1}{2}}. \quad (20)$$

On using the integral

$$\int_{-\infty}^{\infty} d\omega_1' \exp(j\omega_1'\tau_1)\frac{\sin[\omega_1'^2 - (q_1^2 - p_1^2)/4]^{\frac{1}{2}}R_1}{[\omega_1'^2 - (q_1^2 - p_1^2)/4]^{\frac{1}{2}}} = \begin{cases} \pi J_0\{[(q_1^2 - p_1^2)/4]^{\frac{1}{2}}(R_1 - \tau_1)^{\frac{1}{2}}\}, & -R_1 \le \tau_1 \le R_1, \\ 0, & \text{otherwise}, \end{cases} \quad (21)$$

[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1962).

and dropping terms that always vanish, we have[5]

$$G(\mathbf{R},\tau)=\frac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}}\exp(-q_1\tau_1/2+p_1Z_1/2)\times\left\{\frac{\delta(R_1-\tau_1)}{R_1}+\frac{(\sigma'/2\epsilon')}{(R_1-\tau_1)^{\frac{1}{2}}}\times J_1[(\sigma'/2\epsilon')(R_1-\tau_1)^{\frac{1}{2}}]u(R_1-\tau_1)\right\},\quad(22)$$

where $u(x)$ is the unit step function, and $[(q_1^2-p_1^2)/4]^{\frac{1}{2}}$ has been reduced to $\sigma'/2\epsilon'$ by using (15) and (16).

Similarly, the solution of (13) for $n'\beta>1$ is

$$G(\mathbf{R},\tau)=\frac{1}{(2\pi)^4(\mu'\epsilon')^{\frac{1}{2}}}\exp(q_1\tau_1/2-p_1Z_1/2)\times\int\frac{\exp(j\mathbf{k}_1'\cdot\mathbf{R}_1+j\omega_1'\tau_1)\,dk_1'\,d\omega_1'}{\omega_1'^2-k_{1z}'^2+k_x^2+k_y^2-(q_1^2-p_1^2)/4}.\quad(23)$$

Integration with respect to k_{1z}' yields

$$G(\mathbf{R},\tau)=\begin{cases}\dfrac{1}{(2\pi)^4(\mu'\epsilon')^{\frac{1}{2}}}\exp(q_1\tau_1/2-p_1Z_1/2)\\ \quad\times\displaystyle\int\frac{\exp(jk_xX+jk_yY+j\omega_1'\tau_1)\sin[k_x^2+k_y^2+\omega_1'^2-(q_1^2-p_1^2)/4]^{\frac{1}{2}}Z_1}{[k_x^2+k_y^2+\omega_1'^2-(q_1^2-p_1^2)/4]^{\frac{1}{2}}}\,dk_x\,dk_y\,d\omega_1',\quad Z_1>0,\\ 0,\quad\text{otherwise,}\end{cases}\quad(24)$$

where

$$X=x-x',\quad Y=y-y'.$$

Introducing $X=\rho\cos\phi$, $Y=\rho\sin\phi$ and $k_x=k_\rho\cos\chi$, $k_y=k_\rho\sin\chi$, and integrating with respect to χ, we obtain

$$G(\mathbf{R},\tau)=\frac{1}{(2\pi)^2(\mu'\epsilon')^{\frac{1}{2}}}\exp(q_1\tau_1/2-p_1Z_1/2)\int_{-\infty}^{\infty}\int_0^{\infty}\frac{J_0(k_\rho\rho)\sin Z_1[k_\rho^2+\omega_1'^2-(q_1^2-p_1^2)/4]^{\frac{1}{2}}}{[k_\rho^2+\omega_1'^2-(q_1^2-p_1^2)/4]^{\frac{1}{2}}}\times\exp(j\omega_1'\tau_1)k_\rho\,dk_\rho\,d\omega_1'.\quad(25)$$

This, in turn, yields

$$G(\mathbf{R},\tau)=\begin{cases}\dfrac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}}\exp(q_1\tau_1/2-p_1Z_1/2)\left\{\dfrac{\delta(R_1'-\tau_1)}{R_1'}+\dfrac{\delta(R_1'+\tau_1)}{R_1'}+\dfrac{(\sigma'/2\epsilon')}{(R_1'-\tau_1)^{\frac{1}{2}}}\right.\\ \qquad\left.\times J_1[(\sigma'/2\epsilon')(R_1'-\tau_1)^{\frac{1}{2}}][u(R_1'-\tau_1)-u(-R_1'-\tau_1)]\right\},\quad Z_1^2\geq X^2+Y^2,\\ 0,\quad\text{otherwise,}\end{cases}\quad(26)$$

where

$$R_1'=(Z_1^2-X^2-Y^2)^{\frac{1}{2}}.$$

The support of $\delta(R_1'+\tau_1)$ and $u(-R_1'-\tau_1)$ for R_1', $\tau_1\geq 0$ is $R_1'+\tau_1=0$, which is satisfied only when $\tau=0$ and

$$R=\left[\frac{1}{n'^2}\left(\frac{n'^2-\beta^2}{1-\beta^2}\right)Z^2-X^2-Y^2\right]^{\frac{1}{2}}=0.^6$$

For $\tau\neq 0$, again, we can drop these two terms. There remains

$$G(\mathbf{R},\tau)=\begin{cases}\dfrac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}}\exp(q_1\tau_1/2-p_1Z_1/2)\left\{\dfrac{\delta(R_1'-\tau_1)}{R_1'}+\dfrac{(\sigma'/2\epsilon')}{(R_1'-\tau_1)^{\frac{1}{2}}}J_1[(\sigma'/2\epsilon')(R_1'-\tau_1)^{\frac{1}{2}}]u(R_1'-\tau_1)\right\},\\ \qquad Z_1^2\geq X^2+Y^2,\\ 0,\quad\text{otherwise.}\end{cases}\quad(27)$$

[5] With a choice of $\lambda=|\delta/\alpha|$ in the affine transformation, one can impose the causality condition in the integral. Since the equivalence of the solutions for different values of λ follows from the uniqueness theorem of the partial-differential equation, we conclude that (22) satisfies the causality condition.

[6] $\lambda=|\delta/\alpha|$ is assumed to satisfy the causality condition.

Equations (22) and (27) are independent of λ, because of the uniqueness of the solution of the partial differential equation. To explain this point, we consider the wavefront of (22). Using (9) and (10), we have, for $n'\beta < 1$,

$$R_1'^2 - \tau_1^2 = \frac{1}{4a\mu'\epsilon'}[(\lambda - 1)\tau + (\alpha\lambda - \delta)Z]^2 + X^2 + Y^2 - \frac{1}{4a\mu'\epsilon'}[(\lambda + 1)\tau + (\alpha\lambda + \delta)Z]^2$$
$$= -\frac{1}{4a\mu'\epsilon'}[\tau^2 + 2\Omega\tau Z + (\Omega^2 - a^2\mu'\epsilon')Z^2] + X^2 + Y^2, \quad (28)$$

which is independent of λ, as expected. A similar expression can be derived for $n'\beta > 1$ by using (27). The spheroidal shape of the wavefront of (22), i.e., (28), was discussed in great detail by Besieris and Compton.[1] One can easily show that the wavefront of (27) is a hyperboloid.

To see that (22) reduces to that derived by Compton[7] for the lossless case. With $\sigma' = 0$, $p_1 = q_1 = 0$. Thus, (22) becomes

$$G(\mathbf{R}, \tau) = \frac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}}\frac{\delta(R_1 - \tau_1)}{R_1}. \quad (29)$$

Choosing $\lambda = |\delta/\alpha|$, i.e., applying the transformation (A6) and (A7), and using the identity $\delta(ax) = 1/|a|\,\delta(x)$, we have

$$G(\mathbf{R}, \tau) = \frac{-1}{4\pi n'}\left(\frac{n'^2 - \beta^2}{1 - \beta^2}\right)\frac{1}{R_{10}}\,\delta\left[\tau - \frac{R_{10}}{c}\left(\frac{n'^2 - \beta^2}{1 - \beta^2}\right)\right], \quad (30)$$

where

$$R_{10} = \left[\frac{(n'^2 - \beta^2)}{n'^2(1 - \beta^2)}\left(Z - \frac{n'^2 - 1}{n'^2 - \beta^2}v\tau\right)^2 + X^2 + Y^2\right]^{\frac{1}{2}}. \quad (31)$$

Equation (30) is exactly the formula Compton derived.

We then proceed to show that with the choice of $\lambda = |1 - n'\beta|/|1 + n'\beta|$ (see the Appendix), (22) reduces to that obtained by applying the Lorentz transformation to the Green's function in a stationary lossy medium. Using (A8), (A9), (15), and (16), we obtain

$$p_1 = 0, \quad q_1 = \sigma'\mu' c/n'. \quad (32)$$

Substitution of (A8), (A9), and (32) into (22) gives

$$G(\mathbf{R}, \tau) = \frac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}}\exp(-\sigma'\tau'/2\epsilon') \times \left\{\frac{\delta(R' - c\tau'/n')}{R'} + \frac{(\sigma'/2\epsilon')}{(R' - c\tau'/n')^{\frac{1}{2}}} \times J_1[(\sigma'/2\epsilon')(R' - c\tau'/n')^{\frac{1}{2}}]u(R' - c\tau'/n')\right\}, \quad (33)$$

where

$$\tau' = \gamma(\tau - vZ/c^2), \quad (34)$$
$$Z' = \gamma(Z - v\tau), \quad (35)$$

and

$$R' = (X^2 + Y^2 + Z'^2)^{\frac{1}{2}}. \quad (36)$$

One immediately recognizes that (33) is the time-dependent Green's function for electromagnetic waves in a stationary lossy medium, i.e., the solution of the well-known wave equation we began with, which is

$$\left(\nabla'^2 - \frac{n'^2\partial^2}{c^2\partial t'^2} - \frac{\sigma' n'^2}{\epsilon' c^2}\frac{\partial}{\partial t'}\right)G = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'). \quad (37)$$

Therefore, we arrive at the conclusion that for $n'\beta < 1$, the simplest way of obtaining the time-dependent Green's function is by applying the Lorentz transformation to that in a stationary lossy medium. Note that the same transformation gives the correct result for $n'\beta > 1$ (Cerenkov radiation)[8] if one replaces (36) by

$$R' = (-X^2 - Y^2 + Z'^2)^{\frac{1}{2}}. \quad (38)$$

Lastly, we let $\lambda = 1$. Again (A4), (A5), (15), and (16) yield

$$p_1 = \pm\sigma'\mu'\gamma v\,|a|^{\frac{1}{2}}, \quad (39)$$
$$q_1 = \pm\sigma'\mu'\gamma\,|a|^{\frac{1}{2}}/(\mu'\epsilon')^{\frac{1}{2}}, \quad (40)$$

where the upper sign is for $n'\beta < 1$ and the lower sign for $n'\beta > 1$. Substituting (A4), (A5), (39), and (40) into (22) and (26), we have, for $n'\beta < 1$,

$$G(\mathbf{R}, \tau) = \frac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}} \times \exp[(-\sigma'\gamma/2\epsilon')(\tau + \Omega Z) + \sigma'\mu'\gamma v a Z/2] \times \left[\frac{\delta[R_{11} - (\tau + \Omega Z)/(a\mu'\epsilon')^{\frac{1}{2}}]}{R_{11}} + \frac{(\sigma'/2\epsilon')}{[R_{11} - (\tau + \Omega Z)/(a\mu'\epsilon')^{\frac{1}{2}}]^{\frac{1}{2}}} \times J_1\{(\sigma'/2\epsilon')[R_{11} - (\tau + \Omega Z)/(a\mu'\epsilon')^{\frac{1}{2}}]^{\frac{1}{2}}\} \times u[R_{11} - (\tau + \Omega Z)/(a\mu'\epsilon')^{\frac{1}{2}}]\right], \quad (41)$$

where

$$R_{11} = (aZ^2 + X^2 + Y^2)^{\frac{1}{2}}; \quad (42)$$

[7] R. T. Compton, J. Math. Phys. 7, 2145 (1966).

[8] For $\tau \neq 0$.

for $n'\beta > 1$,

$$G(\mathbf{R},\tau) = \begin{cases} \dfrac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}} \exp\left[(-\sigma'\gamma/2\epsilon')(\tau + \Omega Z) + \sigma'\mu' vaZ/2\right]\Big[\dfrac{\delta[R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]}{R'_{11}} \\ \qquad + \dfrac{\delta[R'_{11} + (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]}{R'_{11}} + \dfrac{(\sigma'/2\epsilon')}{[R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]^{\frac{1}{2}}} \\ \qquad \times J_1\{(\sigma'/2\epsilon')[R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon|^{\frac{1}{2}}]^{\frac{1}{2}}\}\{u[R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}] \\ \qquad - u[-R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]\}\Big], \quad |a|\, Z^2 \geq X^2 + Y^2, \\ 0, \quad \text{otherwise} \end{cases} \tag{43}$$

where

$$R'_{11} = (|a|\, Z^2 - X^2 - Y^2). \tag{44}$$

Before taking the Fourier transform of (41) and (43) to yield the time-harmonic Green's function, notice that (26) instead of (27) should be used for $n'\beta > 1$, because the value of the function at $t = 0$ contributes to its Fourier transform for all frequencies. We then write (41) and (43) as follows:

$$G(\mathbf{R},\tau) = \frac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}} \exp\left[-(\sigma'\gamma/2\epsilon')(\tau + \Omega Z) + \sigma'\mu'\gamma vaZ/2\right] \frac{1}{R_{11}} \frac{\partial}{\partial R_{11}} \left[J_0\{(\sigma'/2\epsilon')[R_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]^{\frac{1}{2}}\}\right.$$
$$\left. \times\, u[R_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]\right], \tag{45}$$

for $n'\beta < 1$;

$$G(\mathbf{R},\tau) = \begin{cases} \dfrac{-1}{4\pi(\mu'\epsilon')^{\frac{1}{2}}} \exp\left[-(\sigma'\gamma/2\epsilon')(\tau + \Omega Z) + \sigma'\mu'\gamma vaZ/2\right] \dfrac{1}{R'_{11}} \dfrac{\partial}{\partial R'_{11}} \left[J_0\{(\sigma'/2\epsilon')[R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]^{\frac{1}{2}}\}\right. \\ \qquad \left. \times\, \{u[R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}] - u[-R'_{11} - (\tau + \Omega Z)/|a\mu'\epsilon'|^{\frac{1}{2}}]\}\right], \quad |a|\, Z^2 \geq X^2 + Y^2, \\ 0, \quad \text{otherwise}, \end{cases} \tag{46}$$

for $n'\beta > 1$.

Since R_{11} and R'_{11} in (45) and (46) do not depend on τ, it is easy to perform the Fourier transform to give the time-harmonic Green's function. The result of such a calculation is

$$G(\mathbf{R},\omega) = \frac{-a^{\frac{1}{2}}}{4\pi R_{11}} \exp\left(-\sigma'\gamma\Omega/2 + \sigma'\mu'\gamma va/2\right)Z \exp j[(\omega - j\sigma'\gamma/2\epsilon')^2 - (\sigma'/2\epsilon')^2/|a\mu'\epsilon'|^{\frac{1}{2}}]^{\frac{1}{2}}(R_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z),$$

for $n'\beta < 1$; (47)

$$G(\mathbf{R},\omega) = \begin{cases} \dfrac{-|a|^{\frac{1}{2}}}{2\pi R'_{11}} \exp\left(-\sigma'\gamma\Omega/2\epsilon' + \sigma'\mu'\gamma va/2\right)Z \\ \qquad \times \cos\left[(\omega - j\sigma'\gamma/2\epsilon')^2 - (\sigma'/2\epsilon')^2/|a\mu'\epsilon'|^{\frac{1}{2}}]^{\frac{1}{2}} \cdot (R'_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z\right), \quad |a|\, Z^2 \geq X^2 + Y^2, \\ 0, \quad \text{otherwise}, \end{cases} \tag{48}$$

for $n'\beta > 1$.

Separating the real and imaginary parts in the exponentials, we finally obtain

$$G(\mathbf{R},\omega) = \frac{-a^{\frac{1}{2}}}{4\pi R_{11}} \exp\left(-\sigma'\gamma\Omega/2\epsilon' + \sigma'\mu'\gamma va/2\right)Z \exp -\{j[(P^2 + Q^2)^{\frac{1}{2}} + P]^{\frac{1}{2}}$$
$$+ [(P^2 + Q^2)^{\frac{1}{2}} - P]^{\frac{1}{2}}\}(R_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z) \tag{49}$$

for $n'\beta < 1$;

$$G(\mathbf{R},\omega) = \begin{cases} \dfrac{-|a|^{\frac{1}{2}}}{2\pi R'_{11}} \exp\{(-\sigma'\gamma\Omega/2\epsilon' + \sigma'\mu'\gamma va/2)Z - [(P^2 + Q^2)^{\frac{1}{2}} - P]^{\frac{1}{2}}(R'_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z)\} \\ \qquad \times \cos\left[(P^2 + Q^2)^{\frac{1}{2}} + P\right]^{\frac{1}{2}}(R'_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z), \quad |a|\, Z^2 \geq X^2 + Y^2, \\ 0, \quad \text{otherwise}, \end{cases} \tag{50}$$

for $n'\beta > 1$, where

$$P = |\omega^2 - (\sigma'\gamma/2\epsilon')^2 - (\sigma'/2\epsilon')^2/|a\mu'\epsilon'|^{\frac{1}{2}}| \quad (51)$$

and

$$Q = \sigma'\gamma\omega/\epsilon'. \quad (52)$$

With $\sigma' = 0$, (49) and (50) become

$$G(\mathbf{R}, \omega) = \frac{-a^{\frac{1}{2}}}{4\pi R_{11}} \exp -j(R_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z)\omega, \quad (53)$$

for $n'\beta < 1$;

$$G(\mathbf{R}, \omega) = \begin{cases} \dfrac{-|a|^{\frac{1}{2}}}{2\pi R'_{11}} \cos \omega(R'_{11}\,|a\mu'\epsilon'|^{\frac{1}{2}} - \Omega Z), \\ \qquad\qquad |a|\,Z^2 \geq X^2 + Y^2, \\ 0, \quad \text{otherwise}, \end{cases} \quad (54)$$

for $n'\beta > 1$.

Equations (53) and (54) can then be compared with the formulas given by Lee and Papas.[2,9]

APPENDIX

The affine transformation

$$\xi = t + \alpha z,$$
$$\eta = t + \delta z,$$

where

$$\alpha = \Omega + a(\mu'\epsilon')^{\frac{1}{2}}, \quad \delta = \Omega - a(\mu'\epsilon')^{\frac{1}{2}},$$

transforms the operator

$$\frac{1}{a}\frac{\partial^2}{\partial z^2} - 2\frac{\Omega}{a}\frac{\partial^2}{\partial z \partial t} + \left(\frac{\Omega^2}{a} - \mu'\epsilon' a\right)\frac{\partial^2}{\partial t^2} + \cdots \quad (A1)$$

into

$$-4\mu'\epsilon' a \frac{\partial^2}{\partial \xi \partial \eta} + \cdots. \quad (A2)$$

Another transformation,

$$t_1 = \pm \frac{1}{2\,|a\mu'\epsilon'\lambda|^{\frac{1}{2}}}(\lambda\xi \pm \eta)$$
$$= \frac{\pm 1}{2\,|a\mu'\epsilon'\lambda|^{\frac{1}{2}}}[(\lambda \pm 1)t + (\lambda\alpha \pm \delta)z], \quad (9)$$

$$z_1 = \pm \frac{1}{2\,|a\mu'\epsilon'\lambda|^{\frac{1}{2}}}(\lambda\xi \mp \eta)$$
$$= \frac{\pm 1}{2\,|a\mu'\epsilon'\lambda|^{\frac{1}{2}}}[(\lambda \mp 1)t + (\lambda\alpha \mp \delta)z], \quad (10)$$

brings (A2) into

$$\epsilon_1\left(-\frac{\partial^2}{\partial t_1^2} + \frac{\partial^2}{\partial z_1^2}\right) + \cdots$$
$$= \frac{\epsilon_1}{(\mu'\epsilon')^{\frac{1}{2}}}\,\delta(\boldsymbol{\rho} - \boldsymbol{\rho}')\delta(z_1 - z_1')\delta(t_1 - t_1'). \quad (A3)$$

$1/2\,|a\mu'\epsilon'\lambda|^{\frac{1}{2}}$ is the scale factor used to give the magnitude of the coefficients of (A3) unity and ϵ_1 is given by (15). Also, in (9) and (10), the signs should be chosen so that t_1 increases with t.

These transformations are described in most books on partial differential equations. The only difference is the introduction of the parameter λ, of which we would like to discuss three particular choices.

Consider $\lambda = 1$. Equations (9) and (10) reduce to

$$\begin{cases} t_1 = \dfrac{1}{|a\mu'\epsilon'|^{\frac{1}{2}}}(t + \Omega z), & (A4) \\ z_1 = \pm |a|^{\frac{1}{2}}\, z. & (A5) \end{cases}$$

This choice of λ is used to derive the time-harmonic Green's function.

Setting $\lambda = |\delta/\alpha|$ in (9) and (10), we have

$$\begin{cases} t_1 = \left(\dfrac{1 - \beta^2}{n'^2 - \beta^2}\right)^{\frac{1}{2}} ct, & (A6) \\ z_1 = \dfrac{1}{n'}\left(\dfrac{n'^2 - \beta^2}{1 - \beta^2}\right)^{\frac{1}{2}}\left[z - \dfrac{(n'^2 - 1)vt}{(n'^2 - \beta^2)}\right]. & (A7) \end{cases}$$

Finally, for $\lambda = |1 - n'\beta|/|1 + n'\beta|$, (9) and (10) become

$$\begin{cases} t_1 = \dfrac{c}{n'(1 - \beta^2)^{\frac{1}{2}}}(t - vz/c^2), & (A8) \\ z_1 = \dfrac{1}{(1 - \beta^2)^{\frac{1}{2}}}(z - vt). & (A9) \end{cases}$$

Aside from a minor factor (c/n'), (A8) and (A9) is the familiar Lorentz transformation.

[9] There are some notational differences between Ref. 2 and our work, e.g., a (Ref. 2) $= |a|^{\frac{1}{2}}$. To compare (53) and (54) with the formulas in Ref. 2, we give the notational conversions as follows:

$$k \text{ (Ref. 2)} = |a|^{\frac{1}{2}} n' k_0 = |a|^{\frac{1}{2}} n'\omega/c = |a|^{\frac{1}{2}}(\mu'\epsilon')^{\frac{1}{2}}\omega,$$

$$bk \text{ (Ref. 2)} = k\,|a|^{\frac{1}{2}}\beta\gamma^2[(n'^2 - 1)/n'] = \frac{\beta(n'^2 - 1)\omega}{c(1 - n'^2\beta^2)} = \omega\Omega.$$

Also, the Green's functions are defined with a difference in sign.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Class of Representations of the $IU(n)$ and $IO(n)$ Algebras and Respective Deformations to $U(n, 1)$, $O(n, 1)$

A. CHAKRABARTI
Centre de Physique Théorique de l'Ecole Polytechnique, 17 rue Descartes, 75 Paris, France

(Received 15 December 1967)

We define directly the matrix elements of the generators of the algebra of $U(n) \times I_{2n}$ on a chosen basis. This construction, though naturally infinite-dimensional, has a very close formal resemblance (interpretable, if so desired, in terms of a suitably defined "contraction" procedure) to the Gel'fand–Zetlin (GZ) representation for $U(n + 1)$. The representations we obtain are characterized by $(n - 1)$ integers and one continuous parameter. We then exploit the formal analogy with the GZ pattern in order to prove the necessary commutation relations and to derive the explicit expressions for some invariants. However, direct alternative methods are indicated where they are useful. Our representation is easily enlarged to that of $(U(n) \otimes U(1)) \times I_{2n}$, which we use, together with a deformation formula to obtain a class of representations of the $U(n, 1)$ algebra. The irreducible components are characterized by $(n - 1)$ integers and two continuous parameters. We compare our deformation formula with that of Rosen and Roman. We indicate briefly the typical difficulties that arise for the case $IU(p, q)$ $(q \geq 1)$. Parallel constructions, finally, are given for the $IO(n)$ and $O(n, 1)$ algebras

1. INTRODUCTION

In Sec. 2, we show how a very simple formal modification of the well-known Gel'fand–Zetlin (GZ) matrix elements[1] for $U(n + 1)$ leads to a representation of the $IU(n)$ algebra, defined in (2.2)–(2.5). The representations obtained are labeled by $(n - 1)$ integers and one continuous parameter, which we can take to be real.

We can easily generalize our algebra to that of

$$(U(n) \otimes U(1)) \times I_{2n} \tag{1.1}$$

and use this as a starting point to obtain a class of representations of the $U(n, 1)$ algebra. The necessary deformation formula and the corresponding explicit results for $U(n, 1)$ representations [labeled by $(n - 1)$ integers and two continuous real parameters] are given in Sec. 3.

The deformation formulas used [(4.3)–(4.5)] *are also applicable to the more general case of*

$$(U(p, q) \otimes U(1)) \times I_{2n} \to U(p, q + 1). \tag{1.2}$$

But then formal difficulties arise if we want to start with a representation which diagonalizes all the Casimir operators of the homogeneous part. This seems to be a general phenomenon (see our discussion of the Poincaré algebra[2] and some references quoted in that paper). In Sec. 5, we examine (briefly) exactly how the difficulties (different from those encountered in Ref. 2) manifest themselves if we start with the discrete Gel'fand–Graev (GG) representation.[3] In order to obtain well-behaved matrix elements corresponding to the deformation (1.2), we should use representations of $IU(p, q)$ which diagonalize the generators of the Abelian subalgebra (or probably at least some of them[4]). But this is a separate problem. In this paper we confine ourselves to representations of $IU(n)$ reduced with respect to those of $U(n)$, in which none of the "translation" generators are diagonal, and we have a maximum number of discrete parameters.

In Sec. 4, we elucidate the close connection of our deformation formula with that proposed by Rosen and Roman.[5] We obtained the deformation formula in the form used in the text, since we have been thinking in terms of contractions and deformations of algebras without altering the number of generators.

For both cases, unitary (Sec. 2) and orthogonal (Sec. 6), we merely indicate the necessary prescriptions for interpreting the passage to the inhomogeneous algebra as a Wigner–Inönu contraction, since, once one notices the needed coefficients [as in (2.15)], the passage to the infinite limit is not really necessary unless one has a definite physical interpretation in view (as for the passage to the Galilei group for $c \to \infty$), which is not our case.

Everywhere, we exploit, as far as possible, the known results for the GZ basis. This, in our opinion, gives a better understanding of the structure of the formulas, both for the homogeneous and inhomogeneous case. To show, however, that other techniques can be introduced usefully, we calculate the second-order invariant $(\Delta_{(2)})$ in two different ways. The second, and the longer, method proves useful later on.

[1] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963); J. G. Nagel and M. Moshinsky, *ibid.* **6**, 682 (1965).

[2] A. Chakrabarti, M. Levy-Nahas, and R. Seneor, J. Math. Phys. **9**, 1274 (1968).

[3] I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR Ser. Mat. **29**, 1329 (1965); I. T. Todorov, Trieste preprint IC/66/71.

[4] R. Mirman, J. Math. Phys. **8**, 57 (1967).

[5] J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966). Some interesting constructions are also given by P. Chand, C. L. Mehta, N. Mukunda, and E. C. G. Sudarshan, Syracuse University Preprint SU-1206-90.

The basic results for the discrete representations for the homogeneous case (particularly the GZ representations for the compact case) are supposed to be known.[1,3] Nevertheless, we recapitulate, very briefly, some results in the Appendix—mainly in order to fix our notations and conventions and, sometimes, for the sake of ready comparison.

2. REPRESENTATIONS OF $IU(n)$

In this section we give the matrix elements of the generators of the algebra of

$$U(n) \times I_{2n} \tag{2.1}$$

acting on a basis to be defined below. In (2.1), $U(n)$ represents the maximal compact subalgebra [with the generators A^i_j $(i, j = 1, \cdots, n)$] and I_{2n} is an Abelian subalgebra formed by the $2n$ generators I^i_{n+1}, I^{n+1}_i $(i = 1, \cdots, n)$ in semidirect product with $U(n)$. (*The index $n+1$ is evidently not necessary here or in the other formulas in this section. But, so far as this paper is concerned, we prefer this notation in order to display, as clearly as possible, certain very useful analogies.*)

The commutation relations are

$$[A^i_j, A^k_l] = \delta^i_l A^k_j - \delta^k_j A^i_l, \tag{2.2}$$

$$[A^i_j, I^k_{n+1}] = -\delta^k_j I^i_{n+1}, \quad [A^i_j, I^{n+1}_k] = \delta^i_k I^{n+1}_j, \tag{2.3}$$

$$[I^i_{n+1}, I^j_{n+1}] = [I^{n+1}_i, I^{n+1}_j] = [I^{n+1}_i, I^j_{n+1}] = 0, \tag{2.4}$$

with

$$(A^i_j)^\dagger = A^j_i, \quad (I^i_{n+1})^\dagger = I^{n+1}_i \quad (i, j, k, l = 1, \cdots, n). \tag{2.5}$$

Let us now define the (infinite-dimensional) basis

$$\left| \begin{array}{ccccccc} h_{2\,n+1} & & h_{3\,n+1} & \cdots & & & h_{n\,n+1} \\ & h_{1n} & & h_{2n} & \cdots & h_{n-1\,n} & & h_{nn} \\ & h_{1\,n-1} & & & \cdots & & h_{n-1\,n-1} & \\ & & & & \cdots & & & \\ & & & & h_{11} & & & \end{array} \right\rangle_{(\kappa)} \tag{2.6}$$

where κ is an arbitrary real number and the integers h satisfy the following inequalities:

$$h_{i\,j+1} \geq h_{ij} \geq h_{i+1\,j+1} \quad (i, j = 1, \cdots, n) \tag{2.7}$$

and

$$h_{1n} \geq h_{2\,n+1} \geq h_{2n} \geq \cdots \geq h_{n\,n+1} \geq h_{nn}. \tag{2.8}$$

The matrix elements of A^i_j acting on this basis are formally identical with the familiar ones of the GZ formalism [Ref. 1, Appendix A].

Those of the I's are defined as follows. The nonzero matrix elements of I^n_{n+1} and I^{n+1}_n are given (in the notation of the Appendix) by

$$\langle h_{jn} - 1 | I^n_{n+1} | h \rangle = \kappa \left[\frac{\prod_{i=2}^{n} (h_{i\,n+1} - h_{jn} - i + j + 1) \prod_{i=1}^{n-1} (h_{i\,n-1} - h_{jn} - i + j)}{\prod_{1 \leq i \leq n; i \neq j} (h_{in} - h_{jn} - i + j + 1)(h_{in} - h_{jn} - i + j)} \right]^{\frac{1}{2}} \tag{2.9}$$

$$= \langle h | I^{n+1}_n | h_{jn} - 1 \rangle^* \quad (j = 1, \cdots, n). \tag{2.10}$$

The matrix elements of I^i_{n+1}, I^{n+1}_i $(i = 1, \cdots, n-1)$ are obtained from (2.3), (2.9), and (2.10).

[Taking κ to be complex amounts to a different phase convention for the I's; the invariants to be evaluated would then be functions of $(\kappa\kappa^*)^{\frac{1}{2}}$ instead of κ^2. We will always choose κ to be real.] Thus we have effectively a representation characterized by the $(n-1)$ integral discrete parameters $h_{i\,n+1}$ $(i = 1, \cdots, n)$ and one real continuous parameter κ.

Let us now compare the matrix elements (2.9), (2.10) with the corresponding ones of A^n_{n+1} and A^{n+1}_n for a (finite-dimensional) GZ basis of $U(n+1)$. The difference arises, evidently, from the absence in (2.6) of the upper and lower bounds

$$h_{1\,n+1}, \quad h_{n+1\,n+1}, \tag{2.11}$$

(which make the GZ basis finite-dimensional) and the suppression in the definition (2.9) of a corresponding factor (always real):

$$[-(h_{1\,n+1} - h_{j\,n+j})(h_{n+1\,n+1} - h_{jn} - n + j)]^{\frac{1}{2}}. \tag{2.12}$$

It is true that this modification changes fundamentally the nature of the basis, releasing the outermost diagonals so that we may have

$$h_{11} \leq h_{12} \leq \cdots \leq h_{1n} \to +\infty,$$
$$h_{11} \geq h_{22} \geq \cdots \geq h_{nn} \to -\infty, \tag{2.13}$$

and, in particular,

$$h_{11} \to +\infty \quad \text{or} \quad -\infty. \tag{2.14}$$

Since the matrix elements are defined explicitly, one can try to verify all the necessary properties directly. But a more economical method is to exploit the known results for the GZ basis in spite of the fundamental differences between the two cases.

What is sufficient for our purpose is that, *as a formal expression* [for $\kappa \neq 0$ and supposing, for the moment, that the I's of (2.9), (2.10) are acting on states whose parameters happen to be in the ranges conforming to those for $U(n+1)$], we have

$$\langle h|\,[A^{n+1}_n, A^n_{n+1}]\,|h\rangle = -\kappa^{-2}(h_{1\,n+1}h_{n+1\,n+1})\langle h|\,[I^{n+1}_n, I^n_{n+1}]\,|h\rangle + (\text{other terms not containing the product } h_{1\,n+1}h_{n+1\,n+1}). \quad (2.15)$$

But we also have

$$[A^{n+1}_n, A^n_{n-1}]\,|h\rangle = (A^n_n - A^{n+1}_{n+1})\,|h\rangle = \left(2\sum_{i=1}^{n} h_{in} - \sum_{i=1}^{n-1} h_{i\,n-1} - \sum_{i=1}^{n+1} h_{i\,n+1}\right)|h\rangle. \quad (2.16)$$

Since the right-hand side of (2.16) does not contain the product $(h_{1\,n+1}h_{n+1\,n+1})$, comparing it with (2.15), we obtain

$$\langle h|\,[I^{n+1}_n, I^n_{n+1}]\,|h\rangle = 0. \quad (2.17)$$

(For the trivial case $\kappa = 0$, the result is again evident.)

Now if we pass out of the previous domains of the parameters, the validity of (2.17) is not affected, since that depends on formal cancelations and not on particular values. *What we have rather to ensure is the essential Hermiticity restriction* (2.10). *But this point is guaranteed through our choice of the extreme factors (namely, those involving* h_{1n+1} *and* $h_{n+1\,n+1}$*) in* (2.12). Hence, in general, we can write

$${}_\kappa\langle h|\,[I^{n+1}_n, I^n_{n+1}]\,|h\rangle_\kappa = 0. \quad (2.17')$$

Once this crucial result is established, the rest follows quite easily. For this it is only necessary to note that the vanishing of the nondiagonal terms in (2.16) does not depend on the factors suppressed (2.12) and that, in such relations as

$$[A^i_{n+1}, A^j_{n+1}] = 0 = [A^{n+1}_i, A^{n+1}_j] \quad (i, j = 1, \cdots, n), \quad (2.18)$$

the coefficients of the factors containing (under square root) $h_{i\,n+1}$ $(i = 1, \cdots, n)$ each vanish separately. The derivation of (2.3) is also now evident from analogy.

Let us note, for subsequent use, that we can immediately enlarge our algebra to that of

$$(U(n) \otimes U(1)) \times I_{2n}, \quad (2.19)$$

by defining (*introducing an extra parameter* ζ):

$$A^{n+1}_{n+1}\,|h\rangle_{(\kappa,\zeta)} = \left(\zeta + \sum_{i=2}^{n} h_{i\,n+1} - \sum_{i=1}^{n} h_{in}\right)|h\rangle_{(\kappa,\zeta)}, \quad (2.20)$$

where ζ *is an arbitrary real number*, if A^{n+1}_{n+1} is to be Hermitic diagonal. Now,

$$[A^{n+1}_{n+1}, A^i_j] = 0 \quad (i, j = 1, \cdots, n),$$
$$[A^{n+1}_{n+1}, I^i_{n+1}] = I^i_{n+1}, \quad [A^{n+1}_{n+1}, I^{n+1}_i] = -I^{n+1}_i. \quad (2.20')$$

The other matrix elements remain unchanged. We could, of course, have absorbed the sum on $h_{i\,n+1}$ in ζ. But we prefer the form (2.20) for the eventual simplicity of certain formulas.

The representation (2.20) has now for the first-order invariant

$$\Delta_{(1)}\,|h\rangle_{(\kappa,\zeta)} \equiv \left(\sum_{i=1}^{n+1} A^i_i\right)|h\rangle_{(\kappa,\zeta)} = \left(\zeta + \sum_{i=2}^{n} h_{i\,n+1}\right)|h\rangle_{(\kappa,\zeta)}. \quad (2.21)$$

We will also need the second- and third-order invariants defined as

$$\Delta_{(2)} = \sum_{i=1}^{n} I^{n+1}_i I^i_{n+1}, \quad (2.22)$$

$$\Delta_{(3)} = \sum_{i,j=1}^{n} I^{n+1}_i A^i_j I^j_{n+1} + A^{n+1}_{n+1}\Delta_{(2)}, \quad (2.23)$$

respectively. The general expressions for the eigenvalues of these operators will also be denoted by $\Delta_{(1)}$, $\Delta_{(2)}$, and $\Delta_{(3)}$, respectively.

For the calculation of these eigenvalues, we can again very usefully employ a technique similar to that of (2.15).

For example, the second-order Casimir operator of $U(n+1)$ in the GZ representation is given by[6]

$${}_{U(n+1)}\langle h|\left(\sum_{i,j=1}^{n+1} A^i_j A^j_i\right)|h\rangle_{U(n+1)} = \sum_{i=1}^{n+1} h_{i\,n+1}(h_{i\,n+1} + n + 2 - 2i). \quad (2.24)$$

Comparing [as for (2.15)] the coefficients of

$$h_{1\,n+1} \quad h_{n+1\,n+1}$$

on both sides of (2.24), we obtain

$$2h_{1\,n+1}h_{n+1\,n+1} - h_{1\,n+1}h_{n+1\,n+1}\left(\frac{2\Delta_{(2)}}{\kappa^2}\right) = 0. \quad (2.25)$$

[The first term comes from $(A^{n+1}_{n+1})^2$.] Hence,

$$\Delta_{(2)} = \kappa^2. \quad (2.26)$$

This expression is the same for the algebras (2.1) and (2.19).

[6] A. M. Peremolov and V. S. Popov, Sov. J. Nucl. Phys. **3**, 676 819 (1966); **5**, 489 (1967).

Other techniques are, however, possible; for future use we indicate a relatively simple method.

As is well known[6] the evaluation of the invariants is greatly facilitated for the compact case by the use of "maximal" (or "minimal") states, namely, the states for which the variable parameters attain their maximal (or minimal) values. In our case, strictly speaking, there are no such states, the representation being of infinite dimensions. Nevertheless, we can attain a relative simplicity by considering, for example, such a "pseudominimal" state as,

$$\left|\begin{matrix} & h_{2\,n+1} & & h_{3\,n+1} & \cdots & & h_{n\,n+1} & \\ h_{2\,n+1} & & h_{3\,n+1} & & \cdots & h_{n\,n+1} & & h_{nn} \\ & h_{3\,n+1} & & & \cdots & & h_{nn} & \\ & & & & \cdots & & & \\ & & & & h_{nn} & & & \end{matrix}\right\rangle_{(\kappa,\zeta)}. \tag{2.27}$$

(The corresponding "pseudomaximal" state can also be used.)

Acting on such a state, I^n_{n+1} has only one nonzero matrix element corresponding to

$$h_{nn} \to (h_{nn} - 1). \tag{2.28}$$

In fact, denoting (for our immediate purpose) such a state by

$$\left|\begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right\rangle, \tag{2.29}$$

we have (in an obvious notation):

$$\Delta_{(2)} = \left\langle\begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right| \left(\sum_{i=1}^{n} I_i^{n+1} I_{n+1}^i\right) \left|\begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right\rangle$$

$$= \left\langle\begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right| I_n^{n+1} I_{n+1}^n + \sum_{i=1}^{n-1} (A_i^n I_n^{n+1} - I_n^{n+1} A_i^n)$$

$$\times (I_{n+1}^n A_n^i - A_n^i I_{n+1}^n) \left|\begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right\rangle$$

$$= \left\langle\begin{matrix} h_{nn} - 1 \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right| I_{n+1}^n \left|\begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right\rangle^2$$

$$\times \left[1 + \left\langle\begin{matrix} h_{nn} - 1 \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right| \left(\sum_{i=1}^{n-1} A_i^n A_n^i\right) \left|\begin{matrix} h_{nn} - 1 \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix}\right\rangle\right]. \tag{2.30}$$

Continuing in this fashion, we find that we have finally to evaluate

$$\langle I_{n+1}^n \rangle_{(m)}^2 \left[1 + \langle A_n^{n-1} \rangle_{(m)}^2 \times \left[1 + \cdots [1 + \langle A_3^2 \rangle^2 [1 + \langle A_2^1 \rangle^2]] \cdots\right]\right], \tag{2.31}$$

where

$$\langle A_2^1 \rangle_{(m)} \equiv \left\langle\begin{matrix} h_{n\,n+1} & h_{nn} - 1 \\ & h_{nn} - 1 \end{matrix}\right| A_2^1 \left|\begin{matrix} h_{n\,n+1} & h_{nn} - 1 \\ & h_{nn} \end{matrix}\right\rangle, \tag{2.32}$$

and so on. Namely, we start for the A's each time with a true minimal state modified only by $(h_{nn} - 1)$ at the extreme right; whereas for I^n_{n+1}, it is the matrix element corresponding to (2.28). These matrix elements are sufficiently simple, so that, starting with the innermost term and continuing outwards, after an easy calculation we obtain

$$\Delta_{(2)} = (h_{2\,n+1} - h_{nn} + n)\langle I_{n+1}^n \rangle_{(m)}^2 \tag{2.33}$$

$$= \kappa^2. \tag{2.34}$$

We have taken the trouble of rederiving (2.26) by a longer method for the reason that the factorization (2.33) will be useful later on when the I's will be replaced by different operators.

Let us now evaluate $\Delta_{(3)}$ in a similar fashion. We

note that

$$\left\langle \begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix} \right| \left(\sum_{i,j=1}^{n} I_i^{n+1} A_j^i I_{n+1}^j \right) \left| \begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix} \right\rangle$$

$$= \left\langle \begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix} \right| I_n^{n+1} \left(A_n^n - 2 \left(\sum_{i=1}^{n-1} A_i^n A_n^i \right) + \left(\sum_{i,j=1}^{n-1} A_i^n A_j^i A_n^j \right) \right) I_{n+1}^n \left| \begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix} \right\rangle \tag{2.35}$$

$$= \langle I_{n+1}^n \rangle_{(m)}^2 [\langle A_n^n \rangle_{(m)} + \langle A_n^{n-1} \rangle_{(m)}^2 [\langle A_{n-1}^{n-1} \rangle_{(m)} - 2 + \cdots + \langle A_2^1 \rangle_{(m)}^2 [\langle A_1^1 \rangle - 2(n-1)] \cdots] \tag{2.36}$$

$$= \kappa^2 (h_{nn} - n). \tag{2.37}$$

This gives

$$\Delta_{(3)} = {}_{(\kappa,\zeta)}\langle h | \left(\sum_{i,j=1}^{n} I_i^{n+1} A_j^i I_{n+1}^j + A_{n+1}^{n+1} \Delta_{(2)} \right) | h \rangle_{(\kappa,\zeta)}$$

$$= \left\langle \begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix} \right|_{(\kappa,\zeta)} \left(\sum_{i,j=1}^{n} I_i^{n+1} A_j^i I_{n+1}^j + A_{n+1}^{n+1} \Delta_{(2)} \right) \times \left| \begin{matrix} h_{nn} \\ h_{nn} \\ \cdot \\ \cdot \\ \cdot \\ h_{nn} \end{matrix} \right\rangle_{(\kappa,\zeta)}$$

$$= \kappa^2 ((h_{nn} - n) + (\zeta - h_{nn}))$$

$$= \kappa^2 (\zeta - n). \tag{2.38}$$

The fact that $\Delta_{(3)}$ does not contain $h_{i\ n+1}$ $(i = 2, \cdots, n)$ is a consequence of our definition (2.20).

It should be noted that for algebra (2.1) [*as well as for* (2.9)] we can define $\Delta_{(3)}$ as

$$\Delta_{(3)} = \sum_{i,j=1}^{n} I_i^{n+1} A_j^i I_{n+1}^j - \left(\sum_{i=1}^{n} A_i^i \right) \Delta_{(2)} \tag{2.39}$$

$$= -\kappa^2 \left(\sum_{i=2}^{n} h_{i\ n+1} + n \right). \tag{2.40}$$

In fact [for (2.9)], replacing A_{n+1}^{n+1} in (2.23) by

$$\frac{1}{2} \left(A_{n+1}^{n+1} - \sum_{i=1}^{n} A_i^i \right) = A_{n+1}^{n+1} - \frac{1}{2} \left(\sum_{i=1}^{n+1} A_i^i \right), \tag{2.41}$$

we merely subtract $\frac{1}{2}\Delta_{(1)}\Delta_{(2)}$ from the previous definition. In the following sections we will use the form (2.38).

As a final remark we may add the hint that, if one is interested in interpreting the results of Sec. 2 as a contraction process, in the GZ basis one should make

$$h_{1\ n+1} \to +\infty, \quad h_{n+1\ n+1} \to -\infty,$$

while

$$(h_{1\ n+1} + h_{n+1\ n+1})$$

remains finite.

3. THE DEFORMATION $[U(n) \otimes U(1)] \times I_{2n} \to U(n, 1)$

Our starting point is the algebra of

$$(U(n) \otimes U(1)) \times I_{2n}, \tag{3.1}$$

whose representations we have discussed in Sec. 2.

Let us now define[7]

$$A_{n+1}^i = \pm [\Delta, I_{n+1}^i] + i\epsilon I_{n+1}^i, \tag{3.2}$$

$$A_i^{n+1} = \pm [\Delta, I_i^{n+1}] + i\epsilon I_i^{n+1}, \tag{3.3}$$

where ϵ is *real and arbitrary*, and [with the definitions (2.22), (2.23)]:

$$\Delta \equiv \frac{1}{2\Delta_{(2)}^{\frac{1}{2}}} \left[\left(\sum_{i,j=1}^{n} A_j^i A_i^j \right) + \left(\frac{\Delta_{(3)}}{\Delta_{(2)}} + n \right) A_{n+1}^{n+1} \right]. \tag{3.4}$$

It is to be noted that, for a nontrivial representation of the I's, we always have

$$\Delta_{(2)} > 0. \tag{3.5}$$

We obtain [with the convention of choosing the + sign in (4.3), (4.4)]:

$$A_{n+1}^i = \frac{1}{\Delta_{(2)}^{\frac{1}{2}}} \left[-\left(\sum_{j=1}^{n} A_j^i I_{n+1}^j \right) + \frac{\Delta_{(3)}}{2\Delta_{(2)}} I_{n+1}^i \right] + i\epsilon I_{n+1}^i \tag{3.6}$$

$$= -(A_i^{n+1})^\dagger. \tag{3.7}$$

It may be verified [with the help of (2.2)–(2.5)] that now we have

$$[A_{n+1}^i, A_{n+1}^j] = [A_i^{n+1}, A_j^{n+1}] = 0 \tag{3.8}$$

and

$$[A_i^{n+1}, A_{n+1}^j] = A_i^j - \delta_i^j A_{n+1}^{n+1}. \tag{3.9}$$

[7] Often the deformation formulas are written in a form which is essentially what we obtain by dividing (3.2), (3.3) by ϵ [see Ref. 2 and also M. Levy-Nahas, J. Math. Phys. **8**, 1211 (1967)]. Since our purpose is to reproduce exactly the algebra of $U(n, 1)$ and not to study it as a perturbation, the forms (3.2) and (3.3) are the ones suitable for us.

In other words, the generators

$$A^i_j \quad (i, j = 1, \cdots, n+1)$$

now generate the algebra of [see (A1)–(A3)]:

$$U(n, 1).$$

In particular, if we start with the explicit matrix elements of Sec. 2, we obtain at once those for the deformed representation, whose representations are now labeled by

(i) $(n-1)$ integers $h_{i\,n+1} \quad (i = 2, \cdots, n)$

and

(ii) two continuous real parameters ζ and ϵ.

[For our present purpose, we can always consider $\kappa = +1$, as κ is normalized away in (3.6) through the substitution $\epsilon \to \epsilon\kappa$. Since ϵ is as yet arbitrary, this involves no loss of generality.]

We can denote such a basis of $U(n, 1)$ as

$$\left| \begin{array}{ccccccc} & h_{2\,n+2} & & \cdots & & h_{n\,n+1} & \\ h_{1n} & & h_{2n} & \cdots & h_{n-1\,n} & & h_{nn} \\ & h_{1\,n-1} & & \cdots & & h_{n-1\,n-1} & \\ & & & \cdots & & & \\ & & & h_{11} & & & \end{array} \right\rangle_{(\zeta,\epsilon)}. \tag{3.10}$$

The nonzero matrix elements of A^n_{n+1} and A^{n+1}_n can now be shown to be given by

$$\langle h_{jn} - 1 | A^n_{n+1} | h \rangle = (\tfrac{1}{2}(\zeta - n) + i\epsilon - (h_{jn} - j)) \langle h_{jn} - 1 | I^n_{n+1} | h \rangle \tag{3.11}$$

$$= -\langle h | A^{n+1}_n | h_{jn} - 1 \rangle^* \quad (j = 1, \cdots, n). \tag{3.12}$$

The proof is very simple on using (A11) to evaluate the first term of Δ in (3.2) and (3.4). From the result (3.11) we can construct all the matrix elements of A^i_{n+1}, A^{n+1}_i $(i = 1, \cdots, n)$.

The first-order invariant remains, evidently, the same as (2.21), namely

$$\Delta_{(1)} |h\rangle_{(\zeta,\epsilon)} = \left(\zeta + \sum_{i=2}^{n} h_{i\,n+1}\right) |h\rangle_{(\zeta,\epsilon)}. \tag{3.13}$$

Let us now evaluate $\Delta_{(2)}$.

We have

$$\Delta_{(2)} = \left(\sum_{i,j=1}^{n+1} A^i_j A^j_i\right) \tag{3.14}$$

$$= \left(\sum_{i,j=1}^{n} A^i_j A^j_i\right) + (A^{n+1}_{n+1})^2 + \sum_{i=1}^{n} (A^{n+1}_i A^i_{n+1} + A^i_{n+1} A^{n+1}_i) \tag{3.15}$$

$$= 2\left(\sum_{i=1}^{n} A^{n+1}_i A^i_{n+1}\right) + \left(\sum_{i=1}^{n} A^i_j A^j_i\right) + (A^{n+1}_{n+1})^2 + (n+1)A^{n+1}_{n+1} - \Delta_{(1)}. \tag{3.16}$$

We will now consider the diagonal matrix elements of these terms acting on a pseudominimal state (2.27). The first term can be evaluated just as in (2.30)–(2.33). A difference arises, since now (*considering first the case* $\epsilon = 0$) we have instead, taking account of (3.7),

$$-2(h_{2\,n+1} - h_{nn} + n)\langle A^n_{n+1}\rangle^2_{(m)} \tag{3.17}$$

$$= -2(\tfrac{1}{2}(\zeta - n) - (h_{nn} - n))^2. \tag{3.18}$$

The second term is just the $\Delta_{(2)}$ for $U_{(n)}$ which, for this state, comes out as

$$\left\{\sum_{i=2}^{n} h_{i\,n+1}(h_{i\,n+1} + n + 1 - 2(i-1)) + h_{nn}(h_{nn} + n + 1 - 2n)\right\}. \tag{3.19}$$

Now substituting for the diagonal terms in (3.16) and adding them to (3.18) and (3.19), we obtain (*for* $\epsilon = 0$)

$$\Delta_{(2)} = \sum_{i=2}^{n} h_{i\,n+1}(h_{i\,n+1} + n + 2 - 2i) + \tfrac{1}{2}(\zeta^2 - n^2). \tag{3.20}$$

For

$$\epsilon \neq 0,$$

we note that

$$A^{n+1}_i A^i_{n+1} = ([\Delta, I^{n+1}_i] + i\epsilon I^{n+1}_i)([\Delta, I^i_{n+1}] + i\epsilon I^i_{n+1}) = [\Delta, I^{n+1}_i][\Delta, I^i_{n+1}] - \epsilon^2 I^{n+1}_i I^i_{n+1}. \tag{3.21}$$

Hence, we finally obtain

$$\Delta_{(2)} |h\rangle_{(\zeta,\epsilon)} = \left\{\sum_{i=2}^{n} h_{i\,n+1}(h_{i\,n+1} + n + 2 - 2i) + \tfrac{1}{2}(\zeta^2 - 4\epsilon^2 - n^2)\right\} |h\rangle_{(\zeta,\epsilon)}. \tag{3.22}$$

An analogous exercise can be performed for $\Delta_{(3)}$ also. However, we will not attempt, in this article, a systematic explicit evaluation of the invariants.

In order to compare with familiar things, let us now consider a very simple and somewhat special case, namely the representation of

$$U(1, 1)$$

obtained by our technique.

The starting point in the representation of

$$(U(1) \otimes U(1)) \times I_2, \tag{3.23}$$

given (*for* $\kappa = 1$) by

$$A_1^1 |h_{11}\rangle_{(\zeta)} = h_{11} |h_{11}\rangle_{(\zeta)},$$
$$A_2^2 |h_{11}\rangle_{(\zeta)} = (\zeta - h_{11}) |h_{11}\rangle_{(\zeta)}, \quad (3.24)$$

$$I_2^1 |h_{11}\rangle_{(\zeta)} = |h_{11} - 1\rangle_{(\zeta)}, \quad I_1^2 |h_{11}\rangle_{(\zeta)} = |h_{11} + 1\rangle_{(\zeta)}. \quad (3.25)$$

We obtain, for $U(1, 1)$,

$$A_2^1 |h_{11}\rangle_{(\zeta,\epsilon)} = (\tfrac{1}{2}(\zeta - 1) + i\epsilon - (h_{11} - 1)) |h_{11} - 1\rangle_{(\zeta,\epsilon)}, \quad (3.26)$$

$$A_1^2 |h_{11}\rangle_{(\zeta,\epsilon)} = -(\tfrac{1}{2}(\zeta - 1) - i\epsilon - (h_{11})) |h_{11} + 1\rangle_{(\zeta,\epsilon)}. \quad (3.27)$$

Defining

$$(K_1 + iK_2) \equiv K_+ = iA_1^2, \quad (K_1 - iK_2) \equiv K_- = iA_2^1,$$
$$K_3 \equiv K_0 = \tfrac{1}{2}(A_1^1 - A_2^2), \quad (3.28)$$

we obtain the Hermitian generators K_1, K_2, K_3 satisfying the $SU(1, 1)$ algebra[8,9]

$$[K_1, K_2] = -iK_3, \quad [K_3, K_1] = iK_2,$$
$$[K_2, K_3] = iK_1, \quad (3.29)$$

with

$$[-\tfrac{1}{2}(K_+K_- + K_-K_+) + (K_0)^2] |h_{11}\rangle_{(\zeta,\epsilon)} = -(\epsilon^2 + \tfrac{1}{4}) |h_{11}\rangle_{(\zeta,\epsilon)}. \quad (3.30)$$

Thus, so far as this subalgebra is concerned, ζ does not appear in the invariant, but only in the matrix elements in the combination

$$m \equiv \tfrac{1}{2}(2h_{11} - \zeta). \quad (3.31)$$

Thus we obtain the following continuous representation of $SU(1, 1)$[10]:

$$K_\pm |m\rangle_\epsilon = \pm \frac{i}{2}(2m \pm 1 \pm i2\epsilon) |m \pm 1\rangle_\epsilon, \quad (3.32)$$

$$K_0 |m\rangle_\epsilon = m |m\rangle_\epsilon. \quad (3.33)$$

The restriction of m to integral or half-integral values can only come through global properties.

Another particularly simple case (for arbitrary n) is obtained when

$$h_{i\,n+1} = 0 \quad (i = 1, \cdots, n) \quad (3.34)$$

with the states

$$\begin{array}{ccccccccccc}
 & 0 & & 0 & & & & 0 & & 0 & \\
h_{1n} & & 0 & & \cdots & & & & 0 & & h_{nn} \\
 & h_{1\,n-1} & & 0 & & \cdots & & 0 & & h_{n-1\,n-1} & \\
 & & & & & \cdots & & & & & \\
 & & & & h_{13} & & 0 & & h_{33} & & \\
 & & & & & h_{12} & & h_{22} & & & \\
 & & & & & & h_{11} & & & & \\
 & & & & & & & & & & (\zeta, \epsilon)
\end{array} \quad (3.35)$$

The only nonzero matrix elements of A_{n+1}^n are now given by

$$\langle h_{1n} - 1| A_{n+1}^n |h\rangle = (\tfrac{1}{2}(\zeta - n) + i\epsilon - (h_{1n} - 1)) \times \left(\frac{(h_{1\,n-1} - h_{1n})(h_{n-1\,n-1} - h_{1n} - n)}{(h_{nn} - h_{1n} - n + 2)(h_{nn} - h_{1n} - n + 1)}\right)^{\frac{1}{2}} \quad (3.36)$$

and

$$\langle h_{nn} - 1| A_{n+1}^n |h\rangle = (\tfrac{1}{2}(\zeta - n) + i\epsilon - (h_{nn} - n)) \times \left(\frac{(h_{1\,n-1} - h_{nn} + n - 1)(h_{n-1\,n-1} - h_{nn} + 1)}{(h_{1n} - h_{nn} + n)(h_{1n} - h_{nn} + n - 1)}\right)^{\frac{1}{2}}. \quad (3.37)$$

The first two invariants are now simply

$$\Delta_{(1)} = \zeta \quad (3.38)$$

and

$$\Delta_{(2)} = \tfrac{1}{2}(\zeta^2 - 4\epsilon^2 - n^2). \quad (3.39)$$

4. COMPARISON WITH THE ROSEN-ROMAN CONSTRUCTION

Let us again start with the algebra (2.2)–(2.5) of

$$U(n) \times I_{2n}.$$

Defining (in terms of an arbitrary real parameter λ)

$$B_{n+1}^n = \sum_{i=1}^n A_i^n I_{n+1}^i + \lambda I_{n+1}^n, \quad (4.1)$$

we have

$$[B_{n+1}^n, (B_{n+1}^n)^\dagger] = \left(\sum_{i=1}^n I_i^{n+1} I_{n+1}^i\right)(A_n^n + 2\lambda) + \left(\sum_{i,j=1}^n I_i^{n+1} A_j^i I_{n+1}^j\right). \quad (4.2)$$

Starting from this result, it is easy to verify that, instead of enlarging the algebra first [as in (2.19)–(2.21)] and then utilizing (3.2)–(3.7) for the deformation, we can also start directly with the algebra of

$$U(n) \times I_{2n},$$

[8] This and more the general cases (relevant for us) of $SU(p, 1)$ have been studied from the algebraic point of view by L. C. Biedenharn, J. Nuyts, and N. Straumann, Ann. Inst. Henri Poincaré **3**, 13 (1965); L. C. Biedenharn, *Non-Compact Groups in Particle Physics*, Yutze Chow, Ed. (W. A. Benjamin, Inc., New York, 1966), p. 23.

[9] We may note, however, the point that in our case the continuous parameters (in contrast to the discrete ones) always occur outside the square root in the matrix elements. Hence, evidently without detailed calculations, we can say that tests involving the nonnegativity of norms such as are performed in Ref. 8 can impose no restrictions on them. In our representation the distinction between the two types of parameters is thus particularly clear from the beginning. Explicit examples in Sec. 6 show how, by suitably redefining the phases, we can obtain real matrix elements.

[10] It has been pointed out to the author by Monique Lévy-Nahas that (as might have been expected for this particular case) the same form can also be obtained by calculating $O(2, 1)$ with the corresponding deformation formula for the orthogonal groups.

and define (for the moment putting $\epsilon = 0$):

$$A^n_{n+1} = \frac{1}{\Delta^{\frac{1}{2}}_{(2)}} B^n_{n+1} = -(A^{n+1}_n)^\dagger \tag{4.3}$$

and

$$A^{n+1}_{n+1} = -\frac{1}{\Delta^{\frac{1}{2}}_{(2)}} \left\{ \left(\sum_{i,j=1}^{n} I^{n+1}_i A^i_j I^j_{n+1} \right) + 2\lambda \Delta_{(2)} \right\}. \tag{4.4}$$

Since ζ in (2.20) and λ in (4.1)–(4.4) are as yet both *arbitrary*, it is easy to see that no essential differences arise, as compared to the construction of Sec. 3. For this we need only ascertain that the commutation relations of

$$\left(\sum_{i,j=1}^{n} I^{n+1}_i A^i_j I^j_{n+1} \right) \tag{4.5}$$

correspond to those of

$$-A^{n+1}_{n+1} \tag{4.6}$$

of (2.20), which implies that the matrix elements (diagonal) of (4.5) are of the form

$$(\text{some invariant term}) + \left(\sum_{i=1}^{n} h_{in} \right). \tag{4.7}$$

Now coming to the (Rosen–Roman) construction,[5] upon translating their notation into ours through the relations

$$\begin{aligned} E_{ij} &= (A^i_j - A^j_i), \qquad F_{ij} = i(A^i_j + A^j_i), \\ Q_i &= (I^i_{n+1} - I^{n+1}_i), \quad R_i = i(I^i_{n+1} + I^{n+1}_i), \end{aligned} \tag{4.8}$$

we find that their result corresponds to the choice [in (4.1)–(4.4)] of *the particular value*

$$\lambda = n/2. \tag{4.9}$$

Besides this, as compared to Sec. 3, it is implied that

$$\epsilon = 0. \tag{4.10}$$

We prefer to display the roles of these two continuous parameters (ζ, ϵ) as in Sec. 3.

As already noted in the Introduction, our deformation formula is applicable to the more general case of $U(p, q)$, though here we utilize it only for the explicit construction of $U(n, 1)$ representations.

Let us finally add a remark about the alternatives

$$\epsilon = \pm 1 \tag{4.11}$$

given by RR [see their equations (121)–(124)].

Taking as an example the simple case discussed in (3.36)–(3.45), we note that they correspond to

$$\epsilon = -1. \tag{4.12}$$

Corresponding to

$$\epsilon = +1, \tag{4.13}$$

instead of (3.40) and (3.41) with

$$L_+ = A^2_1, \quad L_- = A^1_2, \quad L_3 = \tfrac{1}{2}(A^1_1 - A^2_2), \tag{4.14}$$

we would have the familiar rotation algebra

$$[L_i, L_j] = +i\epsilon_{ijk} L_k, \tag{4.15}$$

with L_3 *Hermitic* and L_1, L_2 *anti-Hermitic*, since

$$A^2_1 = -(A^1_2)^+.$$

This is a general phenomenon for the choice (4.13). When integrable, such algebras should leave to non-unitary representations obtained by some "suitably defined" [11] continuation of the unitary representations for the case (4.12).

In Sec. 3, we have considered only the one-way process leading to the required Hermiticity conditions.

5. REMARKS ABOUT $IU(p, q)$

We will now discuss the difficulties that prevent a construction similar to that of Sec. 2, when the starting point is the GG bases[3] for $U(p, q + 1)$ instead of $U(n + 1)$.

Let us start by considering a very simple, though rather special, case.

For $IU(2)$ we have constructed the basis (with $\kappa = 1$, say):

$$\left| \begin{matrix} & h_{23} & \\ h_{12} & & h_{22} \\ & h_{11} & \end{matrix} \right\rangle, \tag{5.1}$$

with, for example,

$$\langle h_{12} - 1 | \, I^2_3 \, | h \rangle = \left(\frac{(h_{23} - h_{12})(h_{11} - h_{12})}{(h_{22} - h_{12})(h_{22} - h_{12} - 1)} \right)^{\frac{1}{2}}. \tag{5.2}$$

Now considering the case $IU(1, 1)$, say, in particular, the representation of $U(1, 1)$ (see Appendix) with

$$\alpha_+ = 0,$$

with the bases

$$\left| \begin{matrix} h_{12} & h_{22} & \\ & & h_{11} \end{matrix} \right\rangle, \tag{5.3}$$

we find that the required Hermiticity restrictions (along with the commutation relations) can only be maintained if we replace h_{23} in (5.1) ($h_{12} \geq h_{23} \geq h_{22}$) by a correspondingly shifted parameter. Thus, we

[11] For examples of such techniques, though not exactly for these cases, see J. Kuriyan, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **9**, 2100 (1968). Also, W. J. Holman III and L. C. Biedenharn, Jr., Ann. Phys. N.Y. **39**, 1 (1966).

may write, for example,

$$\left| \begin{matrix} h_{13} & & \\ & h_{12} \quad h_{22} & \\ & & h_{11} \end{matrix} \right\rangle, \tag{5.4}$$

with

$$\langle h_{12}-1|\, I_3^2\, |h\rangle = \left(-\frac{(h_{13}-h_{12}+1)(h_{11}-h_{12})}{(h_{22}-h_{12})(h_{22}-h_{12}-1)}\right)^{\frac{1}{2}}, \tag{5.5}$$

and corresponding expressions for the other matrix elements. [Note also the extra minus sign as compared to (5.2).]

But now the change in the system of inequalities is such that the term

$$(h_{22}-h_{12}) \tag{5.6}$$

in the denominator is no longer canceled out in time as h_{12} is decreased step by step towards h_{22}. Such a difficulty arises for the other matrix elements and is, in fact, quite a general one for a noncompact homogeneous part. Thus our previous construction fails.

Let us consider now the case $q \geq 2$. Starting from a representation of $U(p, q+1)$ with $p+q=n$, we find that the shifts in the inequalities bring together, in the GG patterns, either one pair (for $\alpha_+ = 0$ or $\alpha_- = 0$) or two pairs (for $\alpha_+ = 1, \cdots, p-1$) of the parameters $h_{i\,n+1}$.

Thus, for example, for $U(2, 2)$ we have (writing explicitly only the two top rows):

for $\alpha_+ = 0$:

$$\begin{matrix} h_{14} & [h_{24}h_{34}] & h_{44} \\ h_{13} & h_{23} & h_{33}, \end{matrix} \tag{5.7}$$

for $\alpha_+ = 1$:

$$\begin{matrix} [h_{14}h_{24}] & [h_{34}h_{44}] \\ h_{13} & h_{23} & h_{33}, \end{matrix} \tag{5.8}$$

and for $\alpha_+ = 2$:

$$\begin{matrix} h_{14} & [h_{24}h_{34}] & h_{44} \\ h_{13} & h_{23} & h_{33} \end{matrix} \tag{5.9}$$

In each case the Hermiticity restrictions can be met only by suppressing such paired terms. This, however, removes essential barriers between the parameters of the next row; hence the construction fails.

It is amusing to note that, for the cases

$$\alpha_+ = 1, \cdots, p-1, \tag{5.10}$$

on *suppressing both the pairs* we could have obtained (if everything remained well defined) representations of $IU(p, q)$ with

$$\Delta_{(2)} = 0. \tag{5.11}$$

Such representation would have been rigid under such a deformation as (4.1), like the case of zero-mass discrete spin representations of the Poincaré algebra considered by us elsewhere.[2]

The special case

$$q = 1$$

is again best discussed by starting with a particular case, namely

$$U(2, 1) \to IU(2). \tag{5.12}$$

Corresponding to (5.7)–(5.9), we now have

$\alpha_+ = 0$:

$$\begin{matrix} [h_{13}h_{23}] & h_{33} \\ h_{12} & h_{22}, \end{matrix} \tag{5.13}$$

$\alpha_+ = 1$:

$$\begin{matrix} \{h_{13}h_{23}h_{33}\} \\ h_{12} \quad h_{22}, \end{matrix} \tag{5.14}$$

$\alpha_+ = 2$:

$$\begin{matrix} h_{13} & [h_{23}h_{33}] \\ h_{12} & h_{22} \end{matrix} \tag{5.15}$$

The above patterns again illustrate a general feature, namely, for $U(p, q)$ with

$$p = n-1, \quad q = 1.$$

We have a grouped pair for the extreme cases ($\alpha_+ = 0$ or $\alpha_- = 0$) and a grouped triplet for all other cases. Now, indeed, we can define the matrix elements of I_{n+1}^n by suppressing [as in (2.12), but leaving in the minus sign] the factors corresponding to the grouped pair (or one pair chosen from the triplet). In each case, however, what we obtain is not essentially different from the construction of Sec. 2.

6. $O(n+1) \to IO(n) \to O(n, 1)$

In this section we give the construction analogous to the preceeding ones for the case of orthogonal algebras. Since the techniques are quite similar, and even simpler in certain respects, we will content ourselves with briefly stating the results.

So far as $O(n)$ is concerned, we adopt (with minor changes) the conventions of Pang and Hecht[12] and we define the $IO(n)$ algebra through the following commutation relations for the generators (all *skew*

[12] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967). This paper contains references to other sources.

symmetric and Hermitian):

$$[J_{i_1j_1}, J_{i_2j_2}] = i(\delta_{i_1i_2}J_{j_1j_2} + \delta_{j_1j_2}J_{i_1i_2} - \delta_{i_1j_2}J_{j_1i_2} - \delta_{j_1i_2}J_{i_1j_2}), \quad (6.1)$$

$$[J_{ij}, I_{n+1\,k}] = i(\delta_{ik}I_{n+1\,j} - \delta_{jk}I_{n+1\,i}),$$
$$[I_{n+1\,k_1}, I_{n+1\,k_2}] = 0 \quad (i, j, k = 1, \cdots, n). \quad (6.2)$$

In order to construct the bases, in contrast to the unitary case, we need now suppress *only one* parameter (the leading one) of the GZ basis for $O(n+1)$ and, of course, the corresponding factors in the matrix elements. [e.g., in the following formula (6.5), the index β starts from 2 instead of from 1, as in formula (5.46) of Ref. (12).] If we want to use the language of contraction, we should start by making $h_{n,1} \to \infty$.

A. $n = 2k$

We parametrize the basis vectors of $IO(2k)$ as

$$|h\rangle_\kappa \equiv \begin{bmatrix} & h_{2k+1\,2} & \cdots & h_{2k+1\,k-1} & h_{2k+1\,k} \\ h_{2k\,1} & h_{2k\,2} & \cdots & h_{2k\,k-1} & h_{2k\,k} \\ h_{2k-1\,1} & h_{2k-2\,2} & \cdots & h_{2k-1\,k-1} & \\ \cdot & \cdot & \cdots & h_{2k-2\,k-1} & \\ \cdot & & & & \\ \cdot & & & & \\ h_{51}h_{52} & & & & \\ h_{41}h_{42} & & & & \\ h_{31} & & & & \\ h_{21} & & & & \end{bmatrix}_\kappa, \quad (6.3)$$

where κ is a continuous real parameter and the $(k-1)$ discrete ones $(h_{2k+1\,2} \cdots, h_{2k+1\,k})$ satisfy the inequalities

$$h_{2k\,1} \geq h_{2k+1\,2} \geq h_{2k\,2} \geq \cdots \geq h_{2k+1\,k} \geq h_{2k\,k} \geq -h_{2k+1\,k}. \quad (6.4)$$

The key matrix elements for the I's (from which the others can be deduced) are given by

$$\langle h_{2k\,j} + 1 | I_{2k+1\,2k} | h\rangle = \langle h | I_{2k+1\,2k} | h_{2k\,j} + 1\rangle^*$$
$$= \frac{i\kappa}{2}\left[\frac{\prod_{\beta=2}^{k}(l_{2k+1\,\beta} - l_{2k\,j} - 1)(l_{2k+1\,\beta} + l_{2k\,j}) \prod_{\alpha=1}^{k-1}(l_{2k-1\,\alpha} - l_{2k\,j} - 1)(l_{2k+1\,\beta} + l_{2\,kj})}{\prod_{\alpha\neq j}^{k}(l_{2k\,\alpha}^2 - l_{2k\,j}^2)(l_{2k\,\alpha}^2 - (l_{2k\,j}+1)^2)}\right]^{\frac{1}{2}}$$
$$= -\langle h_{2k\,j} + 1 | I_{2k\,2k+1} | h\rangle, \quad (6.5)$$

where

$$l_{2k\,\alpha} \equiv h_{2k\,\alpha} + k - \alpha, \quad l_{2k-1\,\alpha} \equiv h_{2k-1\,\alpha} + k - \alpha. \quad (6.6)$$

B. $n = 2k - 1$

The basis vectors are now

$$|h\rangle_\kappa \equiv \begin{bmatrix} & h_{2k\,2} & \cdots & h_{2k\,k-1} & h_{2k\,k} \\ h_{2k-1\,1} & h_{2k-1\,2} & \cdots & h_{2k-1\,k-1} & \\ h_{2k-2\,1} & h_{2k-2\,2} & \cdots & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ h_{41} & h_{42} & & & \\ h_{31} & & & & \\ h_{21} & & & & \end{bmatrix}_\kappa, \quad (6.7)$$

where

$$h_{2k-1\,1} \geq h_{2k\,2} \geq \cdots h_{2k\,k-1} \geq h_{2k-1\,k-1} \geq |h_{2k\,k}|. \tag{6.8}$$

The key matrix elements are now given [as compared to (5.44) and (5.45) of Ref. 12] by

$$\langle h_{2k-1\,j}+1|\, I_{2k-1\,2k}\,|h\rangle = -i\kappa\left[\frac{\prod_{\beta=2}^{k}(l_{2k\,\beta}^2 - l_{2k-1\,j}^2)\prod_{\alpha=1}^{k-1}(l_{2k-2\,\alpha}^2 - l_{2k-1\,j}^2)}{l_{2k-1\,j}^2(4l_{2k-1\,j}^2-1)\prod_{\alpha\neq j}^{k-1}(l_{2k-1\,\alpha}^2 - l_{2k-1\,j}^2)((l_{2k-1\,\alpha}^2-1)^2 - l_{2k-1\,j}^2)}\right]^{\frac{1}{2}}, \tag{6.9}$$

$$\langle h|\, I_{2k-1\,2k}\,|h\rangle = \kappa\frac{\prod_{\beta=2}^{k} l_{2k\,\beta}\prod_{\alpha=1}^{k-1} l_{2k-2\,\alpha}}{\prod_{\alpha=1}^{k-1} l_{2k-1\,\alpha}(l_{2k-1\,\alpha}-1)}. \tag{6.10}$$

In both cases (A) and (B), the second-order Casimir operator is given by

$$\Delta_{(2)}\,|h\rangle = \left(\sum_{k=1}^{n} I_{n+1\,k}I_{n+1\,k}\right)|h\rangle = \kappa^2\,|h\rangle. \tag{6.11}$$

The deformation corresponding to that of Sec. 3 is now given by the relatively simple and well-known[5,13] formula, namely

$$J_{m,n+1} = +\frac{i}{\Delta_{(2)}^{\frac{1}{2}}}[\Delta, I_{m\,n+1}] + \lambda I_{m\,n+1}, \tag{6.12}$$

where

$$\Delta = \left(\sum_{i<j=1}^{n} J_{ij}^2\right) \tag{6.13}$$

and λ is a real continuous parameter.

As in Sec. 3, we can (without loss of generality) put

$$\kappa = 1$$

in the formulas for $IO(n)$, and consider the generators

$$J_{j\,n+1} = \sum_{i=1}^{n}(I_{i\,n+1}J_{ij} + J_{ij}I_{n+1}) + \lambda I_{j\,n+1} \qquad (j = 1, \cdots, n) \tag{6.14}$$

acting on the basis

$$|h\rangle_\lambda. \tag{6.15}$$

The $O(n, 1)$ algebra is now given by replacing the third equation in (6.1) by

$$[J_{n+1\,k_1}, J_{n+1\,k_2}] = -iJ_{k_1k_2}. \tag{6.2'}$$

The explicit expressions for the matrix elements of $J_{m\,n+1}$ are again easily obtained by noting that,[6,12] for $n = 2k$,

$$\Delta = \sum_{i=1}^{k} h_{2k\,i}^2 + \sum_{i=1}^{k-1}(2k-2i)h_{2k\,i} \tag{6.16}$$

and, for $n = 2k-1$,

$$\Delta = \sum_{i=1}^{k-1} h_{2k-1\,i}^2 + \sum_{i=1}^{k-1}(2k-2i-1)h_{2k-1\,i}. \tag{6.17}$$

[13] T. O. Philips and E. P. Wigner (to be published).

Finally, from (6.5), (6.9), (6.10), (6.16), and (6.17) we obtain

$$\langle h_{2k\,j}+1|\, J_{2k+1\,2k}\,|h\rangle_\lambda = \{-i(2h_{2k\,j}+2k-2j+1)+\lambda\} \times \langle h_{2k\,j}+1|\, I_{2k+1\,2k}\,|h\rangle \qquad \text{[putting } \kappa = 1 \text{ in (6.5)]}, \tag{6.18}$$

$$\langle h_{2k-1\,j}+1|\, J_{2k-1\,2k}\,|h\rangle = \{-i(2h_{2k-1\,j}+2k-2j)+\lambda\} \times \langle h_{2k-1\,j}+1|\, I_{2k-1\,2k}\,|h\rangle \qquad \text{[putting } \kappa = 1 \text{ in (6.9)]}, \tag{6.19}$$

and

$$\langle h|\, J_{2k-1\,2k}\,|h\rangle_\lambda = \lambda\langle h|\, I_{2k-1\,2k}\,|h\rangle \qquad \text{[putting } \kappa = 1 \text{ in (6.10)]}. \tag{6.20}$$

We now briefly consider two simple but important particular cases:

$$n = 3, 4;$$

namely, the homogeneous Lorentz group, the de Sitter group (and their corresponding contracted Euclidean groups).

The first thing to be noted is that the phase convention, adopted as being rather suitable for the general case, is not the one most familiar for $n = 3$. In fact, for $O(3, 1)$, making trivial changes in order to introduce more familiar symbols,

$$|h\rangle \equiv \left|\begin{matrix} & j_0 \\ j & \\ m & \end{matrix}\right\rangle_\lambda, \tag{6.21}$$

with

$$J_3\,|h\rangle \equiv J_{12}\,|h\rangle = m\,|h\rangle, \tag{6.22}$$

$$(J_1 \pm iJ_2)\,|h\rangle = \mp i(J_{13} \pm iJ_{23})\,|h\rangle = \mp i((j \mp m)(j \pm m + 1))^{\frac{1}{2}}\left|\begin{matrix} & j_0 \\ j & \\ m \pm 1 & \end{matrix}\right\rangle_\lambda, \tag{6.23}$$

and

$$K_3|h\rangle \equiv J_{34}|h\rangle = (j+1-i\lambda) \times \left(\frac{(j_0^2-(j+1)^2)(m^2-(j+1)^2)}{(j+1)^2(4(j+1)^2-1)}\right)^{\frac{1}{2}} \times \left|\begin{matrix} j+1 \\ m \end{matrix}\right\rangle_{\lambda}^{j_0} + \lambda\frac{j_0 m}{j(j+1)}\left|\begin{matrix} j \\ m \end{matrix}\right\rangle_{\lambda}^{j_0} + (j+i\lambda)\left(\frac{(j_0^2-j^2)(m^2-j^2)}{j^2(4j^2-1)}\right)^{\frac{1}{2}} \times \left|\begin{matrix} j-1 \\ m \end{matrix}\right\rangle_{\lambda}^{j_0}. \quad (6.24)$$

(The corresponding formulas for I_{34} need hardly be written down separately.) The two invariants are given by putting $K_1 \equiv J_{14}$ and $K_2 \equiv J_{24}$:

$$(\mathbf{K}^2 - \mathbf{J}^2)|h\rangle = (\lambda^2 - j_0^2 + 1)|h\rangle, \quad \mathbf{K}\cdot\mathbf{J}|h\rangle = \lambda j_0 |h\rangle. \quad (6.25)$$

The formula (6.24) has been utilized by Joos[14] [formula (4.22) of Ref. 14] and differs through a phase convention from a more familiar version.[7,14] Thus we note that the well-known continuous parameter λ characterizing the irreducible representations of HLG corresponds exactly to the deformation parameter introduced in (6.12). (The above particular case has been discussed[15] with the additional restriction $\lambda = 0$).

Let us now consider the case of $O(4, 1)$. Ström[16] has studied a certain basis of $O(4, 1)$ which can be contracted directly to the angular-momentum basis of the Poincaré group. He obtained it by transforming the usual discrete basis reduced by diagonalizing the $O(3) \otimes O(3)$ subgroup. Our formalism (6.18) for this particular case directly gives the required basis, apart from a difference in phase convention. In order to make this more explicit, starting with the basis

$$|h\rangle \equiv \left|\begin{matrix} & h_{52} & \\ h_{41} & & h_{42} \\ & h_{31} & \\ & h_{21} & \end{matrix}\right\rangle_{\lambda}, \quad (6.26)$$

let us redefine the phase by introducing

$$|h\rangle' = \left(\frac{\left(h_{11}+\frac{1-i\lambda}{2}\right)\left(h_{42}-\frac{1+i\lambda}{2}\right)}{\left(h_{41}+\frac{1+i\lambda}{2}\right)\left(h_{42}-\frac{1-i\lambda}{2}\right)}\right)^{\frac{1}{2}}|h\rangle. \quad (6.27)$$

The new phase makes the matrix elements all real and we have

$$\langle h_{41}+1|J_{45}|h\rangle' = \langle h|J_{45}|h_{41}+1\rangle' = \left((h_{41}+1)(h_{41}+2)+\left(\frac{1+\lambda^2}{4}\right)\right)^{\frac{1}{2}} \times \left(\frac{(h_{52}-h_{41}-1)(h_{52}+h_{41}+2)(h_{31}-h_{41}-1)(h_{31}+h_{41}+2)}{(h_{42}^2-(h_{41}+1)^2)(h_{42}^2-(h_{41}+2)^2)}\right)^{\frac{1}{2}} \quad (6.28)$$

and

$$\langle h_{42}+1|J_{45}|h\rangle' = \langle h|J_{45}|h_{42}+1\rangle' = \left(h_{42}(h_{42}+1)+\frac{1+\lambda^2}{4}\right)^{\frac{1}{2}} \times \left(\frac{(h_{52}-h_{42})(h_{52}+h_{42}+1)(h_{31}-h_{42})(h_{31}+h_{42}+1)}{((h_{41}+1)^2-h_{42}^2)((h_{41}+1)^2-(h_{42}+1)^2)}\right)^{\frac{1}{2}}. \quad (6.29)$$

Now substituting

$$\tfrac{1}{4}(1+\lambda^2) \equiv \sigma, \quad h_{52} \equiv S, \quad h_{41}+1 \equiv l, \quad h_{42} \equiv n, \quad h_{31} \equiv j, \quad h_{21} \equiv m, \quad (6.30)$$

we obtain the basis of Ström, with the two invariants

$$\Omega = -s(s+1)+2+\sigma$$

and

$$\Omega' = -s(s+1)\sigma. \quad (6.31)$$

Thus we see that, after contraction, $h_{52}(\equiv S)$ plays the role of the spin of the irreducible representation of the Poincaré algebra. (The discussion in Sec. 4 of Ref. 2 may also be compared in this connection.)

Also to be noted is the fact that, as compared to Ref. 16 we do not get the full range of σ, since in our case

$$\sigma = \tfrac{1}{4}(1+\lambda^2) \geq \tfrac{1}{4}. \quad (6.32)$$

7. CONCLUSION

As is well known, during the last few years, a very considerable amount of attention has been devoted to the possible applications of noncompact groups in particle physics. The last-named of the Refs. 8 and the lectures of Nambu and the following ones in the

[14] H. Joos, Fortschr. Physik **10**, 65 (1962).
[15] A. Sankaranarayanan, Nuovo Cimento **52A**, 91 (1967). See also: J. Math. Phys. **9**, 611 (1968)).
[16] S. Ström, Arkiv Fysik **30**, 455 (1965).

Rochester Proceedings[17] contain a fair amount of references to the original papers.

Even in molecular quantum mechanics[18] applications of $IO(4)$ and $O(4, 1)$ have been studied. [The possibilities of $O(4, 1)$ are not, of course, limited to internal symmetries only.[13]]

The intriguing possibility, of course, is that of classifying "towers" of multiplets corresponding to the successive irreducible representations of the homogeneous compact subgroup through the action of the noncompact generators which act as transition operators.

Since, for physical applications, it is often interesting to have as many discrete quantum numbers as possible, we have explicitly constructed such bases for $IU(n)$ and $IO(n)$, which have then been related, through suitable deformation formulas, to $U(n, 1)$ and $O(n, 1)$, respectively.

In this paper, however, we have not studied the problems of integrability and explicit construction of finite transformation matrix elements.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor I. T. Todorov and Monique Levy-Nahas for their interest in the problem and for useful discussions. The author is indebted to Professor L. Michel for a final discussion of the manuscript.

APPENDIX

The generators A^i_j of $U(p, q)$ $(p \geq q,\ p + q = n)$ satisfy the commutation relations

$$[A^i_j, A^k_l] = \delta^i_l A^k_j - \delta^k_j A^i_l \quad (i, j, k, l = 1, \cdots, n), \tag{A1}$$

along with the relations

$$(A^i_j)^\dagger = +(A^j_i), \quad \text{for} \quad (i, j) \leq p \quad \text{or} \quad > p, \tag{A2}$$

$$= -(A^j_i), \quad \text{for} \quad i \leq p, \quad j > p \quad \text{or} \quad i > p, \quad j \leq p. \tag{A3}$$

For the case

$$q = 0, \quad p = n, \tag{A4}$$

an orthonormalized basis is given by the Gel'fand–Zetlin (GZ) patterns

$$|h\rangle \equiv \begin{vmatrix} h_{1n} & h_{2n} & \cdots & h_{n-1\,n} & h_{nn} \\ & h_{1\,n-1} & \cdots & h_{n-1\,n-1} & \\ & & \cdots & & \\ & & h_{12} \quad h_{22} & & \\ & & h_{11} & & \end{vmatrix}, \tag{A5}$$

with the inequalities

$$h_{i\,j-1} \geq h_{ij} \geq h_{i+1\,j+1}. \tag{A6}$$

The key matrix elements are given by

$$A^k_k |h\rangle = \left(\sum_{i=1}^{k} h_{ik} - \sum_{i=1}^{k-1} h_{ik-1}\right) |h\rangle \tag{A7}$$

and

$$\langle h_{jk} - 1|\, A^k_{k+1} |h\rangle = \left[- \frac{\prod_{i=1}^{k+1} (h_{i\,k+1} - h_{jk} - i + j + 1) \prod_{i=1}^{k-1} (h_{i\,k-1} - h_{jk} - i + j)}{\prod_{1 \leq i \leq k, i \neq j} (h_{ik} - h_{jk} - i + j + 1)(h_{ik} - h_{jk} - i + j)} \right]^{\frac{1}{2}} \quad (j = 1, \cdots, k), \tag{A8}$$

where

$$|h_{jk} - 1\rangle$$

denotes the state differing from $|h\rangle$ only through the change $h_{jk} \to h_{jk} - 1$. The phase is chosen so that (A8) is positive. From (A8), together with (A1) and (A2), all other matrix elements can be obtained.

The explicit values of the Casimir operators of degree m defined as

$$\sum_{i_1, \cdots, i_m = 1}^{n} A^{i_1}_{i_2} A^{i_2}_{i_3} \cdots A^{i_m}_{i_1} \tag{A9}$$

are known.[6] We need only

$$\left(\sum_{i=1}^{n} A^i_i\right) |h\rangle = \left(\sum_{i=1}^{n} h_{in}\right) |h\rangle \tag{A10}$$

and

$$\left(\sum_{i,j=1}^{n} A^i_j A^j_i\right) |h\rangle = \left(\sum_{i=1}^{n} h_{in}(h_{in} + n + 1 - 2i)\right) |h\rangle. \tag{A11}$$

For the noncompact case $(q \geq 1)$ the bases for the discrete Gel'fand–Graev (GG) representations are obtained simply by changing the system of inequalities (A6) for the previous case [see (A13) and (A14)].

The formal expression for the matrix elements and invariants remain unchanged, but we have to note the altered phases implied through the changed inequalities. The changes in the inequalities are indicated by corresponding shifts of the relative positions of the parameters in the pattern.

For each pair of nonnegative integers α_+, α_-

[17] *Proceedings of the* 1967 *International Conference on Particles and Fields* (Interscience Publishers, Inc., New York, 1967).

[18] C. E. Wulfman and Y. Takahata, J. Chem. Phys. **47**, 488 (1967).

satisfying

$$\alpha_+ + \alpha_- = p \quad (\alpha_+ = 0, 1, \cdots, p), \tag{A12}$$

an inequivalent representation is obtained by changing (only) the inequalities corresponding to the first α_+ and the last α_- elements of each of the top q rows as follows:

$$h_{1j} \geq h_{1\,j+1} + 1 \geq h_{2j} \geq h_{2\,j+1} + 1 \geq \cdots \geq h_{\alpha+j} \geq h_{\alpha+j+1} + 1 \tag{A13}$$

and

$$h_{j-\alpha_-+2\,j+1} \geq h_{j-\alpha_-+1\,j+1} \geq \cdots \geq h_{j+1\,j+1} \geq h_{jj} + 1. \tag{A14}$$

The patterns induced by these shifts will be clearer on referring to the examples discussed in Sec. 5.

As regards the phases, for our purposes, we need only note that now the right-hand side of (A8) is pure imaginary for

$$k = p \tag{A15}$$

and real (positive or negative) otherwise.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Master Analytic Representation: Reduction of $O(2, 1)$ in an $O(1, 1)$ Basis

JACOB G. KURIYAN*†
Institute for Advanced Study, Princeton, New Jersey

N. MUKUNDA‡§ AND E. C. G. SUDARSHAN
Syracuse University, Syracuse, New York

(Received 21 December 1967)

We display the reduction of the pseudo-orthogonal group $O(2, 1)$ with respect to a noncompact $O(1, 1)$ basis. After the explicit solution is obtained, we rederive the results using the method of master analytic representations.

1. INTRODUCTION

Lie groups and Lie algebras have become increasingly familiar to particle physicists. Conservation laws and symmetries have been studied in terms of invariance and noninvariance groups. The central idea that is exploited in these applications is the assumption that the analytic properties of amplitudes have their counterpart in the analytic properties of the representations of Lie algebras.

We have studied the representation theory of Lie algebras in terms of analytic representations. Specifically, we wished to show that every linear representation of a (locally compact) Lie algebra is a special case of a master analytic representation; that the unitary representation of any of the Lie groups with this Lie algebra is a specialization of the master analytic representation (MAR).

The theory of the MAR synthesizes all Hermitian representations of the Lie algebra. It also brings out the relation between the representations of two different Lie algebras whose complex extensions are isomorphic. This is, therefore, an elegant and powerful method for finding the unitary representations of various noncompact groups.

Elsewhere,[1] we have illustrated the technique by finding the representations of some pseudo-orthogonal groups.

When a noncompact group is such that its maximal compact subgroup labels the states within a UIR uniquely, we believe that the MAR method is quite straightforward, and it is not too difficult to see why it works. We, however, believe that this method is quite general and fundamental and is applicable to many other groups as well. In particular, one could reduce UIR's of a noncompact group with respect to a noncompact subgroup. In this direction we have made a beginning by reducing representation of $O(2, 1)$ with respect to $O(1, 1)$. Throughout the paper we use, as far as possible, only infinitesimal-operator techniques. A difficult problem is to find out when a representation of the Lie algebra permits exponentiation to provide a representation of the group. We do

* Research supported by the National Science Foundation.
† Present address: UCLA, Los Angeles, California.
‡ Present address: Tata Institute of Fundamental Research, Bombay, India.
§ Supported by USAEC.

[1] J. G. Kuriyan, N. Mukunda, and E. C. Sudarshan, Commun. Math. Phys. **8**, 204 (1968).

satisfying

$$\alpha_+ + \alpha_- = p \quad (\alpha_+ = 0, 1, \cdots, p), \tag{A12}$$

an inequivalent representation is obtained by changing (only) the inequalities corresponding to the first α_+ and the last α_- elements of each of the top q rows as follows:

$$h_{1j} \geq h_{1\,j+1} + 1 \geq h_{2j} \geq h_{2\,j+1} + 1 \geq \cdots \geq h_{\alpha+j} \geq h_{\alpha+j+1} + 1 \tag{A13}$$

and

$$h_{j-\alpha_-+2\,j+1} \geq h_{j-\alpha_-+1\,j+1} \geq \cdots \geq h_{j+1\,j+1} \geq h_{jj} + 1. \tag{A14}$$

The patterns induced by these shifts will be clearer on referring to the examples discussed in Sec. 5.

As regards the phases, for our purposes, we need only note that now the right-hand side of (A8) is pure imaginary for

$$k = p \tag{A15}$$

and real (positive or negative) otherwise.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Master Analytic Representation: Reduction of $O(2, 1)$ in an $O(1, 1)$ Basis

JACOB G. KURIYAN*†
Institute for Advanced Study, Princeton, New Jersey

N. MUKUNDA‡§ AND E. C. G. SUDARSHAN
Syracuse University, Syracuse, New York

(Received 21 December 1967)

We display the reduction of the pseudo-orthogonal group $O(2, 1)$ with respect to a noncompact $O(1, 1)$ basis. After the explicit solution is obtained, we rederive the results using the method of master analytic representations.

1. INTRODUCTION

Lie groups and Lie algebras have become increasingly familiar to particle physicists. Conservation laws and symmetries have been studied in terms of invariance and noninvariance groups. The central idea that is exploited in these applications is the assumption that the analytic properties of amplitudes have their counterpart in the analytic properties of the representations of Lie algebras.

We have studied the representation theory of Lie algebras in terms of analytic representations. Specifically, we wished to show that every linear representation of a (locally compact) Lie algebra is a special case of a master analytic representation; that the unitary representation of any of the Lie groups with this Lie algebra is a specialization of the master analytic representation (MAR).

The theory of the MAR synthesizes all Hermitian representations of the Lie algebra. It also brings out the relation between the representations of two different Lie algebras whose complex extensions are isomorphic. This is, therefore, an elegant and powerful method for finding the unitary representations of various noncompact groups.

Elsewhere,[1] we have illustrated the technique by finding the representations of some pseudo-orthogonal groups.

When a noncompact group is such that its maximal compact subgroup labels the states within a UIR uniquely, we believe that the MAR method is quite straightforward, and it is not too difficult to see why it works. We, however, believe that this method is quite general and fundamental and is applicable to many other groups as well. In particular, one could reduce UIR's of a noncompact group with respect to a noncompact subgroup. In this direction we have made a beginning by reducing representation of $O(2, 1)$ with respect to $O(1, 1)$. Throughout the paper we use, as far as possible, only infinitesimal-operator techniques. A difficult problem is to find out when a representation of the Lie algebra permits exponentiation to provide a representation of the group. We do

* Research supported by the National Science Foundation.
† Present address: UCLA, Los Angeles, California.
‡ Present address: Tata Institute of Fundamental Research, Bombay, India.
§ Supported by USAEC.

[1] J. G. Kuriyan, N. Mukunda, and E. C. Sudarshan, Commun. Math. Phys. **8**, 204 (1968).

not address ourselves to this problem. We do hope to demonstrate, however, the simplicity and the consequent ease with which some of these groups can be handled.

Various authors have investigated the representation theory of groups which make use of some of the ideas mentioned above; particular mention must be made of the work of Barut and Fronsdal, Hermann, Holman and Biedenharn, and Fronsdal.[2] The results of this theory have been used in other investigations by one or another of the present authors of this paper,[3] and the theory was announced more than two years ago at the Third Coral Gables Conference[4]; an account was given in a thesis by one of us (J. G. K.).[5]

The reduction of noncompact groups with respect to noncompact subgroups has been carried out by Mukunda, Ström, and Sciarrino and Toller,[6] using global techniques. Our approach is in the spirit of Joos'[7] classic work on the Poincaré group. Subsequent to the completion of our work, Barut and Philips[8] have extended our techniques to study (global) representations of the group $O(2, 1)$. Much more recently Itzykson[9] has attacked the same problem as ours using very similar methods.

The plan of the paper is as follows. In Sec. 2 we review the salient points of the method of MAR by applying it to the group $SO(3)$ and deriving the UIR of $SO(2, 1)$. Section 3 deals with the solution of the reduction of $SO(2, 1)$ in an $SO(1, 1)$ basis. For those who have lost sight of the method of MAR, in Sec. 4 we derive the essential results of Sec. 3 by using the recipe given in Sec. 1. We conclude the paper with a discussion.

2. UIR OF $SO(2, 1)$ OBTAINED FROM THE UIR OF $SO(3)$

We recapitulate the recipe prescribed by the method of MAR. First carry out the Weyl's unitary trick to obtain the generators of $O(2, 1)$ from those of $O(3)$. Then analytically continue the matrix elements into regions in which the corresponding operators are Hermitian.

The simplest application of the method of MAR is to $SO(3)$. The commutation relations are well known and the matrix elements of J_0 and J_+ are

$$\langle m|\, J_0\, |m\rangle = m,$$
$$\langle m \pm 1|\, J_\pm\, |m\rangle = \{(j + \tfrac{1}{2})^2 - (m \pm \tfrac{1}{2})^2\}, \quad (2.1)$$

with $2j$ a nonnegative integer.

Weyl's unitary trick gives the prescription for the generators of $O(2, 1)$ (distinguished by primes):

$$J_0 \to J_0' = J_0,$$
$$J_\pm \to J_\pm' = iJ_\pm, \quad (2.2)$$

so that the matrix elements of J_0' and J_+' are

$$\langle m|\, J_0'\, |m\rangle = m,$$
$$\langle m \pm 1|\, J_+'\, |m\rangle = \{(m \pm \tfrac{1}{2})^2 - (j + \tfrac{1}{2})^2\}^{\frac{1}{2}}. \quad (2.3)$$

To get Hermitian representations with

$$(J_0')^\dagger = J_0', \quad (J_\pm')^\dagger = J_\mp, \quad (2.4)$$

we search for the domain of m and the values of the parameter j, considering the matrix elements (2.3) as analytic functions of its variables. We get the following classes of representations:

(i) D_j^+ $\quad 2j + 1$ nonnegative integer; $m \geq j + 1$,
(ii) D_j^- $\quad 2j + 1$ nonnegative integer; $m \leq -j - 1$,
(iii) C_σ^0 $\quad j = -\frac{1}{2} + i\sigma$, $\sigma \geq 0$ m integral,
(iv) $C_\sigma^{\frac{1}{2}}$ $\quad j = -\frac{1}{2} + i\sigma$, $\sigma > 0$ m half integral,
(v) E_s $\quad j = -\frac{1}{2} + s$, $\frac{1}{2} > s > 0$ m integral,
(vi) I $\quad j = 0$, $m = 0$.

Thus we obtain all the different classes of representations that Bargmann[10] first derived by analyzing the master analytic function in Eq. (2.3), which are subject to conditions 2.4 (Hermiticity or unitarity correlation).

3. REPRESENTATIONS OF $O(2, 1)$ IN AN $O(1, 1)$ BASIS

In this section we present a direct construction of Hermitian representations of the Lie algebra of the group $O(2, 1)$ in a basis in which the (noncompact) generator of the $O(1, 1)$ subgroup is diagonal. This is an interesting problem in itself, with many "peculiar" features and so we shall deviate from the pattern set in Sec. 2 and attempt to solve this directly. In Sec. 4 we follow the usual spirit of MAR and deduce the representations of $O(2, 1)$ in an $O(1, 1)$ basis from a knowledge of the representations of $O(2, 1)$ in an $O(2)$ basis.

[2] A. O. Barut and C. Fronsdal, Proc. Roy. Soc. **A287**, 532 (1965); R. Hermann, Commun. Math. Phys. **3**, 75 (1966); W. J. Holman and L. C. Biedenharn, Ann. Phys. (N.Y.) **39**, 1 (1966); C. Fronsdal, *Proceedings of the IAEA Conference, Trieste, 1965* (IAEA, Vienna, 1965).

[3] J. G. Kuriyan and E. C. G. Sudarshan, Phys. Letters **21**, 106 (1966); N. Mukunda, *Conference on Non-compact Groups in Particles*, Y. Chow, Editor (W. A. Benjamin, Inc., New York, 1966).

[4] E. C. G. Sudarshan in *Proceedings of the Third Coral Gables Conference 1966*, A. Perlmutter, J. Wojtaszek, E. C. G. Sudarshan, and B. Kursunoglu, Eds. (W. H. Freeman and Company, San Francisco, 1966).

[5] J. G. Kuriyan, Ph.D. thesis, Syracuse University, 1966.

[6] N. Mukunda, J. Math. Phys. **9**, 50, 417 (1968); S. Ström, Arkiv Fysik **34**, 215 (1967); A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).

[7] H. Joos, Fortschr. Physik **10**, 65 (1962).

[8] A. O. Barut and R. N. Philips, Commun. Math. Phys. **8**, 52 (1968).

[9] C. Itzykson, Stanford Linear Acceleration Center preprint, 1968.

[10] V. Bargmann, Ann. Math. **48**, 568 (1947).

The Lie algebra of $O(2, 1)$ has three independent elements J_1, J_2, J_3 obeying the commutation rules

$$[J_1, J_2] = -iJ_3, \tag{3.1a}$$
$$[J_2, J_3] = iJ_1, \tag{3.1b}$$
$$[J_3, J_1] = iJ_2. \tag{3.1c}$$

A representation of the J_j by Hermitian operators would lead to a unitary representation of the group $O(2, 1)$. The operator Q given by

$$Q = J_3^2 - J_1^2 - J_2^2 \tag{3.2}$$

commutes with the J_j and so reduces to a real number in every Hermitian irreducible representation of the J_j.

We want to introduce a complete set of orthonormal eigenvectors for the operator J_2 in the space of a representation of the J_j. Naturally we used to know the nature of the eigenvalue spectrum of J_2. To this end, let us rewrite (3.1) in terms of the *Hermitian* operators $J_\pm$:

$$J_\pm = J_1 \pm J_3. \tag{3.3}$$

Then (3.1) reads

$$[J_2, J_\pm] = \pm iJ_\pm, \tag{3.4a}$$
$$[J_+, J_-] = 2iJ_2. \tag{3.4b}$$

The Hermitian operators J_2 and J_+ form a subalgebra of the $O(2, 1)$ Lie algebra. An irreducible Hermitian representation of all the J_j may be expected to be reducible with respect to the subalgebra generated by J_2 and J_+. Imagine this further reduction has been carried out. Within an irreducible representation of J_2 and J_+, what can be said about the spectrum of eigenvalues of J_2 and J_+? First we find easily that for all real α,

$$\exp(-i\alpha J_2) J_+ \exp(i\alpha J_2) = e^\alpha J_+, \tag{3.5}$$

so that the eigenvalue spectrum of J_+ consists of all real positive or of all real negative numbers. One can then consider a Hermitian operator $\ln J_+$ *or* $\ln(-J_+)$, depending on whether J_+ is positive- or negative-semidefinite. Then assuming $\ln J_+$ to be Hermitian, say, we find

$$\exp(i\alpha \ln J_+) J_2 \exp(-i\alpha \ln J_+) = J_2 + \alpha. \tag{3.6}$$

The spectrum of J_2 then consists of all real numbers from $-\infty$ to $+\infty$. This then is the situation within a subspace irreducible under J_2 and J_+ alone.

It is natural then to introduce a basis of eigenvectors of J_2 as follows:

$$J_2 |\lambda; r\rangle = \lambda |\lambda; r\rangle; \quad \langle \lambda'; r' \mid \lambda; r\rangle = \delta_{r'r}\delta(\lambda' - \lambda);$$
$$-\infty < \lambda, \quad \lambda' < \infty. \tag{3.7}$$

What we have to discover is how often a given eigenvalue λ appears, or how many irreducible representations of J_2 and J_+ are needed to synthesize our irreducible representation of J_2, J_+, and J_-. The label r corresponds to this "multiplicity." It is clear though that the range of values of r is independent of the particular eigenvalue λ.

At this point we must comment on the structure of the commutation rules (3.4a). Taken literally, they seem to say, for example, that the state

$$J_+ |\lambda; r\rangle \tag{3.8}$$

is an eigenstate of J_2 with eigenvalue $\lambda + i$. This is impossible since J_2 is a Hermitian operator. We infer that it is not possible to apply the operators $J_\pm$ to the vectors $|\lambda; r\rangle$. The solution to this problem is the following. We must remember that in any case the states $|\lambda; r\rangle$ are "ideal" vectors, which do not represent normalizable vectors in Hilbert space. Omitting for the moment the index r, a normalizable vector $|\phi\rangle$ is really a linear combination of the form

$$|\phi\rangle = \int_{-\infty}^{\infty} d\lambda\, \phi(\lambda) |\lambda\rangle. \tag{3.9}$$

The wavefunction $\phi(\lambda)$ is normalizable in the sense

$$\|\phi\|^2 \equiv \int_{-\infty}^{\infty} d\lambda\, |\phi(\lambda)|^2 < \infty \tag{3.10}$$

and the total Hilbert space is made up of *all* vectors $|\phi\rangle$ with (Lebesgue) square-integrable wavefunctions $\phi(\lambda)$. Now the generators J_2, $J_\pm$ are, in general, unbounded operators and each one has a corresponding domain of vectors $|\phi\rangle$ on which it is defined. For example, J_2 can only act on a vector $|\phi\rangle$ if, in addition to $\phi(\lambda)$, even $\lambda\phi(\lambda)$ is square-integrable. [In this sense, (3.7) is quite formal.] Among all wavefunctions $\phi(\lambda)$, those that J_+ can act upon are characterized as follows: $\phi(\lambda)$ should be the boundary value of an analytic function of λ, such that $f(\lambda)\phi(\lambda - i)$ is also a square-integrable wavefunction:

$$\int_{-\infty}^{\infty} |\phi(\lambda - i)|^2 |f(\lambda)|^2 d\lambda < \infty. \tag{3.11}$$

Here, $f(\lambda)$ is a function to be determined, and which plays the role of the matrix element of J_+. Thus for a vector in the domain of J_+, the wavefunction $\phi(\lambda)$ determines, via analytic continuation, a unique new wavefunction $f(\lambda)\phi(\lambda - i)$, and

$$J_+ |\phi\rangle = J_+ \int_{-\infty}^{\infty} d\lambda\, \phi(\lambda) |\lambda\rangle = \int_{-\infty}^{\infty} d\lambda\, f(\lambda)\phi(\lambda - i) |\lambda\rangle. \tag{3.12}$$

Assuming that a wavefunction $\phi(\lambda)$ is such that both J_2J_+ and J_+J_2 may be applied to it, one can explicitly verify the validity of (3.4a). A similar situation exists

for J_-. [We should also remark that all three operators J_2, J_+, J_- possess a common dense domain of vectors on which all the commutation rules (3.4) may be verified.] The function $f(\lambda)$ plays the role of a matrix element of J_+.

Just as the use of the ideal vectors $|\lambda\rangle$ conveys in a formal but succinct manner the fact that the Hilbert space consists of *all* square-integrable wavefunctions $\phi(\lambda)$, it is natural to introduce a formalism that conveys the information of the last paragraph [and especially (3.12)] in an equally compact and elegant manner. We introduce, then, new "states" labeled as follows:

$$|\lambda \pm i\rangle, \tag{3.13}$$

and write

$$\begin{aligned} J_+ |\phi\rangle &\equiv \int_{-\infty}^{\infty} d\lambda\, f(\lambda)\phi(\lambda - i)\, |\lambda\rangle \\ &= \int_{-\infty}^{\infty} d\lambda\, f(\lambda + i)\phi(\lambda)\, |\lambda + i\rangle, \\ J_- |\phi\rangle &\equiv \int_{-\infty}^{\infty} d\lambda\, g(\lambda + i)\phi(\lambda + i)\, |\lambda\rangle \\ &= \int_{-\infty}^{\infty} d\lambda\, g(\lambda)\phi(\lambda)\, |\lambda - i\rangle, \\ J_+ |\lambda\rangle &= f(\lambda + i)\, |\lambda + i\rangle, \\ J_- |\lambda\rangle &= g(\lambda)\, |\lambda - i\rangle. \end{aligned} \tag{3.14}$$

The use of such objects as $|\lambda \pm i\rangle$ is only a convenient way of representing equations like (3.12). However, their use turns out to have a practical advantage as well.

We have yet to discover the range of the index r in the states $|\lambda; r\rangle$. For the moment we continue to omit this index. Let us now apply the commutation rule (3.4b) to a state $|\lambda\rangle$. Using (3.14) we get

$$f(\lambda)g(\lambda) - f(\lambda + i)g(\lambda + i) = 2i\lambda \tag{3.15}$$

with the solution

$$f(\lambda)g(\lambda) = a_0 - \lambda(\lambda - i), \tag{3.16}$$

where a_0 is a constant. If we use the notation $j(j + 1)$ to denote the eigenvalue of Q, we find

$$Q = j(j + 1) = -a_0 \tag{3.17}$$

and then

$$\begin{aligned} f(\lambda)g(\lambda) &= -j(j + 1) - \lambda(\lambda - i) \\ &= -(j + \tfrac{1}{2})^2 - (\lambda - \tfrac{1}{2}i)^2; \end{aligned} \tag{3.18}$$

j is in general a complex variable such that $j(j + 1)$ is real.

An interesting feature of this problem is that (3.4b) leads only to the product $f(\lambda)g(\lambda)$, and $f(\lambda)$ cannot be related to $g(\lambda)$ in any other way. This is because each of the operators J_+ and J_- is Hermitian. If we had been working in a basis of eigenstates of the compact generator J_3, then the corresponding raising and lowering operators $J_1 \pm iJ_2$ are Hermitian conjugates of one another. In that case, the commutation rule analogous to (3.4b),

$$[J_1 + iJ_2, J_1 - iJ_2] = -2J_3 \tag{3.19}$$

evaluated between eigenstates of J_3, leads to an equation which determines the absolute magnitudes of the matrix elements of $J_1 \pm iJ_2$. [This is also familiar from the treatment of representations of $O(3)$.] In the present case, we are free to choose $f(\lambda)$ in any way we please; then $g(\lambda)$ is fixed by (3.18). The only condition we must obey is that J_+ and J_- are Hermitian; the scalar product with respect to which they must be Hermitian has already been specified in Eqs. (3.10) and (3.7).

Let us write $b^2 = -(j + \frac{1}{2})^2$ and consider first the case where both $f(\lambda)$ and $g(\lambda)$ are linear in λ:

$$f(\lambda) = b + \lambda - \tfrac{1}{2}i; \quad g(\lambda) = b - \lambda + \tfrac{1}{2}i. \tag{3.20}$$

According to (3.14), the operators J_+, J_-, J_2 acting on a wavefunction $\phi(\lambda)$ may then be represented as

$$\begin{aligned} J_2 &= \lambda, \\ J_+ &= (b + \lambda - \tfrac{1}{2}i) \exp\left(-i\frac{\partial}{\partial\lambda}\right) \\ &= \exp\left(-i\frac{\partial}{\partial\lambda}\right)(b + \lambda + \tfrac{1}{2}i), \\ J_- &= (b - \lambda - \tfrac{1}{2}i) \exp\left(i\frac{\partial}{\partial\lambda}\right) \\ &= \exp\left(i\frac{\partial}{\partial\lambda}\right)(b - \lambda + \tfrac{1}{2}i). \end{aligned} \tag{3.21}$$

Working purely formally, $\exp(\pm i\,\partial/\partial\lambda)$ is a Hermitian operator, and $J_\pm$ will also be Hermitian if and only if

$$b^* = b. \tag{3.22}$$

Consequently, such a choice for $f(\lambda)$ and $g(\lambda)$ can lead to Hermitian J_+ and J_- only if j is a complex number of the form

$$j = -\tfrac{1}{2} + ib, \quad b \text{ real}, \tag{3.23}$$

and

$$Q = -\tfrac{1}{4} - b^2 \leq -\tfrac{1}{4}. \tag{3.24}$$

We are then dealing with the continuous nonexceptional series of unitary representations of $O(2, 1)$.

The problem of the index r, or the problem of multiplicity of eigenvalues of J_2, is solved by trying to construct the eigenfunctions of the compact generator J_3, whose eigenvalue spectrum is known to be discrete. For this purpose, we pass to the description of states by wavefunctions $\psi(x)$, where x is a real variable

related to λ, via a Fourier transformation:

$$\psi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{i\alpha\lambda} \phi(\lambda)\, d\lambda,$$
$$\|\psi(x)\|^2 = \int_{-\infty}^{\infty} |\psi(x)|^2\, dx. \tag{3.25}$$

In the x language we have

$$J_2 = -i\frac{\partial}{\partial x},$$
$$J_+ = e^{-x}\left(b + \tfrac{1}{2}i - i\frac{\partial}{\partial x}\right), \tag{3.26}$$
$$J_- = e^{x}\left(b + \tfrac{1}{2}i + i\frac{\partial}{\partial x}\right).$$

Let us look for eigenfunctions of $J_3 = \frac{1}{2}(J_+ - J_-)$ for an eigenvalue m:

$$-i\cosh x \frac{\partial}{\partial x}\psi_m(x) - (b + \tfrac{1}{2}i)\sinh x \psi_m(x) = m\psi_m(x); \tag{3.27}$$

the solution turns out to be

$$\psi_m(x) = \frac{1}{(2\pi)^{\frac{1}{2}}}\left(\frac{1 + ie^{x}}{1 - ie^{x}}\right)^{m} (\cosh x)^{-\frac{1}{2}+ib}. \tag{3.28}$$

Restricting ourselves now to single-valued representations of $O(2, 1)$, m is an integer, positive, negative, or zero. The question now is this: how often must each eigenvalue λ of J_2 appear in order that J_3 have one eigenvector for each integer m as eigenvalue, and such that eigenvectors of J_3 for distinct eigenvalues be orthogonal? We can explicitly compute the scalar product of two wavefunctions $\psi_m(x)$ and $\psi_{m'}(x)$:

$$\int_{-\infty}^{\infty} dx\, \psi_m^*(x)\psi_{m'}(x), \tag{3.29}$$

and we find that this expression is of the form $\delta_{m,m'}$ only if m and m' are both even integers or both odd integers! Thus the set of functions

$$\psi_{2n}(x), \quad n = 0, \pm 1, \pm 2, \cdots, \tag{3.30}$$

by itself forms a complete orthonormal basis for the Hilbert space of square-integrable functions of n; and the same is true for the set of functions

$$\psi_{2n+1}(x), \quad n = 0, \pm 1, \pm 2, \cdots. \tag{3.31}$$

This shows that if we assume that every eigenvalue λ of J_2 occurs only once [in a representation of the continuous nonexceptional series of $O(2, 1)$], we have a contradiction since we end up with the wrong spectrum of eigenvalues for the compact generator J_3. But it is quite clear that this situation can be remedied as follows. We define the eigenfunctions of J_3 to be two-rowed column vectors, each element being made up of a function of x:

$$\Psi_m = \begin{pmatrix} \psi_m(x) \\ (-1)^m \psi_m(x) \end{pmatrix}; \quad m = 0, \pm 1, \pm 2, \cdots. \tag{3.32}$$

By definition the Ψ_m are to be a basis for the Hilbert space of a representation of J_1, J_2, J_3. Since each of the sets of wavefunctions (3.30) and (3.31) forms a complete orthonormal system for the space of square-integrable functions of x, it is clear that every column vector of the form

$$\Phi = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix},$$
$$\|\Phi\|^2 \equiv \int_{-\infty}^{\infty} dx(|\phi_1(x)|^2 + |\phi_2(x)|^2) < \infty, \tag{3.33}$$

where $\phi_1(x)$ and $\phi_2(x)$ are chosen quite independently of one another, can be expanded as a linear combination of the Ψ_m. And one can see that one now has

$$(\Psi_m, \Psi_{m'}) = \delta_{mm'}; \quad m, m' = 0, \pm 1, \pm 2, \cdots. \tag{3.34}$$

The requirement that J_3 have the right spectrum of eigenvalues led to the fact that we have to consider a Hilbert space of wavefunctions of the type (3.33). The variable x is related by Fourier transformation to λ, which is the eigenvalue of J_2. It follows that in representations of J_1, J_2, and J_3 corresponding to the continuous nonexceptional series, every eigenvalue λ of J_2 appears twice; the multiplicity index r has two values. This is in agreement with the observation of Bargmann.[10] Corresponding to the two values of the multiplicity index r, the expressions (3.26) have to be modified by writing the generators as two-dimensional matrices in addition to being linear differential operators in x. The appropriate expressions have been derived elsewhere,[11] and here we quote the results:

$$J_2 = -i\frac{d}{dx} \otimes \sigma_3,$$
$$J_1 = \left[i\sinh x \frac{d}{dx} + i(\tfrac{1}{2} - ib)\cosh x\right] \otimes \sigma_3, \tag{3.35}$$
$$J_3 = \left[-i\cosh x \frac{d}{dx} - i(\tfrac{1}{2} - ib)\sinh x\right] \otimes \mathbf{1}.$$

To summarize the above discussion, the factorization of Eq. (3.18) so as to yield the simplest possible expressions for the functions $f(\lambda)$ and $g(\lambda)$ led to operators $J_\pm$ which were Hermitian only when the parameter b was real. This corresponded exactly

[11] See N. Mukunda, Ref. 6.

to the continuous nonexceptional representations of $O(2, 1)$. The fact that $f(\lambda)$ and $g(\lambda)$ were linear in λ meant that in trying to solve the eigenvalue equation for the compact generator J_3, (3.27), we had a first-order differential equation. The condition that the eigenvalues of J_3 differ from one another by integers, and not by even integers, led to the doubling of the spectrum of the noncompact generator J_2.

There are two other simple choices for $f(\lambda)$ and $g(\lambda)$ which could be made, namely

$$f(\lambda) = \lambda(\lambda - i) + j(j+1); \quad g(\lambda) = -1 \tag{3.36}$$

and

$$f(\lambda) = -\lambda(\lambda - i) - j(j+1); \quad g(\lambda) = +1. \tag{3.37}$$

The operators J_2, and J_+, J_- will then be

$$\begin{aligned} J_2 &= \lambda, \\ J_+ &= \pm[\lambda(\lambda - i) + j(j+1)] \exp\left(-i\frac{\partial}{\partial\lambda}\right) \\ &= \pm\exp\left(-i\frac{\partial}{\partial\lambda}\right)[\lambda(\lambda + i) + j(j+1)], \\ J_- &= \mp\exp\left(i\frac{\partial}{\partial\lambda}\right). \end{aligned} \tag{3.38}$$

The over-all $\pm$ signs in J_+ and $\mp$ signs in J_- correspond, respectively, to (3.36) and (3.37). These choices for $f(\lambda)$ and $g(\lambda)$ are suited to a description of the discrete classes of UIR's of $O(2, 1)$ as we show now.

We first demonstrate that the choice (3.36) leads to the discrete representations $D^{(+)}$ of $O(2, 1)$. The eigenvalue equation for the operator J_3 is

$$\begin{aligned} J_3\psi(x) = -\tfrac{1}{2}e^{-x}\frac{d}{dx}\left(\frac{d\psi}{dx} - \psi\right) \\ + \tfrac{1}{2}j(j+1)e^{-x}\psi + \tfrac{1}{2}e^{x}\psi = m\psi. \end{aligned} \tag{3.39}$$

If we introduce the variable $z = 2e^x$, and the function $\phi(z) = 2\psi(x)/z$, then (3.39) becomes

$$\frac{1}{z^2}\frac{d}{dz}\left(z^2\frac{d\phi}{dz}\right) + \left[\frac{m}{z} - \frac{1}{4} - \frac{j(j+1)}{z^2}\right]\phi(z) = 0. \tag{3.40}$$

This is exactly the form of the radial equation for the bound states of the nonrelativistic hydrogen atom. For any of the following values of j:

$$j = 0, 1, 2, \cdots, \tag{3.41}$$

we know that normalizable solutions of (3.40) exist; for a given j, we find one bound-state type discrete solution for each of the following values of m:

$$m = j+1, j+2, \cdots, \infty. \tag{3.42}$$

It must be kept in mind that in our problem the norm of $\phi(z)$ is given by

$$\|\phi(z)\|^2 = \tfrac{1}{4}\int_0^\infty |\phi(z)|^2 z\, dz \tag{3.43}$$

and this is not the same as the normalization integral for radial wavefunctions of the three-dimensional hydrogen atom. The solutions to (3.40) are, apart from normalization constants,

$$\phi_m(z) = e^{-z/2}z^j L^{2j+1}_{m+j}(z), \quad m = j+1, j+2, \cdots, \infty. \tag{3.44}$$

[The $L^q_p(z)$ are the associated Laguerre polynomials.[12]] Knowing the behavior of $\phi_m(z)$ both near $z = 0$ and $z = \infty$, the basic Eq. (3.40) can be used to show

$$\int_0^\infty \phi_{m'}(z)^*\phi_m(z) z\, dz = 0 \quad \text{if} \quad m \neq m'. \tag{3.45}$$

Thus the operators J_2, $J_\pm$ corresponding to the choice (3.36) lead to the correct spectrum of J_3, and give rise to the discrete representations of $O(2, 1)$ of the type $D^{(+)}_k$. If instead we use (3.37), then the eigenvalue equation for J_3 for any eigenvalue m is the same as (3.40) but with m replaced by $-m$. This time, bound-state type solutions exist provided

$$m = -j-1, -j-2, \cdots, -\infty \tag{3.46}$$

and we are led to the discrete representations of the type $D^{(-)}_k$. In both cases, $D^{(+)}_k$ and $D^{(-)}_k$, we see that each eigenvalue λ of J_2 appears exactly once.

Now we present a construction of the exceptional class of UIR's of $O(2, 1)$ in an $O(1, 1)$ basis. In this class of representations, the Casimir invariant Q has the form

$$Q = -\tfrac{1}{4} + k^2; \quad -\tfrac{1}{4} < Q < 0, \quad 0 < k < \tfrac{1}{2}. \tag{3.47}$$

As in Sec. 3, we begin with a set of eigenvectors of J_2:

$$\begin{aligned} J_2|\lambda\rangle = \lambda|\lambda\rangle; \quad \langle\lambda'|\lambda\rangle = \delta(\lambda' - \lambda); \\ -\infty < \lambda, \quad \lambda' < \infty, \end{aligned} \tag{3.48}$$

and applying the operators $J_\pm = J_1 \pm J_3$ to these vectors we arrive at the following equation for the functions $f(\lambda)$ and $g(\lambda)$:

$$f(\lambda)g(\lambda) = -(\lambda - \tfrac{1}{2}i)^2 - k^2. \tag{3.49}$$

The operators $J_\pm$ would be given by

$$J_+ = \exp\left(-i\frac{\partial}{\partial\lambda}\right)f(\lambda + i); \quad J_- = \exp\left(i\frac{\partial}{\partial\lambda}\right)g(\lambda). \tag{3.50}$$

However, we must obey the condition that J_+ and J_-

[12] For example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Co., New York, 1952).

are separately Hermitian operators. If, analogous to the nonexceptional continuous representation treated in Sec. 3, we attempt to express $f(\lambda)$ and $g(\lambda)$ as quantities linear in λ by writing, say,

$$f(\lambda) = \lambda - \tfrac{1}{2}i + ik, \quad g(\lambda) = -(\lambda - \tfrac{1}{2}i - ik), \tag{3.51}$$

then clearly the Hermiticity of J_+ and J_- is violated. It is desirable, nonetheless, to have both f and g linear in λ. We can reconcile these two requirements by working with an *indefinite-metric* space: we introduce a triplet of Pauli matrices τ_α, $\alpha = 1, 2, 3$, we double the spectrum of J_2 and write

$$\begin{aligned} f(\lambda) &= \lambda - \tfrac{1}{2}i + ik\tau_1, \\ g(\lambda) &= -(\lambda - \tfrac{1}{2}i) + ik\tau_1, \\ J_2 |\lambda; a\rangle &= \lambda |\lambda; a\rangle, \\ \langle \lambda'; a' | \lambda; a\rangle &= \delta(\lambda' - \lambda)(\tau_3)_{a'a}. \end{aligned} \tag{3.52}$$

Then J_+ and J_- are given by

$$\begin{aligned} J_+ &= \exp\left(-i\frac{\partial}{\partial\lambda}\right) f(\lambda + i) \\ &= e^{-x}\left[-i\frac{\partial}{\partial x} + \frac{i}{2} + ik\tau_1\right], \\ J_- &= \exp\left(i\frac{\partial}{\partial\lambda}\right) g(\lambda) = e^{x}\left[i\frac{\partial}{\partial x} + \frac{i}{2} + ik\tau_1\right]. \end{aligned} \tag{3.53}$$

[Subscripts $a, b, c', b', \cdots$ will be used to denote rows and columns associated with the matrices τ_α.] Because of the indefinite metric introduced by the matrix τ_3 above, the operators J_+ and J_- are Hermitian with respect to this metric. [We should call them pseudo-Hermitian.] It should be emphasized that the doubling of the spectrum of J_2 introduced above is not the same as the possible need for doubling the spectrum of J_2 *within* a UIR of $O(2, 1)$ belonging to the continuous exceptional family. Whether or not this latter doubling is called for has to be investigated. The doubling introduced above is just so that J_+ and J_- may be represented by linear differential operators in x, and so that at the same time they may be (pseudo) Hermitian with respect to the appropriate metric. If the spectrum of J_2 within a UIR of $O(2, 1)$ is covered twice, this will certainly have nothing to do with an indefinite metric.

Again we compute the eigenfunctions of J_3, and see whether we can find an orthonormal family of such eigenfunctions, with the eigenvalues being all integers, positive, negative, and zero. Since on the one hand the spectrum of J_3 within a UIR of $O(2, 1)$ is *simple*, and on the other hand we have explicitly introduced a doubling of states via the indefinite metric above, we would expect to find two eigenvectors of J_3 for each eigenvalue m:

$$J_3\Psi_{m,a} = m\Psi_{m,a}; \quad m = 0, \pm1, \pm2, \cdots, \quad a = 1, 2 \tag{3.54}$$

with the property

$$(\Psi_{m',a'}, \Psi_{m,a}) = \delta_{m'm}(\tau_3)_{a'a}. \tag{3.55}$$

We first compute the eigenfunctions of J_3, taking J_3 from (3.53),

$$J_3\psi_m \equiv -i\left[\cosh x\frac{\partial}{\partial x} + \tfrac{1}{2}\sinh x + k\tau_1 \sinh x\right]\psi_m = m\psi_m.$$

We find indeed two independent solutions which we choose to be

$$\begin{aligned} \psi_{m,1} &= [\cosh x]^{-\frac{1}{2}}\left(\frac{1 + i\sinh x}{\cosh x}\right)^m \\ &\quad \times \begin{pmatrix} (\cosh x)^{-k} + (\cosh x)^{+k} \\ (\cosh x)^{-k} - (\cosh x)^{k} \end{pmatrix}, \\ \psi_{m,2} &= [\cosh x]^{-\frac{1}{2}}\left(\frac{1 + i\sinh x}{\cosh x}\right)^m \\ &\quad \times \begin{pmatrix} (\cosh x)^{-k} - (\cosh x)^{k} \\ (\cosh x)^{-k} + (\cosh x)^{k} \end{pmatrix}. \end{aligned} \tag{3.56}$$

The column vectors appearing in these wavefunctions are vectors in the space of the τ_α matrices. When we compute the inner products

$$(\psi_{m',a'}, \psi_{m,a}) = \int_{-\infty}^{\infty} dx\, \psi^\dagger_{m',a'}(x)\tau_3\psi_{m,a}(x)$$

of these wavefunctions, however, we find

$$\begin{aligned} &(\psi_{m',1}, \psi_{m,2}) = (\psi_{m',2}, \psi_{m,1}) = 0; \\ &(\psi_{m',1}, \psi_{m,1}) = -(\psi_{m',2}, \psi_{m,2}) \\ &\quad = \begin{cases} 4\pi, & \text{if } m = m', \\ \dfrac{4e^{i(m-m')\pi/2}}{i(m - m')}[1 - (-1)^{m+m'}], & \\ & \text{if } m \neq m'. \end{cases} \end{aligned} \tag{3.57}$$

Thus the expected relations (3.55) hold only for odd values of m, or only for even values of m, but not jointly for both. This is exactly the situation encountered in our analysis of the continuous nonexceptional UIR's. Again we introduce a second doubling of the spectrum of J_2. We use the Pauli matrices σ_a to describe

this degree of freedom. Thus we may define the generators to be

$$J_2 = -i\frac{\partial}{\partial x}\cdot\sigma_3,$$
$$J_3 = -i\left[\cosh x\frac{\partial}{\partial x} + \tfrac{1}{2}\sinh x + k\tau_1\sinh x\right],$$
$$J_1 = i\left[\sinh x\frac{\partial}{\partial x} + \tfrac{1}{2}\cosh x + k\tau_1\cosh x\right]\sigma_3. \tag{3.58}$$

The eigenfunctions of J_3 are

$$\Psi_{m,1} = \frac{1}{2(2\pi)^{\frac{1}{2}}}[\cosh x]^{-\frac{1}{2}}\left(\frac{1+i\sinh x}{\cosh x}\right)^m \times \begin{pmatrix}(\cosh x)^{-k}+(\cosh x)^{k}\\(\cosh x)^{-k}-(\cosh x)^{k}\end{pmatrix}\otimes\begin{pmatrix}1\\(-1)^m\end{pmatrix},$$
$$\Psi_{m,2} = \frac{1}{2(2\pi)^{\frac{1}{2}}}[\cosh x]^{-\frac{1}{2}}\left(\frac{1+i\sinh x}{\cosh x}\right)^m \times \begin{pmatrix}(\cosh x)^{-k}-(\cosh x)^{k}\\(\cosh x)^{-k}+(\cosh x)^{k}\end{pmatrix}\otimes\begin{pmatrix}1\\(-1)^m\end{pmatrix}. \tag{3.59}$$

In each of these wavefunctions, the first column vector is in the space of the τ_α, while the second is in the space of the σ_α. They obey

$$(\Psi_{m'a'}, \Psi_{ma}) \equiv \int_{-\infty}^{\infty} dx\Psi^{\dagger}_{m'a'}(x)\tau_3\Psi_{ma}(x) = \delta_{m'm}(\tau_3)_{a'a}. \tag{3.60}$$

The doubling of the spectrum of J_2 associated with the matrices σ_α represents a true doubling within a UIR of $O(2,1)$. Again within an exceptional UIR of $O(2,1)$, for each eigenvalue λ, J_2 has two independent eigenvectors. The generators (3.58) are, at the same time, linear differential operators as well as four-dimensional matrices. They give rise to a reducible representation of $O(2,1)$. The exceptional UIR of $O(2,1)$ corresponding to the parameter k appears once with states of positive norm, and once with states of negative norm. The reduction of this representation of $O(2,1)$ into two irreducible parts is *not* equivalent to diagonalizing the metric operator τ_3, since J_1 and J_3 involve the operator τ_1. However, we can show explicitly that if we start with a vector Φ and apply to it the generators J_+, J_-, and J_2 repeatedly, we then obtain an irreducible subspace all of whose elements have positive, negative, or zero norm according as Φ has positive, negative, or zero norm. It is easiest to show this using the eigenfunctions of J_3. We find that the raising and lowering operators $J_1 \pm iJ_2$, with respect to the eigenvalues of J_3, act on $\Psi_{m,a}$ as follows:

$$(J_1 + iJ_2)\Psi_{m,a} = \sum_{a'} i[(m+\tfrac{1}{2}) + k\tau_1]_{aa'}\Psi_{m+1,a'},$$
$$(J_1 - iJ_2)\Psi_{m,a} = \sum_{a'} (-i)[(m-\tfrac{1}{2}) - k\tau_1]_{aa'}\Psi_{m-1,a'}. \tag{3.61}$$

[Thus, the vectors $\Psi_{m,1}$ by themselves are not invariant under the action of the generators!] If we now construct a sequence of vectors Φ_m, one for each value of m,

$$\Phi_m = \sum_a \phi_a(m)\Psi_{m,a}, \tag{3.62}$$

and demand that $J_1 \pm iJ_2$ acting on Φ_m give, respectively, some multiples of $\Phi_{m\pm1}$, we find

$$i[(m+\tfrac{1}{2}) + k\tau_1]\phi(m) = c_m\phi(m+1),$$
$$-i[(m-\tfrac{1}{2}) - k\tau_1]\phi(m) = d_{m-1}\phi(m-1), \tag{3.63}$$
$$c_m d_m = (m-\tfrac{1}{2})^2 - k^2 > 0.$$

Here c_m and d_m are nonvanishing complex numbers. The norms of the Φ_m then obey

$$(\Phi_{m+1}, \Phi_{m+1}) = \phi(m+1)^{\dagger}\tau_3\phi(m+1) = |c_m|^{-2}[(m+\tfrac{1}{2})^2 - k^2]\phi(m)^{\dagger}\tau_3\phi(m) = |c_m|^{-2}[(m+\tfrac{1}{2})^2 - k^2](\Phi_m, \Phi_m). \tag{3.64}$$

Thus the vectors Φ_m appearing in such a sequence, span an irreducible subspace under $O(2,1)$; and all of them have norms of the same sign. (In particular, all their norms might vanish.) For example, for the two choices

$$\Phi_0 = \Psi_{0,1}$$

and

$$\Phi_{0'} = \Psi_{0,2},$$

we obtain two sequences Φ_m and Φ'_m; the former have all positive norms, the latter all negative norms. [Φ_m *do not* coincide with $\Psi_{m,1}$ for all m!] This accomplishes the reduction of the original representation of $O(2,1)$ into two irreducible parts.

It seems as if the freedom of choosing S (or σ) to be positive or negative in the expression $j = -\frac{1}{2} + S$ (or $j = -\frac{1}{2} + i\sigma$) for the continuous exceptional (or nonexceptional) class of representations of $O(2,1)$ is related to the feature of the doubling of the spectrum explained earlier. A careful analysis of $O(4,1)$ in an $O(3,1)$ basis ought to test the validity of this conjecture.

4. $O(2,1)$ IN AN $O(1,1)$ BASIS VIA MAR

Here we would like to indicate how the method of MAR can be used to obtain the representations of $O(2,1)$ in an $O(1,1)$ basis from a knowledge of the

representations of $O(2, 1)$ in an $O(2)$ basis—and thus obtain Eq. (3.18) which was explicitly derived from first principles in the main body of the text. More specifically, we shall use the method of MAR and Eq. (2.3) to obtain Eq. (3.18).

We define

$$\begin{aligned} N_1 &\equiv J_1', \\ N_2 &\equiv iJ_3', \\ N_3 &\equiv iJ_2', \end{aligned} \tag{4.1}$$

and observe that N_1, N_2, N_3 generate an $O(2, 1)$ group which leaves $-N_1^2 - N_2^2 + N_3^2$ invariant. Our aim is to diagonalize N_2.

The eigenstate $|m\rangle$ of J_3' with eigenvalue m is now an eigenstate of N_2, which we call $|\lambda\rangle'$, with eigenvalues $\lambda \equiv im$. That is

$$N_2 |\lambda\rangle' \equiv iJ_3' |m\rangle = im |m\rangle \equiv \lambda |\lambda\rangle', \tag{4.2}$$

where we define

$$\lambda \equiv im$$

and

$$|\lambda\rangle' \equiv |m\rangle. \tag{4.3}$$

The raising and lowering operators defined by

$$N_\pm \equiv J_1' \pm iJ_2' \equiv J_\pm' \tag{4.4}$$

change the (eigen-) state of J_2' with eigenvalue $m = -i\lambda$ to a state of eigenvalue $m \pm 1 = -i(\lambda \pm i)$

Therefore we can use the notation of Sec. 3, and proceed in a cavalier fashion, to define $f(\lambda)$ and $g(\lambda)$:

$$\begin{aligned} N_+ |\lambda\rangle' &= f(\lambda + i) |\lambda + i\rangle', \\ N_- |\lambda\rangle' &= g(\lambda) |\lambda - i\rangle', \end{aligned} \tag{4.5}$$

such that

$$N_+ N_- |\lambda\rangle' = f(\lambda) g(\lambda) |\lambda\rangle'. \tag{4.6}$$

The lhs of Eq. (4.6) is thus

$$N_+ N_- |\lambda\rangle' = J_+' J_-' |m\rangle \tag{4.7a}$$

$$= (m - \tfrac{1}{2})^2 - (j + \tfrac{1}{2})^2 |m\rangle \tag{4.7b}$$

$$= (-i\lambda - \tfrac{1}{2})^2 - (j + \tfrac{1}{2})^2 |\lambda\rangle' \tag{4.7c}$$

$$= -(\lambda + \tfrac{1}{2}i)^2 - (j + \tfrac{1}{2})^2 |\lambda\rangle', \tag{4.7d}$$

where to obtain (4.7a) and (4.7c) we have used Eqs. (4.3) and (4.4) and to obtain (4.7b) we have used Eq. (4.3).

We have, on comparing (4.7) with Eq. (4.6), $(f\lambda)g(\lambda) = -(\lambda + \frac{1}{2}i)^2 - (j + \frac{1}{2})^2$ which is Eq. (3.18). This serves to illustrate the power of the method of MAR and renders the claims of general validity of this principle more plausible—at least to the discerning reader.

5. DISCUSSION

We had asserted that the method of MAR is not only useful for the purpose of reducing noncompact groups with respect to its maximal compact subgroups, but also to reduce noncompact groups with respect to its noncompact subgroups. To render this assertion plausible, we reduced $O(2, 1)$ with respect to an $O(1, 1)$ subgroup. Since this problem at the time of writing this paper had been handled only with global techniques, we analyzed this problem in great detail (Sec. 3)—and later (in Sec. 4) obtained the same results in a MAR.

We have not attempted to formulate the prescription in those cases where a state labeling problem exists. We hope that when the state labeling problem is solved one could guarantee the method of MAR to those cases as well.

The crucial fact that is exploited in the whole approach is that there exists a master analytic function which describes the representations of groups that have the same complex extension—and once this is determined, the representations are obtained after some algebraic manipulations. The implication is that a student, armed with the matrix elements of the generators of the group $SU(n)$ and $SO(n)$ that are tabulated in Gel'fand and Tseitlin,[13] can obtain after some trivial manipulations the matrix elements of the generators of the groups such as $SU(n - 1, 1)$, $SO(n - 1, 1)$[14] [and perhaps even $SU(n - 2, 2)$, $SO(n - 2, 2)$!]. Then these can be analyzed to obtain the representations of the group in question.

ACKNOWLEDGMENTS

The authors would like to thank Professor A. P. Balachandran, Professor A. J. Macfarlane, Professor J. Nilsson, and Professor L. O'Raifeartaigh for many useful discussions.

[13] I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR **71**, 1017 (1950).

[14] Degenerate representations of groups such as $O(p, q)$ reduced with respect to $O(p) \otimes O(q)$ and $U(p, q)$ reduced with respect to $U(p) \otimes U(q)$ have been obtained by the Trieste group, J. Fischer, J. Niederle, and R. Raçzka, International Center for Theoretical Physics Preprint IC/65/63, 1965; R. Raçzka, IC/65/80, 1965; and *Elementary Particle Theories* (Springer-Verlag, Berlin, 1966). MAR can be used directly to obtain these results as well.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Auxiliary Variables in Statistical Mechanics: Variational Principle

GERALD SPEISMAN
Florida State University, Tallahassee, Florida

(Received 5 July 1967)

The variational principle of Bohm and Pines for the ground-state energy of the electron gas with uniform neutralizing positive-charge background, employing auxiliary variables, is reviewed as an illustration of the past use of auxiliary variables and as an example of the type of physical system to which the variational principle of this paper can be applied. We then develop this variational principle for the logarithm of the partition function of a physical system at nonzero temperatures, employing auxiliary variables. The variational principle contains a trial Hamiltonian H' in an extended Hilbert space. For H' equal to its optimal value H, the variational expression $\ln Q'$ is equal to the logarithm of the partition function $\ln Q$. For $H' \neq H$, it is shown that $\ln Q \geq \ln Q'$ for $H' - H$ sufficiently small or temperatures sufficiently high, or for sufficiently low temperatures when an additional assumption is made, which reduces to one made by Bohm and Pines when applied to the electron gas. The variational expression $\ln Q'$ contains more complicated trace formulas than are usually encountered in quantum statistical mechanics; one possible method of evaluation is sketched leading to a simpler approximate formula for $\ln Q'$. Corrections to the variational approximation for $\ln Q$ are provided by the second- and higher-order terms in a certain perturbation expansion of $\ln Q$. The variational principle developed here implies the variational principle of Bohm and Pines in the zero-temperature limit; in the "no auxiliary variable" limit it reduces to a modified form of Peierls' variational theorem. It is shown how the variational principle can be applied to any physical system containing charged particles in which the long-range collective effects of the Coulomb interaction are important.

This paper is divided into three sections. The variational principle for the logarithm of the partition function, employing auxiliary variables, which is the main contribution of the paper, is developed in Sec. II. In Sec. I we discuss auxiliary variables and statistical mechanics preparatory to Sec. II. In Sec. III we examine the results of the preceding sections.

I. PRELIMINARY REMARKS

When Bohm and Pines[1] set out to compute the ground-state energy of an electron gas in the presence of a uniform neutralizing positive charge background, they were faced with the mathematical and physical problem of dealing in a fairly accurate manner with the long-wavelength Fourier components of the electrostatic interaction between electrons. The mathematical problem was that these Fourier components caused straightforward perturbation theory to diverge. The physical problem was that plasma modes of oscillation were known to exist in an ionized gas which is, to a good approximation, an electron gas with neutralizing positive charge background.

Bohm and Pines were interested in the electron-gas problem as a mathematically simpler model of the conduction electrons in a metallic crystal, which is well approximated by the conduction electrons moving in a neutralizing lattice of positive charge distribution. Because they were aware that straightforward perturbation theory would not work, and they felt that plasma modes made an important contribution to the ground-state energy of an electron gas at the densities found for conduction electrons in metals, Bohm and Pines added extra degrees of freedom to the physical system by introducing auxiliary variables that more directly represented the plasma modes than did the individual electron coordinates. They did this in essentially the following way:

The physical system under consideration is that of N electrons confined to a region of volume V in the presence of a neutralizing uniform positive charge density. N and V are large but not independent, since the electron density N/V is to have a prescribed value. N and V are to be taken large enough so that when appropriate we can let N and V approach infinity, holding N/V fixed, with negligible error.

One physically reasonable method of confining the N electrons is to place them inside a cubical box with impenetrable, perfectly conducting walls. A less physically obvious but equivalent and mathematically simpler procedure is to confine the electrons to a cubical box of volume V subject to periodic boundary conditions on the nonrelativistic wavefunction and its normal gradient. The electrostatic potential of each electron must also satisfy the same boundary conditions. If we neglect spin interactions, then, as the nonrelativistic Hamiltonian operator of the physical system, we obtain

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m} + 2\pi\epsilon^2 V^{-1} \sum_{\mathbf{k}\neq 0} k^{-2}(p_{\mathbf{k}}^{\dagger}\rho_{\mathbf{k}} - N), \quad (1)$$

where $\mathbf{p}_i$ is the vector-momentum operator of the ith

[1] D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953).

electron, $p_i^2 = \mathbf{p}_i \cdot \mathbf{p}_i$, m is the mass of the electron, ϵ the magnitude of its charge,

$$\mathbf{k} = 2\pi V^{-\frac{1}{3}}[l\mathbf{i}_1 + m\mathbf{i}_2 + n\mathbf{i}_3], \tag{2}$$

where $\mathbf{i}_1$, $\mathbf{i}_2$, and $\mathbf{i}_3$ are unit vectors parallel, respectively, to three mutually perpendicular edges of the cubical box, l, m, and n are any integers, and the operator

$$\rho_{\mathbf{k}} \equiv \sum_{j=1}^{N} e^{-i\mathbf{k}\cdot\mathbf{r}_j}, \tag{3}$$

where $\mathbf{r}_j$ is the position operator of the jth electron. $\rho_{\mathbf{k}}^\dagger$ is the adjoint of $\rho_{\mathbf{k}}$.

If we apply an external electrostatic field, derivable from a potential $\phi(\mathbf{r})$, to the system of N electrons plus neutralizing positive charge background, we obtain the Hamiltonian operator

$$H_\phi = H - \epsilon\sum_{j=1}^{N}\phi(\mathbf{r}_j) + \epsilon NV^{-1}\iiint_{\text{cube}}\phi(\mathbf{r})\,dx\,dy\,dz + (8\pi)^{-1}\iiint_{\text{cube}}(\nabla\phi(\mathbf{r}))^2\,dx\,dy\,dz. \tag{4}$$

In the above equation, H is the Hamiltonian operator of Eq. (1), the second term on the right-hand side is the electrostatic energy of the N electrons, the third term is the electrostatic energy of the positive charge background and the fourth term is the electrostatic-field energy.

If we take the electrostatic potential of the form

$$\phi(\mathbf{r}) = 2\pi^{\frac{1}{2}}V^{-\frac{1}{2}}\sum_{\substack{\mathbf{k}\\0<k<k_c}} k^{-1}\beta_{\mathbf{k}}^* e^{i\mathbf{k}\cdot\mathbf{r}} + \phi_0, \tag{5}$$

where $\beta_{\mathbf{k}}$ and ϕ_0 are independent of $\mathbf{r}$, $\phi_0^* = \phi_0$,

$$\beta_{\mathbf{k}}^* = \beta_{-\mathbf{k}}, \tag{6}$$

and k_c is a positive value of k that we are free to choose, then H_ϕ of Eq. (4) becomes, using Eqs. (1), (5), and (6),

$$H_\phi = H\{\beta_{\mathbf{k}}\} \equiv \sum_{i=1}^{N}\frac{p_i^2}{2m} + 2\pi\epsilon^2V^{-1}\sum_{k\geq k_c}\sum_{\mathbf{k}} k^{-2}(\rho_{\mathbf{k}}^\dagger\rho_{\mathbf{k}} - N) + \tfrac{1}{2}\sum_{\substack{\mathbf{k}\\0<k<k_c}}[(\beta_{\mathbf{k}} - 2\pi^{\frac{1}{2}}V^{-\frac{1}{2}}\epsilon k^{-1}\rho_{\mathbf{k}})^\dagger \times (\beta_{\mathbf{k}} - 2\pi^{\frac{1}{2}}V^{-\frac{1}{2}}\epsilon k^{-1}\rho_{\mathbf{k}}) - 4\pi\epsilon^2NV^{-1}k^{-2}]. \tag{7}$$

What Bohm and Pines[2] did was to introduce new complex-coordinate operators $q_{\mathbf{k}}$ and their canonically conjugate momentum operators $\pi_{\mathbf{k}}$. Thus

$$[q_{\mathbf{k}'}, q_{\mathbf{k}}] = [\pi_{\mathbf{k}'}, \pi_{\mathbf{k}}] = 0,$$
$$[\pi_{\mathbf{k}'}, q_{\mathbf{k}}] = i\hbar\delta_{\mathbf{k}'\mathbf{k}}, \tag{8}$$

where $\delta_{\mathbf{kk}} = 1$, $\delta_{\mathbf{k}'\mathbf{k}} = 0$ for $\mathbf{k}' \neq \mathbf{k}$, and $[A, B]$ denotes the commutator AB–BA of the operators A and B. The operators were further taken to satisfy the relations:

$$q_{\mathbf{k}}^\dagger = q_{-\mathbf{k}}, \quad \pi_{\mathbf{k}}^\dagger = \pi_{-\mathbf{k}}. \tag{9}$$

The operators $q_{\mathbf{k}}$, $\pi_{\mathbf{k}}$ were taken to commute with the coordinate, momentum, and spin operators of each electron. The Hilbert space upon which these new operators, plus the original electron operators, operate was taken as one in which the position and z component of spin operators for each electron, plus each operator $q_{\mathbf{k}}$, were a complete set of commuting operators, i.e., the simultaneous eigenvectors of these operators are nondegenerate and form a complete set. In this "extended" Hilbert space, Bohm and Pines defined an "extended" Hamiltonian operator H_{ext} by replacing the complex numbers $\beta_{\mathbf{k}}$, $\beta_{\mathbf{k}}^* = \beta_{-\mathbf{k}}$, in Eq. (7) by the operators $\pi_{\mathbf{k}}$, $\pi_{\mathbf{k}}^\dagger = \pi_{-\mathbf{k}}$ [we have used Eqs. (6) and (9)]. Thus,

$$H_{\text{ext}} \equiv H\{\pi_{\mathbf{k}}\}, \tag{10}$$

where $H\{\beta_{\mathbf{k}}\}$ is defined in Eq. (7).

It is possible to establish the following results, based on Eq. (10), and the fact that $H\{\beta_{\mathbf{k}}\}$ of Eq. (7) is the Hamiltonian operator for a physical system and must, therefore, possess a complete orthonormal set of eigenvectors:

(A) H_{ext} possesses a complete set of eigenvectors;

(B) Every eigenvalue of H_{ext} is an eigenvalue of $H\{\beta_{\mathbf{k}}\}$ for some value of $\{\beta_{\mathbf{k}}\}$, and for every value of $\{\beta_{\mathbf{k}}\}$ every eigenvalue of $H\{\beta_{\mathbf{k}}\}$ is an eigenvalue of H_{ext}.

If $E_0\{\beta_{\mathbf{k}}\}$, denotes the ground-state (lowest) energy of the Hamiltonian operator $H\{\beta_{\mathbf{k}}\}$, then what Bohm and Pines[1,3] assumed was that

$$E_0\{\beta_{\mathbf{k}}\} \geq E_0\{0\}, \quad \forall\ \{\beta_{\mathbf{k}}\} \neq \{0\}. \tag{11}$$

Stating Eq. (11) in physical terms:

(C) *The ground-state energy of the physical system in the presence of any external electrostatic field, whose potential is of the form of Eq.* (5), *is assumed to be never less than the ground-state energy in the absence of the external electrostatic field.*

From (A), (B), and Eq. (11) we then obtain as the standard variational principle for the lowest eigenvalue of the Hermitian operator H_{ext}:

$$\langle\text{ext}|\,H_{\text{ext}}\,|\text{ext}\rangle/\langle\text{ext}\,|\,\text{ext}\rangle \geq E_0, \tag{12}$$

where $|\text{ext}\rangle$ is any nonzero vector in the extended Hilbert space, and $E_0 = E_0\{0\}$ is the ground-state

[2] D. Bohm, K. Huang, and D. Pines, Phys. Rev. **107**, 71 (1957).

[3] The actual assumption made (p. 614 of Ref. 1) was that the ground state of H_{ext} was nondegenerate. This assumption was then used to show that ground-state energy of H_{ext} was $E_0\{0\}$, from which Eq. (11) follows.

energy of the N-electron system being considered (with no external electrostatic field present).

If the response of the electron gas in its ground state to the applied electrostatic field, neglecting effects from the surface of the cubical volume V, is linear, homogeneous, and isotropic, we then obtain as the electrostatic energy

$$\frac{1}{(8\pi)}\iiint_{\text{cube}} \mathbf{E}(\mathbf{r})\cdot\mathbf{D}(\mathbf{r})\,dx\,dy\,dz = \frac{1}{2}\sum_{\mathbf{k}\ 0<k<k_c}\frac{\beta_{\mathbf{k}}^{*}\beta_{\mathbf{k}}}{\epsilon(k)}, \quad (13)$$

where $\mathbf{E}(\mathbf{r})$ and $\mathbf{D}(\mathbf{r})$ are the macroscopic electric-field and displacement vectors, respectively, and $\epsilon(k)$ is the static dielectric constant for an electrostatic field of wavenumber vector $\mathbf{k}$.

When Eq. (13) is valid we see that statement (C) is valid if and only if the static dielectric constant $\epsilon(k)$ is nonnegative for $0 < k < k_c$. In the random-phase or equivalent approximation the electrostatic energy of the electron gas in its ground state always has the form given in Eq. (13) with $\epsilon(k) > 0$ for $k > 0$.[4] As the density of the electron gas increases, the random-phase approximation becomes more and more accurate, so that assumption (C) is valid for the electron gas at high densities.

Statement (C), or the nonnegativeness of $\epsilon(k)$ in Eq. (13) for the case of a linear, homogeneous, isotropic, electrostatic response, is a statement of stability of the zero-field ground state of the physical system. If (C) did not hold it would be energetically possible to have a spontaneous transition from the zero-field ground state to some nonzero electrostatic-field ground state.

Equation (12) is a variational principle for the lowest energy E_0 of the physical system in terms of an operator and vectors in the extended Hilbert space. This enables us to use trial vectors $|\text{ext}\rangle$ that have no counterparts $|\ \rangle$ in the physical Hilbert space, and yet yield a lower upper limit to E_0 than can be obtained from the usual variational principle

$$\langle|H|\rangle/\langle\ |\ \rangle \geq E_0, \quad (14)$$

with computationally tractable choices of the vectors $|\ \rangle$.

The operators $q_{\mathbf{k}}$ and $\pi_{\mathbf{k}}$ introduced by Bohm and Pines[2] are non-Hermitian operators related by Eq. (9). For our purposes it is more convenient to deal with independent Hermitian canonically conjugate operators. The operators $X_{\mathbf{k}}$, $Y_{\mathbf{k}}$ defined below have such properties:

$$X_{\mathbf{k}} \equiv \tfrac{1}{2}[q_{\mathbf{k}} + q_{-\mathbf{k}} + i(q_{\mathbf{k}} - q_{-\mathbf{k}})],$$
$$Y_{\mathbf{k}} \equiv \tfrac{1}{2}[\pi_{\mathbf{k}} + \pi_{-\mathbf{k}} - i(\pi_{\mathbf{k}} - \pi_{-\mathbf{k}})]. \quad (15)$$

[4] J. Lindhard, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **28**, 57 (1954).

Equation (15) can be inverted to give

$$q_{\mathbf{k}} = \tfrac{1}{2}[X_{\mathbf{k}} + X_{-\mathbf{k}} - i(X_{\mathbf{k}} - X_{-\mathbf{k}})],$$
$$\pi_{\mathbf{k}} = \tfrac{1}{2}[Y_{\mathbf{k}} + Y_{-\mathbf{k}} + i(Y_{\mathbf{k}} - Y_{-\mathbf{k}})]. \quad (16)$$

In this paper we extend the use of auxiliary variables of the type employed by Bohm and Pines to obtaining thermodynamic properties of a system in thermodynamic equilibrium at temperature T, with $T > 0$. More specifically, we will obtain a variational principle which in the limit $T \to 0$ implies Eq. (12) when applied to the electron gas with a uniform neutralizing positive charge background.

We begin by reviewing some basic concepts of statistical mechanics and thermodynamics. A physical system that is confined to a finite region of space possesses a complete orthonormal set $|i\rangle$ of energy eigenvectors with corresponding energy eigenvalues E_i, where i takes on a set of integer values. If this system is in thermodynamic equilibrium, then the eigenvalues E_i are independent of time and the probability P_i of finding the system in the ith energy eigenstate is proportional to the Boltzmann factor, $\exp(-E_i/kT)$, where k is Boltzmann's constant and T is the temperature. Thus,

$$P_i = Q^{-1}\exp(-E_i/kT). \quad (17)$$

Since we must have $\Sigma_i P_i = 1$, we see that

$$\Sigma_i \exp(-E_i/kT)$$

must converge and that

$$Q = \Sigma_i \exp(-E_i/kT). \quad (18)$$

The quantity Q is called the partition function. The Helmholtz function (or free energy),

$$A \equiv U - TS, \quad (19)$$

where U is the internal energy and S the entropy, is related to Q by the simple relation

$$A = -kT\ln Q. \quad (20)$$

It is a fundamental theorem of thermodynamics that all thermodynamic properties of a physical system can be expressed in terms of the Helmholtz function, and thus, from Eq. (20), in terms of the partition function.

For the subsequent discussion, it is convenient to define

$$\beta = 1/kT. \quad (21)$$

Then Eq. (18) for the partition function Q becomes

$$Q = \Sigma_i e^{-\beta E_i}. \quad (22)$$

This completes the preliminary remarks.

II. VARIATIONAL PRINCIPLE

The main purpose of this paper is to develop a variational principle for the logarithm of the partition function $\ln Q$ in terms of operators in an extended Hilbert space $\mathcal{H}'$. This extended Hilbert space arises from the physical Hilbert space $\mathcal{H}$ through the introduction of auxiliary operators similar to the operators $X_{\mathbf{k}}$, $Y_{\mathbf{k}}$ of Eq. (15). We proceed as follows.

Our first task is to construct an extended Hilbert space analogous to the one used by Bohm and Pines[1] for the electron gas, to define canonically conjugate operators analogous to $X_{\mathbf{k}}$, $Y_{\mathbf{k}}$ of Eq. (15), and to define an Hermitian operator analogous to H_{ext} of Eq. (10) in the extended Hilbert space.

If H_0 is the Hamiltonian operator of the physical system being considered, then if the system is in thermodynamic equilibrium H_0 must be independent of time. Since the vectors $|i\rangle$ are energy eigenvectors [cf. discussion preceding Eq. (17)], we have

$$H_0 |i\rangle = E_i |i\rangle. \tag{23}$$

Let $y_1, \cdots, y_g$ be a set of g real variables. Choose an Hermitian operator $H(y_1, \cdots, y_g)$ for each value of $y_1, \cdots, y_g$ such that

$$H(0, \cdots, 0) = H_0. \tag{24}$$

The operators H_0, $H(y_1, \cdots, y_g)$ and the vectors $|i\rangle$ are all in the physical Hilbert space $\mathcal{H}$. The Hermitian operators $H(y_1, \cdots, y_g)$ are analogous to the operators $H\{\beta_{\mathbf{k}}\}$ of Eq. (7).

Now take any complete orthonormal set $|\alpha\rangle$ in the physical Hilbert space $\mathcal{H}$, where α takes on a set of integer values. Consider the set of all functions of α and the real variables $x_1, \cdots, x_g$. This set of functions forms a linear vector space $\mathfrak{V}$. We define operators $X_1, \cdots, X_g$, $Y_1, \cdots, Y_g$ in $\mathfrak{V}$ as follows: For any function $f(\alpha, x_1, \cdots, x_g)$,

$$X_i f(\alpha, x_1, \cdots, x_g) \equiv x_i f(\alpha, x_1, \cdots, x_g), \qquad i = 1, 2, \cdots, g, \tag{25}$$

and for any function $f(\alpha, x_1, \cdots, x_g)$ possessing first partial derivatives with respect to $x_1, \cdots, x_g$,

$$Y_j f(\alpha, x_1, \cdots, x_g) \equiv i\hbar(\partial/\partial x_j) f(\alpha, x_1, \cdots, x_g), \qquad j = 1, 2, \cdots, g. \tag{26}$$

It then follows that

$$[X_i, X_j] = [Y_i, Y_j] = 0, \quad [Y_j, X_k] = -i\hbar\delta_{jk}, \tag{27}$$

where $\delta_{kk} = 1$, $\delta_{jk} = 0$ for $j \neq k$, and $[A, B] \equiv AB - BA$, when operating on any function of $\alpha, x_1, \cdots, x_g$ possessing continuous second partial derivatives with respect to $x_1, \cdots, x_g$. If we require that all functions $f(\alpha, x_1, \cdots, x_g)$ satisfy the relation

$$\Sigma_\alpha \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(\alpha, x_1, \cdots, x_g)|^2 \, dx_1 \cdots dx_g < \infty,$$

then we obtain a Hilbert space $\mathcal{H}_\infty$ analogous to the extended Hilbert space used by Bohm and Pines for the electron gas. The operators X_i, Y_i are canonically conjugate operators in this Hilbert space analogous to the operators $X_{\mathbf{k}}$, $Y_{\mathbf{k}}$ of Eq. (15) for the electron gas.

The Hilbert space $\mathcal{H}_\infty$ is not suitable for the variational principle to be developed in this paper. This is because the eigenvalues of the operators Y_i are continuous, so that a projection operator onto the linear manifold of all simultaneous eigenvectors of Y_i, $i = 1, 2, \cdots, g$, with eigenvalues zero does not exist in $\mathcal{H}_\infty$. To remedy this situation we define the extended Hilbert space $\mathcal{H}'$ as the set of all functions of $\alpha, x_1, \cdots, x_g, f(\alpha, x_1, \cdots, x_g)$, that are periodic in x_i with period L_i, $i = 1, 2, \cdots, g$, and satisfy the relation

$$\Sigma_\alpha \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \cdots \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} |f(\alpha, x_1, \cdots, x_g)|^2 \times dx_g \cdots dx_2 \, dx_1 < \infty. \tag{28}$$

We denote such a function $f(\alpha, x_1, \cdots, x_g)$ as $|'\rangle$, where the prime denotes the vector is in the Hilbert space $\mathcal{H}'$. The inner product of two functions $f_1(\alpha, x_1, \cdots, x_g)$, $f_2(\alpha, x_1, \cdots, x_g)$ in $\mathcal{H}'$ is defined as follows:

$$\langle' 1 | ' 2\rangle \equiv \Sigma_\alpha \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \cdots \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} f_1^*(\alpha, x_1, \cdots, x_g) \times f_2(\alpha, x_1, \cdots, x_g) \, dx_g \cdots dx_2 \, dx_1. \tag{29}$$

From the definition of $\mathcal{H}'$ and Eq. (26) we see that Y_i, $i = 1, 2, \cdots, g$, are operators in the Hilbert space $\mathcal{H}'$. Examination of Eq. (25) shows that the operators X_i are not in $\mathcal{H}'$, and we do not use these operators.

We define an Hermitian operator analogous to H_{ext} of Eq. (10) as follows: The functions of $\alpha, x_1, \cdots, x_g$: $|' \beta, n_1, \cdots, n_g\rangle$

$$\equiv \left[\prod_{j=1}^{g} L_j\right]^{-\frac{1}{2}} \delta_{\beta\alpha} \exp\left[2\pi i \sum_{j=1}^{g} (n_j x_j / L_j)\right], \tag{30}$$

where $n_1, \cdots, n_g$ take on all integer values independent of each other, β takes on the same set of integer values as α does and $\delta_{\alpha\alpha} = 1$, $\delta_{\beta\alpha} = 0$ for $\beta \neq \alpha$, are a complete orthonormal set in the Hilbert space $\mathcal{H}'$, so that

$$\langle' \beta', n_1', \cdots, n_g' | ' \beta, n_1, \cdots, n_g\rangle = \delta_{\beta'\beta} \prod_{j=1}^{g} \delta_{n_j' n_j}, \tag{31}$$

where $\delta_{\beta\beta} = \delta_{n_j n_j} = 1$ and $\delta_{\beta'\beta} = \delta_{n_j' n_j} = 0$ for $\beta' \neq \beta$, $n_j' \neq n_j$. We define the operator H in the extended Hilbert space $\mathcal{H}'$ as one whose matrix elements

$$\langle' \beta', n_1', \cdots, n_g'| H |' \beta, n_1, \cdots, n_g\rangle \equiv \langle\beta'| H(hn_1/L_1, \cdots, hn_g/L_g) |\beta\rangle \prod_{j=1}^{g} \delta_{n_j' n_j}. \quad (32)$$

From the above equation and the fact that $H(y_1, \cdots, y_g)$ are Hermitian operators, it follows that H is an Hermitian operator. H is analogous to the operator H_{ext} of Eq. (10) for the electron gas.

Our next task is to formulate the variational principle for $\ln Q$, where Q is given in Eqs. (18) or (22). From Eqs. (26) and (30) we see that

$$Y_i |' \beta, n_1, \cdots, n_g\rangle = (hn_i/L_i) |' \beta, n_1, \cdots, n_g\rangle, \quad i = 1, 2, \cdots, g, \quad (33)$$

so that the operators Y_i are Hermitian, i.e.,

$$Y_i^\dagger = Y_i. \quad (34)$$

We define the operator Λ in the Hilbert space $\mathcal{H}'$ as follows:

$$\Lambda |' \beta, n_1, \cdots, n_g\rangle \equiv \prod_{j=1}^{g} \delta_{n_j 0} |' \beta, n_1, \cdots, n_g\rangle, \quad \forall\ \beta, \quad n_1, \cdots, n_g, \quad (35)$$

where $\delta_{00} = 1$ and $\delta_{n_j 0} = 0$ for $n_j \neq 0$. From the above equation we can show that

$$\Lambda^\dagger = \Lambda, \quad \Lambda^2 = \Lambda. \quad (36)$$

Geometrically, Λ is the projection operator onto the linear manifold of simultaneous eigenvectors of $Y_1, \cdots, Y_g$ with eigenvalues zero.

The operator Λ is a function of the operators $Y_1, \cdots, Y_g$, as can be seen from Eqs. (33) and (35). An explicit form for this function is as follows:

$$\Lambda = \left[\prod_{j=1}^{g} L_j\right]^{-1} \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \cdots \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} \times \exp\left(i\hbar^{-1} \prod_{j=1}^{g} \xi_j Y_j\right) d\xi_g \cdots d\xi_2\, d\xi_1. \quad (37)$$

This can be shown by applying both sides of the above equation to the complete orthonormal set of vectors $|' \beta, n_1, \cdots, n_g\rangle$ and using Eqs. (33) and (35).

The variational principle for $\ln Q$ is as follows: If H' is any Hermitian operator in the extended Hilbert space $\mathcal{H}'$, i.e.,

$$H'^\dagger = H', \quad (38)$$

for which

$$\ln Q' \equiv \ln \text{Tr}\{\Lambda e^{-\beta H'}\} - \left[\int_0^\beta \text{Tr}\{\Lambda e^{-\gamma H'}(H - H') \times e^{-(\beta-\gamma)H'}\}\, d\gamma / \text{Tr}\{\Lambda e^{-\beta H'}\}\right] \quad (39)$$

exists, where Tr denotes the trace operation in $\mathcal{H}'$, then

$$\ln Q \geq \ln Q', \quad (40)$$

for $\Lambda H' = H'\Lambda$, or $H - H'$ small, or β small. The validity of Eq. (40) for large β is investigated separately. The operator H' plays the role of a trial Hamiltonian in the variational principle, Eqs. (39) and (40).

We now consider some preliminary relations that are to be used in the proof of Eq. (40).

We define vectors $|' i, n_1, \cdots, n_g\rangle$ in the extended Hilbert space $\mathcal{H}'$ as follows:

$$|' i, n_1, \cdots, n_g\rangle = \Sigma_\alpha \langle\alpha \mid i\rangle |' \alpha, n_1, \cdots, n_g\rangle. \quad (41)$$

The vectors $|i\rangle$ and $|\alpha\rangle$ are completely orthonormal sets in the physical Hilbert space $\mathcal{H}$, with $|i\rangle$ satisfying Eq. (23). The vectors $|\alpha\rangle$ were introduced prior to Eq. (25). Eq. (41) can be inverted to yield

$$|' \beta, n_1, \cdots, n_g\rangle = \Sigma_i \langle i \mid \beta\rangle |' i, n_1, \cdots, n_g\rangle. \quad (42)$$

From Eq. (31) and (41) we get

$$\langle' i', n_1', \cdots, n_g' \mid' i, n_1, \cdots, n_g\rangle = \delta_{i'i} \prod_{j=1}^{g} \delta_{n_j' n_j}. \quad (43)$$

In obtaining the last two equations we have used the completeness and orthonormality of the vector sets $|i\rangle$ and $|\alpha\rangle$. Since the vectors $|' \beta, n_1, \cdots, n_g\rangle$ are a complete set, Eq. (42) establishes the vectors $|' i, n_1, \cdots, n_g\rangle$ as a complete set. Equation (43) is a statement of the orthonormality of $|' i, n_1, \cdots, n_g\rangle$.

From Eqs. (35) and (41) we get

$$\Lambda |' i, n_1, \cdots, n_g\rangle = \prod_{j=1}^{g} \delta_{n_j 0} |' i, n_1, \cdots, n_g\rangle, \quad (44)$$

where $\delta_{00} = 1$ and $\delta_{n_j 0} = 0$ for $n_j \neq 0$, and from Eq. (23), (24), (32), (41), and the fact that the vectors $|\alpha\rangle$ and $|' \alpha, n_1 \cdots, n_g\rangle$ are complete orthonormal sets in the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ respectively, we get

$$H |' i, 0, \cdots, 0\rangle = E_i |' i, 0, \cdots, 0\rangle, \quad (45)$$

where the eigenvalues E_i are the same as in Eq. (23).

We make extensive use of properties of the trace operation in proving Eq. (40). We write down explicitly the following inequality:

For any operator A:

$$\text{Tr}\{A^\dagger A\} > 0, \quad \text{for} \quad A \neq 0, \quad \text{Tr}\{0\} = 0, \quad (46)$$

where 0 denotes the zero operator in the Hilbert space $\mathcal{H}'$.

Next we obtain an expression for the partition function Q in terms of operators in the extended

Hilbert space $\mathcal{H}'$. From Eqs. (22), (43)–(45) we get

$$Q = \mathrm{Tr}\,\{\Lambda e^{-\beta H}\}. \tag{47}$$

From Eqs. (43)–(45) we can show that

$$\Lambda H = H\Lambda. \tag{48}$$

From Eq. (48) it follows that Λ commutes with any function of the operator H, in particular,

$$\Lambda e^{-sH} = e^{-sH}\Lambda, \quad \text{for} \quad s \geq 0. \tag{49}$$

To complete the relations preliminary to proving Eq. (40) we define the operators

$$H(\epsilon) \equiv H' + \epsilon(H - H') \tag{50}$$

for real values of ϵ in the range $0 \leq \epsilon \leq 1$. From the Hermiticity of H and H' it then follows that $H(\epsilon)$ is Hermitian for all ϵ. The only restriction we put on the operator H' introduced in Eq. (38) is that

$$f(\epsilon) \equiv \mathrm{Tr}\,\{\Lambda e^{-\beta H(\epsilon)}\} \tag{51}$$

should exist and possess a continuous second derivative with respect to ϵ for $0 \leq \epsilon \leq 1$. It follows from the reality of β, the Hermiticity of $H(\epsilon)$, and Eq. (51) that $f(\epsilon)$ is real for $0 \leq \epsilon \leq 1$. Equation (50) can be rewritten as

$$H(\epsilon) = H - (1 - \epsilon)(H - H'). \tag{52}$$

Applying Taylor's theorem to $\ln f(\epsilon)$ we get

$$\begin{aligned}\ln f(1) = \ln f(0) &+ \{f'(0)/f(0)\} \\ &+ \int_0^1 (1 - \epsilon)[\{f''(\epsilon)/f(\epsilon)\} \\ &- \{f'(\epsilon)/f(\epsilon)\}^2]\, d\epsilon,\end{aligned} \tag{53}$$

where $f'(\epsilon)$ and $f''(\epsilon)$ denote the first and second derivatives, respectively, of $f(\epsilon)$ with respect to ϵ.

From Eqs. (50)–(52) we get

$$f'(\epsilon) = -\int_0^\beta \mathrm{Tr}\,\{\Lambda e^{-\gamma H(\epsilon)}(H - H')e^{-(\beta-\gamma)H(\epsilon)}\}\, d\gamma, \tag{54}$$

$$\begin{aligned}f''(\epsilon) = 2\int_0^\beta \int_0^\gamma \mathrm{Tr}\,\{&\Lambda e^{-\delta H(\epsilon)}(H - H') \\ &\times e^{-(\gamma-\delta)H(\epsilon)}(H - H')e^{-(\beta-\gamma)H(\epsilon)}\}\, d\delta\, d\gamma.\end{aligned} \tag{55}$$

Combining Eqs. (39), (47), (53), and (54) we get

$$\begin{aligned}\ln Q = \ln Q' + \int_0^1 (1 - \epsilon)[\{f''(\epsilon)/f(\epsilon)\} \\ - \{f'(\epsilon)/f(\epsilon)\}^2]\, d\epsilon.\end{aligned} \tag{56}$$

Using Eqs. (51), (54), and (55) we can show that

$$\begin{aligned}[2/f(\epsilon)]\int_0^\beta \int_0^\gamma \mathrm{Tr}\,\{&\Lambda e^{-\delta H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \\ &\times e^{-(\gamma-\delta)H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \\ &\times e^{-(\beta-\gamma)H(\epsilon)}\}\, d\delta\, d\gamma \\ &= \{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2.\end{aligned} \tag{57}$$

We now proceed directly to the proof of Eq. (40). First we consider the case $\Lambda H' = H'\Lambda$. From Eqs. (48) and (50) we get, for $\Lambda H' = H'\Lambda$,

$$\Lambda H(\epsilon) = H(\epsilon)\Lambda.$$

It follows from the above that, for $\beta \geq \gamma \geq \delta \geq 0$,

$$\begin{aligned}&\mathrm{Tr}\,\{\Lambda e^{-\delta H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \\ &\quad \times e^{-(\gamma-\delta)H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}]e^{-(\beta-\gamma)H(\epsilon)}\} \\ &= \mathrm{Tr}\,\{\Lambda e^{-\frac{1}{2}(\beta-\gamma+\delta)H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \\ &\quad \times e^{-(\gamma-\delta)H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \\ &\quad \times e^{-\frac{1}{2}(\beta-\gamma+\delta)H(\epsilon)}\}.\end{aligned}$$

Since H and H' are Hermitian operators, β and $f(\epsilon)$ are real and the operator Λ satisfies Eq. (36), the right-hand side of the above equation can be expressed as $\mathrm{Tr}\,\{A^\dagger A\}$, where

$$\begin{aligned}A \equiv e^{-\frac{1}{2}(\gamma-\delta)H(\epsilon)}&[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\} \\ &\times e^{-\frac{1}{2}(\beta-\gamma+\delta)H(\epsilon)}\Lambda.\end{aligned}$$

From the above results and Eqs. (46), (57) it follows that

$$\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2 \geq 0 \quad \text{for} \quad \Lambda H' = H'\Lambda. \tag{58}$$

Next we consider the case in which $H - H'$ is small. For $\Lambda H' = H'\Lambda$, Eq. (58) is valid, so we need only consider the case in which $\Lambda H' \neq H'\Lambda$. We see from Eq. (52) that for $H - H'$ small[5]

$$H(\epsilon) = H + O([H - H']) = H(1) + O([H - H']).$$

It then follows from the above and Eqs. (51), (54), and (55) that[5]

$$\begin{aligned}&\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2 \\ &\quad = \{f''(1)/f(1)\} - \{f'(1)/f(1)\}^2 + O([H - H']^3).\end{aligned} \tag{59}$$

We also have from Eqs. (52) and (57) that

$$\begin{aligned}&\{f''(1)/f(1)\} - \{f'(1)/f(1)\}^2 \\ &= [2/f(1)]\int_0^\beta \int_0^\gamma \mathrm{Tr}\,\{\Lambda e^{-\delta H}[H - H' + \{f'(1)/\beta f(1)\}] \\ &\quad \times e^{-(\gamma-\delta)H}[H - H' + \{f'(1)/\beta f(1)\}]e^{-(\beta-\gamma)H}\}\, d\delta\, d\gamma.\end{aligned}$$

Letting $\delta' = \gamma - \delta$ in the integral over δ, using Eqs. (36) and (49), performing an integration by parts in the integral over γ, and using the fact that H and H'

[5] The expression $O([H - H']^n)$, $n = 1, 2, \cdots$, is not intended to necessarily mean that $[H - H']^n$ appears in the expression being considered. The precise meaning is that if $H' = H - \lambda V$, with λ a positive number, then $O([H - H']^n)$ means $O(\lambda^n)$ for small λ.

are Hermitian operators and β and $f(\epsilon)$ are real we get

$$\{f''(1)/f(1)\} - \{f'(1)/f(1)\}^2 = [2/f(1)]\int_0^\beta (\beta - \gamma)\,\mathrm{Tr}\,\{A^\dagger A\}\,d\gamma, \quad (60)$$

where

$$A \equiv e^{-\frac{1}{2}\gamma H}[H - H' + \{f'(1)/\beta f(1)\}]\Lambda e^{-\frac{1}{2}(\beta-\gamma)H}.$$

We note that $A = 0$ implies $H'\Lambda = H\Lambda + \{f'(1)/\beta f(1)\}\Lambda$. Taking the adjoint of the above equation and noting that H, H', and Λ are Hermitian operators, β and $f(\epsilon)$, defined in Eq. (51), are real, and Eq. (48) holds we get

$$\Lambda H' = H'\Lambda.$$

Since we are assuming $\Lambda H' \neq H'\Lambda$ we must have $A \neq 0$. Then we get from Eqs. (46) and (60) that

$$\{f''(1)/f(1)\} - \{f'(1)/f(1)\}^2 > 0, \quad \text{for} \quad \Lambda H' \neq H'\Lambda. \quad (61)$$

From Eqs. (51), (52), (54), and (55) we see that[5] $\{f''(1)/f(1)\} - \{f'(1)/f(1)\}^2$ is $O([H - H']^2)$. Thus we can neglect terms $O([H - H']^3)$ in comparison to $\{f''(1)/f(1)\} - \{f'(1)/f(1)\}^2$ for small $H - H'$. This fact, together with Eqs. (59) and (61) gives us

$$\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2 > 0, \quad \text{for} \quad \Lambda H' \neq H'\Lambda \quad (62)$$

and $H - H'$ small.

Next consider the case in which β is small. Since Eq. (58) is valid for $\Lambda H' = H'\Lambda$, we need only consider the case $\Lambda H' \neq H'\Lambda$. Letting $\delta = s\beta$, $\gamma = t\beta$ in Eq. (57) we get

$$\beta^{-2}f(\epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2] = 2\int_0^1\int_0^t \mathrm{Tr}\,\{\Lambda e^{-\beta s H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \times e^{-\beta(t-s)H(\epsilon)}[H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \times e^{-\beta(1-t)H(\epsilon)}\}\,ds\,dt.$$

If we define the operators

$$K(\beta, s, t, \epsilon) \equiv \{\exp[\tfrac{1}{2}\beta(1 - s - t)H(\epsilon)]\}(H - H') \times \exp[-\tfrac{1}{2}\beta(1 - s - t)H(\epsilon)], \quad (63)$$

then we have

$$\beta^{-2}f(\epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2] = 2\int_0^1\int_0^t \mathrm{Tr}\,\{\Lambda(\exp[-\tfrac{1}{2}\beta(1 + s - t)H(\epsilon)]) \times [K + \{f'(\epsilon)/\beta f(\epsilon)\}] \times (\exp[-\beta(t - s)H(\epsilon)])[K + \{f'(\epsilon)/\beta f(\epsilon)\}] \times \exp[-\tfrac{1}{2}\beta(1 + s - t)H(\epsilon)]\}\,ds\,dt. \quad (64)$$

From Eq. (63) we see that

$$\lim_{\beta\to 0} K(\beta, s, t, \epsilon) = H - H'. \quad (65)$$

If we let $K = H - H'$ in the right-hand side of Eq. (64) and use Eqs. (36), (38) and the fact that H is Hermitian and $f'(\epsilon)/\beta f(\epsilon)$ is real, then we get

$$\{\beta^{-2}f(\epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2]\}_{K=H-H'} = 2\int_0^1\int_0^t \mathrm{Tr}\,\{A^\dagger A\}\,ds\,dt,$$

where

$$A \equiv (\exp[-\tfrac{1}{2}\beta(t - s)H(\epsilon)]) \times [H - H' + \{f'(\epsilon)/\beta f(\epsilon)\}] \times (\exp[-\tfrac{1}{2}\beta(1 + s - t)H(\epsilon)])\Lambda.$$

We see that

$$[A]_{s=0,t=1} = 0 \Rightarrow H'\Lambda = H\Lambda + \{f'(\epsilon)/\beta f(\epsilon)\}\Lambda.$$

Using reasoning similar to that following Eq. (60) we get

$$\Lambda H' = H'\Lambda.$$

Since we are considering the case $\Lambda H' \neq H'\Lambda$, $[A]_{s=0,t=1} = 0$ is impossible, so that $[A]_{s=0,t=1} \neq 0$, which implies from Eq. (46) that

$$[\mathrm{Tr}\,\{A^\dagger A\}]_{s=0,t=1} > 0.$$

If we assume that $\mathrm{Tr}\,\{A^\dagger A\}$ is a continuous function of s and t for $0 \leq s \leq t \leq 1$, then the above inequality and Eq. (46) lead to

$$2\int_0^1\int_0^t \mathrm{Tr}\,\{A^\dagger A\}\,ds\,dt > 0,$$

so that

$$\{\beta^{-2}f(\epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2]\}_{K=H-H'} > 0. \quad (66)$$

If we assume that

$$\beta^{-2}f(\epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2] \sim \{\beta^{-2}f(\epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2]\}_{K=\lim_{\beta\to 0} K}, \quad (67)$$

then we get, from Eq. (65)–(67) and the fact that $\beta \equiv 1/kT > 0$ and $f(\epsilon) > 0$ from Eq. (51), that

$$\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2 > 0, \quad \text{for} \quad \Lambda H' \neq H'\Lambda \quad (68)$$

and β small.

From Eqs. (58), (62), and (68) we get

$$\int_0^1 (1 - \epsilon)[\{f''(\epsilon)/f(\epsilon)\} - \{f'(\epsilon)/f(\epsilon)\}^2]\,d\epsilon \geq 0, \quad (69)$$

for $\Lambda H' = H'\Lambda$, or $H - H'$ small, or β small. Finally, by combining Eqs. (56) and (69) we obtain Eq. (40) for $\Lambda H' = H'\Lambda$, or $H - H'$ small, or β small.

We now investigate the conditions under which Eq. (40) is valid for large β. In order for the partition function Q of Eqs. (18) or (23) to exist we see that the energy eigenvalues E_i of the physical system must have a lowest value E_0 with a finite degeneracy. It then follows from Eq. (22) that

$$\lim_{\beta\to\infty}\beta^{-1}\ln Q = -E_0. \tag{70}$$

We consider first the case in which the Hermitian operator H' of Eq. (38) possesses a complete orthonormal set $|'j\rangle$ of eigenvectors, where j takes on a set of integer values. Then we generalize our result. Thus,

$$H'|'j\rangle = E_j'|'j\rangle, \quad \langle' k|'j\rangle = \delta_{kj}, \tag{71}$$

where $\delta_{jj}=1$, $\delta_{kj}=0$ for $k\neq j$, and

$$\mathrm{Tr}\{\Lambda e^{-\beta H'}\} = \Sigma_j\langle' j|\Lambda|'j\rangle e^{-\beta E_j'}. \tag{72}$$

Assume that there exists a lowest value of E_j', E_0', such that for some j for which $E_j'=E_0'$ $\Lambda|'j\rangle\neq 0$. Then, since $\langle' j|\Lambda|'j\rangle = \langle' j|\Lambda|\Lambda|'j\rangle$ from Eq. (36), we see that $\langle' j|\Lambda|'j\rangle\geq 0$, with $\langle' j|\Lambda|'j\rangle>0$ for $\Lambda|'j\rangle\neq 0$. Using Eqs. (39), (70)–(72) we get

$$\lim_{\beta\to\infty}\beta^{-1}[\ln Q - \ln Q'] = \sum_{E_j'=E_k'=E_0'}\sum_j\sum_k \langle' j|\Lambda|'k\rangle\langle' k|H-E_0|'j\rangle \Big/ \sum_{E_j'=E_0'}\sum_j\langle' j|\Lambda|'j\rangle. \tag{73}$$

Consider the matrix

$$M_{jk}\equiv\langle' j|\Lambda|'k\rangle, \quad E_j'=E_0', \quad E_k'=E_0'.$$

Since Λ is a positive Hermitian operator, M_{jk} is a positive Hermitian matrix. We can choose the eigenvectors of H' corresponding to the eigenvalue E_0' so that M_{jk} is diagonal, i.e.,

$$\langle' j|\Lambda|'k\rangle = \lambda_j\delta_{jk}. \tag{74}$$

From the remark following Eq. (72) we see that

$$\lambda_i\geq 0, \quad \text{with} \quad \lambda_i>0 \quad \text{for some} \quad i. \tag{75}$$

Combining Eqs. (73) and (74), we get

$$\lim_{\beta\to\infty}\beta^{-1}[\ln Q - \ln Q'] = \sum_{E_i'=E_0'}\lambda_i\langle' i|H-E_0|'i\rangle \Big/ \sum_{E_i'=E_0'}\lambda_i. \tag{76}$$

Any vector $|'\rangle$ in the extended Hilbert space $\mathcal{H}'$ can be expanded in terms of the complete orthonormal set $|'\beta, n_1, \cdots, n_g\rangle$ of Eq. (30) as follows:

$$|'\rangle = \Sigma_{\beta,n_1,\cdots,n_g}\langle\beta', n_1,\cdots,n_g|'\rangle\,|'\beta, n_1,\cdots,n_g\rangle. \tag{77}$$

From Eqs. (31), (32), (77) and the fact that $|\beta\rangle$ is an orthonormal set we get

$$\langle'|H-E_0|'\rangle = \Sigma_{n_1,\cdots,n_g}\langle n_1,\cdots,n_g|H(hn_1/L_1,\cdots,hn_g/L_g) - E_0|n_1,\cdots,n_g\rangle, \tag{78}$$

where

$$|n_1,\cdots,n_g\rangle \equiv \Sigma_\beta\langle'\beta, n_1,\cdots,n_g|'\rangle\,|\beta\rangle. \tag{79}$$

From Eq. (24) and the fact that E_0 is the lowest eigenvalue of H_0 we see that, for any vector $|\ \rangle$ in the physical Hilbert space $\mathcal{H}$,

$$\langle|H(0,\cdots,0)-E_0|\rangle\geq 0. \tag{80}$$

If we also assume that, for the operators $H(y_1,\cdots,y_g)$ in the physical Hilbert space $\mathcal{H}$, defined after Eq. (23),

$$\langle|H(y_1,\cdots,y_g)-E_0|\rangle\geq 0, \quad \forall\ (y_1,\cdots,y_g)\neq(0,\cdots,0), \tag{81}$$

for any vector $|\ \rangle$ in $\mathcal{H}$, then we get from Eqs. (78), (80), and (81) that

$$\langle'|H-E_0|'\rangle\geq 0, \quad \text{for any} \quad |'\rangle, \tag{82}$$

if Eq. (81) holds in the extended Hilbert space $\mathcal{H}'$. By considering small variations from a vector $|'\rangle$ that produces equality in Eq. (82) we can show that

$$\langle'|H-E_0|'\rangle = 0 \quad \text{only if} \quad H|'\rangle = E_0|'\rangle. \tag{83}$$

Combining Eqs. (75), (76), (82), and (83) we see that

$$\lim_{\beta\to\infty}\beta^{-1}[\ln Q - \ln Q'] > 0 \tag{84}$$

if Eq. (81) holds and no linear combination of $|'j\rangle$ for $E_j'=E_0'$ is an eigenvector of H. Equation (84) establishes Eq. (40) for large β.

Equation (84) can be generalized to the case in which the eigenvectors $|'j\rangle$ do not form a complete set, but must be augmented by continuum eigenvectors to obtain a complete set. If all continuum eigenvalues are greater than the E_0' defined after Eq. (72), then Eq. (73) can be shown to hold as before, and the remaining steps in the derivation of Eq. (84) are the same as before.

The important point here is that Eq. (81) must be assumed. For the case of the electron gas Eq. (81) is equivalent to Eq. (11), whose validity we saw had to be assumed to use Eq. (12) to compute the ground-state energy of the electron gas.

If E_0' is a nondegenerate eigenvalue of H', with $|'j_0\rangle$ the corresponding normalized eigenvector, then, noting Eq. (70), Eq. (73) reduces to

$$\lim_{\beta\to\infty}\beta^{-1}\ln Q' = -\langle' j_0|H|'j_0\rangle/\langle' j_0|'j_0\rangle.$$

We also note from the discussion following Eq. (72) that we must have $\Lambda\,|'j_0\rangle \neq 0$. The restriction on $|'j_0\rangle$ for $E'_j = E'_0$ of Eq. (84) reduces in this case to $|'j_0\rangle$ not being an eigenvector of H. Noting the above results and Eq. (70) we see that Eq. (84) reduces to the following:

$$\lim_{\beta\to\infty} \beta^{-1} \ln Q = -E_0 > -\langle' j_0|\, H\,|'j_0\rangle / \langle' j_0\,|\,'j_0\rangle, \quad (85)$$

if Eq. (81) holds, $\Lambda\,|'j_0\rangle \neq 0$, and $|'j_0\rangle$ is not an eigenvector of H.

Since Eq. (85) is unaltered when $|'j_0\rangle$ is replaced by $C\,|'j_0\rangle$, where C is any nonzero complex constant, we can remove the restriction that $\langle' j_0\,|\,'j_0\rangle = 1$. The inequality of Eq. (85) is just the standard variational principle for the lowest eigenvalue of H when $|'j_0\rangle$ is not an eigenvector of H; this lowest eigenvalue of H is equal to E_0, the lowest eigenvalue of H_0 when Eq. (81) is assumed to be true. This completes our investigation of the validity of the variational principle for $\ln Q$, Eqs. (39) and (40).

III. CONCLUDING REMARKS

Just how useful is the variational principle [Eqs. (39) and (40)] that we have derived? This question can be divided into two subquestions: (i) Assuming we can evaluate $\ln Q'$ of Eq. (39), how useful is Eq. (40)? (ii) How can we evaluate $\ln Q'$? We answer (i) first. Equation (40) would be most useful if it held for all values of the trial Hamiltonian H'. Then we would choose a set of trial Hamiltonians depending on certain parameters for which $\ln Q'$ can be evaluated, maximize $\ln Q'$ with respect to these parameters, and obtain the highest lower bound to $\ln Q$ obtainable from the trial Hamiltonians. Our results are not inconsistent with Eq. (40) holding for all values of H', and it may hold in some applications of the auxiliary-variable method, but we cannot rely on Eq. (40) being valid for all values of H' in general. Our variational principle would be least useful if $\ln Q$ were just a stationary value of $\ln Q'$ for $H' = H$. Then all we could say was that, for H' a good approximation to H, $\ln Q'$ would be a very good approximation to $\ln Q$ with an error that was[5] $O([H - H']^2)$, but we could not say whether $\ln Q'$ was greater than or less than $\ln Q$. The actual usefulness of Eq. (40) lies somewhere between these two extremes. We can proceed as if Eq. (40) did hold for all values of H' and find that choice of H' that maximizes $\ln Q'$. We must then examine this choice of H' to see if it appeared to be a good approximation to H. If this is the case we can not only say that $\ln Q'$ is a very good approximation to $\ln Q$, but could also be reasonably certain that $\ln Q'$ was a lower bound to $\ln Q$.

Now we turn our attention to the evaluation of $\ln Q'$ of Eq. (39). The detailed evaluation of $\ln Q'$ clearly depends on the specific problem to which the auxiliary variable method is applied. All we do here is sketch one possible method of evaluation, stating the assumptions we make but not examining their validity, to show that the variational principle, Eqs. (39) and (40), is not devoid of computational usefulness. We begin by substituting Eq. (37) into Eq. (39) to obtain

$$\begin{aligned}\ln Q' = \ln \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \cdots \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} \mathrm{Tr} \left\{ \left[\exp\left(i\hbar^{-1} \sum_{j=1}^{g} \xi_j Y_j \right) \right] \right. \\ \left. \times e^{-\beta H'} \right\} d\xi_g \cdots d\xi_2\, d\xi_1 - \sum_{j=1}^{g} \ln L_j \\ - \left[\int_0^{\beta} \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} \mathrm{Tr} \left\{ \left[\exp\left(i\hbar^{-1} \right. \right. \right. \right. \\ \left. \left. \left. \left. \times \sum_{j=1}^{g} \xi_j Y_j \right) \right] e^{-\gamma H'} (H - H') e^{-(\beta-\gamma)H'} \right. \\ \times d\xi_g \cdots d\xi_2\, d\xi_1\, d\gamma \Big/ \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \cdots \\ \times \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} \mathrm{Tr} \left\{ \left[\exp\left(i\hbar^{-1} \sum_{j=1}^{g} \xi_j Y_j \right) \right] e^{-\beta H'} \right\} \\ \left. \times d\xi_g \cdots d\xi_2\, d\xi_1 \right]. \end{aligned} \quad (86)$$

Next we examine the dependence of the function

$$f(\xi_1, \cdots, \xi_g) \equiv \mathrm{Tr} \left\{ \left[\exp\left(i\hbar^{-1} \sum_{j=1}^{g} \xi_j Y_j \right) \right] e^{-\beta H'} \right\} \quad (87)$$

on $\xi_1, \cdots, \xi_g$.

This function has the property that

$$\begin{aligned} &|f(\xi_1, \cdots, \xi_g)| < |f(0, \cdots, 0)|, \quad \text{for} \\ &\quad (\xi_1, \cdots, \xi_g) \neq (0, \cdots, 0) \quad \text{and} \\ &-\tfrac{1}{2}L_1 \leq \xi_1 \leq \tfrac{1}{2}L_1, \cdots, -\tfrac{1}{2}L_g \leq \xi_g \leq \tfrac{1}{2}L_g. \end{aligned} \quad (88)$$

This can be seen by evaluating the trace on the right-hand side of Eq. (87) in terms of the complete orthonormal set $|'\beta, n_1, \cdots, n_g\rangle$ of Eq. (30), using Eq. (33).

Equation (88) suggests the following method for evaluating the integral

$$I_1 \equiv \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1} \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} \cdots \int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} f(\xi_1, \cdots, \xi_g)\, d\xi_g \cdots d\xi_2\, d\xi_1$$

which appears in Eq. (86):

$|f(\xi_1, \cdots, \xi_g)|$ assumes its maximum value in the range of integration at $(\xi_1, \cdots, \xi_g) = (0, \cdots, 0)$ and nowhere else. Let us assume that we obtain the major contribution to the integral I_1 from values of $(\xi_1, \cdots, \xi_g)$ near $(0, \cdots, 0)$, and that for these values we can, to good accuracy, expand $\ln f(\xi_1, \cdots, \xi_g)$

in a Taylor series about $(0, \cdots, 0)$, keeping only the first few terms, i.e.,

$$\ln f(\xi_1, \cdots, \xi_g) \approx \ln f(0, \cdots, 0) + i\sum_{j=1}^{g} a_j\xi_j - \frac{1}{2}\sum_{j,k=1}^{g} b_{jk}\xi_j\xi_k. \quad (89)$$

From Eq. (87) and (89) we see that

$$a_j = \hbar^{-1}\,\mathrm{Tr}\,\{Y_j e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\}, \quad (90)$$

$$b_{jk} = \hbar^{-2}[\mathrm{Tr}\,\{Y_jY_k e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\} - \mathrm{Tr}\,\{Y_j e^{-\beta H'}\} \times \mathrm{Tr}\,\{Y_k e^{-\beta H'}\}/(\mathrm{Tr}\,\{e^{-\beta H'}\})^2]. \quad (91)$$

From Eq. (91) we see that a_j is real for all j, so that the effect on $f(\xi_g, \cdots, \xi_g)$ of nonzero values of a_j is, from Eq. (89), to produce a phase variation in $f(\xi_1, \cdots, \xi_g)$ with no effect on its magnitude. Such a phase variation tends to cause cancellation of the contribution to the integral over $f(\xi_1, \cdots, \xi_g)$, I_1, from values of $(\xi_1, \cdots, \xi_g)$ near $(0, \cdots, 0)$ and is undesirable. We therefore assume that

$$a_1 = a_2 = \cdots = a_g = 0. \quad (92)$$

From Eqs. (90)–(92) we see that the matrix b_{jk} is real, symmetric, and positive-definite, i.e.,

$$b_{jk}^* = b_{jk} = b_{kj}, \quad \text{matrix } b_{jk} > 0. \quad (93)$$

Combining Eqs. (89), (92), and (93) we get for sufficiently large $L_1, L_2, \cdots, L_g$ that

$$I_1 \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(0, \cdots, 0) \times \exp\left(-\frac{1}{2}\sum_{j,k=1}^{g} b_{jk}\xi_j\xi_k\right) d\xi_g \cdots d\xi_2\, d\xi_1 = [(2\pi)^g/\det\{b_{jk}\}]^{\frac{1}{2}} f(0, \cdots, 0)$$

or, using Eq. (87),

$$\int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1}\int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2}\cdots\int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g}\mathrm{Tr}\left\{\left[\exp\left(i\hbar^{-1}\sum_{j=1}^{g}\xi_jY_j\right)\right]e^{-\beta H'}\right\} \times d\xi_g \cdots d\xi_2\, d\xi_1 \approx [(2\pi)^g/\det\{b_{jk}\}]^{\frac{1}{2}}\,\mathrm{Tr}\,\{e^{-\beta H'}\}, \quad (94)$$

where b_{jk} are given by Eq. (91).

We now turn our attention to the evaluation of the integral which appears in Eq. (86):

$$I_2(\gamma) \equiv \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1}\int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2}\cdots\int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g} g(\xi_1, \cdots, \xi_g; \gamma) \times d\xi_g \cdots d\xi_2\, d\xi_1,$$

where $0 \le \gamma \le \beta$ and

$$g(\xi_1, \cdots, \xi_g; \gamma) \equiv \mathrm{Tr}\left\{\left[\exp\left(i\hbar^{-1}\sum_{j=1}^{g}\xi_jY_j\right)\right]e^{-\gamma H'}(H - H')e^{-(\beta-\gamma)H'}\right\}. \quad (95)$$

From Eqs. (87) and (95) we get

$$g(0, \cdots, 0; \gamma)/f(0, \cdots, 0) = \mathrm{Tr}\,\{(H - H')e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\}. \quad (96)$$

We assume that the main contribution to the integral $I_2(\gamma)$ comes from the same region considered for I_1 of values of $(\xi_1, \cdots, \xi_g)$ near $(0, \cdots, 0)$. We further assume that $g(\xi_1, \cdots, \xi_g; \gamma)/f(\xi_1, \cdots, \xi_g)$ is a slowly varying function of $\xi_1, \cdots, \xi_g$ in this region so that

$$g(\xi_1, \cdots, \xi_g; \gamma)/f(\xi_1, \cdots, \xi_g) \approx g(0, \cdots, 0; \gamma)/f(0, \cdots, 0).$$

From the above and Eq. (96) we get

$$I_2(\gamma) \approx [\mathrm{Tr}\,\{(H - H')e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\}]I_1,$$

or, using Eq. (87) and (95),

$$\int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1}\int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2}\cdots\int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g}\mathrm{Tr}\left\{\left[\exp\left(i\hbar^{-1}\sum_{j=1}^{g}\xi_jY_j\right)\right] \times e^{-\gamma H'}(H - H')e^{-(\beta-\gamma)H'}\, d\xi_g \cdots d\xi_2\, d\xi_1 \approx [\mathrm{Tr}\,\{(H - H')e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\}] \times \int_{-\frac{1}{2}L_1}^{\frac{1}{2}L_1}\int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2}\cdots\int_{-\frac{1}{2}L_g}^{\frac{1}{2}L_g}\mathrm{Tr}\left\{\left[\exp\left(i\hbar^{-1} \times \sum_{j=1}^{g}\xi_jY_j\right)\right]e^{-\beta H'}\right\} d\xi_g \cdots d\xi_2\, d\xi_1. \quad (97)$$

Substituting Eqs. (94) and (97) into Eq. (86) we finally obtain the result we set out to achieve:

$$\ln Q' \approx \ln \mathrm{Tr}\,\{e^{-\beta H'}\} - \tfrac{1}{2}\ln\det\{b_{jk}\} - \sum_{j=1}^{g}\ln([2\pi]^{-\frac{1}{2}}L_j) - \beta[\mathrm{Tr}\,\{(H - H')e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\}], \quad (98)$$

where b_{jk} are given by Eq. (91).

What we have done in going from Eqs. (39) and (40) to the approximate equation (98) is to approximate the effect of the projection operator Λ in Eq. (39). It is the presence of this projection operator that makes the trace formulas in Eq. (39) more complicated than the trace formulas usually encountered in quantum statistical mechanics. The trace formulas in Eq. (98), on the other hand, are of the types usually encountered in quantum statistical mechanics. Explicitly, $\mathrm{Tr}\,\{e^{-\beta H'}\}$ is the partition function for a system with Hamiltonian operator H' at temperature $T = 1/k\beta$, and b_{jk} and

$$\mathrm{Tr}\,\{(H - H')e^{-\beta H'}\}/\mathrm{Tr}\,\{e^{-\beta H'}\}$$

are expectation values of the operators $\hbar^{-2}Y_jY_k$ and $(H - H')$, respectively.

The first and last terms on the right-hand side of Eq. (98) are just the right-hand side of Eq. (39) with the replacement $\Lambda \to 1$, i.e., differences between the physical and extended Hilbert spaces are completely ignored. The middle terms on the right-hand side of Eq. (98) are corrections arising from these differences between the physical and extended Hilbert spaces. The validity of the assumptions used to derive Eq. (98) and the accuracy of this approximate equation depends on the specific problem being dealt with, and is not considered further in this paper.

The usefulness of a variational principle is enhanced if corrections to the variational result can be obtained by some nonvariational methods, such as perturbation theory. By expanding $\ln f(\epsilon)$, where $f(\epsilon)$ is defined in Eqs. (50) and (51), as a power series in ϵ and setting $\epsilon = 1$ we see from Eqs. (47), (50), and (51) that we obtain a perturbation expansion for $\ln Q$ in which H' is the unperturbed Hamiltonian operator in the extended Hilbert space and $H - H'$ is the perturbation on it. The first two terms on the right-hand side of Eq. (53) are the zero- and first-order perturbations. Combining Eqs. (50), (51), and (54) we see that these terms are just $\ln Q'$ of Eq. (39), the variational approximation to $\ln Q$ [see Eq. (40)]. Therefore, the second- and higher- order terms in the perturbation expansion of $\ln Q$ are corrections to the variational result.

We can see the relationship of the variational principle, Eqs. (39) and (40), to earlier variational principles by considering limiting cases of Eq. (40). We have already considered the zero-temperature ($\beta \to \infty$) limit in Sec. II, Eqs. (70)–(77). For the case of the electron gas, Eq. (81) is equivalent to Eq. (11) and Eq. (85) to Eq. (12) when we take the limit as the auxiliary-variable periods $L_1, L_2, \cdots, L_g \to \infty$.

Another limiting case of Eqs. (39) and (40) of interest is the "no auxiliary variable" limit. In this limit the extended Hilbert space $\mathcal{H}'$ is identical to the physical Hilbert space $\mathcal{H}$, $H(y_1, \cdots, y_g) = H(0, \cdots, 0) = H_0$, using Eq. (24), and $\Lambda = 1$, using Eq. (35) with

$$|' \beta, n_1, \cdots, n_g\rangle = |' \beta, 0, \cdots, 0\rangle = |\beta\rangle.$$

Thus Eq. (40) holds in all cases, since we always have $\Lambda H' = H'\Lambda$, and the proof of Eq. (40) reduces to the proof of Eq. (58) plus Eq. (56). Equation (39) reduces to the following:

$$\ln Q' = \ln \operatorname{Tr} \{e^{-\beta H'}\} - \beta \operatorname{Tr} \{(H - H')e^{-\beta H'}\}/\operatorname{Tr} \{e^{-\beta H'}\}.$$

Combining the above with Eq. (40) we obtain

$$\ln Q \geq \ln \operatorname{Tr} \{e^{-\beta H'}\} - \beta \operatorname{Tr} \{(H - H')e^{-\beta H'}\}/\operatorname{Tr} \{e^{-\beta H'}\} \quad (99)$$

as the "no auxiliary-variable" limit of the variational principle developed in this paper, Eqs. (39) and (40). Equation (99) is a modified form of Peierls' variational theorem.[6]

The variational principle, Eqs. (39) and (40), can be used to compute the logarithm of the partition function, $\ln Q$, of any physical system containing charged particles in which the long-range collective effects of the Coulomb interaction are important, such as the electron gas, as follows: Let us first confine the physical system to a cubical box of volume V subject to periodic boundary conditions, just as we did for the electron gas in Sec. I. Next, obtain the Hamiltonian operators for our system in the presence of external electrostatic fields of the form given in Eqs. (5) and (6), being sure to include the electrostatic-field energy in the cubical box. These Hamiltonian operators will be functions of the complex parameters $\{\beta_k\}$ of Eqs. (5) and (6). Now introduce the parameters

$$y_{\mathbf{k}} \equiv \tfrac{1}{2}[\beta_{\mathbf{k}} + \beta_{-\mathbf{k}} - i(\beta_{\mathbf{k}} - \beta_{-\mathbf{k}})]. \quad (100)$$

Solving the above equations for $\beta_{\mathbf{k}}$ we get

$$\beta_{\mathbf{k}} = \tfrac{1}{2}[y_{\mathbf{k}} + y_{-\mathbf{k}} + i(y_{\mathbf{k}} - y_{-\mathbf{k}})]. \quad (101)$$

From Eqs. (6) and (100) we can show that the parameters $y_{\mathbf{k}}$ are real and independent. From Eq. (101) we see that the Hamiltonian operators of the physical system in the presence of the external electrostatic fields of Eq. (5) and (6) are functions of the independent real parameters $\{y_{\mathbf{k}}\}$:

$$H\{y_{\mathbf{k}}\}.$$

Finally, we identify the parameters $y_1, \cdots, y_g$ and the Hermitian operators $H(y_1, \cdots, y_g)$ of Eq. (24) and the discussion preceding Eq. (24) with $\{y_{\mathbf{k}}\}$ and $H\{y_{\mathbf{k}}\}$, respectively, i.e.,

$$y_1, \cdots, y_g \leftrightarrow \{y_{\mathbf{k}}\}, \quad H(y_1, \cdots, y_g) \leftrightarrow H\{y_{\mathbf{k}}\}. \quad (102)$$

The above identification enables us to apply the results of Sec. II to the specific physical system we have considered. If we assume (C) is valid for our physical system, then Eq. (81) holds and, according to (84), the variational principle, Eqs. (39) and (40), are valid for low temperatures.

[6] H. Falk, Physica **29**, 1114 (1963).

The application of the variational principle, Eqs. (39) and (40), to the electron gas at nonzero temperatures is presently being studied by L. Chan and the author. An investigation of the zero-temperature limit of Eqs. (39) and (40), namely Eq. (85), as applied to the electron gas at metallic densities was carried out by Bhatia and the author.[7]

Note added in proof: A recently completed Ph.D. dissertation (Lorenzo C. Chan, "Computational Feasibility of a Variational Principle in an Extended Hilbert Space," Florida State University, 1968) contains results of the study by L. Chan and the author mentioned in the previous paragraph. Numerical results were obtained for some physically reasonable choices of the trial Hamiltonian H'. To obtain a result of sufficient accuracy to be useful, however, one would have to choose a more optimal H' and then compute perturbative corrections to the resulting variational bound to the partition function.

[7] M. S. Bhatia and G. Speisman, Phys. Rev. **136**, A362 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

First-Order Phase Transitions in Quantum-Coulomb Plasmas

GIDEON CARMI
Belfer Graduate School of Science, Yeshiva University, New York

(Received 3 January 1968)

A simplified model is suggested for the understanding (in principle) of the mechanism of a phase transition in a Coulomb system on a uniform neutralizing background. Quantum theory is taken into account only in so far as it provides discrete collective energy levels and, possibly, fermion statistics for the calculation of the collective modes. Other features stemming from the uncertainty relations are supposed to be irrelevant to the mechanism of the phase transition. (This is qualitatively justified in the text which goes with Fig. 1.) This procedure provides an "effective quantum Hamiltonian" $H_{\rm qu}$, which incorporates the relevant quantum features and from which one can calculate the (approximate) quantum partition function using classical methods:

$$Z^{\rm Qu}_{\rm approx} = \frac{1}{N!\,h^N}\int\cdots\int \exp\left(-\beta H_{\rm qu}\right) d^3x_1\cdots d^3p_N .$$

In the evaluation of this integral we use the same approximation in which the plasma modes are collective, viz., the RPA (random phase approximation). Because of the freezing-out of the collective degrees of freedom at the relevant temperatures, and because the number of these degrees of freedom changes with temperature and density, $H_{\rm qu}$ describes a system with variable degrees of freedom, which seems to provide the mechanism for the phase transition. Preliminary numerical evaluations indicate a phase transition at $T = 0$ for an electron plasma at $r_s = 7.6$, i.e., just below the region of metallic densities. This is shown by finding a concave region in the free energy as a function of the density. The ground is then prepared for numerical evaluation of the transition at $T \neq 0$. For white dwarfs we find a transition temperature of $10^7 \sim$ K°.

1. INTRODUCTION

Current methods for calculating melting temperatures and densities for real metals,[1] for electron plasmas on a uniform positive background,[2–10] or ion plasmas on a uniform negative background[6,11,12] are, at best, of semiphenomenological nature.

In order to derive a first-order phase transition from first principles, one should start out from the calculation of the partition function and the free energy; assuming a single homogeneous phase, a region of concavity in the free energy as a function of the density then indicates an instability towards the formation of another phase. In these calculations it becomes necessary to work quantum mechanically.

Quantum-mechanical calculations for the partition function of a Coulomb system have been carried out

[1] D. Pines, *Elementary Excitations in Solids* (W. A. Benjamin, Inc., New York, 1963), p. 34.
[2] E. P. Wigner, Trans. Faraday Soc. **34**, 678 (1938).
[3] R. A. Coldwell-Horsfall and A. A. Maradudin, J. Math. Phys. **1**, 395 (1960).
[4] D. Noziéres and D. Pines, Phys. Rev. **111**, 442 (1958).
[5] F. W. de Wette, Phys. Rev. **135**, A287 (1964).
[6] H. M. Van Horn, Phys. Rev. **157**, 342 (1967).
[7] S. G. Brush, H. L. Sahlin, and E. Teller, J. Chem. Phys. **45**, 2101 (1966).
[8] S. Gartenhaus and G. Stranahan, Phys. Rev. Letters **14**, 341 (1965); **15**, 621 (1965).
[9] Y. Osaka, J. Phys. Soc. Jap. **22**, 1513 (1967).
[10] N. Wiser and M. H. Cohen (private communication).
[11] H. M. Van Horn, Astrophys. J. **151**, 227 (1968).
[12] L. Mestel and M. A. Ruderman (manuscript in preparation).

The application of the variational principle, Eqs. (39) and (40), to the electron gas at nonzero temperatures is presently being studied by L. Chan and the author. An investigation of the zero-temperature limit of Eqs. (39) and (40), namely Eq. (85), as applied to the electron gas at metallic densities was carried out by Bhatia and the author.[7]

Note added in proof: A recently completed Ph.D. dissertation (Lorenzo C. Chan, "Computational Feasibility of a Variational Principle in an Extended Hilbert Space," Florida State University, 1968) contains results of the study by L. Chan and the author mentioned in the previous paragraph. Numerical results were obtained for some physically reasonable choices of the trial Hamiltonian H'. To obtain a result of sufficient accuracy to be useful, however, one would have to choose a more optimal H' and then compute perturbative corrections to the resulting variational bound to the partition function.

[7] M. S. Bhatia and G. Speisman, Phys. Rev. **136**, A362 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

First-Order Phase Transitions in Quantum-Coulomb Plasmas

GIDEON CARMI
Belfer Graduate School of Science, Yeshiva University, New York

(Received 3 January 1968)

A simplified model is suggested for the understanding (in principle) of the mechanism of a phase transition in a Coulomb system on a uniform neutralizing background. Quantum theory is taken into account only in so far as it provides discrete collective energy levels and, possibly, fermion statistics for the calculation of the collective modes. Other features stemming from the uncertainty relations are supposed to be irrelevant to the mechanism of the phase transition. (This is qualitatively justified in the text which goes with Fig. 1.) This procedure provides an "effective quantum Hamiltonian" $H_{\rm qu}$, which incorporates the relevant quantum features and from which one can calculate the (approximate) quantum partition function using classical methods:

$$Z^{\rm Qu}_{\rm approx} = \frac{1}{N!\,h^N}\int\cdots\int \exp\left(-\beta H_{\rm qu}\right) d^3x_1\cdots d^3p_N .$$

In the evaluation of this integral we use the same approximation in which the plasma modes are collective, viz., the RPA (random phase approximation). Because of the freezing-out of the collective degrees of freedom at the relevant temperatures, and because the number of these degrees of freedom changes with temperature and density, $H_{\rm qu}$ describes a system with variable degrees of freedom, which seems to provide the mechanism for the phase transition. Preliminary numerical evaluations indicate a phase transition at $T = 0$ for an electron plasma at $r_s = 7.6$, i.e., just below the region of metallic densities. This is shown by finding a concave region in the free energy as a function of the density. The ground is then prepared for numerical evaluation of the transition at $T \neq 0$. For white dwarfs we find a transition temperature of $10^7 \sim$ K°.

1. INTRODUCTION

Current methods for calculating melting temperatures and densities for real metals,[1] for electron plasmas on a uniform positive background,[2–10] or ion plasmas on a uniform negative background[6,11,12] are, at best, of semiphenomenological nature.

In order to derive a first-order phase transition from first principles, one should start out from the calculation of the partition function and the free energy; assuming a single homogeneous phase, a region of concavity in the free energy as a function of the density then indicates an instability towards the formation of another phase. In these calculations it becomes necessary to work quantum mechanically.

Quantum-mechanical calculations for the partition function of a Coulomb system have been carried out

[1] D. Pines, *Elementary Excitations in Solids* (W. A. Benjamin, Inc., New York, 1963), p. 34.
[2] E. P. Wigner, Trans. Faraday Soc. **34**, 678 (1938).
[3] R. A. Coldwell-Horsfall and A. A. Maradudin, J. Math. Phys. **1**, 395 (1960).
[4] D. Noziéres and D. Pines, Phys. Rev. **111**, 442 (1958).
[5] F. W. de Wette, Phys. Rev. **135**, A287 (1964).
[6] H. M. Van Horn, Phys. Rev. **157**, 342 (1967).
[7] S. G. Brush, H. L. Sahlin, and E. Teller, J. Chem. Phys. **45**, 2101 (1966).
[8] S. Gartenhaus and G. Stranahan, Phys. Rev. Letters **14**, 341 (1965); **15**, 621 (1965).
[9] Y. Osaka, J. Phys. Soc. Jap. **22**, 1513 (1967).
[10] N. Wiser and M. H. Cohen (private communication).
[11] H. M. Van Horn, Astrophys. J. **151**, 227 (1968).
[12] L. Mestel and M. A. Ruderman (manuscript in preparation).

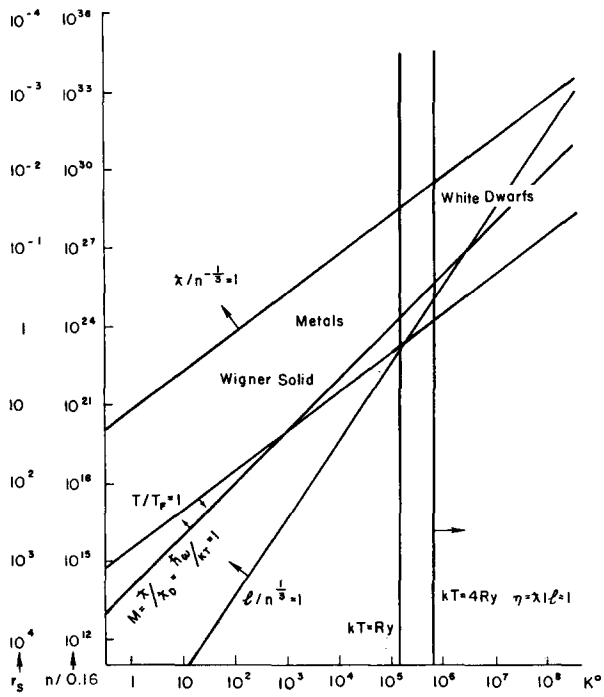


FIG. 1. Density n divided by 0.16 or r_s (interparticle spacing in units of the Bohr radius) vs the temperature, for an electron system on a uniform positive background, on a logarithmic scale. The value 1 for various dimensionless constants divides the plane into regions of different physical character. The arrow on the line $x = 1$ shows the direction in which x grows.

by various authors using different methods and approximations.[13–20] The discussion of their work in Sec. 3 below will indicate that none of these treatments is suitable for investigating the question at hand: either the approximations used exclude the region where the phase transition is expected to occur or the treatment is confined to Boltzmann statistics (see Sec. 3 about the sensitivity of the transition to the statistics).

Recently the present author[21] has sketched a method for calculating the quantum partition function and the free energy within certain approximations. In the present paper we carry out this method in further detail. Within the region of interest in the n, T plane (see Fig. 1), the (Fermi) statistics and the discreteness of the collective energy levels are the relevant quantum features to be taken into account; the so-called "diffraction features" (i.e., the interference of particle wavefunctions) do not play any role except to provide the necessary wavefunction overlap in order for the exclusion principle and the collective oscillator wavefunctions to come into play. Thus, we borrow from quantum theory these two features in order to construct an "effective quantum Hamiltonian" H_{qu}, with which we calculate the partition function classically. This latter calculation is further simplified by the use of the random phase approximation (RPA). The actual calculations are quite cumbersome, particularly the determination of the dielectric constant $\delta(k, \omega, T)$ at any temperature and of the k range of collectivity of ρ_k, the Fourier components of the density. One of the reasons for the complexity is the use of the fermion distribution function in the calculation of δ.

The ultimate goal of these calculations is threefold:

(a) To suggest a physical mechanism for the role of quantum effects (in particular, the discreteness of the collective levels and the statistics) in the mechanism of a phase transition;

(b) To find the phase transitions of the *electron plasma* (i.e., electrons on a uniform positive background) for the range of metallic densities ($r_s = 2$–6)[22] and beyond ($r_s = 6$–104) and for a range of temperature T from 0°K to above the boiling point of ordinary metals ($T \sim 10^4$ °K);

(c) To find the phase transitions in *ion plasmas* as expected in white dwarfs ($\rho = 10^6$–$10^7 gr/cm^2$, $T \sim 10^7$ °K), in *hydrogen plasmas* as found in Jupiter ($\rho \sim 1.5$–$15\ gr/cm^3$, $T \sim 10^2$–10^3 °K from the surface inwards), and possibly also for prewhite dwarf configurations.

2. SURVEY OF SOME CURRENT RESULTS ON THE MELTING AND VAPORIZATION OF METALS AND PLASMAS AT CONDITIONS PREVAILING ON EARTH, IN WHITE DWARFS, AND ON JUPITER

A. Metals and Plasmas on Earth

Current theories of *melting of metals*[1] are based on Lindemann's semiphenomenological assumption that melting occurs when the amplitude of thermal vibration of the ions approaches a certain fraction δ of the interionic distance. δ, which is taken from experiment, is, e.g., $\approx \frac{1}{4}$ for alkali metals. Thus, melting is considered as a case of broken long-range

[13] E. W. Montroll and J. C. Ward, Phys. Fluids **1**, 55 (1958).
[14] (a) H. E. De Witt, J. Math. Phys. **3**, 1216 (1962); (b) **7**, 616 (1966).
[15] W. T. Grandy, Jr., and F. Mohling, Ann. Phys. (N.Y.) **34**, 424 (1965).
[16] A. A. Vedenov and A. L. Larkin, Zh. Eksp. Teor. Fiz. **36**, 1133 (1959) [Sov. Phys.—JETP **9**,806 (1959)].
[17] G. Kelbg, Ann. der Phys. (Leipzig) **9**, 159, 168 (1962); **12**, 219, 354 (1953); G. Kelbg and H. J. Hoffmann, *ibid.* **14**, 310 (1964); H. J. Hoffmann and G. Kelbg, *ibid.* **17**, 356 (1966); **19**, 186 (1967).
[18] T. Morita, Progr. Theor. Phys. (Kyoto) **22**, 757 (1959).
[19] H. Hertzheim, Ann. der Phys. (Leipzig) **19**, 380 (1967).
[20] T. Dunn and A. A. Broyles, Phys. Rev. **157**, 156 (1967).
[21] G. Carmi, *Proceedings of the Latin American Summer School of Physics 1965* (Gordon and Breach, Science Publ., New York, 1967), p. 90.

[22] r_s is the interelectronic mean distance in units of the Bohr radius, i.e., $4\pi r_s^3 a_0^3/3 = n^{-1}$, where n is the density ($n = 0.16 \times 10^{24} r_s^{-3}$).

order in the *ions*; the electrons for this purpose may be approximated by a uniform negative background.

On the other hand, a model for metals which has received extensive attentions over the past fifteen years is the *electron* plasma, i.e., a collection of electrons on a uniform neutralizing positive background. For low densities, the Coulomb forces rather than the exclusion principle dominate the motions, and this led Wigner[2] to postulate that at $T = 0$ and low density the electrons will form a perfect lattice, the "electron solid," whereas at high densities they form a degenerate electron gas.

There have recently been several attempts to find theoretical estimates for the phase transitions which presumably should exist between these solid and gaseous extremes of the electron plasma. Coldwell, Horsfall, and Maradudin,[3] Noziéres and Pines,[4] and de Wette[5] have considered *pressure melting* at $T = 0$ for the electron solid on the basis of Lindemann's or similar phenomenological criteria, in which case the vibration amplitudes are determined by the zero-point motions rather than by the thermal motions. They have arrived at melting density points corresponding to r_s values ranging from $r_s = 6$ to $r_s = 104$. Van Horn[6] has improved de Wette's method and arrived at $r_s \approx 27 \pm 20$. In a later publication,[11] Van Horn attempted to estimate the *temperature melting* of a plasma. He used Lindemann's and de Wette's criteria to obtain lower bounds for the melting point of the parameter $\Gamma = e^2Z^2/RkT$, where $R \approx n_{\text{ion}}^{-\frac{1}{3}}$, obtaining $\Gamma_{\text{melt}} \geq 32$. He also obtained an approximate upper bound $\Gamma_{\text{melt}} \leq 126$ from the Monte Carlo calculations of Brush, Sahlin, and Teller.[7] This is in agreement with an estimate $\Gamma_{\text{melt}} \approx 64$ by Mestel and Rudermann.[12]

During the same period, some authors have been reportion a *liquid–gas phase transition* for the electron plasma at $r_s \approx 6 \pm 1$. These again are *pressure transitions* at $T = 0$ and they are called "liquid–gas transitions" rather than "solid–liquid transitions" for essentially no better reason than that previous estimates for the melting point had come up with much larger r_s values. Thus, Gartenhaus and Stranahan[8] reported a transition in the Hartree–Fock approximation at $r_s \approx 6$; Osaka[9] reported a transition at $r_s \geq 6.4$, and Wiser and Cohen[10] reported a transition at $r_s \approx 5$. It is noteworthy that these authors associate a higher density with the liquid phase and a lower density with the gaseous phase. Our general physical picture will support this point of view.

In order to clear up the situation, replace the semi-empirical ansatz (such as Lindemann's and de Wette's) by first principles, drive up the theory from the pressure transition at $T = 0$ into the more general temperature transition at $T \neq 0$, and prepare a framework in which charged Boson plasmas (such as in He or C white dwarfs) can be treated equally well, it would seem highly desirable to calculate the first-order phase transitions from first principles. This will be done in Sec. 5.

B. Crystallization of White Dwarfs, Hydrogen Plasmas on Jupiter, and Related Phenomena

At the high densities occurring in a white dwarf ($\rho \approx 10^6$–10^7 g/cm^3, $n_{\text{el}} \approx 10^{30}$–$10^{32}$, $r_s = 0.01$–0.02), the electrons make a highly degenerate, uniform electron gas.[23] The ions, on the other hand, form a nearly rigid lattice[24] or are just about to solidify into one.[11]

There exist scaling relationships[6] between the electron plasma on a uniform positive background and the ion plasma on a uniform negative background which enable the results discussed in the previous section to be transcribed to the ion plasma. Thus, one obtains[6] pressure solidification of the ion plasma at densities of the order of magnitude of white dwarf densities. Order of magnitude considerations[11] also enable estimating the heat of crystallization which is released in the $T \neq 0$ regime, when the thermal and Coulomb energies are comparable. The heat of crystallization turns out to be of the order of KT (where $T \sim 10^7$ °K is the temperature of the white dwarf) and its release causes[11] a decrease in the rate of evolution of the white dwarf. This seems to explain,[11] qualitatively at least, the existence[25] of two distinct sequences in the H–R diagrams of white dwarfs.

Obviously, a more accurate theory for the crystallization process and for the calculation of crystallization temperatures and densities would be highly desirable. Furthermore, such a theory should be equally applicable to fermions, bosons, and particles obeying Boltzmann statistics.

It may be that the *electron* plasma will show some kind of phase change when $\hbar\omega_p \sim E_F$, where ω_p is the electron-plasma frequency. This indicates a density of $n_{\text{el}} \sim 10^{26}$. Also, *both* plasmas may show liquid–gas transitions in addition to solidification. All these transitions may show up in various stages of evolution of stars off the main sequence (e.g., during the degenerate core phase of a red giant), and the associated latent heat release may have sizeable effects on their history of evolution.

One of the interesting plasmas to consider is the hydrogen plasma on Jupiter. This is a rather warm

[23] E. E. Salpeter, Australian J. Phys. **7**, 373 (1954).
[24] E. E. Salpeter, Astrophys. J. **134**, 669 (1961).
[25] O. J. Eggen and J. L. Greenstein, Astrophys. J. **141**, 83 (1965).

($T \sim$ 100–1000°K) fermion plasma at relatively low density ($\rho \sim$ 1.5–15 g/cm^3), and the usual plasma models, which lean either on a metal on earth ($T \ll T_F$) or on a classical plasma ($T \gg T_F$), are insufficient, since in the case discussed here we have $T \sim T_F$.

The preliminary numerical calculations performed on the improved model[25] were in fact carried out in the regime $T \sim T_F$.

3. DISCUSSION OF CURRENT CALCULATIONS OF THE PARTITION FUNCTION OF A QUANTUM-COULOMB SYSTEM

Various treatments have been given for calculating the (grand) partition function and the equation of state of a quantum-mechanical Coulomb system at nonzero temperatures (see Montroll and Ward,[13] De Witt,[14] Grandy and Mohling,[15] Vedenov and Larkin,[16] Kelbg,[17] Morita,[18] Hertzheim,[19] and Dunn and Broyles[20]).

These treatments differ from each other in the approximations used and in the range of validity of their results. To obtain a clearer graphical view of the situation, we first consider the density–temperature plane (n, T) for the electron plasma. In particular, we will look for the dimensionless constants which are relevant for the electron plasma and find out how their values divide up the T, n plane into regions of different physical characteristics (Fig. 1).

The dimensionless constants for a plasma can be obtained as ratios of constants of the dimension of length. These latter are (m = electron mass, $\beta = 1/\kappa T$, $\omega_p^2 = 4\pi e^2 n/m$):

$n^{-\frac{1}{3}}$ is the mean interparticle distance;

$l = e^2\beta$ (Landau length) is the mean distance of mutual approach (at which $\bar{E}_{\text{kin}} = E_{\text{pot}}$);

$\lambda\!\!\!^{-} = h/p = h/(2\pi m\kappa T)^{\frac{1}{2}}$ is the thermal de Broglie wavelength of a particle.

When the particles are quasi-free, as when the interaction is strongly inhibited by the exclusion principle, their wavefunctions are plane waves to a good approximation. When T is high enough for most scattering states to be available, the wavefunction will be a wave packet of smallest possible extension for a given momentum, in order to reduce the (positive) interaction energy. $\lambda\!\!\!^{-}$ then gives, on the average, the extension of these wave packets.

$\lambda_F = (8\pi/3n)^{\frac{1}{3}}$ is the Fermi length (radius of the sphere of screening by exclusion);

$\lambda_D = (\bar{v}^1)^{\frac{1}{2}}/\omega_p = (KT/4\pi ne^2)^{\frac{1}{2}}$ is the Debye screening radius (screening by Coulomb correlation).

The following ratios of these lengths are physically meaningful.

Classical: (1) $l/n^{-\frac{1}{3}} = e^2\beta n^{\frac{1}{3}}$. The Debye–Huckel theory (DH), which is the classical limit of a quantum plasma in the "ring approximation,[13]" is a low-n limit. Since $e^2\beta n^{\frac{1}{3}}$ is the only classical dimensionless parameter which can be constructed from m, e, n, β, the DH theory may be considered as an expansion in $e^2\beta n^{\frac{1}{3}}$. (Although one sums to any order of e^2 in the ring approximation!)

(2) $\Lambda = e^2\beta\lambda_D^{-1} = (4\pi)^{\frac{1}{2}}e^3\beta^{\frac{3}{2}}n^{\frac{1}{2}}$. This parameter is proportional to $(ln^{\frac{1}{3}})^{\frac{3}{2}}$ and is therefore not an independent parameter.

(3) $\mathcal{E}^{-1} = 4\pi n\lambda_D^3/3$, the number of particles within the screening sphere. $\mathcal{E} = 3(4\pi)^{\frac{1}{2}}\beta^{\frac{3}{2}}n^{\frac{1}{2}} = 3\Lambda$ is again not an independent parameter.

Quantum: (4) $\mu = \lambda\!\!\!^{-}/\lambda_D = \hbar\omega p/\kappa T = 2\pi^{\frac{1}{2}}m^{-\frac{1}{2}}\hbar e\beta n^{\frac{1}{2}}$. This is our most important parameter. When $\kappa T \ll \hbar\omega_p$ (i.e., $\mu \gg 1$), the discrete collective plasmon levels $(n + \frac{1}{2})\hbar\omega p$ will be unexcited; the collective modes will be "frozen out" and the Hamiltonian will reduce to its noncollective part, which will serve as an effective Hamiltonian H_{eff}, capable of having a phase transition (see below). At $\kappa T > \hbar\omega_p$, the collective levels may be considered as a continuum and classical considerations apply. $\kappa T_p = \hbar\omega_p$ defines the plasma temperature T_p. The region above the line $\mu = 1$ in the (T, n) plane is the region where quantum theory shows up through effects stemming from the discreteness of the collective levels.

(5) $\eta = \lambda\!\!\!^{-}/l = \hbar(2m)^{-\frac{1}{2}}\beta^{-\frac{1}{2}}e^{-2}$. When $\eta = 1$, $\kappa T = \kappa T_d = 2me^4/\hbar^2 = 4Ry$, giving $T_d \approx 6.3 \times 10^5$ °K. When the temperature is very high ($T > T_d$), $\lambda\!\!\!^{-} > l$, i.e., wave packets overlap and the diffraction phenomena become prominent, contrary to superficial intuition. When $T < T_d$, the minimum extension $\lambda\!\!\!^{-}$ of the wave packet of the electron is less than the mean distance of mutual approach and thus quantum diffraction, it seems, need not occur. However, above the line $l/n^{-\frac{1}{3}} = 1$ the actual interparticle distance $n^{-\frac{1}{3}}$ is less than the distance l of dynamical approach. Thus, above $l/n^{-\frac{1}{3}} = 1$, the criterion for diffraction is $\lambda\!\!\!^{-}/n^{-\frac{1}{3}} > 1$.

(6) $\lambda\!\!\!^{-}/n^{-\frac{1}{3}} = 1.13 \times 10^{-7}n^{\frac{1}{3}}/T^{\frac{1}{2}}$. Above the line $\lambda\!\!\!^{-}/n^{-\frac{1}{3}} = 1$ quantum diffraction occurs for any T, and to the right of the line $T = 6.3 \times 10^5$ °K it occurs also everywhere below the line $\lambda\!\!\!^{-}/n^{-\frac{1}{3}} = 1$, because particles approach each other dynamically to within their de Broglie wavelengths.

(7) $\tau = \kappa T/\mathcal{E}_F = 0.8 \times 10^{11}Tn^{-\frac{2}{3}}$ measures the degeneracy ($\tau \ll 1$) of the electron gas. At $\tau = 1$, $T = T_F \equiv 1.2 \times 10^{-11}n^{\frac{2}{3}}$. In the trapezoidal region between the lines $\lambda\!\!\!^{-}/n^{-\frac{1}{3}} = 1$ and $T/T_F = 1$, and to the left of the line $T = 6.3 \times 10^5$ °K (which includes the region of interest for phase transitions at metallic

densities), the wavefunctions would not overlap in the mean if they would assume their minimum wave packet size given by $\lambda\!\!\!^{-} = \hbar/\langle p\rangle$. It might thus seem that, because of this lack of overlap, the exclusion principle would have no chance to come into play.

However, a degenerate situation is energetically not unfavorable above the line $T/T_F = 1$ (i.e., for $\tau < 1$), because then (a) the temperature is sufficiently low so that the thermal motion will not stir up the particles from this distribution in which (b) interactions are strongly inhibited by the exclusion principle and the energy is thus reduced. Furthermore, because of the inhibition of interactions, it does not cost the particles any potential energy to spread out to plane waves, whose overlap enables the exclusion principle to come into play.

We now return to the discussion of current treatments of the calculation of the partition function by the authors cited.

Most of these authors[13–16,18] apply perturbation methods of quantum field theory to obtain a diagrammatic expansion of the (grand) partition function. In the work of Montroll and Ward,[13] this results in a generalization to quantum systems of Mayer's classical cluster integral theory. In particular, quantum-mechanical "ring diagrams" (characterized by n particles interacting n times and exchanging momentum q at each interaction) dominate; neglecting the remaining diagrams again corresponds to the RPA. For a Coulomb system, in particular, they obtain the Debye–Huckel results in the classical limit (low density n) and the Gell-Mann–Brueckner correlation energy at the low temperature limit $T = 0$ (high n).

De Witt[14] extended these results for the Coulomb system to any T and n, both within the ring approximation[14a] and beyond.[14b] However, he confined himself to Boltzmann statistics. This is a definite drawback from our point of view, since stability and state of phase of a Coulomb system depend rather sensitively on the statistics. Thus, a system of charged bosons or Boltzmann particles at $T \approx 0$ on a uniform neutralizing background would collapse [as can be read off from the equations of state, e.g., Eq. (61),[14a] and from the fact that bosons and boltzons have the same ground state]. Fermion statistics would provide the necessary positive stabilizing pressure (see Dyson and Lenard[26]). Since the existence of stable phases is sensitive to the statistics, it is also evident that the question of phase transitions will be sensitive to the statistics.

Grandy and Mohling,[15] who work from a master-graph formulation by Mohling,[27] take into account the (fermion or Bose) statistics, but they work only in the high-T–low-n range, i.e., in the lower right-hand corner of Fig. 1, where phase transitions are not expected.

Vedenov and Larkin[16] work out the extreme cases $T \ll me^4/\hbar^2$ and $T \gg me^4/\hbar^2$, i.e., the regions in the far left or right in Fig. 1, which are again beyond the range of interest.

Kelbg, in a series of articles,[17] finds an "effective quantum potential" v_{qu} which is so tailored that a *classical* calculation of the partition function with this potential gives the quantum result. However, he too neglects the statistics; in addition, his method is valid for high T only.

Similarly, Morita[18] and Hertzheim[19] work out the cases of high T only.

The work of Dunn and Broyles[20] claims to apply to any T and n; furthermore, it takes into account the statistics. However, they apply an approximation, which has no convincing justification in the relevant regions of the T, n plane. This is the neglect of coupling between diffraction and symmetry effects. Since by diffraction effects they mean any effects which result from the uncertainty relation, the discreteness of collective levels is also included here. It is well known that the collective behavior is quite sensitive to the statistics; for example, the uppermost k_c for the collectivity of ρ_k is $\omega_p/\bar{v}$, $[\bar{v} = (2\kappa T/m)^{\frac{1}{2}}]$ for Boltzmann statistics and $\sim\omega_p/v_F$ for Fermi statistics, where $\bar{v}/v_F \ll 1$ for $T \ll T_F$. Thus, the "diffraction" and symmetry phenomena are strongly coupled, and the region of occurrence of both in the T, n plane nearly coincides (and includes the region of interest).

In conclusion, none of the treatments discussed seems to be suitable for investigating the question of phase transitions in an electron plasma; also for Boltzmann statistics none except one[14] applies to the T, n range of interest. Here we give an alternative calculation also for the latter case, in order to emphasize the role of the discreteness of the collective levels.

4. QUALITATIVE PICTURE FOR THE MECHANISM OF THE PHASE TRANSITION

Our picture is based on the "freezing out," at $\kappa T \ll \hbar\omega_p$, of the collective modes ρ_k. Here

$$\rho_k = \sum_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} = \sum_{\lambda} c^{+}_{\lambda-\mathbf{k}} c_{\lambda} \tag{1}$$

is the $\mathbf{k}$th Fourier component of the density $\rho(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i)$ and the c_k are the electron annihilation

[26] F. J. Dyson and A. Lenard, J. Math. Phys. **9**, 698 (1968).

[27] F. Mohling, Phys. Rev. **122**, 1043, 1062 (1961).

operators. The canonically conjugate variables are ($k = 0 \cdots k_c$)

$$\varphi_k = \alpha_k \rho_k, \quad \chi_k = -i\beta_k \sum_j (\mathbf{k}\cdot\mathbf{p}_j) e^{i\mathbf{k}\cdot\mathbf{x}_j},$$
$$(\chi_k, \varphi_{k'})_{\text{P.b.}} = \delta_{kk'} \quad \text{(in the RPA)}, \tag{2}$$
$$\alpha_k = (4\pi e^2/k^2)^{\frac{1}{2}}, \quad \beta_k = (4\pi e^2 k^2)^{-\frac{1}{2}} N^{-1},$$

where P.b. indicates Poisson bracket. The Hamiltonian of the system is (Ω = volume)

$$H = \sum_i \frac{p_i^2}{2m} + \frac{2\pi e^2}{\Omega} \sum_{k\neq 0} \frac{1}{k^2} (\rho_k \rho_{-k} - N), \tag{3}$$

where the $k = 0$ term was cancelled by the contribution from the uniform positive background. The particle variables $\mathbf{x}_i$, $\mathbf{p}_i$ can be split[25,28–30] into "collective" ($\delta\mathbf{x}_i$, $\delta\mathbf{p}_i$) and "individual" ($\mathbf{X}_i$, $\mathbf{P}_i$) parts, $\mathbf{x}_i = \mathbf{X}_i + \delta\mathbf{x}_i$, $\mathbf{p}_i = \mathbf{P}_i + \delta\mathbf{p}_i$, such that $\mathbf{X}_i$, $\mathbf{P}_i$ contribute identically[31] zero to the collective variables φ_k and χ_k:

$$\bar{\varphi}_k \equiv \alpha_k \sum_i e^{-i\mathbf{k}\cdot\mathbf{X}_i} \equiv 0,$$
$$\bar{\chi}_k \equiv -i\beta_k \sum_j (k\cdot P_j) e^{i\mathbf{k}\cdot\mathbf{X}_i} \equiv 0. \tag{4}$$

Equations (4) hold as identities in the RPA if one chooses

$$X_i = x_i - i\sum_{k=0}^{k_c} \frac{\mathbf{k}}{Nk^2} \rho_k e^{i\mathbf{k}\cdot\mathbf{x}_i} = x_i - \delta x_i,$$
$$P_i = p_i - i\sum_{k=0}^{k_c} \alpha_k \mathbf{k} \chi_k e^{-i\mathbf{k}\cdot\mathbf{x}_i} = p_i - \delta p_i. \tag{5}$$

It can easily be seen that H can be written as

$$H = \left[\frac{1}{2m}\sum_i P_i^2 + \frac{2\pi e^2}{\Omega}\sum_{k>k_c} \frac{1}{k^2}\rho_k\rho_{-k}\right] + \left[\frac{1}{2}\sum_{k=0}^{k_c}(\varphi_k^*\varphi_k + \omega_p^2 \chi_k^*\chi_k)\right]$$
$$= H_{\text{ind}} + H_{\text{coll}}, \tag{6}$$

where we have used the identity (in the RPA)

$$\frac{1}{2m}\sum p_i^2 = \sum_i \frac{1}{2m}\left(p_i - i\sum^{k_c} \alpha_k \mathbf{k}\chi_k e^{-i\mathbf{k}\cdot\mathbf{x}_i}\right)^2 + \frac{\omega_p^2}{2}\sum^{k_c} \chi_k^*\chi_k \tag{7}$$

and that (using $\bar{\rho}_k = 0$ for $k \leq k_c$ and $\bar{p}_i \equiv P_i$)

$$H_{\text{ind}} = \bar{H} \equiv H(X_i, P_i). \tag{8}$$

Furthermore, in the RPA,

$$(X_i, H_{\text{coll}})_{\text{P.b.}} = (P_i, H_{\text{coll}})_{\text{P.b.}} = (\delta x_i, H_{\text{ind}}) = (\delta p_i, H_{\text{ind}}) = (\chi_k, H_{\text{ind}}) = (\varphi_k, H_{\text{ind}}) = 0,$$

so that

$$\dot{x}_i = (X_i, H) = (X_i, H_{\text{ind}}),$$
$$\delta\dot{x}_i = (\delta x_i, H) = (\delta x_i, H_{\text{coll}}), \tag{9}$$
$$\dot{x}_k = (\chi_k, H) = (\chi_k, H_{\text{coll}})$$

(with corresponding relations for $\dot{P}_i$, $\delta\dot{p}_i$, $\dot{\rho}_k$), which shows that H_{ind} and H_{coll} serve as the effective Hamiltonians for the individual and collective variables, respectively.

Upon quantization, each term in H_{coll} gives the energy levels $(n_k + \frac{1}{2})\hbar\omega_p$. When $\kappa T \ll \hbar\omega_p$, the probability for $n_k \neq 0$ is very small and one may write

$$H_{\text{coll}} = s\tfrac{1}{2}\hbar\omega_p \quad (\kappa T \ll \hbar\omega_p), \tag{10}$$

where s is the number of collective modes,

$$s = \frac{4\pi}{3}\frac{\Omega}{(2\pi)^3} k_c^3. \tag{11}$$

(For $\kappa T \sim \hbar\omega_p$, a treatment in powers of $\mu^{-1} = \kappa T/\hbar\omega_p$ will give higher-order results. We shall indeed include these effects in our formulation, Sec. 6.) We thus are left with an effective Hamiltonian

$$H^{\text{qu}} = \frac{1}{2m}\sum P_i^2 + \frac{2\pi e^2}{\Omega}\sum_{k>k_c} \rho_k^*\rho_k + \tfrac{1}{2}s\hbar\omega_p. \tag{12}$$

The number of effective degrees of freedom which participate in H_{eff} is $3N - s$. In a typical metal at room temperatures, $s/N = 0.15$–0.25.[32]

The crucial point is now that, since k_c in general depends on n and T, H_{eff} *explicitly depends on temperature and density*:

$$H_{\text{eff}} = H_{\text{eff}}(x_i, p_i, T, n).$$

Furthermore, *the number $3N - s$ of effective degrees of freedom depends on T, n.* For example, in the case of Boltzmann statistics, k_c may be taken to be of the order of k_D, the plasma Debye wavenumber, i.e.,

$$k_c\langle v^2\rangle_{\text{av}} = \omega_p^2, \quad \text{or} \quad k_c^B = (2\pi e^2 n/\kappa T)^{\frac{1}{2}}. \tag{13}$$

Thus, in this case, when T is raised or n is lowered, $3N - s$ becomes larger. For Fermi statistics, one may take

$$k_c v_F = \omega_p$$

or

$$k_c^F = \pi^{-\frac{1}{2}}(8\pi/3)^{\frac{1}{3}} e m^{\frac{1}{2}} n^{\frac{1}{6}} = 1.145 e\hbar^{-1} m^{\frac{1}{2}} n^{\frac{1}{6}}. \tag{14}$$

Again, a decrease in n increases $3N - s$.

[28] D. Bohm and G. Carmi, Phys. Rev. **133**, 318A (1964).
[29] G. Carmi and D. Bohm, Phys. Rev. **133**, 330A (1964).
[30] G. Carmi, *Lectures in Theoretical Physics* (Univ. of Colorado Press, Boulder, Colo., 1964), Vol. 7c.
[31] Notice that this separation is done *without* redundant variables and subsidiary conditions; Eqs. (4) hold as *identities*. We here use the notations $\overline{f(x_i, p_i)} \equiv f(X_i, P_i)$.
[32] At $T \neq 0$, there always remains, of course, a (usually very small) finite probability for the excitation of higher collective levels. We shall indeed take it into account in our calculations [(Eq. (20)], but in the region of interest it will give a negligible contribution. Physically, the point is that this probability is much smaller than it would be if the collective levels made a continuum.

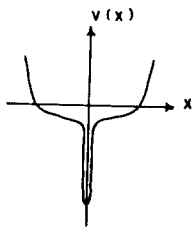


FIG. 2. Classical system with "changing" number of degrees of freedom.

For these considerations to hold, it is essential that when T is raised, a number of collective modes are destroyed *before* they are excited out of their ground state. It can be verified that this is the case in the range of interest.[33]

The situation here is much more drastic than with classical potentials. The change of dimensionality of our system as we change the temperature could be simulated classically by a one-particle system in one dimension in a potential as shown in Fig. 2. In that case, starting from the state in which the particle is deep down in the infinitely narrow "pipeline," the freedom of motion of the particle changes drastically if we give it sufficient potential energy to lift it up to the wide opening.

If the cross section of the narrow part of the potential is of measure zero with respect to the wide part, then there would be a change in dimensionality when the particle is lifted out of the narrow well.

In this case, an infinitesimal increase in kinetic energy ΔE_k cannot be achieved without a finite expenditure of potential energy ΔE_k, however small ΔE_k is, if the system starts deep down the pipeline.[34] This corresponds to a first-order phase transition, in which a finite amount of "latent heat" Q has to be extracted from the surrounding bath (whose temperature was made infinitesimally larger than that of the system in order to raise the latter) and used up as work against the potential energy of the constraints, before one can raise the system to a level where it can have more mobility (higher T). In other words, Q is used to break the constraints which restrict the system to a lower dimensionality, just as the collective modes have to be broken.

To conclude this section we remark that a system whose effective potential depends explicitly on T and n (and in which, moreover, the numbers of degrees of freedom depends on T and n) may in principle show phase transitions *even if the system is one-dimensional* and even if the system is otherwise treated classically.

5. CALCULATIONS

Stability of a phase requires $\partial^2 F/\partial n^2 \geq 0$, where F is the free energy per unit volume. If $\partial^2 F/\partial n^2 \leq 0$, as may occur in a calculation based on the assumption of a single homogeneous phase, the phase will be unstable towards the formation of another phase, and over the region of concavity $F(n)$ has to be replaced by a straight line (double tangent construction) extending from, say, $n = n_1$ to $n = n_2$; the two phases coexist. We have

$$F = -\frac{\kappa T}{\Omega} \ln Z \tag{15}$$

and our task is to find Z.

In the spirit of our approximations, we assume that the system splits into two mutually independent subsystems, the s collective degrees of freedom which are to be treated quantum-mechanically, and the $3N - s$ individual degrees of freedom, which we assume can be treated classically except for the fact that their number $3N - s$ is sensitive to the (Fermi, Bose, or Boltzmann) statistics. Correspondingly, Z splits into a product:

$$Z = Z_{\text{coll}} \cdot Z_{\text{ind}}, \tag{16}$$

$$Z_{\text{coll}} = \sum_{\{n_k\}} \exp\left[-\beta \sum_{k_1}^{k_s} (n_k + \tfrac{1}{2})\hbar\omega_p\right] = \frac{e^{-\frac{1}{2}s\beta\hbar\omega_p}}{(1 - e^{-\beta\hbar\omega_p})^s}. \tag{17}$$

Z_{ind} is a classical integral of $3N - s$ degrees of freedom over $\exp(-\beta H_{\text{ind}})$, where H_{ind} is given by (6), or since within the RPA one has for $k > k_c$, $\bar{\rho} \equiv \rho_k(X_i) = \rho_k(x_i)$:

$$H_{\text{ind}} = \frac{1}{2m}\sum P_i^2 + \frac{2\pi e^2}{\Omega}\sum_{k>k_c} \bar{\rho}_k \bar{\rho}_{-k}. \tag{6'}$$

In writing out Z_{ind}, one has to introduce factors $\delta(\sum e^{-ik\cdot X_i})$, $\delta[\sum (k \cdot P_i)e^{ik\cdot X_i}]$ to ensure that the $x_i p_i$ integration is carried out over the range of definition of the X, P, which are $2N$[35] variables fulfilling the $2s$ identities $\bar{\rho}_k \equiv \bar{\chi}_k \equiv 0$, $k \leq k_c$:

$$Z_{\text{ind}} = \frac{e^{N-s}}{N^{N-s}h^{N-s}} \int dX_1 \cdots dP_N e^{-\beta H_{\text{ind}}} \prod_{k=k_1}^{k_s} \delta(\bar{\rho}_k)\delta(\bar{\chi}_k).$$

The evaluation can be simplified in the RPA by introducing the variables

$$r_k = (\rho_k \rho_{-k})^{\frac{1}{2}}, \quad \vartheta_k = \frac{1}{2i} \ln \frac{\rho_k}{\rho_{-k}} \tag{18}$$

[33] The destruction of a mode ρ_k means that ρ_k no longer obeys a self-determining equation $\ddot{\rho}_k + \omega_k^2 \rho_k = 0$ and one has to go back to the full (exact) equation

$$\ddot{\rho}_k + \sum (k \cdot v_i)^2 e^{-ik\cdot x_i} + \frac{\omega_p^2}{N}\sum_{k'} \frac{k \cdot k''}{k'^2} \rho_{k-k'}\rho_{k'} = 0.$$

The second, "thermal" term, which couples ρ_k to all particle degrees of freedom, shows that this destruction first occurs at highest k's.

[34] It may be assumed that the situation shown in Fig. 2 has been arrived at as the limiting case of a finite-width pipeline with a finite-width particle in it, in such a way that, in the limit, the particle had no "elbowroom" to exercise any motion at all.

[35] To simplify the notation, we henceforth write $3N = N'$ and drop the prime.

and transforming the integration dx_i^3 in (16) to these variables, using the approximations which are customary in similar cases[36] and which are considered as being "reasonable" within the context of the RPA. These approximations consist of

(1) neglecting all but the first N r_k,
(2) considering $r_1 \cdots r_N$ as independent variables,
(3) approximating the Jacobian $J = \partial(x_i)/\partial(r_k, \vartheta_k)$, $dx^N = J \prod_k r_k\, dr_k\, d\vartheta_k$, whose exact form is[36,37]

$$J = \frac{N!}{\Omega^N} \exp\left[-\int_\Omega \rho(\mathbf{x})[\ln \rho(\mathbf{x}) - 1]\, d^3x\right]$$

by the approximate form[38]

$$J = \frac{1}{\pi^N c_0^N} \exp\left(-\frac{1}{N}\sum_{k_x>0} \rho_k \rho_{-k}\right)\left(c_0 = \frac{N}{\Omega}\right). \quad (19)$$

Secondly, we simplify the integration over the P_i by writing[39]

$$\int dP_1 \cdots dP_N \prod \delta(\bar{\chi}_k) e^{-\beta \Sigma P_i^2/2m} = \left(\frac{2\pi m k T}{h^2}\right)^{\frac{1}{2}(N-s)}.$$

This term contributes zero to the calculation of f'' (see below).

[36] J. Percus and G. Yevick, Phys. Rev. **110**, 1 (1958).

[37] Here one has to imagine $\rho(x)$ as expressed in terms of its Fourier components ρ_k and the integration d^3x as carried out.

[38] Although (19) is usually justified (see Ref. 36) by the RPA, it is also closely connected with the assumption of a single homogeneous phase. This can be shown by expanding

$$\ln \rho(\mathbf{x}) = \ln n + \left(\frac{\rho - n}{n}\right) - \frac{1}{2}\left(\frac{\rho - n}{n}\right)^2 + \frac{1}{3}\left(\frac{\rho - n}{n}\right)^3 + \cdots [0 \,..\, \rho \,..\, 2n],$$

where $n = (1/\Omega)\int \rho\, d^3x$. Since Z is, after the change of variables to r_k, θ_k, an integral over all values of r_k, this would have been, under a change of variables to the function space $\rho(x)$, equivalent to an integration over all possible functions $\rho(x)$. The functions $\rho(x)$ are either of the type $\rho_<$, say, for which $|(\rho - n/n| \ll 1$, in which case one may neglect third- and higher-order terms in the above series, or of the type $\rho_>$, for which this is not so. If the $\rho_>$ are neglected, one easily obtains (19), using Paraseval's identity. There will be some range of T and n in which the contributions of the functions $\rho_>$ to Z_{ind} will be negligible. Because of $|(\rho - n)/n| \ll 1$, we shall call this range the range of existence of a single phase. In this range the use of (19) is justified. In the range of which the contribution of the $\rho_>$ is not negligible, the use of (19) is not justified. If one nevertheless uses (19) there, the mistake made thereby may show up by giving an $F(n)$ with a concave portion. It has not been proved that this will *always* show up this way. However, it seems reasonable to conjecture that whenever a concave part of $F(n)$ has shown up, it is because the $\rho_<$ were not negligible.

[39] In the RPA, one can in principle define $3N - s$ functions $\pi_r(\mathbf{x}_i, \mathbf{p}_i)$ such that

$$\sum_{i=1}^{N} P_i^2 = \sum_{r=1}^{3N-s} \pi_r^2,\ dP_1 \cdots dP_N \delta(\chi_1) \cdots \delta(\chi_1) = d\pi_1 \cdots d\pi_{3N-s}.$$

This is because the χ_k are linear in the P_i and contain phase factors $e^{-ki \cdot x_i}$, in which the x_i are (in the RPA) uncorrelated to the P_i.

With these approximations, Z becomes

$$Z = \frac{e^{-\frac{1}{2}s\beta\hbar\omega_p}}{(1 - e^{-\beta\hbar\omega_p})^s}\left(\frac{2\pi m\kappa T}{h^2}\right)^{\frac{1}{2}(N-s)} \frac{2^{N-s} e^{N-s}}{c_0^{N-s} N^{N-s}} \times \int r_{s+1}\, dr_{s+1} \cdots r_N\, dr_N \times \exp\left[-\beta \sum_{n=s+1}^{N} \frac{1}{2}\left(\frac{4\pi e^2}{\Omega k_n^2} + \frac{1}{N}\right) r_{k_n}^2\right]. \quad (20)$$

We remark in passing that for most cases of interest one may approximate the first factor in (20) by $\exp[-\frac{1}{2}s\beta\hbar\omega_p]$. This assumes that the collective oscillators are all in their ground state, even when the temperature is a few thousand degrees. For example, taking Na as a typical case, with $c_0 \sim 10^{21}$ cm^{-3}, one has $\omega_p^2 = 4\pi e^2 c_0/m \sim 3 \times 10^{30}$, i.e., $\hbar\omega_p \sim 1.7 \times 10^{-12}$ erg, which corresponds to a temperature of $\sim$17,000°K. At, say, 6000°K, the relative probability, $\exp[-(1 + \frac{1}{2})\hbar\omega_p/\kappa T]/\exp[-\frac{1}{2}\hbar\omega_p/\kappa T]$, of finding the collective oscillator in its first excited state is $\exp[-\hbar\omega_p/\kappa T] = 5 \times 10^{-2}$. Similarly, one finds that the *total* probability of finding the oscillator in *any* excited state is only by a factor $1 + 5 \times 10^{-2}$ larger than the probability for the first level. Thus the probability of finding it in the ground state is 95%. As can be verified *a posteriori*, the mistake made in neglecting this deviation from 100% produces a mistake in the curve $F(c_0)$ which is less than the deviation of the concave part from the straight-line construction. Thus we are assured that we do not affect the question of the existence or nonexistence of the phase transition if we assume that the collective modes are all in their ground states.

Denoting by σ_N^2 the mean-square deviation of the Gaussian integrand by (20),

$$\sigma_n^2 = \frac{1}{\beta \dfrac{4\pi e^2}{\Omega k_n^2} + \dfrac{1}{N}} = \frac{k_n^2}{k_n^2 + k_D^2} N, \quad (21)$$

and using $\int_0^\infty r \exp(-r^2/2\sigma^2)\, dr = \sigma^2$, the integration in (20) be carried out to give[40]

$$Z = Z_{\text{coll}} \cdot A \cdot \frac{1}{c_0^{N-s}} \frac{1}{N^{N-s}} \left(\prod_{n=s+1}^{N} \frac{k_n^2}{k_n^2 + k_D^2}\right) N^{N-s}, \quad (22)$$

where

$$A = \left(\frac{2\pi m\kappa T}{h^2}\right)^{\frac{1}{2}(N-s)} e^{N-s} 2^{N-s}. \quad (23)$$

[40] To avoid confusion, we denote the average density N/Ω in this calculation by c_0. Note again the connection (11) between s and k_c and the obvious notation $k_s = k_c$.

For $B = \ln \prod_n [k_n^2/(k_n^2 + k_D^2)]$ one obtains

$$\frac{1}{4\pi} B = \frac{n_N^3}{3} \ln \frac{n_N^2}{n_N^2 + n_D^2} - \frac{n_c^2}{3} \ln \frac{n_c^2}{n_c^2 + n_D^2} - \tfrac{2}{3} n_c^2 (n_N - n_c) + \tfrac{2}{3} n_c^3 \left(\arctan \frac{n_N}{n_D} - \arctan \frac{n_c}{n_D} \right), \quad (24)$$

where

$$n_r = \frac{\Omega^{\frac{1}{3}}}{2\pi} k_r = \text{integer}. \quad (25)$$

For F this gives

$$F = -(\kappa T/\Omega) \ln Z = -(\kappa T/\Omega) \ln Z_{\text{coll}} - (\kappa T/\Omega) \ln A + (\kappa T/\Omega)(N - s) \ln c_0 + (\kappa T/\Omega) B. \quad (26)$$

F is a function of c_0 and T both directly and through s and k_D.

A. Boltzmann Statistics

For Boltzmann statistics $k_c (\equiv k_s) = k_D$ [Eq. (13)] and we have

$$n_D^2 = \frac{\Omega^{\frac{2}{3}}}{(2\pi)^2} \frac{\omega_p^2}{\langle v^2 \rangle_{\text{av}}} = \frac{1}{\pi} \Omega^{\frac{2}{3}} \frac{e^2 c_0}{\kappa T}, \quad (27)$$

$$n_N = \Omega^{\frac{1}{3}} (\tfrac{3}{4}\pi)^{\frac{1}{3}} c_0^{\frac{1}{3}}, \quad (28)$$

$$\gamma^{\frac{1}{3}} \equiv \frac{n_D^2}{n_N^2} = \frac{1}{\pi} \left(\frac{4\pi}{3} \right)^{\frac{2}{3}} e^2 c_0^{\frac{1}{3}} \beta \equiv Q^{-\frac{1}{3}} c_0^{\frac{1}{3}}, \quad (29)$$

where

$$Q = c_0/\gamma = \pi^3 (\tfrac{3}{4}\pi)^2 e^{-6} (\kappa T)^3,$$

or

$$\gamma = c_0/1.51 \times 10^{56} (\kappa T)^3 = 2.57 \times 10^{-9} c_0/T^3 \quad (29')$$

and

$$s/N = n_D^3/n_N^3 = \gamma^{\frac{1}{2}}. \quad (30)$$

We introduce the notation

$$f = \pi^{-3} \left(\frac{4\pi}{3} \right)^2 e^6 (\kappa T)^{-4} F = \frac{\gamma}{\kappa T c_0} F, \quad (31)$$

$$\lambda = (2 + \ln 2 - \pi/2) \frac{4\pi}{3} \approx 5, \quad (32)$$

and we obtain (neglecting $\ln A$, which evidently does not contribute to $\partial^2 F/\partial c^2$)

$$-\frac{\kappa T}{\Omega} \ln Z_{\text{coll}} = \frac{1}{2} \frac{s}{\Omega} \hbar\omega_p + \frac{s}{\Omega \kappa T} \ln (1 - e^{-\beta \hbar \omega_p}), \quad (33)$$

$$f_{\text{coll}} = 2.37 e^{-2} \hbar m^{-\frac{1}{2}} (\kappa T)^{\frac{1}{2}} \gamma^2 + (\kappa T)^{-2} \gamma^{\frac{3}{2}} \ln (1 - e^{-\beta \hbar \omega_p}) = 3.26 \times 10^5 (\kappa T)^{\frac{1}{2}} \gamma^2 + (\kappa T)^{-2} \gamma^{\frac{3}{2}} \ln (1 - e^{-\beta \hbar \omega_p}),$$

$$\frac{\gamma}{\kappa T c_0} \cdot \frac{\kappa T}{\Omega} (N - s) \ln c_0 = (\gamma - \gamma^{\frac{2}{3}}) \ln Q\gamma, \quad (34)$$

$$-\frac{\gamma}{\kappa T c_0} \cdot \frac{\kappa T}{\Omega} B = \frac{4\pi}{3} \gamma \ln (1 + \gamma^{\frac{1}{3}}) - \frac{4\pi}{3} \gamma^{\frac{2}{3}} \ln 2 + \frac{8\pi}{3} \gamma^{\frac{4}{3}} - \frac{8\pi}{3} \gamma^{\frac{2}{3}} - \frac{8\pi}{3} \gamma^{\frac{1}{2}} (\arctan \gamma^{-\frac{1}{6}} - \pi/4),$$

$$f = \frac{4\pi}{3} \gamma \ln (1 + \gamma^{\frac{1}{3}}) + \frac{8\pi}{3} \gamma^{\frac{4}{3}} - \gamma^{\frac{2}{3}} \left[\lambda + \frac{8\pi}{3} \arctan \gamma^{-\frac{1}{6}} + \ln Q\gamma + (\kappa T)^{-2} \ln (1 - e^{-\beta \hbar \omega_1}) \right] + 3.26 \times 10^5 \gamma^2 (\kappa T)^{\frac{1}{2}} + \gamma \ln Q\gamma, \quad (35)$$

and

$$\hbar\omega_p = \hbar \left(\frac{4\pi e^2}{m} \right)^{\frac{1}{2}} Q^{\frac{1}{2}} \gamma^{\frac{1}{2}}. \quad (36)$$

This gives

$$f' \equiv \frac{\partial f}{\partial \gamma} = 4.2 \ln (1 + \gamma^{\frac{1}{3}}) + 1.4 \frac{\gamma^{\frac{1}{3}}}{1 + \gamma^{\frac{1}{3}}} + 11.2\gamma^{\frac{1}{3}} + 1 + 2.3 \log_{10} \gamma + 2.3 \log_{10} T + 19.8 - 7.5\gamma^{\frac{1}{2}} - 12.56\gamma^{\frac{1}{2}} \arctan \gamma^{-\frac{1}{6}} - 3.45\gamma^{\frac{1}{2}} [\log_{10} \gamma + \log_{10} T] - 29.7\gamma^{\frac{1}{2}} + 7.3 \times 10^{-3} \gamma T^{-\frac{1}{2}} + \tfrac{3}{2} \gamma^{\frac{1}{2}} (\kappa T)^{-2} \ln (1 - e^{-\beta \hbar \omega_p}), \quad (37)$$

and from this we obtain the equation of state:

$$pv = n \frac{\partial F}{\partial n} = \gamma \frac{\partial F}{\partial \gamma} = \kappa T n \frac{\partial f}{\partial \gamma} = n \Big\{ \kappa T + 11.2 a r_s^{-1} + 1.4 \frac{a r_s^{-1}}{1 + a r_s^{-1}} + 3 \ln b r_s^{-1} + 4.2 \ln (1 + a r_s^{-1}) - [7.5 a^{\frac{3}{2}} r_s^{-\frac{3}{2}} + 12.5 a^{\frac{3}{2}} r_s^{-\frac{3}{2}} \arctan a^{-\frac{1}{2}} r_s^{\frac{1}{2}} + 3.45 a^{\frac{3}{2}} r_s^{-\frac{3}{2}} \ln b r_s^{-1}] + a^3 r_s^{-3} 7.3 \times 10^{-3} T^{\frac{1}{2}} + \tfrac{3}{2} a^{\frac{3}{2}} r_s^{-\frac{3}{2}} (\kappa T)^{-2} \ln (1 - e^{-\beta \hbar \omega_p}) \Big\}, \quad (38)$$

where

$$a = \gamma^{\frac{1}{3}} r_s, \quad b = c_0^{\frac{1}{3}} r_s, \quad n = c_0.$$

From (37) we find that for $\gamma \geq 1$ (which covers the range of interest) the Boltzmann–Coulomb gas is unstable ($f' < 0$) when the temperature is sufficiently low. The values of T at which f' becomes negative for several values of γ are collected in Table I.

Considering the values $\gamma = \text{const}$ as oblique-parallel straight lines in the $\log n/\log T$ plane (Fig. 1), it is easy to find the corresponding range of instability

TABLE I. Instability of the Boltzmann–Coulomb system.

$\gamma^{\frac{1}{3}}$	The value of T below which $f'<0$ (instability)
1	10^{8} °K
10	10^{5} °K
100	10^{2} °K
1000	10^{-1} °K
$\gamma<1$	Stability for all T

in that plane. A further differentiation of (37) gives

$$f''=\frac{\partial^2 f}{\partial\gamma^2}=3.81\gamma^{-\frac{2}{3}}+1.86\frac{\gamma^{-\frac{2}{3}}}{1+\gamma^{\frac{1}{3}}}+\gamma^{-1}+2.1\frac{\gamma^{-\frac{2}{3}}}{1+\gamma^{-\frac{1}{3}}}-\left[\tfrac{3}{4}\gamma^{-\frac{1}{2}}\ln Q\gamma+6.28\gamma^{-\frac{1}{2}}\arctan\gamma^{-\frac{1}{6}}+5.25\gamma^{-\frac{1}{2}}+0.767\frac{\gamma^{-\frac{1}{3}}}{(1+\gamma^{\frac{1}{3}})^2}\right]+7.32\times10^{-3}T^{\frac{1}{2}}+\tfrac{3}{4}(\kappa T)^{-3}\hbar\left(\frac{4\pi e^2}{m}\right)^{\frac{1}{2}}Q^{\frac{1}{2}}\frac{e^{-\beta\hbar\omega_p}}{1-e^{-\beta\hbar\omega_p}}. \tag{39}$$

Since

$$\gamma^{\frac{1}{3}}=\frac{1}{\pi}\left(\frac{4\pi}{3}\right)^{\frac{2}{3}}e^2\beta c_0^{\frac{1}{3}}=0.828e^2\beta c_0^{\frac{1}{3}}, \tag{40}$$

the variable $\gamma^{\frac{1}{3}}$ nearly coincides with the dimensionless parameter $l/n^{-\frac{1}{3}}=e^2\beta c^{\frac{1}{3}}$ introduced in Sec. 2, and the lines $\gamma^{\frac{1}{3}}=\text{const}$ are parallel to the line $ln^{\frac{1}{3}}=1$ of Fig. 1. It is convenient to consider γ and T as independent variables in (39), i.e., we take as frame of reference the T axis and the lines $\gamma^{\frac{1}{3}}=\text{const}$ (parallel to $ln^{\frac{1}{3}}=1$) in Fig. 1. Putting $\ln Q\gamma=\ln\gamma+3\ln T+19.8$ in (39), a straightforward but tedious evaluation shows that $f''>0$ down to the metallic range and somewhat further (e.g., for $\gamma=1$, $f''>0$ whenever $T\geq10^7$). For the white dwarf range, see below.

B. Fermi Statistics

For Fermi statistics $k_c\,(\equiv k_s)=k_c^F$, where k_c^F was defined by

$$k_c^F v_F=\omega_p,$$

or

$$k_c^F=\pi^{-\frac{1}{2}}\left(\frac{8\pi}{3}\right)^{\frac{1}{3}}\frac{em^{\frac{1}{2}}}{\hbar}c_0^{\frac{1}{6}}=1.145\frac{em^{\frac{1}{2}}}{\hbar}c_0^{\frac{1}{6}}, \tag{14}$$

$$n_c^F=\tfrac{1}{2}\Omega^{\frac{1}{3}}\pi^{-\frac{3}{2}}\left(\frac{8\pi}{3}\right)^{\frac{1}{3}}\frac{em^{\frac{1}{2}}}{\hbar}c_0^{\frac{1}{6}}=0.231\Omega^{\frac{1}{3}}em^{\frac{1}{2}}\hbar^{-1}c_0^{\frac{1}{6}}, \tag{41}$$

$$\alpha^{-\frac{1}{3}}\equiv\frac{(n_c^F)^2}{n_N^2}=\frac{2^{\frac{1}{3}}}{3\pi^2}\left(\frac{4\pi}{3}\right)^{\frac{1}{3}}\frac{e^2m}{\hbar^2}c_0^{-\frac{1}{3}}=0.0864\frac{e^2m}{\hbar^2}c^{-\frac{1}{3}},$$

or

$$\alpha=\frac{34\pi^5}{16}\frac{\hbar^6}{e^6m^3}c_0\equiv Q^{-1}c_0;\quad c_0=Q\alpha=0.75\times10^{22}\alpha \tag{42}$$

and

$$\alpha^{\frac{1}{3}}=Q^{-\frac{1}{3}}c_0^{\frac{1}{3}}=\frac{3\pi^2}{2^{\frac{2}{3}}}\left(\frac{3}{4\pi}\right)^{\frac{1}{3}}\frac{\hbar^2}{e^2m}c_0^{\frac{1}{3}}=11.56\frac{\hbar^2}{e^2m}c_0^{\frac{1}{3}}, \tag{43}$$

$$\gamma=\frac{m^3}{3^2\pi^4\hbar^6}(\kappa T)^3\alpha=2.94\times10^{32}T^3\alpha\quad[\text{see (29)}],$$

$$\frac{s}{N}=\left(\frac{n_c^F}{n_N}\right)^3=\alpha^{-\frac{1}{2}}.$$

We introduce

$$f=\frac{\alpha}{\kappa Tc_0}F=\frac{3^4\pi^5}{16\kappa T}\frac{\hbar^6}{e^6m^3}F \tag{44}$$

so that

$$f_{\text{coll}}=-\frac{\alpha}{\Omega c_0}\ln Z_{\text{coll}}$$

$$=\frac{1}{2}\frac{\alpha}{\kappa Tc_0}\frac{s}{\Omega}\hbar\omega_p+\frac{\alpha}{\kappa Tc_0}\frac{s}{\Omega\kappa T}\ln(1-e^{-\beta\hbar\omega_p})$$

$$=\frac{1}{2\kappa T}\left(\frac{16}{3^4\pi^5}\cdot\frac{e^6m^3}{\hbar^6}\right)^{\frac{1}{2}}\hbar\left(\frac{4\pi e^2}{m}\right)^{\frac{1}{2}}\alpha+\frac{1}{(\kappa T)^2}\alpha^{\frac{1}{2}}\ln(1-e^{-\beta\hbar\omega_p})$$

$$=1.69\frac{\alpha}{\kappa T}+\frac{\alpha^{\frac{1}{2}}}{\kappa T}\ln(1-e^{-\beta\hbar\omega_p});$$

$$\frac{\alpha}{\kappa Tc_0}\cdot\frac{\kappa T}{\Omega}(N-s)\ln c_0=(\alpha-\alpha^{\frac{1}{2}})\ln Q\alpha,$$

$$-\frac{\alpha}{\kappa Tc_0}\cdot\frac{\kappa T}{\Omega}B$$

$$=\frac{4\pi}{3}\alpha\ln(1+0.6\times10^{11}T\alpha^{\frac{1}{3}})-\frac{4\pi}{3}\alpha^{-\frac{1}{2}}\alpha\ln(1+0.6\times10^{11}T\alpha^{\frac{2}{3}})+\frac{8\pi}{3}\alpha\cdot\alpha^{-\frac{1}{3}}-\frac{8\pi}{3}\alpha\cdot\alpha^{-\frac{1}{2}}-\frac{8\pi}{3}\alpha\cdot\alpha^{-\frac{1}{2}}\arctan3.96\times10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{6}}+\frac{8\pi}{3}\alpha\cdot\alpha^{-\frac{1}{2}}\arctan3.96\times10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{3}}.$$

Collecting terms, we have (neglecting again $\sim\ln A$)

$$f=1.69\frac{\alpha}{\kappa T}+50.3\alpha+\alpha\ln\alpha+4.19\alpha\ln(1+0.6\times10^{11}T\alpha^{\frac{1}{3}})+8.38\alpha^{\frac{2}{3}}+8.38\alpha^{\frac{1}{2}}\arctan3.96\times10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{3}}-[50.3\alpha^{\frac{1}{2}}+50.3\alpha^{\frac{1}{3}}\ln\alpha+4.19\alpha^{\frac{1}{2}}\ln(1+0.6\times10^{11}T\alpha^{\frac{2}{3}})+8.38\alpha^{\frac{1}{2}}+8.38\alpha^{\frac{1}{2}}\arctan3.9\times10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{6}}]+2.3\frac{\alpha^{\frac{1}{2}}}{(\kappa T)^2}\log_{10}(1-e^{-\beta\hbar\omega_p}) \tag{45}$$

and

$$f' \equiv \frac{\partial f}{\partial \alpha} = 1.235\frac{10^{16}}{T} + 50.3 + 2.3\log_{10}\alpha + 1 + 9.65\log_{10}(1 + 6.6\times 10^{10}T\alpha^{\frac{1}{3}}) + \frac{0.93\times 10^{11}T\alpha^{-\frac{1}{3}}}{1 + 0.6\times 10^{11}T\alpha^{\frac{1}{3}}} + 5.58\alpha^{-\frac{1}{3}} + 4.19\arctan(3.96\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{3}}) + \frac{5.45\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{2}{3}}}{1 + 1.52\times 10^{-11}T^{-1}\alpha^{-\frac{1}{3}}} - \left[\frac{1.14\times 10^{-5}T^{-\frac{1}{2}}\alpha^{-\frac{5}{6}}}{1 + 1.57\times 10^{-11}T^{-1}\alpha^{-\frac{2}{3}}} + 25.15\alpha^{-\frac{1}{2}} + 38.6\alpha^{-\frac{2}{3}}\log_{10}\alpha + 50.3\alpha^{-\frac{2}{3}} + 4.82\alpha^{-\frac{1}{2}}\log_{10}(1 + 0.6\times 10^{11}T\alpha^{\frac{2}{3}}) + \frac{1.86\times 10^{11}T\alpha^{-\frac{1}{3}}}{1 + 0.6 + 10^{11}T\alpha^{\frac{2}{3}}} + 4.19\alpha^{-\frac{1}{2}} + 4.19\alpha^{-\frac{1}{2}}\arctan 3.9\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{6}}\right] + 1.15\frac{\alpha^{-\frac{1}{2}}}{(\kappa T)^2}\log_{10}(1 - e^{-\beta\hbar\omega_p}) + 4.15\times 10^{36}\frac{\alpha}{T^3}\frac{e^{-\beta\hbar\omega_p}}{1 - e^{-\beta\hbar\omega_p}}, \quad (46)$$

which gives the equation of state through

$$pv = n\frac{\partial F}{\partial n} = \alpha\frac{\partial F}{\partial \alpha} = \kappa T n\frac{\partial f}{\partial \alpha} \quad (n - c_0).$$

Thus, for example, at $\alpha = 1$ (i.e., at $c_0 = 0.75\times 10^{22}\,\mathrm{cm}^{-3}$) one has, assuming $T \geq 1$ (in which case terms 8, 9, 10, and 17 are negligible),

$$f'(\alpha = 1) = 30.15 + 1.235\frac{10^{16}}{T} + \frac{1.15}{(\kappa T)^2}\log_{10}(1 - e^{-\beta\hbar\omega_p}) + 4.15\times 10^{36}\frac{\alpha}{T^3}\frac{e^{-\beta\hbar\omega_p}}{1 - e^{-\beta\hbar\omega_p}}. \quad (47)$$

This is positive, i.e., we have stability. Here the second term comes from the zero-point energy of the collective modes, and the last two terms come from the higher levels. Actually, (46) is valid only for $\beta\hbar\omega_p \gg 1$ because of our choice (14) of k_c. Hence the last two terms in (46) and (47) will be negligible. It is interesting to note that for $\alpha = 1$, we would have stability even without the zero-point pressure, but the latter is overwhelmingly large within the range $\beta\hbar\omega_p \gg 1\,[(\hbar\omega/\kappa) \sim 2.10^4\,^\circ\mathrm{K}$ for $\alpha = 1]$.

It is easily seen that f' increases with α and hence the stability holds for all $\alpha > 1$.

As α decreases below 1, the noncollective part passes through zero and becomes negative. The zero-point pressure term decreases too, but remains quite large and positive down to about $\alpha = 10^{-8}$. At that point, $-38.6\alpha^{-\frac{1}{3}}\log_{10}\alpha$ becomes the dominant positive term, which is, however, outweighed, for $\alpha \leqslant 10^{-8}$, by the dominant negative term $(1.86\times 10^{11}T\alpha^{-\frac{1}{3}})/(1 + 0.6\times 10^{11}T\alpha^{\frac{2}{3}})$. Thus, at $\alpha \leqslant 10^{-8}$ (i.e., $c_0 \leqslant 10^{14}\,\mathrm{cm}^{-3}$) and $T \leqslant 1^\circ\mathrm{K}$, our model becomes unstable. At $T \lesssim 1^\circ\mathrm{K}$, relation (46) becomes increasingly more difficult to evaluate by hand.

A further differentiation of (46) gives

$$f'' \equiv \frac{\partial^2 f}{\partial \alpha^2} = \alpha^{-1} + \frac{9 + 10^{10}T\alpha^{-\frac{2}{3}}}{1 + 6.6\times 10^{10}T\alpha^{\frac{1}{3}}} + \frac{2.76\times 10^{-17}T^{-\frac{3}{2}}\alpha^{-2}}{(1 + 1.52 + 10^{-11}T^{-1}\alpha^{-\frac{1}{3}})^2} \times \frac{0.95\times 10^{-5}T^{-\frac{1}{2}}\alpha^{-\frac{11}{6}}}{1 + 1.57\times 10^{-11}T^{-1}\alpha^{-\frac{2}{3}}} + 12.57\alpha^{-\frac{3}{2}} + 25.7\alpha^{-\frac{5}{3}}\log_{10}\alpha + 38.6\alpha^{-\frac{5}{3}} + 2.41\alpha^{-\frac{3}{2}}\log_{10}(1 + 0.6\times 10^{11}T\alpha^{\frac{2}{3}}) + \frac{6.2\times 10^{11}T\alpha^{-\frac{2}{3}}}{1 + 0.6\times 10^{11}T\alpha^{\frac{2}{3}}} + \frac{0.82\times 10^{22}T^2\alpha^{-\frac{2}{3}}}{(1 + 0.6\times 10^{11}T\alpha^{\frac{2}{3}})^2} + 2.1\alpha^{-\frac{3}{2}} + 2.1\alpha^{-\frac{3}{2}}\arctan 3.9\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{1}{6}} + \frac{2.73\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{5}{3}}}{1 + 1.5\times 10^{-11}T^{-1}\alpha^{-\frac{1}{3}}} - \left[\frac{0.31\times 10^{11}T\alpha^{-\frac{4}{3}}}{1 + 0.6\times 10^{11}T\alpha^{\frac{1}{3}}} + \frac{0.2\times 10^{22}T^2\alpha^{-1}}{(1 + 0.6\times 10^{11}T\alpha^{\frac{1}{3}})^2} + 1.86\alpha^{-\frac{4}{3}} + 2.1\alpha^{-\frac{3}{2}}\arctan(3.96\times 10^{-6}T^{\frac{1}{2}}\alpha^{-\frac{1}{3}}) + \frac{5.54\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{11}{6}}}{1 + 15.7\times 10^{-12}T^{-1}\alpha^{-\frac{2}{3}}} + \frac{3.6\times 10^{-6}T^{-\frac{1}{2}}\alpha^{-\frac{5}{3}}}{1 + 1.52\times 10^{-11}T^{-1}\alpha^{-\frac{1}{3}}} + \frac{1.19\times 10^{-16}T^{-\frac{3}{2}}\alpha^{-\frac{5}{2}}}{(1 + 1.57\times 10^{-11}T^{-1}\alpha^{-\frac{2}{3}})^2} + 16.3\alpha^{-\frac{5}{3}} + \frac{9.3\times 10^{10}\alpha^{-\frac{5}{6}}T}{1 + 6.6\times 10^{10}T\alpha^{\frac{2}{3}}}\right] + \text{terms from higher collective levels.} \quad (48)$$

At $T = 0$,

$$f''_{T=0} = \alpha^{-1} + 15.87\alpha^{-\frac{3}{2}} + 22.3\alpha^{-\frac{5}{3}} + 25.7\alpha^{-\frac{5}{3}}\log_{10}\alpha - 1.86\alpha^{-\frac{4}{3}}. \quad (49)$$

From (48) one obtains $f''_{T=0} \geq 0$ for $\alpha^{-1} \leq \alpha_0^{-1} \leq 17.03 \pm 0.02$ and $f''_{T=0} \leq 0$ for $\alpha^{-1} \geq \alpha_0^{-1}$. In terms of the density, a phase transition therefore begins as one *comes down from larger density to*

TABLE II. Curvature of a fermion-Coulomb plasma free energy as a function of the density $n = 0.75 \times 10^{22}\alpha$.

α^{-1}	$f''(\alpha)$	α^{-1}	$f''(\alpha)$	α^{-1}	$f''(\alpha)$
10^6	-10^{12}	17.1	-6.1	15.625	$+32$
10^3	-5×10^6	17	$+2.5$	10	315
10^2	-4.7×10^4	16.9	$+10.81$	1	37
18	-93	16.8	$+12.1$	10^{-3}	2×10^{-3}

$c_{00} = 0.44 \times 10^{21}\,\mathrm{cm}^{-3}$. This corresponds to an r_s value of

$$r_s = 7.63 \pm 0.01. \tag{50}$$

Table II gives a few values of $f''(\alpha)$. We see from this table that f'' remains negative above $r_s = 7.63$, indicating that the coexistence region lasts at least until $c_0 \approx 10^{14}$, i.e., $r_s \approx 100$, which is the limit of validity of (48), the model becoming unstable beyond that point.

The value $r_s = 7.6$ fits in with the pressure transitions at $r_s \approx 6$ reported by Gartenhaus and Stranahan[8] (in the HF approximation), $r_s \geq 6.4$ reported by Osaka,[9] and $r_s \approx 5$ by Wiser and Cohen,[10] all for the electron plasma on uniform positive background at $T = 0$. Their method, however, is applicable to $T = 0$ *only*, whereas ours applies to any temperature for which $\beta\hbar\omega_p \gg 1$ (i.e., up to $\sim 10^4\,^\circ\mathrm{K}$ for metallic densities). If one uses a value of k_c which is valid also for $T \sim T_F$, the method would apply to *any* temperature. This will be the subject of another paper.

Returning to $T \neq 0$ and to (48), we note that when $T \geq 1\,^\circ\mathrm{K}$, terms like the second simplify, because over nearly the entire range of α one may neglect the 1 in the denominator, and terms like the third and the fourth are negligible (and so are the higher collective terms which $\beta\hbar\omega_p \gg 1$). The only temperature dependence of f'' is through the eighth term which contributes $2.41\alpha^{-\frac{3}{2}} \log_{10} T$ (where again we neglect 1 and take $T \geq 1$). Thus, without solving $f'' = 0$ explicitly, we may conclude that the transition density depends logarithmically on the temperature.

The detailed solution of $f''(\alpha, T) = 0$ at $T \neq 0$ requires a computer and will be reported elsewhere.

C. Application to White Dwarfs

A white dwarf may be considered as a relatively hot ($10^7\,^\circ\mathrm{K}$) ion plasma (carbon, say), of number density $\frac{1}{4}10^{30-32}\,\mathrm{cm}^{-3}$. At the temperatures considered, the ions may be taken as Boltzmann particles and (34) and (39) apply. The constants in these equations have to be modified, however, by replacing the electron charge and mass by that of the C^{++++} ion. Thus, the last term (which stems from $\frac{1}{2}s\hbar\omega_p$) has to be multiplied by $m_{ic}^{\frac{1}{2}}/(M_c^{\frac{1}{2}}\nu^2) = 1/2368$, where $\nu = 4$ is the degree of ionization. There are no other mass corrections, and all other charge corrections are incorporated into a change of γ by a multiplicative factor of $\nu^6 = 4096$. Taking, e.g., $c_0 = \frac{1}{4}10^{30}$ and remembering that γ is a function of c_0 and T through (29′), one then has to find a value of T such that the corresponding γ will give $f''(\gamma) = 0$ at the same initial temperature. This temperature will then be the transition temperature.

This procedure can be carried out by hand, by a trial and error method, and gives a transition temperature of

$$T_{\text{transit}} \approx (0.7 \times 10^7 \pm 0.2 \times 10^7)^\circ\mathrm{K},$$

which is of the expected order of magnitude.

The case of the hydrogen plasma on Jupiter requires a k_c^f which is valid in the temperature range $T \sim T_F$, and will be reported elsewhere.

ACKNOWLEDGMENTS

The author wishes to thank Professor John Bardeen for stimulating discussions, Professor H. M. Van Horn for drawing my attention to the case of Jupiter and for letting me have his paper[11] prior to publication, and to Professor M. H. Cohen and Professor N. Wiser for drawing my attention to the possibility of a liquid–vapor transition in plasmas.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Differential Equations for Half-Off-Shell Matrix Elements

M. I. SOBEL*
Theoretical Physics Division, A.E.R.E., Harwell, Berkshire, England

(Received 26 December 1967)

A first-order differential equation is found for the quasiphase parameter which describes the half-off-shell scattering matrix. The quasiphase is given as $\Delta(\infty)$, where $\Delta(r)$ obeys an equation involving the potential but not the wavefunction and $\Delta(0) = 0$. The equation is used to develop low-momentum expansions for Δ, and to derive upper and lower bounds for the expansion parameters.

1. INTRODUCTION

One may solve the problem of potential scattering by eliminating direct reference to the wavefunction $u(r)$ in calculating the phase shift δ.[1,2] In effect one makes a change of variables

$$u(r) = A \sin k[\epsilon(r) + r],$$

where A is independent of r. Physically, $k\epsilon(r)$ is the phase shift that would exist if the potential $V(x)$ were replaced by one equal to $V(x)$ for $x < r$ and equal to zero for $x > r$. One finds a first-order differential equation for $\epsilon(r)$ in terms of $V(r)$, with $u(r)$ eliminated, and the phase shift for the original problem is given by $k\epsilon(\infty)$. Several advantages of this approach, especially in low-energy scattering, are discussed in Ref. 1.

A treatment similar to that in Ref. 1 can be given for the case of the off-energy-shell matrix elements corresponding to the given potential. These matrix elements have been the subject of considerable discussion in low- and intermediate-energy physics in recent years, particularly with reference to the two-nucleon interaction.[3] Although a great deal of information about the two-nucleon interaction has been assembled, this is almost entirely based on elastic-scattering experiments and analysis of these in terms of phase shifts. More complicated processes involving nucleons require knowledge of off-energy-shell matrix elements, or quasiphase parameters,[4] which describe these elements in the same way that phase shifts describe on-shell matrix elements. To extrapolate the matrix element off the energy shell it is necessary to employ some model, generally a potential model, which is fitted to the phase-shift data. In recent years the use of these elements has become important in systems such as infinite nuclear matter,[5] finite nuclei,[6] three-body scattering,[7] and nucleon–nucleon bremsstrahlung.[8] Thus the properties of these elements and the validity of the models must be studied.

Many models of the interaction are local potentials which contain a hard core, a phenomenological intermediate region, and a tail given by the one-pion-exchange potential (OPEP). However it is found[9] that different potentials have the same value for r greater than a certain r', and differ only for r less than r' (although they all fit the scattering data)—and r' is a smaller distance than that at which OPEP dominates. Thus one may say that scattering experiments have probed the potential down to r', but not closer. It is therefore of interest to express the off-shell matrix element in a form which exhibits the dependence on the different regions of r space. This is accomplished by the equation we shall describe which gives the quasiphase Δ as $\Delta(\infty)$, where $\Delta(r)$ obeys a differential equation depending only upon $V(x)$ for $x < r$.

The paper closely parallels Ref. 1. In Sec. 2 we review the results of that paper, derive the equation for $\Delta(r)$, and discuss a low-energy expansion analogous to the effective-range expansion. In Sec. 3 we discuss upper and lower bounds of varying degrees of simplicity for the expansion parameters, and present some calculations for elementary potentials. Section 4 deals with the case of a hard core together with a finite exterior potential.

For simplicity we consider S waves only.

2. DIFFERENTIAL EQUATION AND LOW-ENERGY EXPANSION

Let $V(x)$ be the given potential and consider the potential $V_r(x)$ given by

$$V_r(x) = \begin{cases} V(x), & \text{for } x \geq r, \\ 0, & \text{for } x > r. \end{cases} \qquad (2.1)$$

* Permanent address: Department of Physics, Brooklyn College of City University of New York, Brooklyn, New York.

[1] Roger F. Dashen, J. Math. Phys. **4**, 388 (1963).

[2] S. Franchetti, Nuovo Cimento **6**, 601 (1957); F. Calogero, Nuovo Cimento **27**, 261 (1963).

[3] Rev. Mod. Phys. **39**, No. 3 (1967). This issue contains a report of the Gainsville Conference on the Nucleon–Nucleon Interaction, and a summary of the present state of knowledge in this field.

[4] M. I. Sobel, Phys. Rev. **138B**, 1517 (1965).

[5] K. A. Brueckner and K. S. Masteron, Phys. Rev. **128**, 2267 (1962).

[6] T. T. S. Kuo and G. E. Brown, Nucl. Phys. **85**, 40 (1966).

[7] C. A. Lovelace, in *Strong Interactions and High Energy Physics*, R. G. Moorhouse, Ed. (Plenum Press, Inc., New York, 1964).

[8] A. H. Cromer and M. I. Sobel, Phys. Rev. **152**, 1351 (1966).

[9] H. Feshbach and A. K. Kerman, Comments Nucl. Particle Phys. **4**, 132 (1967).

If $k\epsilon(r)$ is the phase shift due to potential $V_r(x)$, then[1]

$$\epsilon'(r) = -k^{-2}V(r)\sin^2 k[\epsilon(r) + r]. \tag{2.2}$$

Here $k = E^{\frac{1}{2}}$, where E is the energy.[10] From Eq. (2.1) it is obvious that $\epsilon(0) = 0$, so that this boundary condition with Eq. (2.2) determine ϵ everywhere.

The quasiphase Δ for $V(x)$ is given by[4]

$$\Delta = -\frac{1}{k'}\int_0^\infty dx\, V(x)\sin k'x\, v(x), \tag{2.3}$$

where $v(x)$ is the solution of

$$v''(x) + [k^2 - V(x)]v(x) = 0, \tag{2.4}$$

subject to $v(0) = 0$ and $v(x) \sim \sin(kx + \delta)$ as $x \to \infty$. Δ describes the S-wave part of the half-off-shell element of the T matrix

$$\langle \mathbf{k}'|\, T(E)\, |\mathbf{k}\rangle, \quad \text{with} \quad k^2 = E \neq k'^2,$$

and

$$\langle \mathbf{k}'|\, T(E)\, |\mathbf{k}\rangle = -(2\pi^2 k)^{-1}\sum_l (2l+1)\exp[i\delta_l(k)]\Delta_l P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}'). \tag{2.5}$$

As $k' \to k$, $\Delta_l \to \sin\delta_l$.

If in Eq. (2.3) we replace the upper limit by r, the resulting function would still depend on $V(x)$ for all x, because to determine the normalization of $v(x)$ one would still have to integrate out to $x \to \infty$ (or beyond the range of the potential). Instead we consider the quasiphase $\Delta(r)$ which results from the potential $V_r(x)$:

$$\Delta(r) = -\frac{1}{k'}\int_0^\infty dx\, V_r(x)\sin k'x\, v_r(x) = -\frac{1}{k'}\int_0^r dx\, V(x)\sin k'x\, v_r(x). \tag{2.6}$$

To find a first-order equation for $\Delta(r)$ we must differentiate Eq. (2.6) with respect to r, but some care must be taken in describing $v_r(x)$. Suppose $u_r(x)$ is the solution of

$$u_r''(x) + [k^2 - V_r(x)]u_r(x) = 0, \tag{2.7}$$

subject to $u_r(0) = 0$, and $u_r'(0) =$ some arbitrary number (say k). Then $u_r(x) = u(x)$ for $x \leq r$, and $u_r(x) = A_r \sin k[\epsilon(r) + x]$; here $u(x)$ means $u_\infty(x)$. Furthermore, according to the definition of Δ, $v_r(x) = u_r(x)/A_r$. Thus Eq. (2.7) becomes

$$A_r\Delta(r) = -\frac{1}{k'}\int_0^r dx\, V(x)\sin k'x\, u(x), \tag{2.8}$$

with integrand independent of r. Using Eq. (2.2) we find

$$dA_r/dr = -k\epsilon'(r)u(r)\cos k[\epsilon(r) + r]\{\sin k[\epsilon(r) + r]\}^{-2} = k^{-1}V(r)u(r)\cos k[\epsilon(r) + r], \tag{2.9}$$

so that Eq. (2.9) becomes

$$\Delta'(r) = -V(r)\sin k[\epsilon(r) + r]\{(1/k')\sin k'r + (1/k)\cos k[\epsilon(r) + r]\Delta(r)\}. \tag{2.10}$$

With boundary condition $\Delta(0) = 0$, Eqs. (2.10) and (2.2) determine $\Delta(r)$ and $\epsilon(r)$ for all r, and clearly $\Delta = \Delta(\infty)$. As $k' \to k$ we can let $\Delta(r) \to \sin k\epsilon(r)$ and find that Eq. (2.10) reduces to Eq. (2.2), as it should.

The advantages of this method relative to the more common technique of solving the Schrödinger equation and integrating over the potential are similar to the advantages[1] which hold for $\epsilon(r)$. These first-order equations are particularly well suited to numerical calculations: If V does not change sign, then $|\epsilon(r)|$ is monotonically increasing so that it is easier to determine the errors of a numerical calculation. Although $|\Delta(r)|$ is not necessarily monotonically increasing, it is still a slowly varying function, as compared with the wavefunction. Furthermore, the fact that Eq. (2.10) is linear leads, as we will see, to considerable simplifications. Finally, as stated above, $\Delta(r)$ depends only on $V(x)$ for $x < r$. Thus if $V(x)$ can be regarded as "model-independent" for $x > r'$, then, starting with $\epsilon(\infty)$, one can consider $\epsilon(x)$ model-independent for $r' < x < \infty$, and likewise for $\Delta(x)$, i.e., the model-dependent parameter will become $\Delta(r')$ rather than $\Delta(\infty)$ and the two regions of space ($x > r'$ and $x < r'$) are effectively separated.

To simplify further discussion we consider the case of low energies. From Eq. (2.2) one can derive[1] an expansion for $\epsilon(r)$:

$$\epsilon(r) = -\alpha(r) - \beta(r)k^2, \tag{2.11}$$

for small k, where

$$\alpha'(r) = V(r)[\alpha(r) - r]^2, \quad \alpha(0) = 0, \tag{2.12}$$

and a similar equation holds for β. $\alpha(\infty)$ is the scattering length and the effective range is given in terms of $\alpha(\infty)$ and $\beta(\infty)$.

In the case of Eq. (2.10) let us first define

$$\eta(r) = \Delta(r)/[\sin k\epsilon(r)], \tag{2.13}$$

and find that η obeys

$$\eta'(r) = -V(r)[\cos kr + \cot k\epsilon(r)\sin kr] \times [(1/k')\sin k'r - (1/k)\sin kr\eta(r)]. \tag{2.14}$$

[10] We use units such that $\hbar = 2m = 1$.

The boundary value $\eta(0)$ is not obvious, but if we assume[11] that $V(r)$ diverges at $r = 0$ less rapidly than r^{-2}, then we find that Δ and $\sin k\epsilon$ behave in the same way as $r \to 0$ and $\eta(0) = 1$.

Now we expand $\eta(r)$ independently in powers of k and k', using Eq. (2.11) for $\epsilon(r)$. The zeroth-order term is the constant unity, and the next terms are in k^2 and k'^2 with equal and opposite coefficients. Thus,[12]

$$\eta(r) = 1 + (\tfrac{1}{6})\sigma(r)(k'^2 - k^2) + O(k^4) \quad (2.15)$$

and σ obeys

$$\sigma'(r) = U_\sigma(r)[\sigma(r) + r^2] = V(r)[1 - r/\alpha(r)] \times r[\sigma(r) + r^2], \quad \sigma(0) = 0. \quad (2.16)$$

In Sec. 4 we will find a similar equation for σ in the presence of a hard core, with the boundary condition at the core radius r_c. It will be of interest then to expand σ in powers of $(r - r_c)$. We can see to what extent the contribution to σ for r not far from r_c is independent of the parameters describing the potential. For comparison we expand, here, $\sigma(r)$ and $\alpha(r)$ in powers of r. Suppose that, with $\beta < 2$,

$$V(r) = r^{-\beta}(V_0 + V_1 r + \cdots), \quad (2.17)$$

so that, using Eq. (2.12),

$$U_\sigma(r) = r^{-1}[-(3 - \beta) + u_1 r + \cdots]. \quad (2.18)$$

Then we find

$$\sigma(r) = r^2[-(3 - \beta)/(5 - \beta) + \{2/[(6 - \beta)(5 - \beta)]\}u_1 r + \cdots]. \quad (2.19)$$

One might be interested in choosing from among a series of potentials which fit the phase shift (or the scattering length) the one that gives an off-shell element which has the minimum dependence on the potential parameters. Thus, if the potential were fixed for r greater than some value r', one would try to make $\sigma(r)$ for small r least dependent on the potential. Once β is fixed the leading term is independent of V_0. As for the next term, we note that, for example,

$$\text{for } \beta = 0: \quad u_1 = -(4V_0)^{-1}3V_1,$$
$$\text{for } \beta = 1: \quad u_1 = -(3V_0)^{-1}(2V_0^2 + V_1). \quad (2.20)$$

If the potential depends upon a parameter λ one could choose λ such that $\partial u_1(\lambda)/\partial\lambda = 0$. For example, suppose a Yukawa potential $\gamma(e^{-\mu r})/r$ is to be fitted to a positive-scattering length a. This defines a relation $\gamma = \gamma(\mu)$. For this case $u_1 = -\frac{1}{3}(2\gamma - \mu)$, so that one could choose μ such that $\gamma'(\mu) = \frac{1}{2}$. Since the function $\gamma(\mu)$ is roughly a positive quadratic, there will be a solution for some $\mu > 0$.

[11] We continue to make this assumption in the rest of the paper.

[12] The low-energy expansion has been previously derived by other methods; see A. H. Cromer, Rev. Mod. Phys. **39**, 716 (1967).

3. UPPER AND LOWER BOUNDS

A major part of Ref. 1 is devoted to providing upper and lower bounds for $\alpha(r)$ in terms of integrals over $V(r)$. These can in turn be applied to give bounds for $\sigma(r)$, especially since Eq. (2.16) is linear and can be solved by quadratures. Because of the zero of α at $r = 0$ it is convenient to change variables and let $\tau(r) = \sigma(r)\alpha(r)$. It is then easy to show, using Eqs. (2.16) and (2.12), that

$$\tau'(r) = U_\tau(r)[\tau(r) + r^3] = V(r)[\alpha(r) - r][\tau(r) + r^3] \quad (3.1)$$

with $\tau(0) = 0$. Thus,

$$\tau(r) = \int_0^r dy\, y^3 U_\tau(r) \exp\left[\int_y^r dx\, U_\tau(x)\right]. \quad (3.2)$$

In the low-energy expansion of $\Delta(r)$, $-\tau/6$ is the coefficient of the term in kk'^2. Let us first consider the case of the purely repulsive potential. Then from Eq. (2.12) we have

$$0 < \alpha(r) < r. \quad (3.4)$$

More precise bounds are[1]

$$\alpha_<(r) \equiv \int_0^r dy\, y^2 V(y) \exp\left[-2\int_y^r dx\, xV(x)\right] < \alpha(r) < \int_0^r dy\, y^2 V(y) \equiv \alpha_>(r). \quad (3.5)$$

Depending on r, the upper bound may be better in Eq. (3.4) or (3.5). It is obvious that $-r^3 < \tau(r) < 0$, but we obtain the following bounds, useful at $r \to \infty$:

$$\int_0^r dy\, y^3 V(y)[y - \alpha_>(y)] \times \exp\left\{-\int_y^r dx\, V(x)[x - \alpha_<(x)]\right\} < |\tau(r)| < \int_0^r dy\, y^3 V(y)[y - \alpha_<(y)] \times \exp\left\{-\int_y^r dx\, V(x)[x - \alpha_>(x)]\right\}. \quad (3.6)$$

With simple functional forms for $V(r)$, although $\alpha(r)$ [and hence $\tau(r)$] can be found only numerically, one finds that $\alpha_>$ and $\alpha_<$ can often be found in simple analytic forms. Likewise the integrals in Eq. (3.6) may be carried out simply. Even when this is not so it is found that the integrals can be done by expanding in powers of r (and x and y), with results converging so rapidly that they are useful even for finding $\tau(\infty)$. In contrast, experience shows that if we try to solve Eqs. (2.12) and (3.1) by expanding in powers of r the

convergence is very slow. Several simpler and weaker bounds can be obtained from Eqs. (3.5) and (3.6). First we note the simpler lower bound for α:

$$\alpha(r) > \alpha_<(r) > \int_0^r dy\, y^2 V(y) \times \exp\left[-2\int_0^r dx\, xV(x)\right] \equiv \bar{\alpha}_<(r) \quad (3.7)$$

and, from Eq. (3.6), simpler bounds for τ:

$$\int_0^r dy\, y^3 V(y)[y - \alpha_>(y)] \times \exp\left\{-\int_0^r dx\, V(x)[x - \alpha_<(x)]\right\} < |\tau(r)| < \int_0^r dy\, y^3 V(y)[y - \alpha_<(y)]. \quad (3.8)$$

Either Eq. (3.6) or (3.8) can be further simplified by replacing $\alpha_<$ by $\bar{\alpha}_<$ or 0 in various places. To illustrate these results we calculate various bounds for $\tau(r)$ for two examples: (a) the spherical barrier, in which case the exact solution is given analytically, and (b) the Yukawa potential.

(a) Let

$$V(r) = \begin{cases} r_1^{-2}, & r < r_0, \\ 0, & r > r_0. \end{cases} \quad (3.9)$$

Then the exact solution for α is[1]

$$\alpha(r) = r - r_1 \tanh(r/r_1), \quad r < r_0,$$
$$\alpha(r) = \alpha(r_0), \quad r \geq r_0. \quad (3.10)$$

For τ the integrals in Eq. (3.2) can be performed to yield, with $\rho = r/r_1$,

$$|\tau(r)| = -\tau(r) = r_1^3[\rho^3 + 6\rho - (3\rho^2 + 6)\tanh\rho]. \quad (3.11)$$

It is understood that this expression and the bounds we will describe give $\tau(r)$ for $r < r_0$, and that for $r \geq r_0$, $\tau(r) = \tau(r_0)$. The very simplest upper and lower bounds are given by Eq. (3.8) with $\alpha_<$ taken equal to zero. These give

$$\left(\frac{\rho^5}{5} - \frac{\rho^7}{21}\right)\exp\left(-\frac{\rho^2}{2}\right) < \frac{|\tau(r)|}{r_1^3} < \frac{\rho^5}{5}. \quad (3.12)$$

The improved bounds obtained from Eq. (3.6) [with use of equation (3.5)] are also calculated in the form of a series in ρ. These curves, together with some others given in closed form, are shown in Fig. 1. The general conclusion is that the shorter the range of the potential the better are the bounds found. Furthermore, as we should expect, the approximation made in going from Eq. (3.6) to (3.8) is poorer for longer-range potentials.

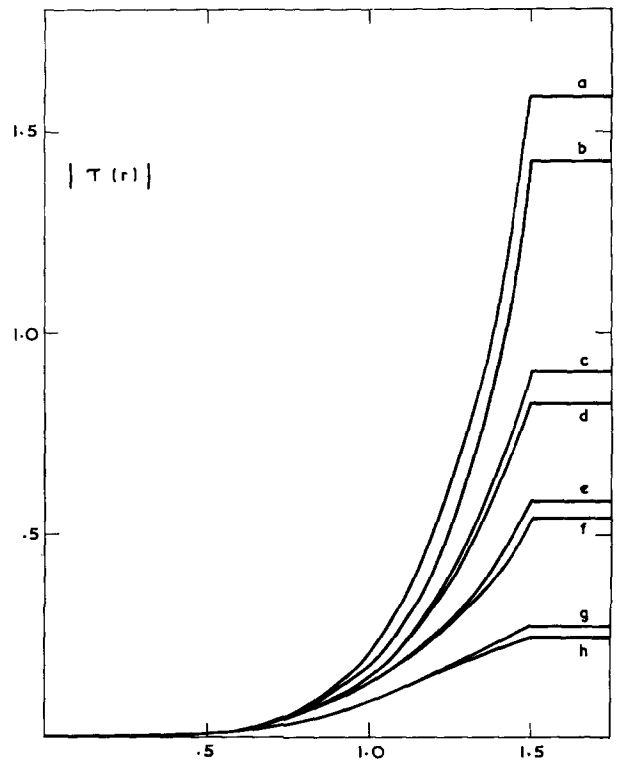


FIG. 1. Upper and lower bounds for the potential of Eq. (3.9). Ordinate in units of r_1^3, abscissa in units of r_1, $r_0/r_1 = 1.5$. a: Upper bound (U.B.), Eq. (3.8), $\alpha_< \to 0$. b: U.B. Eq. (3.8), $\alpha_< \to \bar{\alpha}_<$ of Eq. (3.7). c: U.B. Eq. (3.6). d: Exact equation (3.11). e: Lower bound (L.B.), Eq. (3.6). f: L.B. Eq. (3.6), $\alpha_< \to 0$. g: L.B. Eq. (3.8), $\alpha_< \to \bar{\alpha}_<$. h: L.B. Eq. (3.8), $\alpha_< \to 0$.

(b) As a second example, we calculate the simplest bounds for the Yukawa potential

$$V(r) = (e^{-\sigma r/r})/(rr_1). \quad (3.13)$$

These results, from Eq. (3.8) with $\alpha_< = 0$, are somewhat lengthy and we write only the values at $r \to \infty$:

$$6\frac{r_0^4}{r_1}\exp\left(-\frac{r_0}{r_1}\right)\left(1 - \frac{1}{3}\frac{r_0}{r_1}\right) < |\tau(\infty)| < 6\frac{r_0^4}{r_1}. \quad (3.14)$$

Some curves are shown in Fig. 2. Here, too, it is evident that the bounds are closer for a potential of shorter range.

For the case of a purely attractive potential we find $\alpha(r) < 0$, $U_r > 0$, and so from Eq. (3.2) we have

$$\tau(r) = \int_y^r dy\, y^3 |V(y)|\,(y + |\alpha(y)|) \times \exp\left[\int_y^r dx\, |V(x)|\,(x + |\alpha(x)|)\right]. \quad (3.15)$$

The inequalities in Ref. 1 for α are

$$|\alpha(r)| > \int_0^r dy\, y^2 |V(y)| \exp\left[2\int_y^r dx\, x\, |V(x)|\right] > \int_0^r dy\, y^2 |V(y)|. \quad (3.16)$$

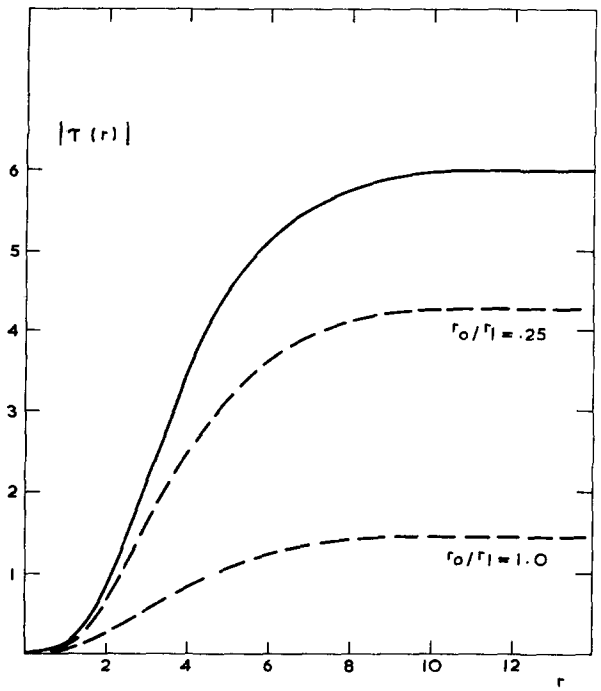


FIG. 2. Upper and lower bounds for the potential of Eq. (3.13). Ordinate in units of r_0^4/r_1, abscissa in units of r_0. Solid line is upper bound, independent of r_0/r_1 in these units. Dashed lines are lower bounds with r_0/r_1 as indicated.

Thus from Eq. (3.15) we obtain only lower bounds for τ. The simplest is found by using $\alpha = 0$ in Eq. (3.15), the next simplest by using the expression on the right of (3.16) for α, and the next by using the expression in the middle of (3.16). In each case, as before, the exponential may or may not be replaced by unity to give a simpler result.

The quantity $\int_0^r dy\, y^2 V(y)$ is the Born-approximation value for α (regardless of the sign of V). Furthermore, if the Born approximation is valid, then $|\alpha(r)| \ll r$, so that in this case

$$\tau(r) = \tau_{\text{Born}}(r) = -\int_0^r dy\, y^4 V(y). \tag{3.17}$$

In the case of repulsive potentials we have seen that $|\tau_{\text{Born}}|$ is a simple upper bound for $-\tau$. Generally if the potential is always of one sign, τ_{Born} is an algebraic lower bound for τ.

It is interesting that the Born-approximation expression for β [of Eq. (2.11)] is given by[1]

$$-\tfrac{1}{3}\int_0^r dy\, y^4 V(y).$$

Since $\beta(\infty)$ is related to the effective range r_0 and the scattering length a the half-off-shell element is given in terms of these on-shell parameters

$$\sigma(\infty) = \tau(\infty)/a = 3\beta(\infty)/a = 3a(\tfrac{1}{2}r_0 - \tfrac{1}{3}a). \tag{3.18}$$

This relation at low energies expresses the well-known fact that in the Born approximation any off-shell matrix element is the Fourier transform of the potential at some momentum transfer and hence equals some on-shell element.

One can derive an expansion for the quasiphase Δ for k small but k' arbitrary. This is the case important from the point of view of double-scattering processes at low energies.[13] One must carry out integrals of half-off-shell t matrices over all values of the intermediate momentum k' with the energy k^2 of the T operator fixed at some (possibly small) value. Let us start with Eq. (2.14) and let

$$\eta(r) = \eta_0(r) + k^2\eta_2(r) + \cdots. \tag{3.19}$$

Then $\eta_0(r)$ obeys

$$\eta_0'(r) = V(r)\left[1 - \frac{r}{\alpha(r)}\right] r\left(\eta_0 - \frac{\sin k'r}{k'r}\right), \quad \eta_0(0) = 1. \tag{3.20}$$

Again we introduce the change of variables, $\xi_0(r) = \eta_0(r)\alpha(r)$ and find

$$\xi_0'(r) = V(r)[\alpha(r) - r][\xi(r) - (1/k')\sin k'r], \quad \xi_0(0) = 0. \tag{3.21}$$

So we find the solution, with $U(r) = V(r)[\alpha(r) - r]$,

$$\eta_0(r) = -\frac{1}{k'\alpha(r)}\int_0^r dy\, U(y)\sin k'y \exp\left[\int_y^r dx\, U(x)\right]. \tag{3.22}$$

One can obtain from this a simple upper bound for η. In the case of repulsive potentials,

$$|\eta_0(r)| < \frac{1}{k'\alpha(r)}\int_0^r dy\, yV(y) \tag{3.23}$$

and

$$\eta_0(\infty) < \frac{1}{k'a}\int_0^\infty dy\, yV(y). \tag{3.24}$$

As before, more stringent bounds can be placed on η for this case and for the case of a purely attractive potential. The equation for $\eta_2(r)$ is more complicated, but one obtains a bound of the form

$$\eta_2(\infty) < \frac{1}{k'a}\int_0^\infty dy\, V(y)g(y), \tag{3.25}$$

where $g(y)$ is independent of k'. Bounds such as these would be useful in approximating the high-momentum part of double-scattering integrals, as mentioned above.

Finally in this section we note that one may study

[13] M. I. Sobel, Phys. Rev. **152**, 1385 (1966).

the dependence of the off-shell element on a parameter λ appearing in the potential. For example, there might be a range of values such that for λ in this range the on-shell parameters are fit, but the off-shell elements vary with λ. In this case, going back to Eq. (3.1) for τ,

$$\left(\frac{\partial\tau}{\partial\lambda}\right)' = \frac{\partial U_\tau}{\partial\lambda}(\tau + r^3) + U_\tau\frac{\partial\tau}{\partial\lambda}, \tag{3.26}$$

and so

$$\frac{\partial\tau}{\partial\lambda}(r) = \int_0^r dy\,\frac{\partial U_\tau}{\partial\lambda}(y)[\tau(y) + y^3]\exp\left[\int_y^r dx\,U_\tau(x)\right]. \tag{3.27}$$

In the case of a repulsive potential, $U_\tau < 0$, $0 < \alpha(r) < r$, and $-r^3 < \tau(r) < 0$, so we have

$$\begin{aligned}\left|\frac{\partial\tau}{\partial\lambda}(r)\right| &< \int_0^r dy\left|\frac{\partial U_\tau}{\partial\lambda}\right| y^3\exp\left[-\int_y^r dx\,|U_\tau(x)|\right]\\ &< \int_0^r dy\,y^4\frac{\partial V}{\partial\lambda}(y)\exp\left[-\int_y^r dx\,xV(x)\right]\\ &< \int_0^r dy\,y^4\frac{\partial V}{\partial\lambda}(y).\end{aligned} \tag{3.28}$$

4. HARD CORE

If the potential is specified as

$$V(r) = \begin{cases}+\infty, & r < r_c,\\ V(r), & r > r_c,\end{cases} \tag{4.1}$$

the differential equation for $\Delta(r)$ is found to be identical to Eq. (2.10) for $r > r_c$. The boundary condition is now at r_c and it is

$$\Delta(r_c) = \Delta_c = -(k/k')\sin k'r_c. \tag{4.2}$$

The equations for $\epsilon(r)$, $\alpha(r)$, $\eta(r)$, $\sigma(r)$, and $\tau(r)$ are the same as in Secs. 2 and 3, for $r > r_c$, with boundary conditions

$$\begin{aligned}&\epsilon(r_c) = -r_c,\\ &\alpha(r_c) = r_c,\\ &\eta(r_c) = \eta_c = \frac{(\sin k'r_c)/k'}{(\sin kr_c)/k},\\ &\sigma(r_c) = -r_c^2,\\ &\tau(r_c) = -r_c^3.\end{aligned} \tag{4.3}$$

Now if we assume $V(r)$ can be expanded as

$$V(r) = V_0 + V_1(r - r_c) + \cdots \tag{4.4}$$

and

$$\sigma(r) = -r_c^2 + \sigma_1(r - r_c) + \sigma_2(r - r_c)^2 + \cdots, \tag{4.5}$$

we find that $\sigma_1 = \sigma_2 = 0$, and the leading terms are

$$\sigma(r) = -r_c^2\left[1 + \tfrac{2}{3}(V_0r_c^2)\left(\frac{r - r_c}{r_c}\right)^3 + \cdots\right]. \tag{4.6}$$

In fact a similar expansion can be found for the complete quasiphase or, in terms of $\eta(r)$,

$$\eta(r) = \eta_c[1 + \tfrac{1}{3}V_0(k'\cot k'r_c - k\cot kr_c) \times (r - r_c)^3 + \cdots]. \tag{4.7}$$

Thus the variation of the quasiphase outside the hard core is slower than one might have expected. If $V(r)$ is only uncertain in some region $r_c < r < r'$ and if $r' - r_c$ is small enough so that only the two terms in Eq. (4.7) need be used, then $\Delta(\infty)$ may be considered a measure of $V_0 = V(r_c)$, i.e., from $\Delta(\infty)$ or $\eta(\infty)$, which might be determined experimentally,[14] one can integrate backwards to find $\eta(r')$, and then determine V_0. For the two-nucleon interaction[9] in $T = 1, J = 0$ states, $(r' - r_c)/r_c \leqslant 0.5$. Of course it should be remembered that r_c is not precisely fixed by phenomenological analysis.

The bounds considered in the previous section can also be applied in the case of the hard core. Let us suppose that outside the repulsive core the potential is attractive. The analog of the bound in Eq. (3.16) is

$$\alpha(r) < r_c + \int_{r_c}^r dy\,yV(y)(y - 2r_c) \times \exp\left[-2\int_y^r dx\,xV(x)\right]. \tag{4.8}$$

The integral may be positive or negative. In the former case this equation is not useful since we always have $\alpha(r) < r_c$. Equation (3.2) for τ becomes

$$\tau(r) = -r_c^3 + \int_{r_c}^r dy\,(y^3 - r_c^3)\,|U_\tau(y)| \times \exp\left[\int_y^r dx\,|U_\tau(x)|\right]. \tag{4.9}$$

Now $|\alpha(r) - r| = |\alpha(r) - r_c| + |r - r_c|$. Thus we always have a lower bound $|\alpha(r) - r| > |r - r_c|$ and, depending on the sign of the integral in equation (4.8), we may obtain an improved bound. In either case we obtain a lower bound for τ using Eq. (4.9). The simplest case, if the exponential is replaced by unity, is

$$\tau > -r_c^3 + \int_{r_c}^r dy(y^3 - r_c^3)(y - r_c)\,|V(y)|. \tag{4.10}$$

[14] Recently there have been attempts to use experimental results on proton–proton bremsstrahlung to determine quasiphases; see A. H. Cromer and M. I. Sobel, Rev. Mod. Phys. **39**, 717 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Matrix Elements in Systems with Nonunitary Symmetry

A. AVIRAN AND J. ZAK*
Department of Physics, Technion — Israel Institute of Technology, Haifa, Israel

(Received 10 April 1968)

The Eckart–Wigner theorem is generalized to include nonunitary groups. The proof is based on the connection between corepresentations of a nonunitary group and the representations of its unitary part. All possible cases of the corepresentations have been considered, and general expressions for matrix elements of operators with given symmetry have been obtained. It has been shown that the antiunitary symmetry leads, in general, to additional connections between different matrix elements.

I. INTRODUCTION

Eckart[1] and Wigner[2] have shown that, when taking into account the symmetry properties of the system (i.e., the symmetry properties of the wavefunctions and the physical operator) with the help of the representation theory, one can write simple relations between matrix elements of operators which have certain similar properties. Their work on this subject culminated in the Eckart–Wigner theorem. Both Eckart and Wigner dealt with physical systems of spherical symmetry. Koster[3] generalized this theorem in order to make it applicable to other groups. The generalization of the Eckart–Wigner theorem given by Koster is as follows:

$$(\varphi_\alpha^i|\, P_\sigma^k \,|\varphi_\beta^j) = A_1 U_{\sigma(\alpha\beta)} + A_2 U_{\sigma+n_k,(\alpha\beta)} + \cdots,$$

where i, j, and k denote irreducible representations of the symmetry group of the system G, $A_1, A_2, \cdots$ are constants called "reduced matrix elements," and U is the matrix which transforms the direct product $\Delta^{i*} \otimes \Delta^j$ into a special reduced form (Δ^i denotes the ith irreducible representation of G). The number of terms on the right-hand side of the above equation is equal to the number of times the irreducible representation Δ^{k*} appears in the reduced form of the direct product.

Koster's work generalized the theorem for all symmetry groups connected with spatial geometry. An additional important symmetry of physical systems is time inversion; the symmetry associated with time cannot be expressed by geometric symmetry groups. For this reason it was necessary to develop the concept of a new type of group, the nonunitary group.

Nonunitary groups and their corepresentations are of great importance in magnetic materials. In such materials the antiunitarity element is a product of time inversion and an element of the space group, which together leave the spin system unchanged. In nonmagnetic materials time inversion is itself a symmetry element. In every case where an antiunitary element is added to the ordinary space group there is a need to deal with corepresentations, and therefore also with the Eckart–Wigner theorem for them.

In Sec. II we generalize the Eckart–Wigner theorem for nonunitary groups; the derivation is parallel to Koster's. Since corepresentations are constructed in three possible ways (according to Wigner's classification) out of the irreducible representations of the unitary subgroup of the nonunitary group, three different cases appear in the generalization.

Relations between matrix elements, due to the addition of the antiunitary element to the symmetry group of the system are derived in Sec. III. Examples are given in Sec. IV.

II. ECKART–WIGNER THEOREM FOR NONUNITARY GROUPS

Consider a nonunitary group G of order N, consisting of unitary and antiunitary operators. These operators are generally denoted by O, while the unitary operators are denoted by u, and the antiunitary ones by a. The irreducible corepresentations of G are denoted by D^i, D^j, D^k, $\cdots$, and their dimensions are $n_i, n_j, n_k, \cdots$, respectively.

The basis functions φ_α^l of the lth irreducible corepresentation ($\alpha = 1, \cdots, n_l$) satisfy the equation

$$O\varphi_\alpha^l = \sum_\delta D^l(O)_{\delta\alpha}\varphi_\delta^l, \tag{1}$$

where $D^l(O)_{\delta\alpha}$ are matrix elements in the lth irreducible corepresentation. Without loss of generality, we can assume these matrices to be unitary.[4]

We are interested in matrix elements of operators with certain transformation properties between the states φ_α^i and φ_β^j.

* Present address: Department of Physics and Francis Bitter National Magnet Laboratory, M.I.T., Cambridge, Massachusetts.

[1] C. Eckart, Rev. Mod. Phys. **2**, 305 (1930).

[2] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).

[3] G. F. Koster, Phys. Rev. **109**, 227 (1958).

[4] J. O. Dimmock, J. Math. Phys. **4**, 1307 (1963).

Let P^k_σ be an operator which transforms according to the kth irreducible corepresentation of G:

$$OP^k_\sigma O^{-1} = \sum_{\rho=1}^{n_k} D^k(O)_{\rho\sigma} P^k_\rho. \tag{2}$$

Consider the matrix element $(\varphi^i_\alpha| P^k_\sigma |\varphi^j_\beta)$, denoted from now on by $P^{ikj}_{\alpha\sigma\beta}$. Using Eqs. (1) and (2), we can write for the unitary operators u:

$$\begin{aligned} P^{ikj}_{\alpha\sigma\beta} &= (\varphi^i_\alpha| P^k_\sigma |\varphi^j_\beta) = (u\varphi^i_\alpha| uP^k_\sigma u^{-1} |u\varphi^j_\beta) \\ &= \sum_{\delta,\rho,\epsilon} D^i(u)^*_{\delta\alpha} D^k(u)_{\rho\sigma} D^j(u)_{\epsilon\beta} (\varphi^i_\delta| P^k_\rho |\varphi^j_\epsilon). \end{aligned}$$

By summing the last equation over u and dividing it by $N/2$, we obtain

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N} \sum_u \sum_{\delta,\rho,\epsilon} D^i(u)^*_{\delta\alpha} D^j(u)_{\epsilon\beta} D^k(u)_{\rho\sigma} P^{ikj}_{\delta\rho\epsilon}. \tag{3}$$

Performance of the parallel procedure for antiunitary operators, use of their property[2]

$$(\psi, \phi) = (a\psi, a\phi)^*,$$

and Eqs. (1) and (2) yield

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N} \sum_a \sum_{\delta,\rho,\epsilon} D^i(a)_{\delta\alpha} D^j(a)^*_{\epsilon\beta} D^k(a)^*_{\rho\sigma} P^{ikj*}_{\delta\rho\epsilon}. \tag{4}$$

In Eqs. (3) and (4) we have products $D^i(u)^*_{\delta\alpha} D^j(u)_{\epsilon\beta}$ and $D^i(a)_{\delta\alpha} D^j(a)^*_{\epsilon\beta}$, which are matrix elements of the direct product $D'(O) = D^i(O)^* \otimes D^j(O)$. In Eq. (3) elements of $D'(O)$ appear:

$$D'(u)_{(\delta\epsilon)(\alpha\beta)} = D^i(u)^*_{\delta\alpha} D^j(u)_{\epsilon\beta}; \tag{5}$$

while in Eq. (4) elements of $D'(O)^*$ appear:

$$D'(a)^*_{(\delta\epsilon)(\alpha\beta)} = D^i(a)_{\delta\alpha} D^j(a)^*_{\epsilon\beta}. \tag{6}$$

In general, $D'(O)$ is reducible. Let us bring it to its reduced form $D(O)$, using a unitary transformation U.

For the unitary part of the group we can write

$$UD'(u)U^\dagger = D(u), \quad U^\dagger D(u) U = D'(u), \tag{7}$$

and for the antiunitary part of the group[4] we have

$$UD'(a)U^{\dagger *} = D(a), \quad U^\dagger D(a) U^* = D'(a). \tag{8}$$

The reduced form we are interested in is

$$D(O) = \begin{pmatrix} D^k(O)^* & & & & 0 \\ & D^k(O)^* & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & D^k(O)^* \\ 0 & & & & & & D^k(O) \\ & & & & & & & \cdot \\ & & & & & & & & \cdot \\ & & & & & & & & & \cdot \end{pmatrix},$$

i.e., the first m_k corepresentations are $D^k(O)^*$, where m_k is the number of times the irreducible corepresentation $D^k(O)^*$ appears in the direct product $D^i(O)^* \otimes D^j(O)$. All the other corepresentations are not equivalent to $D^k(O)^*$ (m_k is given by Karavaev[5]).

The above form can be written as

$$D^i(O)^* \otimes D^j(O)^* = D'(O) \cong m_k D^k(O)^* \dot{+} \text{ other irreducible corepresentations.}$$

$D(O)$ has been chosen in this particular form because $D^k(O)$ characterizes the given operator P^k_σ.

The elements of $D(O)$ in the first $m_k n_k$ rows are

$$\begin{aligned} &D(O)_{\eta\nu} = D^k(O)^*_{\eta\nu}; \quad \eta,\gamma = 1, \cdots, n_k, \\ &D(O)_{\eta\nu} = D^k(O)^*_{\eta-n_k,\nu-n_k}; \quad \eta,\nu = n_k + 1, \cdots, 2n_k, \\ &\cdots \\ &D(O)_{\eta\nu} = D^k(O)^*_{\eta-(m_k-1)n_k,\nu-(m_k-1)n_k}; \\ &\qquad \eta,\nu = (m_k - 1)n_k + 1, \cdots, m_k n_k. \end{aligned} \tag{9}$$

All the others vanish.

Using Eqs. (5) and (7), we obtain

$$\sum_{\eta,\nu} U^*_{\eta(\delta\epsilon)} D(u)_{\eta\nu} U_{\nu(\alpha\beta)} = D'(u)_{(\delta\epsilon)(\alpha\beta)} = D^i(u)^*_{\delta\alpha} D^j(u)_{\epsilon\beta}, \tag{10}$$

and, similarly, from Eqs. (6) and (8) we get

$$\sum_{\eta,\nu} U_{\eta(\delta\epsilon)} D(a)^*_{\eta\nu} U_{\nu(\alpha\beta)} = D'(a)^*_{(\delta\epsilon)(\alpha\beta)} = D^i(a)_{\delta\alpha} D^j(a)^*_{\epsilon\beta}. \tag{11}$$

Substitution of Eqs. (10) and (11) in Eqs. (3) and (4), respectively, yields

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N} \sum_u \sum_{\delta,\rho,\epsilon,\eta,\nu} U^*_{\eta(\delta\epsilon)} D(u)_{\eta\nu} U_{\nu(\alpha\beta)} D^k(u)_{\rho\sigma} P^{ikj}_{\delta\rho\epsilon}, \tag{12}$$

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N} \sum_a \sum_{\delta,\rho,\epsilon,\eta,\nu} U_{\eta(\delta\epsilon)} D(a)^*_{\eta\nu} U_{\nu(\alpha\beta)} D^k(a)^*_{\rho\sigma} P^{ikj*}_{\delta\rho\epsilon}. \tag{13}$$

The sums appearing in Eqs. (12) and (13) vanish for $\eta, \nu > m_k n_k$ because of the orthogonality relations existing for inequivalent irreducible corepresentations (see Ref. 4). Therefore we need consider only the sums with $\eta, \nu = 1, \cdots, m_k n_k$. Equations (12) and (13) do not enable us to use orthogonality relations for an irreducible corepresentation,[4] and we have to exploit the fact that irreducible corepresentations are constructed out of irreducible representations of the unitary subgroup [which will be denoted by $\Delta(u)$],

[5] G. F. Karavaev, Fiz. Tverd. Tela 6, 3676 (1964) [Sov. Phys.—Solid State 6, 2943 (1965)].

according to Wigner's classification.[2] For them we have the well-known orthogonality relations.[6]

Case A: $D^k(O)$ is a corepresentation of the first type[2]:

$$D^k(u) = \Delta^k(u);$$
$$D^k(a) = \Delta^k(aa_0^{-1})\beta = \Delta^k(ua_0a_0^{-1})\beta = \Delta^k(u)\beta.$$

Equation (12) [by using Eq. (9)] then has the form

$$\begin{aligned} P^{ikj}_{\alpha\sigma\beta} &= \frac{2}{N}\sum_u \sum_{\delta,\rho,\epsilon,\eta,\nu} U^*_{\eta(\delta\epsilon)} D(u)_{\eta\nu} U_{\nu(\alpha\beta)} \Delta^k(u)_{\rho\sigma} P^{ikj}_{\delta\rho\epsilon} \\ &= \frac{2}{N}\sum_{\delta,\rho,\epsilon} P^{ikj}_{\delta\rho\epsilon}\Bigg[\sum_{\eta,\nu=1}^{n_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)} \sum_u \Delta^k(u)^*_{\eta\nu}\Delta^k(u)_{\rho\sigma} \\ &\quad + \sum_{\eta,\nu=n_k+1}^{2n_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)} \sum_u \Delta^k(u)^*_{\eta-n_k,\nu-n_k}\Delta^k(u)_{\rho\sigma} \\ &\quad + \cdots + \sum_{\eta,\nu=(m_k-1)n_k+1}^{m_kn_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)} \\ &\quad \times \sum_u \Delta^k(u)^*_{\eta-(m_k-1)n_k,\nu-(m_k-1)n_k}\Delta^k(u)_{\rho\sigma}\Bigg]. \end{aligned} \tag{14}$$

The known orthogonality relations for matrix elements of an irreducible representation of a unitary group

$$\sum_u \Delta^k(u)_{\rho\sigma}\Delta^k(u)^*_{\eta\nu} = \frac{g}{n_k}\delta_{\rho\eta}\delta_{\sigma\nu}, \tag{15}$$

where g is the order of the group (in our case $g = N/2$), enable us to write Eq. (14) in the form

$$\begin{aligned} &P^{ikj}_{\alpha\sigma\beta} \\ &= \frac{1}{n_k}\sum_{\delta,\rho,\epsilon} P^{ikj}_{\delta\rho\epsilon}\Bigg[\sum_{\eta,\nu=1}^{n_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)}\delta_{\rho\eta}\delta_{\sigma\nu} + \cdots \\ &\quad + \sum_{\eta,\nu=(m_k-1)n_k+1}^{m_kn_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)}\delta_{\rho,\eta-(m_k-1)n_k}\delta_{\sigma,\nu-(m_k-1)n_k}\Bigg] \\ &= U_{\sigma(\alpha\beta)}\frac{1}{n_k}\sum_{\delta,\rho,\epsilon} U^*_{\rho(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon} \\ &\quad + U_{\sigma+n_k,(\alpha\beta)}\frac{1}{n_k}\sum_{\delta,\rho,\epsilon} U^*_{\rho+n_k,(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon} \\ &\quad + \cdots + U_{\sigma+(m_k-1)n_k,(\alpha\beta)}\frac{1}{n_k}\sum_{\delta,\rho,\epsilon} U^*_{\rho+(m_k-1)n_k,(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon}. \end{aligned} \tag{16}$$

Substitution of

$$A_{q+1} = \frac{1}{n_k}\sum_{\delta,\rho,\epsilon} U^*_{\rho+qn_k,(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon}, \quad q = 0, \cdots, m_k - 1, \tag{16'}$$

in Eq. (16) yields

$$\begin{aligned} P^{ikj}_{\alpha\sigma\beta} &= A_1 U_{\sigma(\alpha\beta)} + A_2 U_{\sigma+n_k,(\alpha\beta)} \\ &\quad + \cdots + A_{m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}. \end{aligned} \tag{17}$$

Let us now perform similar calculations for the antiunitary part. The relation $D^k(a) = \Delta^k(u)\beta$ defines the matrices of the antiunitary elements; β [or $D^k(a_0)$] is a unitary matrix:

$$\sum_\tau \beta^*_{\tau\sigma}\beta_{\tau\nu} = \sum_\tau (\beta^\dagger)_{\sigma\tau}\beta_{\tau\nu} = \delta_{\sigma\nu}. \tag{18}$$

Using orthogonality relations (15) and the definition of A_{q+1} [Eq. (16′)], one gets the following result from Eq. (13):

$$\begin{aligned} P^{ikj}_{\alpha\sigma\beta} &= A_1^* U_{\sigma(\alpha\beta)} + A_2^* U_{\sigma+n_k,(\alpha\beta)} \\ &\quad + \cdots + A^*_{m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}. \end{aligned} \tag{19}$$

Comparison of Eq. (17) and Eq. (19) (remembering the linear independence of the rows of U, which is a unitary matrix) immediately gives

$$A_p = A_p^*.$$

The Eckart–Wigner theorem for this case is therefore

$$\begin{aligned} P^{ikj}_{\alpha\sigma\beta} &= A_1 U_{\sigma(\alpha\beta)} + A_2 U_{\sigma+n_k,(\alpha\beta)} \\ &\quad + \cdots + A_{m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}, \end{aligned}$$

where the A's are real constants.

This property of the A's is a consequence of the transformation property of the operator (in this case, transformation according to a corepresentation of the first type).

Case B: $D^k(O)$ is a corepresentation of the second type:

$$D^k(u) = \begin{pmatrix} \Delta^k(u) & 0 \\ 0 & \Delta^k(u) \end{pmatrix};$$

$$\begin{aligned} D^k(a) &= \begin{pmatrix} 0 & \Delta^k(aa_0^{-1})\beta \\ -\Delta^k(aa_0^{-1})\beta & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \Delta^k(u)\beta \\ -\Delta^k(u)\beta & 0 \end{pmatrix}. \end{aligned}$$

The dimension of $D^k(O)$ is n_k, and therefore the dimension of $D^k(u)$ is $n_k/2$. The elements of the first m_kn_k rows of the matrix $D(u)$ are

$$\begin{aligned} &D(u)_{\eta\nu} = \Delta^k(u)^*_{\eta\nu}; \quad \eta, \nu = 1, \cdots, n_k/2, \\ &D(u)_{\eta\nu} = \Delta^k(u)^*_{\eta-n_k/2,\nu-n_k/2}; \quad \eta, \nu = n_k/2 + 1, \cdots, n_k, \\ &\quad \cdot \qquad\qquad \cdot \\ &\quad \cdot \qquad\qquad \cdot \\ &\quad \cdot \qquad\qquad \cdot \\ &D(u)_{\eta\nu} = \Delta^k(u)^*_{\eta-(m_k-1)n_k-n_k/2,\nu-(m_k-1)n_k-n_k/2}; \\ &\quad \eta, \nu = (m_k - 1)n_k + n_k/2 + 1, \cdots, m_kn_k. \end{aligned} \tag{20}$$

All the others vanish.

The elements of $D^k(u)$ are

$$\begin{aligned} &D^k(u)_{\alpha\beta} = \Delta^k(u)_{\alpha\beta}; \quad \alpha, \beta = 1, \cdots, n_k/2, \\ &D^k(u)_{\alpha\beta} = \Delta^k(u)_{\alpha-n_k/2,\beta-n_k/2}; \\ &\qquad\qquad \alpha, \beta = n_k/2 + 1, \cdots, n_k. \end{aligned} \tag{21}$$

All the others vanish.

[6] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1962).

(i) For $\sigma = 1, \cdots, n_k/2$, Eq. (12) has the form

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N}\sum_u \sum_{\delta,\epsilon,\eta,\nu} \sum_{\rho=1}^{n_k/2} U^*_{\eta(\delta\epsilon)} D(u)_{\eta\nu} U_{\nu(\alpha\beta)} \Delta^k(u)_{\rho\sigma} P^{ikj}_{\delta\rho\epsilon}$$
$$= \frac{2}{N}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} P^{ikj}_{\delta\rho\epsilon}\Bigg[\sum_{\eta,\nu=1}^{n_k/2} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)} \sum_u \Delta^k(u)^*_{\eta\nu}\Delta^k(u)_{\rho\sigma}$$
$$+ \sum_{\eta,\nu=(n_k/2)+1}^{n_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)}$$
$$\times \sum_u \Delta^k(u)^*_{\eta-n_k/2,\nu-n_k/2}\Delta^k(u)_{\rho\sigma} + \cdots\Bigg]$$

after using Eqs. (20) and (21).

Substitution of the orthogonality relations Eq. (15), where instead of n_k we put $n_k/2$, gives

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} P^{ikj}_{\delta\rho\epsilon}\Bigg[\sum_{\eta,\nu=1}^{n_k/2} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)} \frac{N}{2}$$
$$\times \frac{2}{n_k}\delta_{\rho\eta}\delta_{\sigma\nu} + \cdots\Bigg]$$
$$= \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} P^{ikj}_{\delta\rho\epsilon}[U^*_{\rho(\delta\epsilon)} U_{\sigma(\alpha\beta)}$$
$$+ U^*_{\rho+n_k/2,(\delta\epsilon)} U_{\sigma+n_k/2,(\alpha\beta)} + \cdots]. \quad (22)$$

Defining

$$B_{q+1} = \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} U^*_{\rho+q(n_k/2),(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon}, \quad q = 0, \cdots, 2m_k - 1, \quad (22')$$

we can write Eq. (22) in the form

$$P^{ikj}_{\alpha\sigma\beta} = B_1 U_{\sigma(\alpha\beta)} + B_2 U_{\sigma+(n_k/2),(\alpha\beta)} + B_3 U_{\sigma+n_k,(\alpha\beta)}$$
$$+ \cdots + B_{2m_k} U_{\sigma+m_k n_k-(n_k/2),(\alpha\beta)}. \quad (23)$$

(ii) For $\sigma = n_k/2 + 1, \cdots, n_k$, Eq. (12) becomes [after a procedure parallel to that performed in (i)]

$$P^{ikj}_{\alpha\sigma\beta} = B'_1 U_{\sigma-n_k/2,(\alpha\beta)} + B'_2 U_{\sigma(\alpha\beta)} + B'_3 U_{\sigma+n_k/2,(\alpha\beta)}$$
$$+ \cdots + B'_{2m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}, \quad (24)$$

where

$$B'_{q+1} = \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=n_k/2+1}^{n_k} U^*_{\rho-n_k/2+q(n_k/2),(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon}, \quad q = 0, \cdots, 2m_k - 1. \quad (24')$$

Similarly, for the antiunitary part, Eq. (13), one has the following:

(i') for $\sigma = 1, \cdots, n_k/2$

$$P^{ikj}_{\alpha\sigma\beta} = B'^*_2 U_{\sigma(\alpha\beta)} - B'^*_1 U_{\sigma+n_k/2,(\alpha\beta)} + B'^*_4 U_{\sigma+n_k,(\alpha\beta)}$$
$$- B'^*_3 U_{\sigma+n_k+n_k/2,(\alpha\beta)} + \cdots; \quad (25)$$

(ii') for $\sigma = n_k/2 + 1, \cdots, n_k$

$$P^{ikj}_{\alpha\sigma\beta} = -B^*_2 U_{\sigma-n_k/2,(\alpha\beta)} + B^*_1 U_{\sigma(\alpha\beta)} - B^*_4 U_{\sigma+n_k/2,(\alpha\beta)}$$
$$+ B^*_3 U_{\sigma+n_k,(\alpha\beta)} - \cdots, \quad (26)$$

where B_q and B'_q are defined by Eqs. (22') and (24').

A comparison between Eq. (23) and Eq. (25), and between Eq. (24) and Eq. (26) leads to the following results:

$$B'_1 = -B^*_2, \quad B'_2 = B^*_1, \quad B'_3 = -B^*_4, \quad B'_4 = B^*_3, \cdots,$$

or, in another form,

$$B'_p = (-1)^p B^*_{p-(-1)^p}.$$

As a consequence of the above, the Eckart–Wigner theorem generalized for this case is

$$P^{ikj}_{\alpha\sigma\beta} = B_1 U_{\sigma(\alpha\beta)} + B_2 U_{\sigma+n_k/2,(\alpha\beta)} + B_3 U_{\sigma+n_k,(\alpha\beta)}$$
$$+ \cdots + B_{2m_k} U_{\sigma+m_k n_k-n_k/2,(\alpha\beta)},$$

for

$$\sigma = 1, \cdots, n_k/2;$$

$$P^{ikj}_{\alpha\sigma\beta} = -B^*_2 U_{\sigma-n_k/2,(\alpha\beta)} + B^*_1 U_{\sigma(\alpha\beta)} - B^*_4 U_{\sigma+n_k/2,(\alpha\beta)}$$
$$+ B^*_3 U_{\sigma+n_k,(\alpha\beta)} - \cdots,$$

for

$$\sigma = n_k/2 + 1, \cdots, n_k.$$

Case C: $D^k(O)$ is a corepresentation of the third type:

$$D^k(u) = \begin{pmatrix} \Delta^k(u) & 0 \\ 0 & \Delta^{k'}(u) \end{pmatrix},$$

$$D^k(a) = \begin{pmatrix} 0 & \Delta^k(ua_0^2) \\ \Delta^{k'}(u) & 0 \end{pmatrix}.$$

[$\Delta^k(u)$ and $\Delta^{k'}(u)$ are inequivalent irreducible representations.]

The elements of the first $m_k n_k$ rows of $D(u)$ are

$$D(u)_{\eta\nu} = \Delta^k(u)^*_{\eta\nu}; \quad \eta, \nu = 1, \cdots, n_k/2,$$
$$D(u)_{\eta\nu} = \Delta^{k'}(u)^*_{\eta-n_k/2,\nu-n_k/2}; \quad \eta, \nu = n_k/2 + 1, \cdots, n_k,$$
$$D(u)_{\eta\nu} = \Delta^k(u)^*_{\eta-n_k,\nu-n_k}; \quad \eta, \nu = n_k + 1, \cdots, n_k + n_k/2,$$
$$\vdots$$
$$D(u)_{\eta\nu} = \Delta^{k'}(u)^*_{\eta-(m_k-1)n_k-n_k/2,\nu-(m_k-1)n_k-n_k/2}; \quad \eta, \nu = (m_k - 1)n_k + n_k/2 + 1 \cdots m_k n_k. \quad (27)$$

All the others vanish.

The elements of the matrix $D^k(u)$ are

$$D^k(u)_{\alpha\beta} = \Delta^k(u)_{\alpha\beta}; \quad \alpha, \beta = 1, \cdots, n_k/2,$$
$$D^k(u)_{\alpha\beta} = \Delta^{k'}(u)_{\alpha-n_k/2,\beta-n_k/2}; \quad \alpha, \beta = n_k/2 + 1, \cdots, n_k. \quad (28)$$

All the others vanish.

α. For $\sigma = 1, \cdots, n_k/2$, Eq. (12) becomes

$$P^{ikj}_{\alpha\sigma\beta} = \frac{2}{N}\sum_u \sum_{\rho=1}^{n_k/2} \sum_{\delta,\epsilon,\eta,\nu} U^*_{\eta(\delta\epsilon)} D(u)_{\eta\nu} \Delta^k(u)_{\rho\sigma} U_{\nu(\alpha\beta)} P^{ikj}_{\delta\rho\epsilon}$$
$$= \frac{2}{N}\sum_{\rho=1}^{n_k/2}\sum_{\delta,\epsilon} P^{ikj}_{\delta\rho\epsilon}\Bigg[\sum_{\eta,\nu=1}^{n_k/2} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)} \sum_u \Delta^k(u)^*_{\eta\nu}\Delta^k(u)_{\rho\sigma}$$
$$+ \sum_{\eta,\nu=(n_k/2)+1}^{n_k} U^*_{\eta(\delta\epsilon)} U_{\nu(\alpha\beta)}$$
$$\times \sum_u \Delta^{k'}(u)^*_{\eta-n_k/2,\nu-n_k/2}\Delta^k(u)_{\rho\sigma} + \cdots\Bigg]$$

after using Eqs. (27) and (28).

Use of the orthogonality relations Eq. (15) and the orthogonality relations existing for two inequivalent irreducible representations will bring the last equation to the form

$$P^{ikj}_{\alpha\sigma\beta} = U_{\sigma(\alpha\beta)} \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} U^*_{\rho(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon}$$
$$+ U_{\sigma+n_k,(\alpha\beta)} \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} U^*_{\rho+n_k,(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon} + \cdots. \quad (29)$$

If we define

$$C_{q+1} = \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=1}^{n_k/2} U^*_{\rho+qn_k,(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon}; \quad q = 0, \cdots, m_k - 1, \quad (29')$$

Eq. (29) can be written as

$$P^{ikj}_{\alpha\sigma\beta} = C_1 U_{\sigma(\alpha\beta)} + C_2 U_{\sigma+n_k,(\alpha\beta)}$$
$$+ \cdots + C_{m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}.$$

β. For $\sigma = n_k/2 + 1, \cdots, n_k$, after a procedure parallel to that performed in α, Eq. (12) becomes

$$P^{ikj}_{\alpha\sigma\beta} = C'_1 U_{\sigma(\alpha\beta)} + C'_2 U_{\sigma+n_k,(\alpha\beta)}$$
$$+ \cdots + C'_m U_{\sigma+(m_k-1)n_k,(\alpha\beta)},$$

where, by definition,

$$C'_{q+1} = \frac{2}{n_k}\sum_{\delta,\epsilon}\sum_{\rho=n_k/2+1}^{n_k} U^*_{\rho+qn_k,(\delta\epsilon)} P^{ikj}_{\delta\rho\epsilon},$$
$$q = 0, \cdots, m_k - 1.$$

Again, by carrying out similar calculations for the antiunitary part, one finds

$$C'_p = C^*_p.$$

The Eckart–Wigner theorem generalized for this case is hence

$$P^{ikj}_{\alpha\sigma\beta} = C_1 U_{\sigma(\alpha\beta)} + C_2 U_{\sigma+n_k,(\alpha\beta)}$$
$$+ \cdots + C_{m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}$$
$$\text{for} \quad \sigma = 1, \cdots, n_k/2$$

and

$$P^{ikj}_{\alpha\sigma\beta} = C^*_1 U_{\sigma(\alpha\beta)} + C^*_2 U_{\sigma+n_k,(\alpha\beta)}$$
$$+ \cdots + C^*_{m_k} U_{\sigma+(m_k-1)n_k,(\alpha\beta)}$$
$$\text{for} \quad \sigma = n_k/2 + 1, \cdots, n_k.$$

Let us summarize this section by writing the generalized Eckart–Wigner theorem for all cases:

(A) D^k is a corepresentation of the first type:

$$P^{ikj}_{\alpha\sigma\beta} = A_1 U_{\sigma(\alpha\beta)} + A_2 U_{\sigma+n_k,(\alpha\beta)} + \cdots;$$

(B) D^k is a corepresentation of the second type:

$$P^{ikj}_{\alpha\sigma\beta} = B_1 U_{\sigma(\alpha\beta)} + B_2 U_{\sigma+n_k/2,(\alpha\beta)} + \cdots,$$
$$\sigma = 1, \cdots, n_k/2,$$
$$P^{ikj}_{\alpha\sigma\beta} = -B^*_2 U_{\sigma-n_k/2,(\alpha\beta)} + B^*_1 U_{\sigma(\alpha\beta)}$$
$$- B^*_4 U_{\sigma+n_k/2,(\alpha\beta)} + \cdots,$$
$$\sigma = n_k/2 + 1, \cdots, n_k;$$

(C) D^k is a corepresentation of the third type:

$$P^{ikj}_{\alpha\sigma\beta} = C_1 U_{\sigma(\alpha\beta)} + C_2 U_{\sigma+n_k,(\alpha\beta)} + \cdots,$$
$$\sigma = 1, \cdots, n_k/2,$$
$$P^{ikj}_{\alpha\sigma\beta} = C^*_1 U_{\sigma(\alpha\beta)} + C^*_2 U_{\sigma+n_k,(\alpha\beta)} + \cdots,$$
$$\sigma = n_k/2 + 1, \cdots, n_k,$$

where A_p, B_p, and C_p are defined in Eqs. (16′), (22′), and (29′), respectively.

It is to be mentioned that the Eckart–Wigner theorem is very simple in form when $m_k = 1$. In such a case for unitary groups, there exists a simple proportionality between matrix elements of the given operator, which simplifies very much actual calculations (e.g., the applications of the original theorem). When dealing with nonunitary groups, the case is the same if the given operator transforms according to a corepresentation of the first or the third type.

If the operator transforms according to a corepresentation of the second type, then the matrix element is a sum of two terms:

$$P^{ikj}_{\alpha\sigma\beta} = B_1 U_{\sigma(\alpha\beta)} + B_2 U_{\sigma+n_k/2,(\alpha\beta)},$$

and one does not get the simple proportionality existing for unitary groups.

III. CONNECTIONS BETWEEN MATRIX ELEMENTS

Time inversion or an antiunitary operation connected with it can, in general, influence the number of independent constants necessary to determine matrix elements of symmetric operators.

The investigation of the problem of nonunitary groups, with the help of the corepresentation theory, as observed in the generalization of the Eckart–Wigner theorem in the previous section, enables us to obtain in simple form the connections that arise between matrix elements of operators in the presence of an antiunitary element. It is worth noting here that the bases of corepresentation are built from bases of representations, and therefore we can look upon the

matrix elements $P^{ikj}_{\alpha\sigma\beta}$ as matrix elements calculated for functions belonging to bases of representations.

Substituting $D(a) = D(u)D(a_0)$ in Eq. (4) gives

$$\begin{aligned} P^{ikj}_{\alpha\sigma\beta} &= \sum_{\delta,\rho,\epsilon} D^i(ua_0)_{\delta\alpha} D^j(ua_0)^*_{\epsilon\rho} D^k(ua_0)^*_{\rho\sigma} P^{ikj*}_{\delta\rho\epsilon} \\ &= \sum_{\delta,\rho,\epsilon} \sum_{\lambda,\mu,\nu} D^i(u)_{\delta\lambda} D^i(a_0)_{\lambda\alpha} D^j(u)^*_{\epsilon\mu} \\ &\quad \times D^j(a_0)^*_{\mu\beta} D^k(u)^*_{\rho\nu} D^k(a_0)^*_{\nu\sigma} P^{ikj*}_{\delta\sigma\epsilon} \\ &= \sum_{\lambda,\mu,\nu} D^i(a_0)_{\lambda\alpha} D^j(a_0)^*_{\mu\beta} D^k(a_0)^*_{\nu\sigma} \\ &\quad \times \sum_{\delta,\rho,\epsilon} D^i(u)_{\delta\lambda} D^j(u)^*_{\epsilon\mu} D^k(u)^*_{\rho\nu} P^{ikj*}_{\delta\rho\epsilon}. \end{aligned} \tag{30}$$

With the help of Eq. (3), Eq. (30) will have the form

$$P^{ikj}_{\alpha\sigma\beta} = \sum_{\lambda,\mu,\nu} D^i(a_0)_{\lambda\alpha} D^j(a_0)^*_{\mu\beta} D^k(a_0)^*_{\nu\sigma} P^{ikj*}_{\lambda\nu\mu}, \tag{31}$$

which gives the desired connections.

It is clear that the final form depends on the types of the corepresentations i, j, and k. For example, if the three corepresentations are of the first type:

$$D^i(a_0) = \beta^i, \quad D^j(a_0) = \beta^j, \quad D^k(a_0) = \beta^k,$$

then Eq. (31) has the form

$$P^{ikj}_{\alpha\sigma\beta} = \sum_{\lambda,\mu,\nu} \beta^i_{\lambda\alpha} \beta^{j*}_{\mu\beta} \beta^{k*}_{\nu\sigma} P^{ikj*}_{\lambda\nu\mu}. \tag{32}$$

As another example, let two of the corepresentations be of the first type and the third of the third type:

$$D^i(a_0) = \beta^i; \quad D^j(a_0) = \beta^j; \quad D^k(a_0) = \begin{pmatrix} 0 & \Delta^k(a_0^2) \\ 1 & 0 \end{pmatrix}.$$

Formula (31) then becomes

$$P^{ikj}_{\alpha\sigma\beta} = \sum_{\lambda,\mu} \beta^i_{\lambda\alpha} \beta^{j*}_{\mu\beta} P^{ikj*}_{\lambda,\sigma+n_k/2,\mu}, \qquad \sigma = 1, \cdots, n_k/2,$$

$$P^{ikj}_{\alpha,\sigma+n_k/2,\beta} = \sum_{\lambda,\mu,\nu} \beta^i_{\lambda\alpha} \beta^{j*}_{\mu\beta} \Delta^k(a_0^2)^*_{\nu\sigma} P^{ikj*}_{\lambda\nu\mu}, \qquad \sigma, \nu = 1, \cdots, n_k/2.$$

All other possible cases are discussed by Aviran.[7]

IV. EXAMPLES

A. The Three-Dimensional Rotation Group with Time Inversion (θ)

All the corepresentations of this nonunitary group are of the first type (see Ref. 2, p. 345), and for this reason we shall use here Eq. (32). In this case we always have $m_k = 1$, and therefore we have $P^{ikj}_{\alpha\sigma\beta} = A_1 U_{\sigma(\alpha\beta)}$. Substituting this in Eq. (32), we obtain (remembering that A_1 is real)

$$\begin{aligned} A_1 U_{\sigma(\alpha\beta)} &= \sum_{\lambda,\mu,\nu} \beta^i_{\lambda\alpha} \beta^{j*}_{\mu\beta} \beta^{k*}_{\nu\sigma} A_1^* U^*_{\nu(\lambda\mu)} \\ &= A_1 \sum_{\lambda,\mu,\nu} \beta^i_{\lambda\alpha} \beta^{j*}_{\mu\beta} \beta^{k*}_{\nu\sigma} U^*_{\nu(\lambda\mu)}. \end{aligned}$$

As a result we have the condition imposed on U by time inversion:

$$U_{\sigma(\alpha\beta)} = \sum_{\lambda,\mu,\nu} \beta^i_{\lambda\alpha} \beta^{j*}_{\mu\beta} \beta^{k*}_{\nu\sigma} U^*_{\nu(\lambda\mu)}. \tag{33}$$

As an example of using formula (33), let us prove that matrix elements of the scalar operator p, which anticommutes with time inversion, between eigenfunctions of a defined angular-momentum state are purely imaginary (Ref. 2, p. 346).

To calculate $(\psi^J_\mu | p | \phi^J_\nu)$ it is necessary to know the corresponding U. The connection between U and the matrix S used in Ref. 2 is as follows:

$$U^*_{\delta(\nu+\delta,\nu)} U_{\epsilon(\sigma+\epsilon,\sigma)} = \frac{2k+1}{2i+1} S^*_{i\nu\delta} S_{i\sigma\epsilon} \tag{34}$$

or

$$\frac{1}{2J+1} S^{(J0)}_{J\nu\delta} S^{(J0)}_{J\sigma\epsilon} = U^*_{\delta(\nu+\delta,\nu)} U_{\epsilon(\sigma+\epsilon,\sigma)},$$

where i, j, and k were replaced by J, J, and 0, respectively. δ and ϵ can have the value 0 only, and we get

$$\frac{1}{2J+1} S^{(J0)}_{J\nu 0} S^{(J0)}_{J\sigma 0} = U^*_{0(\nu,\nu)} U_{0(\sigma,\sigma)},$$

and $U_{0(\mu\nu)} = 0$ if $\mu \neq \nu$.

As $S^{(J0)}_{J\nu 0} = 1$, we receive

$$U^*_{0(\nu\nu)} U_{0(\nu\nu)} = \frac{1}{2J+1}.$$

As a result, for the nonvanishing elements of the first row of U we get

$$U_{0(\nu\nu)} = \frac{e^{i\alpha}}{(2J+1)^{\frac{1}{2}}}, \quad \nu = -J, \cdots, +J. \tag{35}$$

We know that $upu^{-1} = p = D^0(u)p$, $D^0(u) = 1$;

$$\theta p \theta^{-1} = -p = D^0(\theta)p; \quad D^0(\theta) = \beta^0 = -1.$$

Substitution of the last expression for $U_{0(\nu\nu)}$ and $\beta^0_{00} = -1$ in (33) gives

$$\begin{aligned} \frac{e^{-i\alpha}}{(2J+1)^{\frac{1}{2}}} = U^*_{0(\lambda\lambda)} &= \sum_{\varphi,\rho} \beta^{J*}_{\varphi\lambda} \beta^J_{\rho\lambda} \beta^0_{00} U_{0(\varphi\rho)} \\ &= -\sum_{\varphi} \beta^{J*}_{\varphi\lambda} \beta^J_{\varphi\lambda} U_{0(\varphi\varphi)} = -\sum_{\varphi} (\beta^J)^\dagger_{\lambda\varphi} \beta^J_{\varphi\lambda} U_{0(\varphi\varphi)} \\ &= -\frac{e^{i\alpha}}{(2J+1)^{\frac{1}{2}}} \sum_{\varphi} (\beta^J)^\dagger_{\lambda\varphi} \beta^J_{\varphi\lambda} = -\frac{e^{i\alpha}}{(2J+1)^{\frac{1}{2}}}. \end{aligned}$$

Hence $e^{i\alpha} = \pm i$, and therefore

$$U_{0(\nu\nu)} = \frac{\pm i}{(2J+1)^{\frac{1}{2}}}.$$

[7] A. Aviran, M.Sc. thesis, Technion — Israel Institute of Technology, 1967.

TABLE I. Representations of the unitary group C_{3v}.

	E	C_3	C_3^2	σ_v	σ_v'	σ_v''
Δ_1	1	1	1	1	1	1
Δ_2	1	1	1	-1	-1	-1
Δ_3	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}\epsilon&0\\0&\epsilon^2\end{pmatrix}$	$\begin{pmatrix}\epsilon^2&0\\0&\epsilon\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon^2\\\epsilon&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon\\\epsilon^2&0\end{pmatrix}$
Δ_4	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-\epsilon&0\\0&-\epsilon^2\end{pmatrix}$	$\begin{pmatrix}\epsilon^2&0\\0&\epsilon\end{pmatrix}$	$\begin{pmatrix}0&i\\i&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon^2 i\\\epsilon i&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon i\\\epsilon^2 i&0\end{pmatrix}$
Δ_5	1	-1	1	i	i	i
Δ_6	1	-1	1	$-i$	$-i$	$-i$

$\epsilon = e^{2\pi i/3}$

The generalized Eckart–Wigner theorem in this case will have the form

$$(\psi_\mu^J|\, p\,|\Phi_\nu^J) = A_1 U_{0(\mu\nu)} = A_1 \frac{\pm i}{(2J+1)^{\frac{1}{2}}}\,\delta_{\mu\nu},$$

where A_1 is real; and this proves that the discussed matrix element is purely imaginary.

B. C_{3v} with Time Inversion

Let us calculate matrix elements, by considering first the corresponding representations and then the corresponding corepresentations, and see the effect of the antiunitary element on them.

The representations of the unitary subgroup C_{3v} are given in Table I.

Using the criteria of Ref. 4, we get the corepresentations given in Table II.

Let us calculate some matrix elements, using the representations of C_{3v}:

$$U(\Delta_3^* \times \Delta_5)U^\dagger = \Delta_4^* \to U = \begin{pmatrix} e^{i\beta} & 0 \\ 0 & -e^{i\beta}\end{pmatrix}.$$

Hence we get

$$(\varphi_1^3|\, P_1^4\,|\varphi_1^5) = AU_{1(11)} = Ae^{i\beta},$$
$$(\varphi_2^3|\, P_2^4\,|\varphi_1^5) = AU_{2(21)} = -Ae^{i\beta},$$

Where β is an arbitrary phase and A is a constant (not necessarily real).

Using the corepresentations, we get

$$U(D_3^* \times D_5)U^\dagger = D_4^* \dot{+} D_4^* \quad \text{(for the unitary elements)},$$
$$U(D_3^* \times D_5)U^{\dagger *} = D_4^* \dot{+} D_4^* \quad \text{(for the antiunitary elements)}.$$

From the unitary part

$$U = \begin{pmatrix} e^{i\delta} & 0 & 0 & 0 \\ 0 & 0 & -e^{i\delta} & 0 \\ 0 & e^{i\Delta} & 0 & 0 \\ 0 & 0 & 0 & e^{i\Delta}\end{pmatrix}.$$

As for this case $m_k = 2$, the general form of U is (see Appendix and Ref. 3)

$$U = \begin{pmatrix} b_{11}e^{i\delta} & b_{12}e^{i\Delta} & 0 & 0 \\ 0 & 0 & -b_{11}e^{i\delta} & b_{12}e^{i\Delta} \\ b_{21}e^{i\delta} & b_{22}e^{i\Delta} & 0 & 0 \\ 0 & 0 & -b_{21}e^{i\delta} & b_{22}e^{i\Delta}\end{pmatrix},$$

where $\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}\end{pmatrix}$ is a unitary matrix.

By demanding

$$U(D_3(\theta)^* \otimes D_5(\theta))U^{\dagger *} = D_4(\theta)^* \dot{+} D_4(\theta)^*,$$

TABLE II. Corepresentations of the unitary subgroup C_{3v} and θ.

	E	C_3	C_3^2	σ_v	σ_v'	σ_v''	θ
D_1	1	1	1	1	1	1	$e^{i\varphi}$
D_2	1	1	1	-1	-1	-1	$e^{i\theta}$
D_3	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}\epsilon&0\\0&\epsilon^2\end{pmatrix}$	$\begin{pmatrix}\epsilon^2&0\\0&\epsilon\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon^2\\\epsilon&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon\\\epsilon^2&0\end{pmatrix}$	$\begin{pmatrix}0&e^{i\psi}\\e^{i\psi}&0\end{pmatrix}$
D_4	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-\epsilon&0\\0&-\epsilon^2\end{pmatrix}$	$\begin{pmatrix}\epsilon^2&0\\0&\epsilon\end{pmatrix}$	$\begin{pmatrix}0&i\\i&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon^2 i\\\epsilon i&0\end{pmatrix}$	$\begin{pmatrix}0&\epsilon i\\\epsilon^2 i&0\end{pmatrix}$	$\begin{pmatrix}0&e^{i\alpha}\\-e^{i\alpha}&0\end{pmatrix}$
D_5	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$

we get

$$U = 2^{-\frac{1}{2}}\begin{pmatrix} e^{i\beta} & -e^{i(\psi-\beta-\alpha)} & 0 & 0 \\ 0 & 0 & -e^{i\beta} & -e^{i(\psi-\beta-\alpha)} \\ \pm ie^{i\beta} & \pm ie^{i(\psi-\beta-\alpha)} & 0 & 0 \\ 0 & 0 & \mp ie^{i\beta} & \pm ie^{i(\psi-\beta-\alpha)} \end{pmatrix},$$

where β is an arbitrary phase, and hence

$$(\varphi_1^3|\, P_1^4\,|\varphi_1^5) = 2^{-\frac{1}{2}}(e^{i\beta}A_1 \pm ie^{i\beta}A_2) = 2^{-\frac{1}{2}}e^{i\beta}(A_1 \pm iA_2),$$

$$(\varphi_2^3|\, P_2^4\,|\varphi_1^5) = 2^{-\frac{1}{2}}(-e^{i\beta}A_1 \mp ie^{i\beta}A_2) = -2^{-\frac{1}{2}}e^{i\beta}(A_1 \pm iA_2),$$

where A_1 and A_2 are real constants.

It might seem that there is no difference between the results obtained by the two different methods. However, we must observe that if we calculated the same matrix elements in representation theory replacing φ_1^5 by φ_1^6, we would again get an arbitrary phase and a constant A' without any connection with A. On the other hand working with corepresentations, one gets

$$(\varphi_1^3|\, P_1^4\,|\varphi_2^5) = -2^{-\frac{1}{2}}e^{i(\psi-\beta-\alpha)}A_1 \pm 2^{-\frac{1}{2}}ie^{i(\psi-\beta-\alpha)}A_2 = 2^{-\frac{1}{2}}e^{i(\psi-\beta-\alpha)}(-A_1 \pm iA_2),$$

$$(\varphi_2^3|\, P_2^4\,|\varphi_2^5) = -2^{-\frac{1}{2}}e^{i(\psi-\beta-\alpha)}A_1 \pm 2^{-\frac{1}{2}}ie^{i(\psi-\beta-\alpha)}A_2 = 2^{-\frac{1}{2}}e^{i(\psi-\beta-\alpha)}(-A_1 \pm iA_2),$$

where the definition $\varphi_2^5 = \varphi_1^6$ was used. Now there is a connection between A and A'. We see therefore that the use of corepresentations leads to a smaller number of independent constants.

APPENDIX

In order to use the generalized Eckart–Wigner theorem, one has to know the matrix U. Since corepresentations are constructed out of representation, Koster's method can be used for finding U. One can use relations (7) together with the conditions imposed by relations (8).

If, for example, D^k is a corepresentation of the first type and $m_k = 1$, Koster's method gives U with an undefined phase factor. Equation (8) turns the phase factor ambiguity into a sign ambiguity only. When Koster's method gives U to within a certain unitary matrix [see Eq. (27) in Ref. 3], Eq. (8) limits the number of different possibilities of this matrix.

The matrix U can be calculated in two ways:

1. By considering the structure of $D(0)$ and $D^k(0)$ it is possible to write equations similar to those written by Koster for the representations of unitary groups.

2. After finding U by Koster's method, one writes the equation

$$UD'(a_0)U^{\dagger *} = m_k D^k(a_0)^* \dotplus \cdots$$

and finds the U which satisfies it. It is sufficient that U satisfies the last equation for Eq. (8) to be satisfied because one can always write $D(a) = D(u)D(a_0)$ and hence

$$D'(u)D'(a_0) = D'(a) = U^{\dagger}D(a)U^* = U^{\dagger}D(u)D(a_0)U^* = U^{\dagger}D(u)UU^{\dagger}D(a_0)U^*,$$

which means that any U satisfying Eq. (8) for the element a_0 will satisfy it for every element a.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Majorana Representations of the Lorentz Group and Infinite-Component Fields

D. Tz. STOYANOV
Joint Institute for Nuclear Research, Dubna, USSR
AND
I. T. TODOROV*
International Centre for Theoretical Physics, Trieste, Italy

(Received 8 November 1967)

A self-contained exposition is given of the theory of infinite-component fields with special emphasis on fields transforming under the Majorana representations of the Lorentz group, for which the scalar vertex function is written down explicitly for particles with arbitrary momenta and spins. The problem of spin and statistics for such fields is analyzed. A class of coupled representations of $SL(2, C)$ is studied, containing unitary as well as nonunitary representations (including the Dirac 4-component spinors), for which invariant first-order equations can be written down for the free field. Most of the results are known (either from old or from recent publications), but are presented here in a unified way, their derivation sometimes being simplified. Among the few new points we mention: (1) The location of singularities of the matrix elements of some infinite-dimensional representations of $SL(2, C)$ for complex values of the group parameters. (2) The construction of an infinite-component local Fermi field transforming under a unitary representation of $SL(2, C)$. (3) the discussion of the quantization of Majorana fields with a proper account of the Fourier components with spacelike momenta.

1. INTRODUCTION

A. Problems of Infinite-Component Quantized Fields

In conventional relativistic quantum field theory, in both the axiomatic and Lagrangian approach, it is assumed that a unitary representation $U(a, A)$ of the covering of the Poincaré group

$$ISL(2, C) \equiv SL(2, C) \cdot T_4$$

[$A \in SL(2, C)$, i.e., $\det A = 1$; $a = (a^0, \mathbf{a}) \in T_4$] is realized in the Hilbert space of states $\mathfrak{H}$ and that the field operators $\psi^\alpha(x)$ transform covariantly under $U(a, A)$:

$$U(a, A)\, \psi^\alpha(x)\, U^{-1}(a, A) = V(A^{-1})^\alpha_\beta\, \psi^\beta(\Lambda x + a). \quad (1.1)$$

Here $\Lambda = \Lambda(A)$ is the proper Lorentz transformation defined by

$$A\sigma_\nu A^* = \sigma_\mu \Lambda^\mu_\nu \quad \text{or} \quad \Lambda^\mu_\nu = \tfrac{1}{2} \operatorname{Tr} (\sigma_\mu A \sigma_\nu A^*); \qquad \mu, \nu = 0, 1, 2, 3 \quad (1.2)$$

(σ_0 is the 2×2 unit matrix, σ_j, $j = 1, 2, 3$, are the Pauli matrices) and $V(A)$ is a finite-dimensional representation of $SL(2, C)$.

It seems at first glance that this last assumption, which asserts that the field has a finite number of components (and hence that V is a finite matrix), is purely technical and has no important physical idea behind it. From a physical point of view the choice of the representation V which occurs in (1.1) is restricted by the existence of discrete symmetries and of covariant forms (Lagrangian, currents).

In a *TCP*-invariant theory we have to introduce, together with each field $\psi(x)$, its Hermitian conjugate $\psi^*(x)$. Usually, it is assumed that a nondegenerate invariant Hermitian form

$$\psi^*(x)\, \beta \psi(x), \quad \beta = \beta^* \quad (1.3)$$

can be written down which implies that the representation V is equivalent to its adjoint:

$$V^{*-1}(A) = \beta\, V(A)\, \beta^{-1}, \quad (1.4)$$

where V^* is defined in some fixed basis by $V^* = \bar{V}^{\mathrm{T}}$ [i.e., $(V^*)^j_i = \bar{V}^j_i$]. If we assume in addition that a space-reflection operation exists in the theory (if a parity has to be assigned to the one-particle states), then we have to require that $V(A)$ is equivalent to its parity conjugate $V(A^{*-1})$,

$$V(A^{*-1}) = SV(A)S^{-1}. \quad (1.5)$$

[We mention that in general $V(A^*) \neq V^*(A)$ so that (1.5) is not a consequence of (1.4).]

None of these restrictions eliminates the infinite-dimensional representations of $SL(2, C)$.

A dropping of the requirement of finite dimensionality of $V(A)$ looks quite attractive from the point of view of interpretation of the present-day experimental data on elementary-particle resonances. The number of resonances, which differ only with respect to spin and parity (having the same internal quantum

* Present address: Institute for Advanced Study, Princeton, N.J. On leave of absence from Joint Institute for Nuclear Research, Dubna, USSR and from the Physical Institute of the Bulgarian Academy of Sciences, Sofia, Bulgaria.

numbers such as charge, hypercharge, and isospin), is so large that it seems more advisable to consider them as a part of an infinite multiplet than to ascribe an independent field to each. There is of course a different and more common possibility, namely, describing all hadrons as bound states of a few underlying fields (say of the quark fields) just as well as, in the case of the hydrogen atom, where one is able to obtain the infinity of spectral lines starting essentially with the two-component electron field. It seems plausible, however, that both descriptions are possible (as suggested recently for the case of the nonrelativistic hydrogen atom). In that case a free infinite-component wave equation for hadrons corresponds to the approximation in which the interaction between quarks (giving rise to the spectrum of hadrons) is taken into account, while the interaction between hadrons is neglected. There is a hope based on some first-order calculations of form factors that a perturbation theory in terms of infinite-component fields might be more practical than the conventional perturbation theory of strong interactions.

It turns out, however, that many important statements of relativistic quantum field theory such as *TCP*, and spin and statistics theorems and usual crossing symmetry crucially depend on the assumption that we are using finite-dimensional representations of $SL(2, C)$ for the field's transformation law.

B. Historical Remarks and References

The development of the idea of infinite multiplets seems to be rather typical in that it is a case of how a work of the thirties, completely overlooked in its time, is rediscovered step by step, first by mathematicians, next by physicists (always independently!). Hence we find it instructive to make a brief digression into the history of the problem.

Infinite-component fields were first studied by Majorana[1] (1932) who introduced the (only) two irreducible representations of $SL(2, C)$ for which an invariant (linear) first-order differential equation can be written. Majorana found the discrete spectrum of this equation and realized that it also possesses a continuous set of solutions with spacelike momenta. This remarkable work of Majorana remained practically unknown until 1966 when Fradkin[2] revived it (on the suggestion of Amaldi), actually translating it into English and placing it in the context of later research. In 1948 Gel'fand and Yaglom[3] (see also the exhaustive reviews in Refs. 4 and 5) rediscovered Majorana's results in the more general framework of the description of all irreducible representations of $SL(2, C)$, but they apparently overlooked the existence of the continuous spectrum of solutions corresponding to spacelike momenta of the infinite system of wave equations. These spacelike solutions have been pointed out by Bargmann.[6] In Ref. 3 it was first observed (though in a different terminology) that the connection between spin and statistics is lost for infinite-component fields.

The present-day interest in infinite multiplets arose mostly in connection with the search for a relativistic generalization of $SU(6)$ (see Refs. 7–11 and further references quoted therein). The breakdown of the spin and statistics theorem has been reemphasized by several authors (see Zumino's contribution in Ref. 9 as well as Refs. 12 and 13). It has been argued in particular[12] that if a free local field transforms under a unitary representation of $SL(2, C)$ then, at least for the case of an index-invariant theory (which is incompatible with a Dirac-type equation), we are forced to use a canonical *commutation* relation (i.e., Bose statistics) both for integer and half-integer spin. It was noted in Ref. 14 that one can consider, instead, the "big" unitary field as a nonlocal collection of conventional finite-component free local fields for which, as we know, the spin and statistics theorem is guaranteed. Examples of fields satisfying Majorana-type equations which can be consistently quantized with anticommutators have been considered in Ref. 15. However, the discussion of this problem was not complete since the rules of quantization of the spacelike

[1] E. Majorana, Nuovo Cimento **9**, 335 (1932).
[2] D. M. Fradkin, Am. J. Phys. **34**, 314 (1966).
[3] I. M. Gel'fand and A. M. Yaglom, Zh. Eksp. Teor. Fiz. **18**, 703, 1096, 1105 (1948).
[4] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon Press Ltd., London, 1963) (in Russian: Fizmatgiz, Moscow, 1958).
[5] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press Ltd., London, 1964) (in Russian: Fizmatgiz, Moscow, 1958).
[6] V. Bargmann, Math. Rev. **10**, 583, 584 (1949).
[7] L. Michel, *Proceedings of the Second Coral Gables Conference on Symmetry Principles at High Energy* (W. H. Freeman and Co., San Francisco, 1965), pp. 331–352.
[8] P. Budini and C. Fronsdal, Phys. Rev. Letters **14**, 968 (1965).
[9] C. Fronsdal and B. Zumino, *Proceedings of the Seminar on High-Energy Physics and Elementary Particles, Trieste, 1965* (International Atomic Energy Agency, Vienna, 1965), pp. 657–664 and 665–678.
[10] C. Fronsdal, P. T. Matthews, V. H. Nguyen, and I. T. Todorov, *Lectures at the International School of Theoretical Physics, Yalta, 1966* (Naukova Dumka, Kiev, 1967).
[11] H. Ruegg, W. Rühl, and T. S. Santhanam, Helv. Phys. Acta **40**, 9 (1967).
[12] G. Feldman and P. T. Matthews, Ann. Phys. (N.Y.) **40**, 19 (1966); Phys. Rev. **151**, 1176 (1966); **154**, 1241 (1967).
[13] R. F. Streater, Commun. Math. Phys. **5**, 88 (1967).
[14] V. H. Nguyen, Bucharest Preprint F.T. 62 (1966); C. Fronsdal, Phys. Rev. **156**, 1653 (1967); V. D. Dao and V. H. Nguyen, Yad. Fiz. **6**, 1861 (1967).
[15] E. Abers, I. T. Grodsky, and R. E. Norton, Phys. Rev. **159**, 1222 (1967).

components of the field were not given. Infinite multiplets have been considered also from different points of view.[16]

A physical understanding of infinite-component fields in terms of composite models has been attempted, particularly looking at the example of the nonrelativistic hydrogen atom.[17–20] As stressed by Budini,[21] this example corresponds to a nonlocal (actually bilocal) infinite-component field except in the case of an infinitely heavy nucleon when all the poles of the Green function are going to infinity.

C. Content of the Present Paper

The main object of the present paper is the study of the peculiar properties of quantized infinite-component fields. Most of these properties are illustrated on the simplest examples of fields transforming according to the Majorana representations of $SL(2, C)$ (but not necessarily satisfying the first-order Majorana equations). The general treatment in Sec. 2C however applies to both finite- and infinite-component fields.

Section 2 is devoted to some mathematical preliminaries about coupled and self-coupled representations of the Lorentz group. It is shown that the self-coupled representations of $SL(2, C)$ are actually representations of the 10-parameter group of real symplectic transformations in four dimension, $Sp(4, R)$. The Lie algebra of the two self-coupled representations $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 0]$ (the Majorana representations) as well as the canonical basis are described in terms of Bose creation and annihilation operators. The well-known results about the classification and description of the irreducible representations of $SL(2, C)$ are summarized in Appendix A. The only nonstandard point in Appendix A is the expression of different 4-vectors as differential operators in the space of function of two complex variables. It is used in Sec. 2C in the analysis of pairs of coupled mutually adjoint as well as irreducible self-adjoint representations of $SL(2, C)$. The content of Appendix B is also related to Sec. 2. The creation and annihilation operators are expressed there as linear functions of z, $\bar{z}$, $\partial/\partial z$, and $\partial/\partial\bar{z}$, where z is a complex variable. In this realization the scalar product in the representation space is defined in terms of an integral over the complex z plane.

In Sec. 3A infinite-component free Bose fields are considered satisfying the Klein–Gordon equation and transforming under an irreducible unitary representation of $SL(2, C)$. We rederive the result of Ref. 12 showing that even in the case when the one-particle states created by such a field have half-integer spin, the field is quantized consistently in terms of local canonical *commutation* (instead of anticommutation) relations. In Sec. 3B we construct an example of a free Fermi field transforming under any of the (unitary) Majorana representations of $SL(2, C)$ and satisfying the Klein–Gordon equation, and discuss the reason of the breakdown of the axiomatic proof of spin and statistics for infinite-component fields. The matrix elements of the Majorana representations for finite Lorentz transformations are evaluated in Sec. 3C and applied in Sec. 3D to first-order calculation for the vertex function between two infinite multiplets and one scalar field.

The complete set of solutions of the Majorana equation (both for the discrete and for the continuous spectrum) is described in a unified way in Sec. 4. The canonical quantization of the free Majorana field is considered in Sec. 4C.

In Appendix C the ladder representations of the conformal group are described in terms of the same two complex variables which are used in Appendix A for the realization of an arbitrary representation of $SL(2, C)$. It is also shown that the analytic continuation of the matrix elements of the ladder representation for complex values of the Lorentz parameters has singularities in the same points as the matrix elements of the irreducible representations of $SL(2, C)$. These are the points corresponding to time reflection.

2. SELF-COUPLED REPRESENTATIONS OF $SL(2, C)$

A. Self-Coupled Representations of $SL(2, C)$ as Representations of $Sp(4, R)$

We use the notation $[l_0, l_1]$ of Ref. 4 for the irreducible representations of $SL(2, C)$ (see Appendix A where all necessary definitions are reproduced). A general (not necessarily irreducible) representation of the Lorentz group is denoted by τ.

A representation $V(A)$ of $SL(2, C)$ (irreducible or not) is called self-coupled if one can write a nondegenerate invariant Hermitian form of the type $i\psi^*\beta L^\mu \partial_\mu \psi$ for a field ψ transforming according to (1.1) (with the given V).[22] It is clear that a representation τ of $SL(2, C)$ is self-coupled if and only if it is

[16] See, e.g., H. Kleinert, Ph.D. thesis, University of Colorado, 1967, where a complete account of recent work of A. O. Barut and the author is given. Nonquantized infinite-component fields are considered in A. Böhm, Syracuse University Preprint SU-1206-125, 1967.

[17] Y. Nambu, Progr. Theoret. Phys. (Kyoto) Suppl. 37 and 38, 368 (1966); Phys. Rev. **160**, 1171 (1967).

[18] C. Fronsdal, Phys. Rev. **156**, 1665 (1967).

[19] A. O. Barut and H. Kleinert, Phys. Rev. **157**, 1180 (1967).

[20] A. O. Barut and H. Kleinert, Phys. Rev. **160**, 1149 (1967).

[21] P. Budini (private communication). See also P. Budini, ICTP Trieste Preprints IC/67/18 and IC/67/80 (1967).

contained in the direct product of τ with the four-vector representation [0, 2].

There exist only two irreducible self-coupled representations[3-5]: $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 0]$. These are exactly the *Majorana representations* (see Refs. 1 and 2). Both of them are infinite-dimensional and unitary (see Sec. 2B and Appendix A). In the representation space X of any of them a 4-vector of operators L^μ can be defined which, in addition to the transformation law

$$V(A)L^\mu V^{-1}(A) = \Lambda^{-1}(A)^\mu_\nu L^\nu, \tag{2.1}$$

or in infinitesimal form [writing $V(A) = \exp\{-i \times \Sigma_{\mu<\nu} S^{\mu\nu}\omega_{\mu\nu}\}$]:

$$[S^{\lambda\mu}, L^\nu] = i(g^{\mu\nu}L^\lambda - g^{\lambda\nu}L^\mu), \tag{2.2}$$

satisfies the commutation relations

$$i[L^\mu, L^\nu] = S^{\mu\nu}. \tag{2.3}$$

Here $S^{\mu\nu}$ are the generators of the homogeneous Lorentz group and $g^{\mu\nu}$ is the metric tensor in Minkowski space ($g^{00} = -g^{kk} = 1$, $k = 1, 2, 3$).

Equations (2.2) and (2.3) coincide with the commutation relations of the Lie algebra of the group of real symplectic transformations in four dimensions $Sp(4, R)$. Indeed, the lowest faithful representation of the commutation relations (2.2) and (2.3) is four-dimensional and may be written in terms of the Dirac matrices

$$L^\mu \to \tfrac{1}{2}\gamma^\mu \equiv s^{4\mu}, \quad S^{\mu\nu} \to (i/4)[\gamma^\mu, \gamma^\nu] \equiv s^{\mu\nu} \tag{2.4}$$

[(2.2) and (2.3) are simple consequences of the anticommutation rule

$$[\gamma^\mu, \gamma^\nu]_+ = 2g^{\mu\nu} \tag{2.5}$$

which defines the Dirac matrices]. It is known that each of the matrices (2.4) satisfies the condition

$$Cs^{ab}C^{-1} = -(s^{ab})^{\mathrm{T}}, \quad a, b = 0, 1, 2, 3, 4, \tag{2.6}$$

where the superscript T stands for transposition and C is the antisymmetric charge-conjugation matrix defined by

$$C\gamma^\mu C^{-1} = -(\gamma^\mu)^{\mathrm{T}}, \quad C^{-1} = C^{\mathrm{T}} = C^* = -C.$$

For the group elements of $Sp(4, R)$ (2.6) gives $CVC^{-1} = (V^{-1})^{\mathrm{T}}$ which implies the conservation of the antisymmetric bilinear form $\xi C\eta$. Furthermore, in the Majorana basis (in which all γ^μ and $s^{\mu\nu}$ are pure imaginary) all group elements V are real matrices, which completes the justification of our terminology.

$Sp(4, R)$ is a covering group of the de Sitter group $SO(3, 2)$, the two groups having the same Lie algebra, defined by the commutation rules:

$$[S^{ab}, S^{cd}] = i(g^{bc}S^{ad} - g^{ac}S^{bd} - g^{bd}S^{ac} + g^{ad}S^{bc}), \tag{2.7}$$

g^{ab} being the metric tensor in five dimensions:

$$g^{00} = g^{44} = -g^{kk} = 1, \quad k = 1, 2, 3,$$
$$g^{ab} = 0, \quad \text{for} \quad a \neq b. \tag{2.8}$$

Majorana representations are defined as such irreducible representations of the real symplectic group $Sp(4, R)$, which are irreducible also with respect to its subgroup $SL(2, C)$.

We mention that (2.3) is just one of the many different ways to close the Lie algebra containing in addition to the Lorentz generators $S^{\mu\nu}$, a 4-vector L^μ. Some other possibilities, which can be relevant in the investigation of reducible self-coupled representations of $SL(2, C)$, have been considered in Ref. 23.

B. Description of Majorana Representations in Terms of Creation and Annihilation Operators

We assume here the Pauli representation of γ matrices in which γ^0 is diagonal

$$\gamma^0 = \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \quad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \tag{2.9}$$

$$C = i\gamma^0\gamma^2 = \begin{pmatrix} 0 & \epsilon \\ \epsilon & 0 \end{pmatrix}, \quad \epsilon = i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -\epsilon^{-1}. \tag{2.10}$$

To describe the Majorana representations of $Sp(4, R)$ we introduce the operators a_α and a^*_α, $\alpha = 1, 2$, satisfying the Bose-type commutation relations

$$[a_\alpha, a_\beta] = [a^*_\alpha, a^*_\beta] = 0,$$
$$[a_\alpha, a^*_\beta] = \delta_{\alpha\beta}, \quad \alpha, \beta = 1, 2. \tag{2.11}$$

Define a pair of four-component operator-valued spinors φ and $\tilde{\varphi}$:

$$\varphi = \begin{pmatrix} a \\ \epsilon^* a^* \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ -a_2^* \\ a_1^* \end{pmatrix},$$

$$\tilde{\varphi} = \varphi^*\gamma^0 = C\varphi = \begin{pmatrix} a^* \\ \epsilon a \end{pmatrix} = \begin{pmatrix} a_1^* \\ a_2^* \\ a_2 \\ -a_1 \end{pmatrix}. \tag{2.12}$$

[22] One could be tempted to say that a field is transforming under a self-coupled representation of $SL(2, C)$ if a first-order invariant equation can be written down for it. This definition, however, is not equivalent to the above because one can write invariant first-order equations for a lot of irreducible representations (for instance, for the four-vector representation [0, 2], one can write the equation $\partial_\mu A^\mu = 0$), while an invariant Lagrangian of the type $\mathfrak{L} = \psi^*\beta(iL^\mu\partial_\mu - \kappa)\psi$ exists only for the two Majorana representations. Such a Lagrangian implies actually the existence of a *complete irreducible set* of invariant first-order equations.

[23] L. Castell, Nuovo Cimento **50**, 945 (1967).

Because of (2.11) they satisfy the commutation rules

$$[\varphi^A, \tilde{\varphi}_B] = \delta^A_B, \quad [\varphi^A, \varphi^B] = C^{AB}, \quad A, B = 1, \cdots, 4. \tag{2.13}$$

The generators of the ladder-type representations of $Sp(4, R)$ are given by[24]

$$S^{ab} = \tfrac{1}{2}\tilde{\varphi} s^{ab} \varphi, \tag{2.14}$$

where s^{ab} are the four-dimensional matrices (2.4). It is easily checked [making use of (2.13) and (2.6)] that S^{ab} fulfill the same commutation relations as s^{ab}

$$[S^{ab}, S^{cd}] = \tfrac{1}{2}\tilde{\varphi}[s^{ab}, s^{cd}]\varphi. \tag{2.15}$$

It is immediately seen that the operators (2.14) are Hermitian because of the identity

$$s^{*ab}\gamma^0 = \gamma^0 s^{ab}, \tag{2.16}$$

so that the corresponding representation of $Sp(4, R)$ is unitary.

The explicit expressions for the generators S^{ab} in terms of the creation and annihilation operators $a^{(*)}$ read as follows:

$$\begin{aligned} M^j &\equiv \tfrac{1}{2}\sum_{k,l} \epsilon_{jkl} S^{kl} = \tfrac{1}{2} a^* \sigma_j a, \\ N^j &\equiv S^{0j} = \frac{i}{4}(a^* \sigma_j \epsilon^{-1} a^* - a \epsilon \sigma_j a) \\ \Big(N^3 &= \frac{i}{2}(a_1 a_2 - a_1^* a_2^*), \\ N_+ &\equiv N^1 + iN^2 = \tfrac{1}{2}(a_1^{*2} + a_2^2)\Big); \end{aligned} \tag{2.17}$$

$$\begin{aligned} L^0 &\equiv S^{40} = \tfrac{1}{2}(a^* a + 1), \\ L^j &\equiv S^{4j} = \tfrac{1}{4}(a^* \sigma_j \epsilon^{-1} a^* + a \epsilon \sigma_j a) \\ (L^3 &= -\tfrac{1}{2}(a_1^* a_2^* + a_1 a_2), \\ L_+ &\equiv L^1 + iL^2 = \tfrac{1}{2}(a_1^{*2} - a_2^2)). \end{aligned} \tag{2.18}$$

[24] The technique of creation and annihilation operators was first applied for the realization of the representations of $SU(2)$ by P. Jordan, Z. Physik **94**, 531 (1935) (see a modern exposition in L. C. Biedenharn, ICTP Trieste Preprint IC/67/52, 1967). For the description of unitary representations of the Lorentz group such a technique was used in P. A. M. Dirac, Proc. Roy. Soc. (London) **A180**, 1 (1942); **183**, 284 (1945). It has been applied in B. Kursunoglu [*Modern Quantum Theory* (W. H. Freeman and Co., San Francisco, 1962), p. 257] for the ladder representations of $U(2, 2)$ (see Appendix C). In the same form as here [for the description of the Majorana representations of SL (2, C)] they have been used in P.A.M. Dirac, J. Math. Phys. **4**, 901 (1963) and in F. Gürsey, *Relativity, Groups and Topology*, C. De Witt and B. De Witt, Ed. (Gordon and Breach Science Publ., Inc., New York, 1964), and more recently in A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1546 (1967) and C. Itzykson, Commun. Math. Phys. **4**, 92 (1967).

The operators $a_\alpha^{(*)} (= a_\alpha$ or $a_\alpha^*)$ should not be confused with particle creation and annihilation operators. They do not depend on coordinates (or momentum) and act only in the space of indices, which in our case is infinite-dimensional. To avoid confusion we use different notations for the vectors in the auxiliary space X and in the Fock space $\mathcal{H}$ of physical states [e.g., $|0) \in X$, $|0\rangle \in \mathcal{H}$].

The representation space X may be spanned by polynomials of a^* acting on an $SU(2)$-invariant vector $|0)$ defined by $a_\alpha |0) = 0$, $\alpha = 1, 2$. We shall use here the Fock variables

$$a_\alpha^* = \xi_\alpha, \quad a_\alpha = \frac{\partial}{\partial \xi_\alpha}, \quad \alpha = 1, 2 \quad (|0) = 1). \tag{2.19}$$

In these variables X is a space of entire analytic functions $f(\xi_1, \xi_2)$ with scalar product

$$(f, g) = \overline{\left[f\left(\frac{\partial}{\partial \xi_1}, \frac{\partial}{\partial \xi_2}\right) g(\xi_1, \xi_2)\right]}_{\xi_1 = \xi_2 = 0}. \tag{2.20}$$

[Another realization of X (corresponding to the Schrödinger picture of quantum mechanics) is given in Appendix B. The scalar product in it is defined by an integral.] One can introduce in X a canonical orthonormal basis of monomials:

$$|s\zeta) = \frac{\xi_1^{s+\zeta} \xi_2^{s-\zeta}}{[(s+\zeta)!\,(s-\zeta)!]^{\frac{1}{2}}}. \tag{2.21}$$

The generators (2.17) and (2.18) act on this basis in the following way:

$$\begin{aligned} M^3 |s\zeta) &= \zeta |s\zeta), \\ M_\pm |s\zeta) &\equiv (M^1 \pm iM^2) |s\zeta) \\ &= [(s \mp \zeta)(s \pm \zeta + 1)]^{\frac{1}{2}} |s\zeta \pm 1), \\ N^3 |s\zeta) &= \frac{i}{2}\{(s^2 - \zeta^2)^{\frac{1}{2}} |s - 1\,\zeta) \\ &\quad - [(s+1)^2 - \zeta^2]^{\frac{1}{2}}\} |s + 1\,\zeta)\}, \\ N_\pm |s\zeta) &= \pm \frac{i}{2}\{[(s \mp \zeta)(s \mp \zeta - 1)]^{\frac{1}{2}} |s - 1\,\zeta \pm 1) \\ &\quad + [(s \pm \zeta + 1)(s \pm \zeta + 2)]^{\frac{1}{2}} |s + 1\,\zeta \pm 1)\}; \end{aligned} \tag{2.22}$$

$$\begin{aligned} L^0 |s\zeta) &= (s + \tfrac{1}{2}) |s\zeta), \\ 2L^3 |s\zeta) &= -\{(s^2 - \zeta^2)^{\frac{1}{2}} |s - 1\,\zeta), \\ &\quad + [(s+1)^2 - \zeta^2]^{\frac{1}{2}} |s + 1\,\zeta)\}, \\ 2L_\pm |s\zeta) &= \pm\{[(s \pm \zeta + 1)(s \pm \zeta + 2)]^{\frac{1}{2}} |s + 1\,\zeta \pm 1) \\ &\quad - [(s \mp \zeta)(s \mp \zeta - 1)]^{\frac{1}{2}} |s - 1\,\zeta \pm 1)\}. \end{aligned} \tag{2.23}$$

We see from (2.22) that the basis (2.21) corresponds to the reduction of the ladder representation with respect to $SU(2)$. The vectors (2.21) have definite spin s and spin projection ζ:

$$M^2 |s\zeta) = s(s+1) |s\zeta), \quad M^3 |s\zeta) = \zeta |s\zeta). \tag{2.24}$$

The basis with these properties is called *canonical basis*. A consequence of (2.22) and (2.23) is that spin

s may be transformed through multiple application of the generators into $s+n$, where n is an integer. Hence, starting with $s=\zeta=0$ we obtain an invariant subspace X_0 of X which contains vectors with integer spin only, and starting with $s=\zeta=\frac{1}{2}$ we obtain another invariant subspace $X_{\frac{1}{2}}$, containing only half-integer spins. Moreover, it is easily checked that X is the direct sum of these two spaces

$$X = X_0 \oplus X_{\frac{1}{2}}. \tag{2.25}$$

It can be verified that (2.22) is a special case of the general formulas (A13) and (A7) (see Appendix A) corresponding to $l_0 l_1 = 0$, $l_0^2 + l_1^2 = \frac{1}{4}$, so we see that our representation splits into the two Majorana representations ($[0, \frac{1}{2}]$ acting in X_0 and $[\frac{1}{2}, 0]$ acting in $X_{\frac{1}{2}}$). This can also be seen by the calculation of the Casimir operators. In general, the Casimir operators of the Lorentz group are connected with the numbers l_0 and l_1 by

$$\mathbf{M}^2 - \mathbf{N}^2 = l_0^2 + l_1^2 - 1, \quad i\mathbf{MN} = l_0 l_1 \tag{2.26}$$

(see Appendix A). On the other hand, a direct calculation using (2.17) gives for the ladder representation

$$\begin{aligned} &\mathbf{MN} = 0, \\ &\mathbf{M}^2 = \frac{a^*a}{2}\left(\frac{a^*a}{2} + 1\right), \\ &\mathbf{N}^2 = \frac{a^*a}{2}\left(\frac{a^*a}{2} + 1\right) + \frac{3}{4}, \end{aligned} \tag{2.27}$$

implying $\mathbf{M}^2 - \mathbf{N}^2 = -\frac{3}{4}$. So we again find the solutions $[l_0 = 0,\ l_1 = \pm\frac{1}{2}]$ and $[l_0 = \pm\frac{1}{2},\ l_1 = 0]$. [We remark that the equivalence relation $[l_0, l_1] \sim [-l_0, -l_1]$ holds for any two irreducible representations of $SL(2, C)$ (cf. Appendix A), so that equivalent representations are contained in each of the above brackets.]

We mention finally that the Lorentz scalar $L^\mu L_\mu$ is also a constant in the ladder representation of $Sp(4, R)$

$$\begin{aligned} L^\mu L_\mu &\equiv (L^0)^2 - \mathbf{L}^2 \\ &= \tfrac{1}{4}\{(a^*a + 1)^2 - (a^*a + 1)^2 - 2\} = -\tfrac{1}{2}. \end{aligned}$$

C. Pairs of Coupled Adjoint Representations and Irreducible Self-adjoint Representations of $SL(2, C)$

The Majorana representations studied in the previous section are of a very special type. We would like to consider here a wider class of admissible representations (in accordance with the general principles stated in Sec. 1B) which includes, among other things, the currently used finite-component tensor fields and the 4-component Dirac field as well as some infinite-component fields.

We shall restrict ourselves to the simplest case when the representation $V(A)$ is either irreducible or a direct sum of two irreducible representations of $SL(2, C)$. Of course, much more complicated reducible representations of $SL(2, C)$ might be of interest as well, for instance representations of some higher group in whose decomposition with respect to $SL(2, C)$ a direct integral of irreducible representations is involved. One such example [namely, the ladder representations of $U(2, 2)$] is considered in Appendix C.

We start with the case of a single irreducible representation $[l_0, l_1]$ satisfying conditions (1.4) and (1.5) of Sec. 1.

Assumption (1.4), i.e., the assumption of existence of a nondegenerate invariant Hermitian form, leads to

$$[l_0, l_1] = \pm[l_0, -\bar{l}_1], \tag{2.28}$$

i.e., either $l_1 = -\bar{l}_1$ (l_1 pure imaginary), or $l_0 = 0$, $l_1 = \bar{l}_1$ (l_1 real) (see Ref. 4 part II). As it should be, all unitary representations [corresponding to positive-definite β] as well as all finite-dimensional tensor representations $[0, n]$ ($n = 1, 2, \cdots$) are included in this class which contains moreover a family of infinite-dimensional real nonunitary representations $[0, l_1]$ with $l_1 > 1$ (l_1 noninteger).

The assumption of self-adjointness [(1.5)] necessary in any theory allowing space reflection [or, alternatively, in any CP-invariant theory of a field, transforming under a real representation of $SL(2, C)$], gives[4]

$$[l_0, l_1] = \pm[l_0, -l_1], \quad \text{or} \quad l_0 l_1 = 0. \tag{2.29}$$

Conditions (2.28) and (2.29) are fulfilled simultaneously in two cases

$$l_1 = 0, \quad l_0 \text{ arbitrary}, \quad \text{or} \quad l_1 = \pm\bar{l}_1, \quad l_0 = 0. \tag{2.30}$$

If we require in addition the reality of the representation $V(A)$ [in other words the existence of a basis in which all matrix elements of $V(A)$ are real; in such a basis one can introduce an invariant notion of Hermitian conjugation] we have to restrict ourselves to integer l_0 in the first case.

Further we proceed to the case of a reducible representation of the type

$$[l_0, l_1] \oplus [l_0', l_1']. \tag{2.31}$$

We shall be interested in the case when the irreducible parts of (2.31) do not fulfil separately (2.30) so that they have to be conjugate to each other. In that case

(1.4) gives

$$[l_0', l_1'] = \pm[l_0, -\bar{l}_1] \tag{2.32}$$

(see Ref. 4, part II Sec. 2.9), while (1.5) leads to

$$[l_0', l_1'] = \pm[l_0, -l_1] \tag{2.33}$$

(see Ref. 4 part II Sec. 2.6). Conditions (2.32) and (2.33) can be fulfilled simultaneously in two cases

$$\left.\begin{aligned} &l_1 = \bar{l}_1 \text{ (i.e., } l_1 \text{ real) } l_0 \text{ arbitrary,} \\ &l_1 = -\bar{l}_1 \text{ (i.e., } l_1 \text{ pure imaginary), } l_0 = 0 \end{aligned}\right\}. \tag{2.34}$$

If we impose in addition the requirement that the two representations are coupled, i.e., that (A17) (see below, Appendix A) takes place, then we find the following sets of admissible pairs:

$$[\tfrac{1}{2}, l_1] + [-\tfrac{1}{2}, l_1], \quad l_1 \text{ real}, \tag{2.35}$$

$$[l_0, \tfrac{1}{2}] + [l_0, -\tfrac{1}{2}], \quad l_0 = 0, \tfrac{1}{2}, 1, \cdots. \tag{2.36}$$

Expressions (2.35) and (2.36) define the simplest representations appropriate to describe fields satisfying first-order equations in a theory of parity-conserving interactions. The Dirac field is contained in the class (2.35) for $l_1 = \pm\frac{3}{2}$. The Majorana fields are also included in (2.35) and (2.36), being the only admissible pairs of equivalent representations and the only unitary representations of this class. Actually, they represent the intersection of (2.35) and (2.36) with (2.30). The infinite-dimensional nonunitary representations used recently in[25] are included in (2.36) for l_0 half-integer.

Using the results of Appendix A, one can construct free field Lagrangians for fields transforming under (2.35) and (2.36) leading to first-order equations

$$\mathfrak{L} = \psi^*(x)\beta(i\Gamma^\mu\partial_\mu - \kappa)\psi(x), \tag{2.37}$$

where ψ and Γ^μ have different meaning in the two cases. For ψ transforming under (2.35) we write

$$\psi_1 = \begin{pmatrix}\varphi_1(x;z)\\ \chi_1(x;z)\end{pmatrix}, \quad \Gamma_1^\mu = \begin{pmatrix} 0 & g^{\mu\mu}z\sigma_\mu\epsilon\dfrac{\partial}{\partial\bar{z}} \\ \bar{z}\epsilon^{-1}\sigma_\mu\dfrac{\partial}{\partial z} & 0 \end{pmatrix}, \tag{2.38}$$

where φ_1 and χ_1 transform under $[\frac{1}{2}, l_1]$ and $[-\frac{1}{2}, l_1]$, respectively, and the continuous spinor-variable $z = (z_1, z_2)$ substitutes the index of the field components (see Appendix A). For ψ transforming under (2.36) we have[26]

$$\psi_2 = \begin{pmatrix}\varphi_2(x;z)\\ \chi_2(x;z)\end{pmatrix},$$

$$\Gamma_2^\mu = \begin{pmatrix} 0 & (|l_0| + \tfrac{1}{2})g^{\mu\mu}z\sigma_\mu\bar{z} \\ -\dfrac{1}{|l_0| + \frac{1}{2}}\dfrac{\partial}{\partial\bar{z}}\sigma_\mu\dfrac{\partial}{\partial z} & 0 \end{pmatrix}, \tag{2.39}$$

where φ_2 and χ_2 transform according to $[l_0, \frac{1}{2}]$ and $[l_0, -\frac{1}{2}]$, respectively. The matrix β in both cases has the form

$$\beta = \sigma_1 \otimes I = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \tag{2.40}$$

I being the unit operator in the corresponding space. The generators of each of the reducible representations (2.35) and (2.36) are given by

$$M^j = \frac{1}{2}\left(z\sigma_j\frac{\partial}{\partial z} - \frac{\partial}{\partial\bar{z}}\sigma_j\bar{z}\right)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},$$

$$N^j = \frac{i}{2}\left(z\sigma_j\frac{\partial}{\partial z} + \frac{\partial}{\partial\bar{z}}\sigma_j\bar{z}\right)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \tag{2.41}$$

[see (A12)]. We mention that because the representations for φ and χ are adjoint, the corresponding matrix elements of M^j and N^j in the canonical basis are related by

$$\begin{pmatrix}\varphi_{s\zeta}\\ 0\end{pmatrix}M^j\begin{pmatrix}\varphi_{s'\zeta'}\\ 0\end{pmatrix} = \begin{pmatrix}0\\ \chi_{s\zeta}\end{pmatrix}M^j\begin{pmatrix}0\\ \chi_{s'\zeta'}\end{pmatrix},$$

$$\begin{pmatrix}\varphi_{s\zeta}\\ 0\end{pmatrix}N^j\begin{pmatrix}\varphi_{s'\zeta'}\\ 0\end{pmatrix} = -\begin{pmatrix}0\\ \chi_{s\zeta}\end{pmatrix}N^j\begin{pmatrix}0\\ \chi_{s'\zeta'}\end{pmatrix}. \tag{2.42}$$

We introduce for each of the cases (2.35) and (2.36) a parity operator by

$$\mathfrak{V}_\alpha(I_s) = \begin{pmatrix} 0 & V_\alpha(I_s) \\ V_\alpha(I_s) & 0 \end{pmatrix}, \quad \alpha = 1, 2, \tag{2.43}$$

where the operators $V_{1,2}$ are given by (A19)–(A22). Using the results of Appendix A we see that Γ_α^μ behave like vectors under the corresponding reflection $\mathfrak{V}_\alpha$:

$$\mathfrak{V}_\alpha(I_s)\Gamma_\alpha^\mu\mathfrak{V}_\alpha^{-1}(I_s) = g^{\mu\mu}\Gamma_\alpha^\mu, \quad \alpha = 1, 2, \tag{2.44}$$

whereas the quantities obtained from Γ_α^μ by changing the relative sign of the nonvanishing (off-diagonal) elements are axial vectors. It is easily seen that the form $\psi^*\beta\psi$ is a scalar, while $\psi^*\sigma_2 \otimes I\psi$ is a pseudoscalar.

[25] C. Fronsdal and R. White, Phys. Rev. **163**, 1835 (1967).

[26] The expressions for Γ_2^μ are simpler in a basis in which $A_s^{l_0 l_1}$ in (A7) is replaced by $[\Gamma(s + 1 - l_1)/\Gamma(s + 1 + l_1)]^{\frac{1}{2}}$; in such a basis the factors $|l_0| + \frac{1}{2}$ in (2.39) do not appear. (The basis used in the paper has the advantage to be defined also for the finite-dimensional representations.)

There is only one common pair of representations of the two classes (2.35) and (2.36), namely the pair

$$[\tfrac{1}{2}, \tfrac{1}{2}] \oplus [-\tfrac{1}{2}, \tfrac{1}{2}] \sim [\tfrac{1}{2}, \tfrac{1}{2}] \oplus [\tfrac{1}{2}, -\tfrac{1}{2}] \quad (2.45)$$

considered in Ref. 25.

3. LOCAL FIELDS WITH AN INFINITELY DEGENERATE MASS LEVEL

A. Quantization of an Irreducible Free Bose Field

It is well known that while the quantization of a free scalar field is obtained by a straightforward application of the canonical Lagrangian formalism, the theory of the higher-rank tensor fields is much more subtle, because of the necessity of introducing supplementary conditions.[27] We shall show here that the quantization of a (free) field, transforming under a real unitary representation of $SL(2, C)$ is just as straightforward and simple as the quantization of a scalar field (which is, by the way, the only finite-component field transforming under an unitary representation of the Lorentz group).

Let $V(A)$ be an irreducible unitary representation of $SL(2, C)$ acting in the Hilbert space X. A field $\varphi(x;f)$ is defined as an operator-valued distribution in x depending linearly on the vector $f \in X$ (and weakly continuous with respect to f). Choosing for f a vector $|l\eta)$ of the canonical basis we obtain, in particular, the field component $\varphi^{l\eta}(x) \equiv \varphi(x; l\eta)$.

A free complex field φ of mass m will be defined by the Lagrangian

$$\mathfrak{L}(x) = :\partial^\mu \varphi^*(x) \partial_\mu \varphi(x): - m^2 :\varphi^*(x)\varphi(x):, \quad (3.1)$$

where : : stands for the normal product and

$$\varphi^*(x)\varphi(y) = \sum_{l\eta} \varphi^*_{l\eta}(x)\varphi^{l\eta}(y)$$

is invariant with respect to pure index transformations. The variational principle with Lagrangian (3.1) leads to the Klein–Gordon equation for each component of the field. Its solution can be decomposed in a sum of nonlocal fields of definite spin:

$$\varphi(x) = \frac{1}{[2(2\pi)^3]^{\frac{1}{2}}} \times \int_{p^0=\omega} \sum_{s=l_0}^{\infty} \sum_{\zeta=-s}^{s} [a_{s\zeta}(\mathbf{p}) e^{-ipx} + b^*_{s-\zeta}(\mathbf{p}) e^{ipx}] \times u_{s\zeta}(\mathbf{p}) \frac{d^3p}{p^0}, \quad (3.2)$$

where

$$\omega = \omega(\mathbf{p}) = (m^2 + \mathbf{p}^2)^{\frac{1}{2}} \quad (3.3)$$

and $u_{s\zeta}(\mathbf{p}) = \{u^{l\eta}_{s\zeta}(\mathbf{p})\}$ is an infinite-component "spinor" corresponding to spin s and spin projection ζ. Let w be the Pauli–Lubanski–Bargmann 4-vector

$$w_p = \tfrac{1}{2}\epsilon_{\lambda\mu\nu\rho} P^\lambda S^{\mu\nu}. \quad (3.4)$$

(ϵ is the completely antisymmetric unit tensor, $\epsilon_{0123} = 1$.) The physical spin operator for a particle of mass m and momentum $\mathbf{p}$ is given by

$$S_j = \frac{1}{m}\left\{w_j + \frac{w_0 p_j}{\omega + m}\right\}. \quad (3.5)$$

In terms of $\mathbf{S}$, the "spinor" $u_{s\zeta}(\mathbf{p})$ is defined by the equations

$$\mathbf{S}^2 u_{s\zeta}(\mathbf{p}) = s(s+1)u_{s\zeta}(\mathbf{p}), \quad S_3 u_{s\zeta}(\mathbf{p}) = \zeta u_{s\zeta}(\mathbf{p}) \quad (3.6)$$

and the normalization condition

$$u_{s\zeta}(0) = |s\zeta).$$

(We have suppressed here for brevity the label $[l_0, l_1]$ of the representation under consideration which is kept fixed). $u^{l\eta}_{s\zeta}(\mathbf{p})$ is given by the matrix elements of the so-called "boost" transformation (cf. Ref. 27). Let B_p be the positive-definite two-by-two matrix corresponding according to (1.2) to the pure Lorentz transformation Λ_p defined by $\Lambda_p(m, \mathbf{0}) = p$. We have

$$B_p = \frac{p^0 + m + p^j\sigma_j}{[2m(p^0 + m)]^{\frac{1}{2}}}, \quad (3.7)$$

and

$$S_j = V(B_p) M_j V^{-1}(B_p),$$

so that the normalized solution of (3.6) is given by

$$u_{s\zeta}(\mathbf{p}) = V(B_p)\,|s\zeta), \quad (3.8)$$

or in components

$$u^{l\eta}_{s\zeta}(\mathbf{p}) = (l\eta|\, V(B_p)\,|s\zeta) = V(B_p)^{l\eta}_{s\zeta}. \quad (3.9)$$

Because of the unitarity of V the spinors (3.9) are orthonormalized for all $\mathbf{p}$:

$$(u_{s\zeta}(\mathbf{p}), u_{s'\zeta'}(\mathbf{p})) = \sum_{l=l_0}^{\infty} \sum_{\eta=-l}^{l} \overline{u^{l\eta}_{s\zeta}(\mathbf{p})} u^{l\eta}_{s'\zeta'}(\mathbf{p}) = \delta_{ss'}\delta_{\zeta\zeta'}. \quad (3.10)$$

Proceeding to the quantization of $\varphi(x)$ we mention that the energy–momentum operator corresponding to the Lagrangian (3.1) is

$$P^\mu = \int_{x^0=t} \left\{ : \frac{\partial \mathfrak{L}}{\partial(\partial_0 \varphi)} \partial^\mu \varphi : + :\partial^\mu \varphi^* \frac{\partial \mathfrak{L}}{\partial(\partial_0 \varphi^*)} : - g^{0\mu}\mathfrak{L} \right\} d^3\mathbf{x} = \sum_{s\zeta} \int_{p^0=\omega} :a^*_{s\zeta}(\mathbf{p}) a_{s\zeta}(\mathbf{p}) + b_{s\zeta}(\mathbf{p}) b^*_{s\zeta}(\mathbf{p}): p^\mu \frac{d^3\mathbf{p}}{p^0} \quad (3.11)$$

[to obtain the expression in the right-hand side we have made use of (3.10)]. The energy is positive if we

[27] The formulation of the theory of particles with arbitrary spin given in S. Weinberg [Phys. Rev. 133, B1318 (1964)], though convenient in a number of cases seems not to be completely satisfactory, e.g., when a minimal electromagnetic interaction for such particles has to be considered.

require the canonical commutation relations

$$[a_{s\zeta}(\mathbf{p}), a^*_{s'\zeta'}(\mathbf{q})] = [b_{s\zeta}(\mathbf{p}), b^*_{s'\zeta'}(\mathbf{q})] = \omega(\mathbf{p})\delta_{ss'}\delta_{\zeta\zeta'}\delta(\mathbf{p}-\mathbf{q}), \quad (3.12)$$

all other commutators vanishing identically.

The same result could be obtained without reference to the positiveness of the energy if we had postulated that a and a^*, and b and b^* satisfy either commutation or anticommutation relations of the type (3.12) and that $\varphi(x)$ and $\varphi^*(y)$ are local (cf. Refs. 12 and 27). To do this we should use once more (3.9) to obtain the completeness relation

$$\sum_{s=l_0}^{\infty}\sum_{\zeta=-s}^{s}\overline{u^{l\eta}_{s\zeta}(\mathbf{p})}u^{l'\eta'}_{s\zeta}(\mathbf{p}) = \delta_{ll'}\delta_{\eta\eta'}, \quad (3.13)$$

which implies the local commutation relations

$$[\varphi(x;f), \varphi(y;g)^*] = (g,f)\frac{1}{i}D_m(x-y) \quad (3.14)$$

(D_m is the Pauli–Jordan function). Thus we are led to canonical *commutation* (rather than anticommutation) relations independently of the value of l_0, i.e., independently of whether the spin is integer or half-integer.

B. Local Fermi Fields Transforming under the Majorana Representations

The discussion in Sec. 3A may give the wrong impression that the unitarity of the representation of the index group implies by itself Bose statistics for the corresponding (local) field (see Ref. 12). We show that this is not true by constructing an explicit example of a free local Fermi field, transforming under any of the Majorana representations.

To do this we define the field ψ by a decomposition of the type (3.2):

$$\psi(x) = \frac{1}{[2(2\pi)^3]^{\frac{1}{2}}}\sum_{s=l_0}^{\infty}(s+\tfrac{1}{2})^{\frac{1}{2}} \times \int_{p^0=\omega}\sum_{\zeta=-s}^{s}[a_{s\zeta}(\mathbf{p})e^{-ipx} + b^*_{s-\zeta}(\mathbf{p})e^{ipx}]u_{s\zeta}(p)\frac{d^3\mathbf{p}}{p^0}, \quad (3.15)$$

and assume that the operators $a^{(*)}$ and $b^{(*)}$ satisfy canonical *anticommutation* relations instead of (3.12). It is clear that the presence of any s-dependent factor [instead of $(s+\frac{1}{2})^{\frac{1}{2}}$] in (3.15) does not affect the Poincaré invariance of the theory. With our particular choice of this factor we obtain the following local anticommutation relations for the field:

$$[\psi(x), \psi(y)]_+ = 0,$$
$$[\psi(x), \psi^*(y)]_+ = \frac{1}{m}L_\mu\frac{\partial}{\partial x_\mu}D_m(x-y), \quad (3.16)$$

L_μ being the Majorana matrices (2.23). To check the second formula (3.16) one has to use (3.8) and the identity

$$\sum_{s=l_0}^{\infty}\sum_{\zeta=-s}^{s}(s+\tfrac{1}{2})V(B_p)\,|s\zeta)\,(s\zeta|\,V(B_p)^* = V(B_p)L_0V(B_p)^* = \frac{1}{m}L_\mu p^\mu.$$

The Hamiltonian corresponding to the field (3.15) is positive

$$H = P^0 = \sum_{s\zeta}\int(a^*_{s\zeta}(\mathbf{p})a_{s\zeta}(\mathbf{p}) + b^*_{s\zeta}(\mathbf{p})b_{s\zeta}(\mathbf{p}))d^3\mathbf{p}.$$

Thus, we have examples of a local Fermi field which exhibits the right connection between spin and statistics for the representation $[\frac{1}{2}, 0]$ and violates this connection for the representation $[0, \frac{1}{2}]$.

From the examples of this and the previous subsections we see that the type of statistics for local infinite-component fields is not determined by the spin content of the field. To see the reason why the axiomatic proof of the spin and statistics theorem[28] is not valid for infinite-component fields we recall one of the main steps in this proof. Due to spectrum conditions and locality the two-point function

$$F_{\alpha\beta}(x) = \langle 0|\,\psi_\alpha\left(\frac{x}{2}\right)\psi^*_\beta\left(-\frac{x}{2}\right)|0\rangle$$

can be continued analytically in the extended tube

$$\mathcal{T} = \{z \in C_4 \mid z^2 \neq b \geq 0\}$$

(this is a consequence of the Bogoliubov–Vladimirov theorem[29]). Moreover, in the case of finite-component fields it is proved that $F_{\alpha\beta}(z)$ is covariant under the proper complex Lorentz transformations. In particular, for the proper complex transformation $\Lambda = -I$, one obtains

$$F_{\alpha\beta}(-z) = (-1)^{2l_0}F_{\alpha\beta}(z). \quad (3.17)$$

We observe that for all known examples of local infinite-component fields the two-point function $F_{\alpha\beta}(z)$ is also covariant under complex Lorentz transformations [though $V^\beta_\alpha(\Lambda)$ may have singularities for some complex Λ as we shall see in Sec. 3C]. Equation (3.17), however, is not automatic for infinite-component fields. We have seen, in particular, that for

[28] See for instance, R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin Inc., New York, 1964) or R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).

[29] N. N. Bogoliubov and V. S. Vladimirov, Nauchn. Dokl. Vyschei Shkoly No. 3, 26 (1958); No. 2, 179 (1959). A generalization of this theorem to functions of several vectors is given in J. Bros, H. Epstein, and V. Glaser, Commun. Math. Phys. 6, 77 (1967).

a field transforming according to any of the Majorana representations, examples exist of free field theories with both even and odd two-point functions. It is exactly at this point where the general proof of the spin and statistics theorem (as well as the proof of *TCP*) does not work in a theory involving infinite-component fields. We remark also that all examples of local fields satisfying spectrum conditions and having the wrong connection between spin and statistics correspond to infinite mass degeneracy with respect to spin. It is worthwhile mentioning that for the case of no mass degeneracy of the one-particle states Epstein[30] has proved *TCP* invariance of the *S*-matrix in a theory of local observables without any assumption about finite-componentness of the underlying field (if any).

C. Matrix Elements of Majorana Representations for Finite Lorentz Transformations

We have seen that the spinors $u_{s\zeta}$ appearing in the canonical decomposition (3.2) [or (3.15)] of an infinite-component field are expressed in terms of the matrix elements of some special Lorentz transformations [see (3.9)]. Matrix elements of more general Lorentz transformations are involved in first-order calculation of the vertex function (see subsection 3.4). General formulas for the matrix elements of $V(A)$ in the canonical basis have been given in Refs. 31 and 32. We reproduce them here for the special case of the Majorana representations for which they are considerably simplified.

First of all we remark that the calculation of $V(A)^{l\eta}_{s\zeta}$ for an arbitrary A of $SL(2, C)$ is reduced to the calculation of these matrix elements for the special Lorentz transformation along the z axis corresponding to the positive-definite diagonal 2×2 matrix

$$D_\lambda = \begin{pmatrix} e^{\lambda/2} & 0 \\ 0 & e^{-\lambda/2} \end{pmatrix}, \quad \lambda \text{ real.} \tag{3.18}$$

Indeed, every matrix A of $SL(2, C)$ can be represented in the form

$$A = U_1 D_\lambda U_2, \tag{3.19}$$

where

$$2 \cosh \lambda = \operatorname{Tr} AA^*,$$

and U_1, $U_2 \in SU(2)$ (see, e.g., Ref. 5 Sec. 11.5). Using the fact that the representation of $SU(2)$ contained in a given representation of $SL(2, C)$ is reduced in the canonical basis and that $V(D_\lambda)$ is diagonal with respect to the pair of indices $\eta\zeta$ we obtain

$$V(A)^{l\eta}_{s\zeta} = \sum_{\zeta'=-r}^{r} D(U_1)^{(l)}_{\eta\zeta'} V(D_\lambda)^{l\zeta'}_{s\zeta'} D(U_2)^{(s)}_{\zeta'\zeta}, \qquad r = \min(l, s), \tag{3.20}$$

where $D(U)^{(s)}$ is the well-known representation of $SU(2)$ (for the explicit expression of $D^{(s)}$ see for instance Ref. 33).

We sketch the derivation of the explicit expression of $V(D_\lambda)^{l\eta}_{s\zeta}$ for the Majorana representations.

We use the realization (2.21) of the canonical basis in terms of the variables ξ_1, ξ_2 and look for the function

$$F_{s\zeta}(\lambda; \xi) = e^{-i\lambda N^3} |s\zeta). \tag{3.21}$$

Differentiation with respect to λ together with (2.17) leads to the following partial differential equation for F:

$$\left\{\frac{\partial}{\partial\lambda} + \frac{1}{2}\left(\xi_1\xi_2 - \frac{\partial^2}{\partial\xi_1\partial\xi_2}\right)\right\} F_{s\zeta}(\lambda; \xi_1, \xi_2) = 0. \tag{3.22}$$

The solution of (3.22) satisfying the initial condition

$$F_{s\zeta}(0; \xi) = [(s + \zeta)!\,(s - \zeta)!]^{-\frac{1}{2}} \xi_1^{s+\zeta} \xi_2^{s-\zeta}, \tag{3.23}$$

is given by

$$\begin{aligned} F_{s\zeta}(\lambda; \xi) = {} & [(s + \zeta)!\,(s - \zeta)!]^{\frac{1}{2}} \\ & \times \sum_{k=0}^{s-|\zeta|} \frac{1}{k!} \left(\tanh\frac{\lambda}{2}\right)^k \left(\cosh\frac{\lambda}{2}\right)^{2(k-s)-1} \\ & \times \frac{\xi_1^{s+\zeta-k}}{(s + \zeta - k)!} \frac{\xi_2^{s-\zeta-k}}{(s - \zeta - k)!} \\ & \times \exp\left\{-\xi_1\xi_2 \tanh\frac{\lambda}{2}\right\}. \end{aligned} \tag{3.24}$$

As it should be, for fixed real λ, $F_{s\zeta}$ is an entire analytic function of ξ. Now, the matrix elements of $V(D_\lambda)$ are obtained by using the orthonormality of the basis (2.21):

$$\begin{aligned} V(D_\lambda)^{l\eta}_{s\zeta} = (l\eta \mid F_{s\zeta}) = {} & \delta^\eta_\zeta \left[\frac{(s - |\zeta|)!\,(l + |\zeta|)!}{(s + |\zeta|)!\,(l - |\zeta|)!}\right]^{\frac{1}{2}} \\ & \times \frac{(-\tanh\frac{1}{2}\lambda)^{l-s}}{(\cosh\frac{1}{2}\lambda)^{2|\zeta|+1}} P^{(l-s, 2|\zeta|)}_{s-|\zeta|}\left(\frac{1}{\cosh\lambda}\right) \end{aligned} \tag{3.25}$$

where $P^{(\alpha,\beta)}_n(x)$ are the Jacobi polynomials

$$P^{(\alpha,\beta)}_n(x) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n + \alpha}{k} \binom{n + \beta}{n - k} (x - 1)^{n-k} (x + 1)^k.$$

[30] H. Epstein, J. Math. Phys. **8**, 750 (1967).

[31] S. Ström, *Lectures in Theoretical Physics* (The Univ. of Colorado Press, Boulder, Colorado, 1964), Vol. VIIA, p. 70–78.

[32] V. D. Dao and V. H. Nguyen, Ann. Inst. H. Poincaré **6**, 17 (1967).

[33] H. Joos, Fortschr, Physik **10**, 65 (1962).

From (3.9) and (3.25) we find the projection operator $\Pi_s(\mathbf{p})$ to a given spin s and $\mathbf{p} = (0, 0, |\mathbf{p}|)$:

$$\begin{aligned}\Pi_s(0,0,|\mathbf{p}|)^{l\eta}_{l'\eta'} &= \sum_{\zeta=-s}^{s} u^{l\eta}_{s\zeta}(0,0,|\mathbf{p}|)\overline{u^{l'\eta'}_{s\zeta}(0,0,|\mathbf{p}|)} \\ &= \delta^{\eta}_{\eta'}\frac{(s-|\eta|)!}{(s+|\eta|)!}\left[\frac{(l+|\eta|)!\,(l'+|\eta|)!}{(l-|\eta|)!\,(l'-|\eta|)!}\right]^{\frac{1}{2}} \\ &\quad\times\frac{(2m)^{2|\eta|+1}(-|\mathbf{p}|)^{l+l'-2s}}{(p^0+m)^{l+l'+2|\eta|+1-2s}} \\ &\quad\times P^{(l-s,2|\eta|)}_{s-|\eta|}\left(\frac{m}{p^0}\right)P^{(l'-s,2|\eta|)}_{s-|\eta|}\left(\frac{m}{p^0}\right). \qquad (3.26)\end{aligned}$$

Actually, (3.26) gives $\Pi_s(\mathbf{p})$ for any $\mathbf{p}$ if η and η' correspond to given helicity (rather than to the third projection of the spin). The helicity basis $h_{s\zeta}(\mathbf{p})$ is defined by

$$h_{s\zeta}(\mathbf{p}) = \sum_{\eta=-s}^{s} D^{(s)}_{\zeta\eta}(U)\,|s\eta),$$

where U is the 2×2 unitary matrix defined by

$$B_p = U^* D_p U,$$

with

$$D_p = \frac{1}{[2m(p^0+m)]^{\frac{1}{2}}}\begin{pmatrix} p^0+m+|\mathbf{p}| & 0 \\ 0 & p^0+m-|\mathbf{p}| \end{pmatrix}.$$

If we put

$$p^0 = m\cosh\lambda, \quad \mathbf{p} = m\sinh\lambda\cdot\mathbf{n},$$
$$n_1 + in_2 = \sin\theta e^{i\varphi}, \quad n_3 = \cos\theta,$$

then $D_p = D_\lambda$ and

$$U = U(\theta,\varphi) = \begin{pmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2}e^{-i\varphi} \\ -\sin\frac{\theta}{2}e^{i\varphi} & \cos\frac{\theta}{2} \end{pmatrix}.$$

From the definition of $V(D_\lambda)$ we have for any unitary representation

$$V(D_\lambda)^{l\zeta}_{s\zeta} = \overline{V(D_{-\lambda})^{s\zeta}_{l\zeta}}. \qquad (3.27)$$

In the Majorana representation, for our choice of the basis, the matrix elements of $V(D_{-\lambda})$ are real so that the bar in (3.27) can be dropped.

It is interesting to mention that the expressions (3.25) for the matrix elements of $V(D_\lambda)$ can be continued analytically for complex values of λ and have pole-type singularities at $\lambda = i\pi(2n+1)$, $n = 0, \pm1, \cdots$ (these are the points corresponding to the reflection of the axes x^0 and x^3). This is not a peculiarity of the Majorana representations only. It is shown in Appendix C, that a pole-type singularity appears at the same point for some reducible unitary representations of $SL(2, C)$.

D. First-Order Calculation for the Vertex Function

We shall consider here as an example the vertex function corresponding to the local interaction Lagrangian

$$\mathfrak{L}_I(x) = g\,{:}\varphi^*(x)\varphi(x)A(x){:}, \qquad (3.28)$$

where φ is either the Bose field (3.2) or the Fermi field (3.15) and A is a Hermitian scalar field. In first order with respect to g the vertex function

$$\langle 0|\, a_{s'\zeta'}(\mathbf{p})\left[\int \mathfrak{L}_I(x)\,d^4x,\, \tilde{A}(p-q)\right]a^*_{s\zeta}(\mathbf{q})\,|0\rangle$$

is proportional to the scalar product

$$(u_{s'\zeta'}(\mathbf{p}), u_{s\zeta}(\mathbf{q})) = \sum_{l\zeta}\overline{u^{l\eta}_{s'\zeta'}}(\mathbf{p})u^{l\eta}_{s\zeta}(\mathbf{q}). \qquad (3.29)$$

We shall evaluate (3.29) for the Majorana representations using (3.9) and (3.25).

We first remark that because of the unitarity of V we have

$$(u_{s'\zeta'}(\mathbf{p}), u_{s\zeta}(\mathbf{q})) = V(B_p^{-1}B_q)^{s'\zeta'}_{s\zeta}, \qquad (3.30)$$

where B_p is given by (3.7) and $B^{-1}_{(p^0,\mathbf{p})} = B_{(p^0,-\mathbf{p})}$, so that the problem of evaluation of (3.29) is reduced to the calculation of a matrix element of V. Choosing the z axis along $\mathbf{p}$ and the y axis orthogonal to $\mathbf{q}$ we can write

$$\begin{aligned} p &= m(\cosh\alpha, 0, 0, \sinh\alpha), \\ q &= m(\cosh\beta, \sinh\beta\sin\varphi, 0, \sinh\beta\cos\varphi). \end{aligned} \qquad (3.31)$$

We calculate $B_p^{-1}B_q$ in terms of α, β, and φ:

$$\begin{aligned} A \equiv B_p^{-1}B_q &= \cosh\frac{\alpha}{2}\cosh\frac{\beta}{2} - \sinh\frac{\alpha}{2}\sinh\frac{\beta}{2}\cos\varphi \\ &\quad + \cosh\frac{\alpha}{2}\sinh\frac{\beta}{2}\sin\varphi\sigma_1 - i\sinh\frac{\alpha}{2}\sinh\frac{\beta}{2}\sin\varphi\sigma_2 \\ &\quad + \left(\cosh\frac{\alpha}{2}\sinh\frac{\beta}{2}\cos\varphi - \cosh\frac{\beta}{2}\sinh\frac{\alpha}{2}\right)\sigma_3. \end{aligned} \qquad (3.32)$$

Introducing the (real) parameter λ by

$$\cosh\lambda = \cosh\alpha\cosh\beta - \sinh\alpha\sinh\beta\cos\varphi = \frac{1}{m^2}pq, \qquad (3.33)$$

we can decompose A in the form

$$A = U_1 D_\lambda U_2^*, \qquad (3.34)$$

where D_λ is given by (3.18) and

$$U_j = \begin{pmatrix} \cos\frac{1}{2}\theta_j & -\sin\frac{1}{2}\theta_j \\ \sin\frac{1}{2}\theta_j & \cos\frac{1}{2}\theta_j \end{pmatrix}, \quad j = 1, 2. \qquad (3.35)$$

The angles θ_j are expressed in terms of α, β, and φ by

$$\tan\frac{\theta_1-\theta_2}{2}=\frac{\sinh\frac{1}{2}\alpha\sinh\frac{1}{2}\beta\sin\varphi}{\cosh\frac{1}{2}\alpha\cosh\frac{1}{2}\beta-\sinh\frac{1}{2}\alpha\sinh\frac{1}{2}\beta\cos\varphi}=\frac{|\mathbf{p}\times\mathbf{q}|}{(p^0+m)(q^0+m)-\mathbf{pq}},$$

$$\tan\frac{\theta_1+\theta_2}{2}=\frac{\cosh\frac{1}{2}\alpha\sinh\frac{1}{2}\beta\sin\varphi}{\cosh\frac{1}{2}\alpha\sinh\frac{1}{2}\beta\cos\varphi-\sinh\frac{1}{2}\alpha\cosh\frac{1}{2}\beta}=\frac{|\mathbf{p}\times\mathbf{q}|}{\mathbf{pq}-(p^0-m)(q^0+m)},\tag{3.36}$$

where $\mathbf{p}\times\mathbf{q}$ is the vector product of $\mathbf{p}$ and $\mathbf{q}$: $(\mathbf{p}\times\mathbf{q})_j=\epsilon_{jkl}p_kp_l$. We mention that for our choice of the coordinate system in the three-dimensional space the matrix A as well as the matrices U_j and D_λ in (3.34) are real.

Now, for the evaluation of (3.30) it is sufficient to apply (3.20) to the matrix (3.34). The matrix elements of $V(D_\lambda)$ for the Majorana representations are given by (3.25) while the $D(U)^{(s)}_{\eta\zeta}$ for a real orthogonal U is expressed by[33]

$$D\begin{pmatrix}\cos\frac{1}{2}\theta & \sin\frac{1}{2}\theta\\ -\sin\frac{1}{2}\theta & \cos\frac{1}{2}\theta\end{pmatrix}^{(s)}_{\eta\zeta}=\left[\frac{(s+\zeta)!\,(s-\zeta)!}{(s+\eta)!\,(s-\eta)!}\right]^{\frac{1}{2}}\times(\cos\tfrac{1}{2}\theta)^{\zeta+\eta}(\sin\tfrac{1}{2}\theta)^{\zeta-\eta}P^{(\zeta-\eta,\zeta+\eta)}_{s-\zeta}(\cos\theta).\tag{3.37}$$

In particular, for $s=0$ and $\frac{1}{2}$ we have[34]:

$$(u_{00}(\mathbf{p}),u_{00}(\mathbf{q}))=V(H_\lambda)^{00}_{00}=\frac{1}{\cosh\frac{1}{2}\lambda}=\left[\frac{2m^2}{pq+m^2}\right]^{\frac{1}{2}}=\left[\frac{4m^2}{4m^2-t}\right]^{\frac{1}{2}},\tag{3.38}$$

$$t=(p-q)^2=2m^2-2pq;$$

$$\begin{aligned}&(u_{\frac{1}{2}\eta}(\mathbf{p}),u_{\frac{1}{2}\zeta}(\mathbf{q}))\\&=V(H_\lambda)^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}D(U_1U_2)^{(\frac{1}{2})}_{\eta\zeta}\\&=\frac{1}{\cosh^2\frac{1}{2}\lambda}\left(\cos\frac{\theta_1-\theta_2}{2}\,\delta_{\eta\zeta}-\sin\frac{\theta_1-\theta_2}{2}\,\epsilon_{\mu\zeta}\right)\\&=\frac{1}{2m}\left(\frac{4m^2}{4m^2-t}\right)^{\frac{3}{2}}[(p^0+m)(q^0+m)]^{-\frac{1}{2}}\\&\quad\times\{[2m^2-\tfrac{1}{2}t+m(p^0+q^0)]\sigma_0-i\mathbf{p}\times\mathbf{q}\boldsymbol{\sigma}\}_{\frac{1}{2}-\eta,\frac{1}{2}-\zeta}.\end{aligned}\tag{3.39}$$

[34] Formula (3.38) coincides with the result of Ref. 25, see also A. O. Barut and H. Kleinert [Phys. Rev. Letters **18**, 754 (1967)], where the scalar product $(u_{ls}(\mathbf{0}),u_{ss}(0,0\,|\mathbf{q}|))$ is calculated for all unitary representations of $SL(2,C)$ of the type $[\frac{1}{2},i\sigma]$. We mention that our formulas are valid for $(\mu_{l\eta}(\mathbf{p}_j),\mu_{s\zeta}(\mathbf{q}))$ for any $\zeta(-s\leq\zeta\leq s)$ and for arbitrary (not necessarily collinear) 3-momenta $\mathbf{p}$ and $\mathbf{q}$. We remark that the matrix multiplying the invariant form factor $[4m^2/4m^2-t]^{\frac{3}{2}}$ coincides with the scalar product of two positive-energy Dirac spinors $\tilde{u}_\eta(\mathbf{p})u_\zeta(\mathbf{q})\equiv u^*_\eta(\mathbf{p})\gamma^0u_\zeta(\mathbf{q})$.

4. QUANTIZATION OF MAJORANA FIELDS

A. Introductory Remarks

The only examples of fields transforming under an irreducible unitary representation of $SL(2,C)$, which can be quantized in terms of anticommutators without getting into contradiction with locality are the two Majorana fields. Each of them is defined as a field transforming under one of the irreducible self-coupled representations and satisfying the Majorana equation

$$(i\partial_\mu L^\mu-\kappa)\psi(x)=0,\tag{4.1}$$

which corresponds to the free-field Lagrangian

$$\mathfrak{L}(x)=\psi^*(x)(i\partial_\mu L^\mu-\kappa)\psi(x).\tag{4.2}$$

We shall see in the next subsection that the Majorana equation has a discrete spectrum of timelike solutions with masses decreasing with the spin and a continuous spectrum of solutions with spacelike momenta. These unrealistic features of the spectrum are not peculiarities of the Majorana equation only, but have a rather general character.[35]

The quantization of the Majorana field will be performed in Sec. 4C. We shall see that the Majorana field can be regarded as a superposition of annihilation operators only and can be quantized by both commutators and anticommutators without being in contradiction with the locality of the field.

B. Complete Set of Solutions of the Majorana Equation

We first review the classical solutions of the Majorana equation in momentum space

$$(L^\mu p_\mu-\kappa)u(\mathbf{p})=0.\tag{4.3}$$

We consider (4.3) as an eigenvalue problem for the energy p^0 for fixed $\mathbf{p}$. We mention that the operator

$$H=(L^0)^{-1}(L^kp^k+\kappa)\tag{4.4}$$

is Hermitian with respect to the scalar product

$$(u,v)_H=(u,L^0v),\tag{4.5}$$

where (u,v) is the (positive-definite) scalar product in the space X, where the unitary Majorana representation acts. We recall that L^0 is a positive operator because of (2.23) and hence it has a positive inverse. It follows that the eigenvectors of H [or otherwise the solutions of (4.3)] corresponding to different eigenvalues of p^0 are orthogonal with respect to the product (4.5).

[35] W. Rühl, Commun. Math. Phys. **6**, 312 (1967).

The simplest way to obtain the spectrum of Eq. (4.3) is to multiply both sides by $L^\mu p_\mu + \kappa$ and to use

$$(L^\mu p_\mu)^2 = \tfrac{1}{4}p^2 - w^2, \tag{4.6}$$

where w is the Pauli–Lubanski–Bargmann 4-vector (3.4). [One way to check (4.6) is to use the explicit expressions (2.17) and (2.18) of L^μ and $S^{\mu\nu}$ in terms of creation and annihilation operators.] Substituting (4.6) in the equation thus obtained we find

$$(\tfrac{1}{4}p^2 - w^2 - \kappa^2)u(\mathbf{p}) = 0. \tag{4.7}$$

We have to consider essentially two different cases depending on the eigenvalue of p^0: the case of timelike and the case of spacelike momentum p. The intermediate case of lightlike momentum can be obtained by going to the limit from either side.

In the case of timelike momenta, where $p^2 > 0$, the little group generated by w_μ is $SU(2)$ and

$$w^2 = -p^2 s(s+1), \tag{4.8}$$

with

$$s = \begin{cases} 0, 1, \cdots, & \text{for } [1_0, 1_1] = [0, \tfrac{1}{2}], \\ \tfrac{1}{2}, \tfrac{3}{2}, \cdots, & \text{for } [1_0, 1_1] = [\tfrac{1}{2}, 0]. \end{cases} \tag{4.9}$$

This immediately leads to the decreasing mass spectrum (for $\kappa > 0$):

$$p^2 = \frac{\kappa^2}{(s+\tfrac{1}{2})^2}. \tag{4.10}$$

Because of the positive-definiteness of L^0 only positive eigenvalues of p^0 appear in this case. It follows in particular that H is semibounded (from above):

$$H \leq \left[\left(\frac{\kappa}{1_0 + \tfrac{1}{2}}\right)^2 + \mathbf{p}^2\right]^{\frac{1}{2}}. \tag{4.11}$$

For p spacelike, the little group conserving p is $SU(1,1) \sim SL(2,R)$ and we have the same formula (4.8) for w^2, but with s in the range

$$s = -\tfrac{1}{2} + i\frac{\sigma}{2}, \quad -\infty < \sigma < \infty. \tag{4.12}$$

These are exactly the values of s involved in the Sommerfeld–Watson integral in Regge theory of complex angular momentum.

The limiting cases $s \to \infty$ ($\sigma \to \infty$) correspond to solutions with lightlike momenta.

Proceeding to the determination of the eigenvectors, we shall discuss here only the case $\kappa > 0$ in terms of the variables ξ_α (2.19). Both this case and the case $\kappa = 0$ are treated in terms of "Schrödinger variables" z and $\bar{z}$ in Appendix B.

It is sufficient to consider the case when the momentum $\mathbf{p}$ is directed along the third axis, the general case being obtained from this through a three-dimensional rotation.

We treat the discrete and the continuous spectrum simultaneously diagonalizing the Hamiltonian (4.4) for a fixed $|\mathbf{p}| \neq 0$; we put

$$p = 2\frac{\kappa}{\lambda}(\beta, 0, 0, 1). \tag{4.13}$$

Substituting (2.18), (2.19), and (4.13) in (4.3) we obtain

$$\left[\frac{\partial^2}{\partial\xi_1\partial\xi_2} + \xi_1\xi_2 + \beta\left(\xi_\alpha\frac{\partial}{\partial\xi_\alpha} + 1\right) - \lambda\right] \times u_\beta\left(0, 0, \frac{\kappa}{\lambda}; \xi\right) = 0. \tag{4.14}$$

To get rid of the degeneracy we require, in addition, that u_β be an eigenvector of the third spin component

$$u_\beta = u_{\beta\zeta}, \quad \left[\frac{1}{2}\left(\xi_1\frac{\partial}{\partial\xi_1} - \xi_2\frac{\partial}{\partial\xi_2}\right) - \zeta\right]u_{\beta\zeta} = 0. \tag{4.15}$$

Further, we make the change of variables

$$\xi_1 = (\tfrac{1}{2}y)^{\frac{1}{2}}e^{i(\varphi/2)}, \quad \xi_2 = (\tfrac{1}{2}y)^{\frac{1}{2}}e^{-i(\varphi/2)}, \tag{4.16}$$

and putting

$$u_{\beta\zeta}\left(0, 0, 2\frac{\kappa}{\lambda}; \xi\right) = f_{\beta|\zeta|}(\lambda, y)e^{i\zeta\varphi}, \tag{4.17}$$

we find the following equation for $f_{\beta|\zeta|}$:

$$\left\{2\frac{d}{dy}\left(y\frac{d}{dy}\right) + \tfrac{1}{2}y - 2\frac{\zeta^2}{y} + \beta\left(2y\frac{d}{dy} + 1\right) - \lambda\right\} \times f_{\beta|\zeta|}(\lambda, y) = 0. \tag{4.18}$$

The solution of (4.18), regular for $y = 0$, is

$$f_{\beta|\zeta|}(\lambda, y) = C_{\lambda\beta|\zeta|}y^{|\zeta|}\exp\{\tfrac{1}{2}[(\beta^2-1)^{\frac{1}{2}} - \beta]y\} \times \Phi\left(\tfrac{1}{2} + |\zeta| - \frac{\lambda}{2(\beta^2-1)^{\frac{1}{2}}}, 1 + 2|\zeta|, -(\beta^2-1)^{\frac{1}{2}}y\right), \tag{4.19}$$

where $\Phi(a, c, z)$ is the confluent hypergeometric function defined by the Kummer series

$$\Phi(a, c, z) = 1 + \frac{a}{c}\frac{z}{1!} + \frac{a(a+1)}{c(c+1)}\frac{z^2}{2!} + \cdots \tag{4.20}$$

and $C_{\lambda\beta|\zeta|}$ is a normalization constant.

The discrete timelike solutions are obtained from (4.19) for

$$\beta = \beta_s = \left[1 + \frac{\lambda^2}{(2s+1)^2}\right]^{\frac{1}{2}},$$

i.e.,

$$(\beta_s^2 - 1)^{\frac{1}{2}} = \frac{\lambda}{2s+1}. \tag{4.21}$$

We find

$$\begin{aligned} f_{s|\zeta|}(\lambda, y) &= C_{\lambda s|\zeta|} y^{|\zeta|} \exp\left\{\frac{1}{2}\left(\frac{\lambda}{2s+1} - \beta_s\right) y\right\} \\ &\quad \times \Phi\left(|\zeta| - s, 1 + 2|\zeta|, -\frac{\lambda y}{2s+1}\right) \\ &= \frac{(s-|\zeta|)!\,(2|\zeta|)!}{(s+|\zeta|)!} C_{\lambda s|\zeta|} y^{|\zeta|} \\ &\quad \times \exp\left\{\frac{1}{2}\left(\frac{\lambda}{2s+1} - \beta_s\right) y\right\} \\ &\quad \times L_{s-|\zeta|}^{2|\zeta|}\left(-\frac{\lambda y}{2s+1}\right), \end{aligned} \tag{4.22}$$

where $L_n^{\alpha}(x)$ are the Laguerre polynomials (B9).

Requiring the normalization

$$(u_{s\zeta}(\mathbf{p}), L^0 u_{s'\zeta'}(\mathbf{p})) = \frac{\omega_s \kappa}{m_s^2} \delta_{ss'} \delta_{\zeta\zeta'} \tag{4.23}$$

with

$$m_s = \frac{\kappa}{s + \frac{1}{2}}, \quad \omega_s = (m_s^2 + \mathbf{p}^2)^{\frac{1}{2}} \left(= \beta_s \frac{\kappa}{\lambda}\right), \tag{4.24}$$

we obtain for the constants C

$$\begin{aligned} C_{\lambda s|\zeta|} &= \frac{1}{\sqrt{2}}\left[\frac{(s+|\zeta|)!}{(s-|\zeta|)!}\right]^{\frac{1}{2}} \left(\frac{2s+1}{\lambda}\right)^{s-|\zeta|} \frac{1}{(2|\zeta|)!} \\ &\quad \times \frac{1}{(1 + \{1 + [(2s+1)\lambda^{-1}]^2\}^{\frac{1}{2}})^{s+\frac{1}{2}}}. \end{aligned} \tag{4.25}$$

With this normalization

$$u_{s\zeta}(\mathbf{0}) = \lim_{\lambda \to \infty} u_{s\zeta}\left(0, 0, 2\frac{\kappa}{\lambda}\right) = |s\zeta) \tag{4.26}$$

[$|s\zeta)$ belongs to the canonical basis (2.21)].

A tedious but straightforward calculation leads to the following formula for the components of the eigenvectors of the discrete spectrum in the canonical basis:

$$\begin{aligned} &(s+\zeta)!\,(s-\zeta)!\,(l+\eta)!\,(l-\eta)!]^{-\frac{1}{2}} u_{s\zeta}^{l\eta}(\mathbf{p}) \\ &= \sum_{ijkn} \frac{(-1)^{n+i} b_s^{2n+s-l} (1 - b_s^2)^{l-n+\frac{1}{2}} q_+^{i+j+k+\zeta-\eta} q_-^{i+j+k} q_3^{2(n-i-j-k)+s-l-\zeta+\eta}}{j!\,k!\,(i+j)!\,(i+k+\zeta-\eta)!\,(n+s-l-j-k)!\,(l-n+\zeta+k-j)!\,(l-n-\zeta-k+j)!\,(n-2i-j-k-\zeta+\eta)!}, \end{aligned} \tag{4.27}$$

where

$$b_s = \frac{|\mathbf{p}|}{\omega_s + m_s}, \quad q_{\pm} = \frac{1}{2|\mathbf{p}|}(p^1 \pm ip^2), \quad q_3 = \frac{p^3}{|\mathbf{p}|},$$

the range of summation being spread over all integers i, j, k, n for which the factorials in the denominator of the right-hand side are finite.

We mention that in general the spinor $u_{s\zeta}(m, \mathbf{p})$ corresponding to spin s, mass m, and three-momentum $\mathbf{p}$ is a function of $\mathbf{p}/m$ only so that the spinors (4.27) are connected to (3.12) by

$$u_{s\zeta}(m_s, \mathbf{p}) = u_{s\zeta}\left(m, \frac{m}{m_s}\mathbf{p}\right).$$

For the continuous spectrum we have to substitute (4.12) in (4.21) thus obtaining

$$\beta \equiv \beta_\sigma = \epsilon\left(1 - \frac{\lambda^2}{\sigma^2}\right)^{\frac{1}{2}}, \quad \epsilon = \pm 1, \quad (\beta^2 - 1)^{\frac{1}{2}} = i\frac{\lambda}{\sigma}. \tag{4.28}$$

We need both signs of β because for spacelike p the sign of the energy is not invariant under proper Lorentz transformations.

To obtain the limit of lightlike momenta ($\sigma \to \infty$, $\beta \to 1$) we use the known result that for $k = \frac{1}{2}c - a \to \infty$ and $|kx|$ bounded

$$\Phi(a, c, x) = \Gamma(c)(kx)^{\frac{1}{2}-\frac{1}{2}c} e^{x/2} J_{c-1}(2\sqrt{kx}) + O(|k|^{-1}),$$

where J_n is the Bessel function (see, for instance, Ref. 36). Applying this to $c = 1 + 2|\zeta|$, $k = -\lambda/2(\beta^2 - 1)^{\frac{1}{2}}$, $kx = -\lambda y/2$, we find

$$\lim_{\beta \to 1} f_{\beta|\zeta|}(\lambda, y) = C_{\lambda|\zeta|}(2|\zeta|)! \left(-\frac{2}{\lambda}\right)^{|\zeta|} e^{-\frac{1}{2}y} J_{2|\zeta|}[(-2\lambda y)^{\frac{1}{2}}] \tag{4.29}$$

with

$$C_{\lambda|\zeta|} = \lim_{\beta \to 1} C_{\lambda\beta|\zeta|}.$$

We shall normalize the eigenfunctions of the continuous spectrum by

$$(u_{\epsilon\sigma\zeta}(\mathbf{p}), L^0 u_{\epsilon'\sigma'\zeta'}(\mathbf{p})) = \frac{\omega_\sigma}{\kappa} \delta_{\epsilon\epsilon'} \delta_{\zeta\zeta'} \delta(\sigma - \sigma'), \tag{4.30}$$

where ϵ is the sign of the energy [see (4.28)] and

$$\omega_\sigma = |p_\sigma^0| = \left[\mathbf{p}^2 - \frac{\kappa^2}{\sigma^2}\right]^{\frac{1}{2}}. \tag{4.31}$$

[36] *The Bateman Manuscript Project: Higher Transcendental Functions Vol. I*, A. Erdélyi, Ed. (McGraw-Hill Book Co., Inc., New York, 1953), Sec. 6.13.2, Eq. (15).

The possibility for such a normalization (with a finite C) is checked directly (cf. Ref. 15 Appendix B). It proves that $u_{\epsilon\sigma\zeta}$ actually belongs to the continuous spectrum of H. We see from (4.31) that for a given $\mathbf{p}$ only those σ of the range (4.12) appear for which

$$|\sigma| \geq \kappa/|\mathbf{p}|. \tag{4.32}$$

The completeness relation for the eigenvectors of the self-adjoint operator H (4.4) may be written in the canonical basis $|l\eta)$ as follows:

$$\sum_{s\zeta} \frac{m_s^2}{\kappa\omega_s} u_{s\zeta}^{l\eta}(\mathbf{p})(\bar{u}_{s\zeta}(\mathbf{p})L^0)_{l'\eta'} + \sum_{\epsilon\zeta} \int_{|\sigma|\geq\kappa/|\mathbf{p}|} \frac{\kappa}{\omega_\sigma} u_{\epsilon\sigma\zeta}^{l\eta}(\mathbf{p})(\bar{u}_{\epsilon\sigma\zeta}(\mathbf{p})L^0)_{l'\eta'}\, d\sigma = \delta_{l'}^{l}\delta_{\eta'}^{\eta}. \tag{4.33}$$

C. Quantization of the Majorana Field

The results of the preceding section allow us to write down the general solution of Eq. (4.1) in the form

$$\psi(x) = \frac{1}{(2\pi)^{\frac{3}{2}}} \left\{ \sum_{s\zeta} \frac{m_s}{\kappa^{\frac{1}{2}}} \int_{p^0=\omega_s} a_{s\zeta}(\mathbf{p}) u_{s\zeta}(\mathbf{p}) e^{-ipx} \frac{d^3p}{p^0} + \kappa^{\frac{1}{2}} \sum_{\epsilon\zeta} \int d\sigma \int_{p^2=-\kappa^2/\sigma^2} b_{\epsilon\sigma\zeta}(\mathbf{p}) u_{\epsilon\sigma\zeta}(\mathbf{p}) e^{-ipx} \frac{d^3p}{|p^0|} \right\}. \tag{4.34}$$

The normalization factors are chosen in such a way that

$$P^\mu = \int_{x^0=t} \frac{\partial\mathfrak{L}}{\partial\left(\frac{\partial\psi}{\partial x^0}\right)} \frac{\partial\psi}{\partial x_\mu}\, d^3\mathbf{x} = \sum_{s\zeta} \int_{p^0=\omega_s} p^\mu a_{s\zeta}^*(\mathbf{p}) a_{s\zeta}(\mathbf{p}) \frac{d^3\mathbf{p}}{p^0} + \sum_{\epsilon\zeta} \int d\sigma \int_{|p^0|=\omega_\sigma} p^\mu b_{\epsilon\sigma\zeta}^*(\mathbf{p}) b_{\epsilon\sigma\zeta}(\mathbf{p}) \frac{d^3\mathbf{p}}{|p^0|}. \tag{4.35}$$

The energy [given by (4.35) for $\mu = 0$] is not positive-definite because of the second term in the right-hand side. It has been argued in Ref. 37 that in general one cannot introduce a covariant 4-momentum operator P^μ with positive P^0 and $P^2 < 0$.

We are evidently free to postulate either canonical commutations or canonical anticommutations for a and b:

$$[a_{s\zeta}(\mathbf{p}), a_{s'\zeta'}^*(\mathbf{q})]_\pm = \omega_s(\mathbf{p})\delta(\mathbf{p}-\mathbf{q})\delta_{ss'}\delta_{\zeta\zeta'},$$
$$[b_{\epsilon\sigma\zeta}(\mathbf{p}), b_{\epsilon'\sigma'\zeta'}^*(\mathbf{q})]_\pm = \omega_\sigma(\mathbf{p})\delta(\mathbf{p}-\mathbf{q})\delta(\sigma-\sigma')\delta_{\zeta\zeta'}\delta_{\epsilon\epsilon'}. \tag{4.36}$$

For both signs these relations imply the canonical equal-time local (anti)commutation relations for the field $\psi(x)$ and its conjugate momentum

$$\frac{\partial\mathfrak{L}}{\partial\left(\frac{\partial\psi}{\partial x^0}\right)} = i\psi^* L^0.$$

Namely, (4.36) together with the completeness relation (4.33) gives

$$[\psi^{l\eta}(t, \mathbf{x}), (\psi^*(t, \mathbf{y})L^0)_{l'\eta'}]_\pm = \delta_{l'}^{l}\delta_{\eta'}^{\eta}\delta(\mathbf{x}-\mathbf{y}). \tag{4.37}$$

This proves in particular the locality of Majorana fields.

Let us mention an important difference between the scalar "tachyon" field satisfying the equation

$$(p^2 + m^2)\tilde{\varphi}(p) = 0$$

(see Ref. 38) and the infinite-component Majorana field. In the case treated in Ref. 38 one does not have a complete set of plane-wave solutions because of the condition

$$\mathbf{p}^2 \geq m^2 > 0 \tag{4.38}$$

for m fixed. It implies the nonlocal character of the "tachyon" field. In our case, instead of (4.38) we have

$$\mathbf{p}^2 \geq \kappa^2/\sigma^2$$

[see (4.32)], but σ varies up to infinity so that all plane waves are admitted. This is a consequence of the fact that we can define a self-adjoint Hamiltonian H [see (4.4)] for any choice of the space momentum $\mathbf{p}$. That is the reason why it is possible to construct a local Majorana field.

We mention that the field (4.34) supplies an example of a local field which annihilates the vacuum. This is not in contradiction with the well-known theorem, asserting that if a local field fulfills Wightman axioms and annihilates the vacuum it vanishes identically,[28] because the spectral conditions are violated in our case.

The Majorana equation (4.1) implies the conservation of the current

$$j^\mu(x) = :\psi^*(x)L^\mu\psi(x):. \tag{4.39}$$

The corresponding conserved charge is positive and gives the particle number (including the number of "particles" with spacelike momentum).

A characteristic feature of the irreducible Majorana field is that it describes (Fermi) particles without antiparticles, so it clearly violates the *TCP* theorem.

[37] M. E. Arons and E. C. G. Sudarshan, Syracuse University preprint, 1967 (unpublished).

[38] G. Feinberg, Phys. Rev. **159**, 1089 (1967).

We can also construct infinite-component Fermi fields which include both particles and antiparticles by using the pairs of coupled representations discussed in Sec. 2C.

5. CONCLUDING REMARKS

Let us try to summarize the situation with infinite-component fields.

These fields supply a description of an infinite set of particles with increasing spin values (either all integers or all half-integers). The spin and statistics theorem as well as the crossing properties connected with it are not a consequence of the general requirements (such as Lorentz invariance with spectral properties of P^μ and locality) as they are in the conventional theory of finite-dimensional fields. However, *TCP* and spin and statistics contradict neither the infinite dimensionality of the field nor the unitarity of the representation of $SL(2, C)$ under which the field transforms. Examples can be constructed both for integer and half-integer spin infinite-component fields which satisfy *TCP* and have the right connection between spin and statistics. All known examples of local infinite-component fields (with a nondegenerate mass spectrum) have some of the peculiar properties of the Majorana field considered in Sec. 4: They satisfy an equation which has spacelike solutions in momentum space and for which the point $p^2 = 0$ is an accumulation point for the timelike spectrum. The existence of spacelike components is a natural and even unavoidable feature of interacting fields,[39] but if it appears for free (asymptotic) fields it violates the spectral properties of the representation of the Poincaré group leading to the existence of faster-than-light particles. It is not clear to us whether or not one can construct a relativistic *interacting* infinite component field (with a non-C-number commutator) which satisfies a linear (free-type) differential equation.

On the other hand, to get a reasonable spin dependence of the mass in the free-field equation, avoiding the spacelike solutions, it seems necessary to introduce a nonlocal term of the type

$$-w^2/p^2 = s(s+1).$$

Such a term may not contradict the locality of the projected finite-component fields with definite spin[14] and, as suggested by composite models, one has actually to expect some nonlocal properties of the infinite-component fields.

It is clear also that conventional local Lagrangian formalism may well not be the best way to introduce infinite-dimensional representations of $SL(2, C)$ in the theory. It might appear more appropriate to start with a current algebra and try to saturate current commutation relations or superconvergent sum rules with an infinite multiplet (cf. Ref. 40) and the suggestion of de Alfaro *et al.* in Ref. 41). The relation between the two approaches is not yet clarified.

In any case, the study of the general framework in which $V(A)$ in (1.1) is not supposed to be finite-dimensional has thrown some new light on the logical structure of quantum field theory and we are persuaded that the infinite-component fields deserve further investigation from both theoretical and practical point of view.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor N. N. Bogoliubov, Professor P. Budini, Professor G. F. Dell'Antonio, Professor C. Fronsdal, Professor I. T. Grodsky, Professor Nguyen van Hieu, Professor Abdus Salam, Professor E. C. G. Sudarshan, and Professor R. P. Zaikov for illuminating discussions at different stages of this work.

The second named author (I. T. T.) is grateful to Professor Abdus Salam and Professor P. Budini and to the IAEA for the hospitality kindly extended to him at the International Centre for Theoretical Physics, Trieste, where the main part of this work was done, and to M. L. Motchane for his kind hospitality at the Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, where the present paper was completed.

APPENDIX A: DESCRIPTION OF THE IRREDUCIBLE REPRESENTATIONS OF $SL(2, C)$ IN TERMS OF HOMOGENEOUS FUNCTIONS

This appendix gives a summary of known results used throughout the paper.

Consider the space $X_{\nu_1\nu_2}$ of one-valued homogeneous functions of two complex variables

$$X_{\nu_1\nu_2} = \{f \,|\, f(\lambda z_1, \lambda z_2) = \lambda^{\nu_1}\lambda^{-\nu_2} f(z_1, z_2)\}. \quad \text{(A1)}$$

The condition of one-valuedness implies that the difference $\nu_1 - \nu_2$ has to be an integer. The group $SL(2, C)$ acts in $X_{\nu_1\nu_2}$ as a transformation group in the

[39] It has been proved that if the Fourier transform of a local field vanishes in a domain of spacelike vectors in momentum space, then the field is a generalized free field. See G. F. Dell'Antonio, J. Math. Phys. **2**, 759 (1961); O. W. Greenberg, J. Math. Phys. **3**, 859 (1962).

[40] G. Cocho, C. Fronsdal, I. T. Grodsky, and R. White, Phys. Rev. **162**, 1662 (1967); R. Delbourgo, M. A. Rashid, Abdus Salam, and J. Strathdee, Phys. Letters **25B**, 475 (1967); C. Fronsdal, ICTP Trieste Preprint IC/67/70 (1967).

[41] V. De Alfaro, S. Fubini, G. Furlan, and C. Rossetti, Torino University preprint 1966.

space of (spinorial) variables $z = (z_1, z_2)$:

$$[V(A)f](z) = f(zA) = f(A^T z). \tag{A2}$$

The representation (A2) of $SL(2, C)$ turns out to be irreducible in $X_{\nu_1\nu_2}$. Its number $[l_0, l_1]$ is related to the degree of homogeneity (ν_1, ν_2) by

$$l_0 = \tfrac{1}{2}(\nu_1 - \nu_2), \quad l_1 = \tfrac{1}{2}(\nu_1 + \nu_2) + 1,$$
$$\nu_1 = l_1 + l_0 - 1, \quad \nu_2 = l_1 - l_0 - 1. \tag{A3}$$

We see that l_0 is always integer or half-integer while l_1 is, in general, an arbitrary complex number. Two representations, $[l_0, l_1]$ and $[l_0', l_1']$ are equivalent if and only if[42]

$$[l_0', l_1'] = [\pm l_0, \pm l_1]$$

(i.e., either $\nu_\alpha' = \nu_\alpha$ or $\nu_\alpha' = -\nu_\alpha - 2$, $\alpha = 1, 2$). (A4)

The representation (A2) is finite-dimensional when ν_1 and ν_2 are both integers of the same sign [its dimensionality being $(\nu_1 + 1)(\nu_2 + 1) = l_1^2 - l_0^2$]. In this case l_0 and l_1 are simultaneously integer or half-integer and $|l_1| \geq |l_0| + 1$. The spin content of the finite-dimensional representations is given by

$$s = |l_0|, |l_0| + 1, \cdots, |l_1| - 1. \tag{A5}$$

If the above condition is not fulfilled, the representation $[l_0, l_1]$ is infinite-dimensional, the sequence (A5) in that case not being limited from above.

To calculate the Casimir operators in terms of l_0 and l_1 we use the formulas

$$\tfrac{1}{4}(\mathbf{M} + i\mathbf{N})^2 = \frac{\nu_1}{2}\left(\frac{\nu_1}{2} + 1\right),$$
$$\tfrac{1}{4}(\mathbf{M} - i\mathbf{N})^2 = \frac{\nu_2}{2}\left(\frac{\nu_2}{2} + 1\right), \tag{A6}$$

which are well known for the finite-dimensional representations of $SL(2, C)$ (they correspond to the labeling of the finite-dimensional representations in terms of two (half-) integers $j_\alpha = \nu_\alpha/2$). This gives

$$\mathbf{M}^2 - \mathbf{N}^2 = l_0^2 + l_1^2 - 1, \quad i\mathbf{M}\mathbf{N} = l_0 l_1,$$

which coincides with (2.26).

The representation $[l_0, l_1]$ is unitary if either $l_1 = i(\sigma/2)$ (σ-real) principal series or $l_0 = 0$, $0 < l_1 \leq 1$ "complementary series."

The canonical basis $|s\zeta)$ in the space $X_{\nu_1\nu_2}$ can be written in the following form, which is equally suited for the unitary and finite-dimensional representations of $SL(2, C)$:

$$|[l_0, l_1]s\zeta) = f_{s\zeta}^{l_0 l_1}(z)$$
$$= A_s^{l_0 l_1}[(2s + 1)(s + \zeta)!\,(s - \zeta)!\,(s + l_0)!\,(s - l_0)!]^{\frac{1}{2}}$$
$$\times (z\bar{z})^{l_1 - s - 1} \sum_k \frac{z_1^{\zeta + l_0 + k} z_2^{s - \zeta - k}(-\bar{z}_1)^k (\bar{z}_2)^{s - l_0 - k}}{(\zeta + l_0 + k)!\,(s - \zeta - k)!\,k!\,(s - l_0 - k)!}, \tag{A7}$$

with

$$A_s^{l_0 l_1} = \left[\frac{\Gamma(s - l_1 + 1)\Gamma(|l_0| + l_1 + 1)}{\Gamma(s + l_1 + 1)\Gamma(|l_0| - l_1 + 1)}\right]^{\frac{1}{2}}$$
$$= \left[\frac{(s - l_1)\cdots(|l_0| + 1 - l_1)}{(s + l_1)\cdots(|l_0| + 1 + l_1)}\right]^{\frac{1}{2}},$$

where the range of summation over k is defined by the condition that all powers in the sum are non-negative:

$$\max(0, -l_0 - \zeta) \leq k \leq \min(s - l_0, s - \zeta).$$

For $l_1 = i\sigma$ (σ real), the functions (A7) are orthonormalized with respect to the scalar product

$$(f, g) = \frac{1}{\pi}\iint_{R_2} \overline{f(x + iy, 1)} g(x + iy, 1)\, dx\, dy.$$

In this basis the unitarity of the representations of both the principle and the supplementary series is manifest, whereas for the finite-dimensional representations the generators N^j are anti-Hermitian. The basis (A7), however, does not exhibit the reality of the representations $[0, l_1]$ for real $l_1 > 1$. For these representations it is convenient to use instead of the complex spinor variables z and $\bar{z}$ the real lightlike 4-vector ξ with components

$$\xi^\mu = z\sigma_\mu \bar{z}, \quad (\xi^2 \equiv (\xi^0)^2 - \boldsymbol{\xi}^2 = 0). \tag{A8}$$

The condition $\xi^2 = 0$ is a simple consequence of the identity

$$\sum_{\mu=0}^{3} g^{\mu\mu}(\sigma_\mu)_{\alpha\dot{\beta}}(\sigma_\mu)_{\gamma\dot{\delta}} = 2(\delta_{\alpha\dot{\beta}}\delta_{\gamma\dot{\delta}} - \delta_{\alpha\dot{\delta}}\delta_{\beta\dot{\gamma}}) = 2\epsilon_{\alpha\gamma}\epsilon_{\dot{\beta}\dot{\delta}}. \tag{A9}$$

In the variables ξ^μ the basis (A7) assumes the form

$$|[0, l_1]s\zeta) = A_s^{0 l_1}[(2s + 1)(s + \zeta)!\,(s - \zeta)!]^{\frac{1}{2}}$$
$$\times \xi_0^{l_1 - s - 1} \sum_{\nu=\zeta_-}^{[s - \zeta/2]} \frac{\xi_+^{\zeta + \nu}(-\xi_3)^{s - \zeta - 2\nu}(-\xi_-)^\nu}{(\zeta + \nu)!\,(s - \zeta - 2\nu)!\,\nu!}, \tag{A10}$$

[42] A complication arises for $\pm l_1 = |l_0| + n$, $n = 1, 2, \cdots$. See I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions, Integral Geometry and Representation Theory* (Academic Press Inc., New York, 1966), Vol. 5, Chap. III, Sec. 3. To avoid it we define in these cases $X_{\nu_1\nu_2}$ to be the set of homogeneous polynomials of z and $\bar{z}$ of degree of homogeneity $|l_1 + l_0| - 1$, $|l_1 - l_0| - 1$.

where $\zeta_\pm = \frac{1}{2}(|\zeta| \pm \zeta)$, $[x]$ is the integer part of x, and

$$\xi_\pm = \tfrac{1}{2}(\xi_1 \pm i\xi_2) = z\sigma_\pm\bar{z},$$

i.e.,

$$\xi_+ = z_1\bar{z}_2, \quad \xi_- = \bar{z}_1 z_2.$$

To obtain a basis in which the reality properties are explicit we have to take the absolute value of the expression under the square root. For the finite-dimensional representations (i.e., for l_1 positive integer) the two bases are connected by

$$|[0, l_1]s\zeta)_F = (-i)^s f^{0l_1}_{s\zeta}(\xi).$$

To prove the identity between (A7) and (A10), i.e., to check the equality

$$s!\sum_{k=\zeta_-}^{s-\zeta_+}\frac{z_1^{\zeta+k}z_2^{s-\zeta-k}(-\bar{z}_1)^k\bar{z}_2^{s-k}}{(\zeta+k)!\,(s-\zeta-k)!\,k!\,(s-k)!} = \sum_{\nu=\zeta_-}^{[s-\zeta/2]}\frac{(z_1\bar{z}_2)^{\zeta+\nu}(z_2\bar{z}_2 - z_1\bar{z}_1)^{s-\zeta-2\nu}(-\bar{z}_1z_2)^\nu}{(\zeta+\nu)!\,(s-\zeta-2\nu)!\,\nu!},$$

one has to use the sum rule

$$\sum_{\mu=0}^{s-\zeta-k}\frac{1}{(s-k-\mu)!\,\mu!\,(2k+\mu+\zeta-s)!} = \frac{s!}{(\zeta+k)!\,(s-\zeta-k)!\,k!\,(s-k)!}.$$

If we put

$$\frac{1}{\xi^0}\xi^3 = \cos\theta, \quad \frac{1}{\xi^0}\xi_\pm = \tfrac{1}{2}\sin\theta e^{\pm i\varphi}$$

we can rewrite (4.10) in the form

$$[0, l_1]s, \zeta) = A_s^{0l_1}\sqrt{2s+1}\left[\frac{(s-|\zeta|)!}{(s+|\zeta|)!}\right]^{\frac{1}{2}}(\xi^0)^{l_1-1}e^{i\zeta\varphi}P_s^{|\zeta|}(\cos\theta), \tag{A10'}$$

where $P_s^{|\zeta|}(\cos\theta)$ are the Legendre functions:

$$P_s^{|\zeta|}(\cos\theta) = (-1)^{s+|\zeta|}\frac{\sin^{|\zeta|}\theta}{2^s s!}\left(\frac{d}{d\cos\theta}\right)^{s+|\zeta|}(1-\cos^2\theta)^s.$$

The basis vectors for the self-coupled representation $[\frac{1}{2}, 0]$ are also related to these functions:

$$^{\frac{1}{2},0}_{s\zeta}(z) = \frac{(s+\zeta)^{\frac{1}{2}}}{s}f^{0,-\frac{1}{2}}_{s-\frac{1}{2}\zeta-\frac{1}{2}}(\xi)z_1 + \frac{(s-\zeta)^{\frac{1}{2}}}{s}f^{0,-\frac{1}{2}}_{s-\frac{1}{2}\zeta+\frac{1}{2}}(\xi)z_2. \tag{A11}$$

We parametrize an arbitrary proper Lorentz transformation Λ in the neighborhood of the identity by

$$\Lambda = \Lambda(\omega) = \exp\left\{-\frac{i}{2}M^{\mu\nu}\omega_{\mu\nu}\right\}$$

where $\omega_{\mu\nu} = -\omega_{\nu\mu}$ are real numbers (only 6 of them being independent) and

$$(M^{\mu\nu})^\kappa_\lambda = i(g^{\mu\kappa}\delta^\nu_\lambda - g^{\nu\kappa}\delta^\mu_\lambda).$$

This parametrization induces a local parametrization $A(\omega)$ of $SL(2, C)$, and, consequently, a local parametrization of each representation $V(A)$ of this group. We define (in accordance with Sec. 2) the generators of an arbitrary representation V of $SL(2, C)$ to be

$$S^{\mu\nu} = i\left[\frac{\partial V(A(\omega))}{\partial\omega_{\mu\nu}}\right]_{\omega_{\mu\nu}=0}.$$

Using (A2) we find the explicit expression for the generators of the representations under consideration

$$M^j = \frac{1}{2}\left(z\sigma_j\frac{\partial}{\partial z} - \frac{\partial}{\partial\bar{z}}\sigma_j\bar{z}\right),$$
$$N^j = \frac{i}{2}\left(z\sigma_j\frac{\partial}{\partial z} + \frac{\partial}{\partial\bar{z}}\sigma_j\bar{z}\right). \tag{A12}$$

In terms of ξ^j or ξ^0, θ, φ we have correspondingly

$$M^j = i\epsilon_{jkl}\xi^l\frac{\partial}{\partial\xi^k}, \quad N^j = i\xi^0\frac{\partial}{\partial\xi^j}, \tag{A12a}$$

$$M^3 = -i\frac{\partial}{\partial\varphi}, \quad M_\pm = e^{\pm i\varphi}\left(\pm\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\varphi}\right), \tag{A12b}$$

$$N^3 = i\left(\cos\theta\xi^0\frac{\partial}{\partial\xi^0} - \sin\theta\frac{\partial}{\partial\theta}\right),$$
$$N_\pm = e^{\pm i\varphi}\left(i\sin\theta\xi^0\frac{\partial}{\partial\xi^0} + i\cos\theta\frac{\partial}{\partial\theta} \mp \frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}\right).$$

From (A7) and (A12) we find the explicit expressions for the action of the generators on the canonical basis of each irreducible representation of $SL(2, C)$. The action of the generators M^j of the three-dimensional rotations coincides with that given by (2.22) for the case of the Majorana representations, while the generators N^j of the pure Lorentz transformations act as follows:

$$N_\pm|[l_0, l_1]s\zeta) = \pm C_s[(s\mp\zeta)(s\mp\zeta-1)]^{\frac{1}{2}}|[l_0, l_1]s-1\,\zeta\pm1) - \frac{il_0l_1}{s(s+1)}[(s\mp\zeta)(s\pm\zeta+1)]^{\frac{1}{2}}|s\,\zeta\pm1) \pm C_{s+1}[(s\pm\zeta+1)(s\pm\zeta+2)]^{\frac{1}{2}} \times|[l_0, l_1]s+1\,\zeta\pm1),$$

$$N^3|[l_0, l_1]s\zeta) = C_s[(s-\zeta)(s+\zeta)]^{\frac{1}{2}}|[l_0, l_1]s-1\,\zeta) - \frac{il_0l_1\zeta}{s(s+1)}|[l_0l_1]s\zeta) - C_{s+1}[(s+\zeta+1)(s-\zeta+1)]^{\frac{1}{2}}|[l_0, l_1]s+1\,\zeta), \tag{A13}$$

where C_s is given by

$$C_s = \frac{i}{s}\left[\frac{(s^2 - l_0^2)(s^2 - l_1^2)}{4s^2 - 1}\right]^{\frac{1}{2}}.$$

As we shall see in Appendix C, formulas (A12) for the generators are closely related to the expression for the generators in the ladder representation of $SU(2, 2)$. This relation helps us to find in the general case four different Hermitian 4-vector operators:

$$\begin{aligned} 2A^{\mu}_{1,2} &= g^{\mu\mu} z\sigma_{\mu}\bar{z} \mp \frac{\partial}{\partial\bar{z}}\sigma_{\mu}\frac{\partial}{\partial z}, \\ 2B^{\mu} &= \bar{z}\epsilon^{-1}\sigma_{\mu}\frac{\partial}{\partial z} + g^{\mu\mu} z\sigma_{\mu}\epsilon\frac{\partial}{\partial\bar{z}}, \\ 2C^{\mu} &= i\left(\bar{z}\epsilon^{-1}\sigma_{\mu}\frac{\partial}{\partial z} - g^{\mu\mu} z\sigma_{\mu}\epsilon\frac{\partial}{\partial\bar{z}}\right), \end{aligned} \tag{A14}$$

each of them satisfying (2.3):

$$i[A^{\mu}_{\alpha}, A^{\nu}_{\alpha}] = i[B^{\mu}, B^{\nu}] = i[C^{\mu}, C^{\nu}] = S^{\mu\nu}. \tag{A15}$$

One can also introduce 4-vectors with commuting (Hermitian) components. We have already used the vector $\xi^{\mu} = z\sigma_{\mu}\bar{z}$ (A8); the same transformation properties are exhibited by

$$\frac{\partial}{\partial\bar{z}}\sigma_{\mu}\frac{\partial}{\partial z}.$$

Two irreducible representations of $SL(2, C)$ are called *coupled* if by acting with some of our vector-operators Z^{μ} in the space $X_{\nu_1\nu_2}$ of one of them we obtain vectors from the second one $X_{\bar{\nu}_1\bar{\nu}_2}$, i.e.,

$$X_{\bar{\nu}_1\bar{\nu}_2} \subset Z^{\mu} X_{\nu_1\nu_2}, \tag{A16}$$

where $Z^{\mu} = A^{\mu}_{\alpha}$, B^{μ}, C^{μ}. It is easily seen that two representations $[l_0, l_1]$ and $[l_0', l_1']$ are coupled if and only if either

$$l_0 = l_0', \qquad l_1 = l_1' \pm 1, \tag{A17}$$

or

$$l_0 = l_0' \pm 1, \quad l_1 = l_1'. \tag{A18}$$

The first possibility is realized when we act in $X_{\nu_1\nu_2}$ with A^{μ}_{α}, the second when we act with B^{μ} or C^{μ}.

On the other hand, because of (A4) there are essentially only two cases in which the two coupled representations are equivalent:

$$\begin{aligned} &l_0' = l_0 = -l_0, \\ &l_1' = l_1 - 1 = -l_1 \Rightarrow [l_0, l_1] = [0, \tfrac{1}{2}], \\ &l_0' = l_0 - 1 = -l_0, \\ &l_1' = l_1 = -l_1 \Rightarrow [l_0, l_1] = [\tfrac{1}{2}, 0]. \end{aligned}$$

Thus, we rediscover the Majorana representations which, as we have seen in Sec. 2, are at the same time irreducible representations of $Sp(4, R)$.

Finally, we shall consider the space-reflection operation $V(I_s)$ in the algebra generated by z, $\bar{z}$, $\partial/\partial z$ and $\partial/\partial\bar{z}$. There exist two different possibilities of defining $V(I_s)$ consistent with the known transformation properties of the generators of the Lorentz group. The first one is

$$V_1(I_s) = \exp\left\{\frac{i\pi}{2}\left(\bar{z}\epsilon^{-1}\frac{\partial}{\partial z} + z\epsilon\frac{\partial}{\partial\bar{z}}\right)\right\}, \tag{A19}$$

with properties

$$\begin{aligned} V_1(I_s) z^{\alpha} V_1^{-1}(I_s) &= i\bar{z}_{\beta}(\epsilon^{-1})^{\beta\alpha}, \\ V_1(I_s)\bar{z}_{\alpha} V_1^{-1}(I_s) &= iz^{\beta}\epsilon_{\beta\alpha}, \\ V_1(I_s)\frac{\partial}{\partial z^{\alpha}} V_1^{-1}(I_s) &= -i\epsilon_{\alpha\beta}\frac{\partial}{\partial\bar{z}_{\beta}}, \\ V_1(I_s)\frac{\partial}{\partial\bar{z}_{\alpha}} V_1^{-1}(I_s) &= -i(\epsilon^{-1})^{\alpha\beta}\frac{\partial}{\partial z^{\beta}}. \end{aligned} \tag{A20}$$

The second is defined by

$$V_2(I_s) = \exp\left\{i\frac{\pi}{2}\left(\frac{\partial^2}{\partial z^{\alpha}\partial\bar{z}_{\alpha}} - z^{\alpha}\bar{z}_{\alpha}\right)\right\} \tag{A21}$$

and has the properties

$$\begin{aligned} &V_2(I_s) z^{\alpha} V_2^{-1}(I_s) = i\frac{\partial}{\partial\bar{z}_{\alpha}}, \qquad V_2(I_s)\bar{z}_{\alpha} V_2^{-1}(I_s) = i\frac{\partial}{\partial z^{\alpha}}, \\ &V_2(I_s)\frac{\partial}{\partial z^{\alpha}} V_2^{-1}(I_s) = i\bar{z}_{\alpha}, \qquad V_2(I_s)\frac{\partial}{\partial\bar{z}_{\alpha}} V_2^{-1}(I_s) = iz_{\alpha}. \end{aligned} \tag{A22}$$

Both definitions lead to the right transformation law of the generators M^j and N^j:

$$V_{1,2} M^j V_{1,2}^{-1} = M^j, \quad V_{1,2} N^j V_{1,2}^{-1} = -N^j. \tag{A23}$$

The fundamental properties (A20) and (A22) are deduced from (A19) and (A21), respectively, by using the operator-calculus formula:

$$e^{-A} B e^{A} = \sum_{n=0}^{\infty}\frac{1}{n!} B_n,$$

with

$$B_0 = B_n, \quad B_{n+1} = [B_n, A]. \tag{A24}$$

Using (A20) and (A22) one easily checks that

$$V_{1,2}^4(I_s) = 1. \tag{A25}$$

The transformation properties of the 4-vectors (A14)

under the two possible space reflections are given by

$$V_1 A^{\mu}_{1,2} V_1^{-1} = g^{\mu\mu} A^{\mu}_{1,2}, \quad V_1 B^{\mu} V_1^{-1} = g^{\mu\mu} B^{\mu}, \quad \text{(A26)}$$

$$V_1 C^{\mu} V_1^{-1} = -g^{\mu\mu} C^{\mu},$$

$$V_2 A^{\mu}_{1,2} V_2^{-1} = \pm g^{\mu\mu} A^{\mu}_{1,2},$$

$$V_2 B^{\mu} V_2^{-1} = g^{\mu\mu} B^{\mu}, \quad V_2 C^{\mu} V_2^{-1} = g^{\mu\mu} C^{\mu}. \quad \text{(A27)}$$

So we see that for each of these reflections three of the quantities A_α, B, and C are vectors and one is an axial vector.

APPENDIX B: "SCHRÖDINGER PICTURE" FOR THE SELF-COUPLED REPRESENTATIONS

Instead of ξ_α used in (2.19) we introduce here one complex variable z setting

$$a_1 = \frac{1}{\sqrt{2}}\left(\bar{z} + \frac{\partial}{\partial z}\right), \quad a_2 = \frac{1}{\sqrt{2}}\left(z + \frac{\partial}{\partial \bar{z}}\right),$$

$$a_1^* = \frac{1}{\sqrt{2}}\left(z - \frac{\partial}{\partial \bar{z}}\right), \quad a_2^* = \frac{1}{\sqrt{2}}\left(\bar{z} - \frac{\partial}{\partial z}\right), \quad \text{(B1)}$$

or, conversely,

$$z = \frac{1}{\sqrt{2}}(a_1^* + a_2), \quad \frac{\partial}{\partial z} = \frac{1}{\sqrt{2}}(a_1 - a_2^*),$$

$$\bar{z} = \frac{1}{\sqrt{2}}(a_1 + a_2^*), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{\sqrt{2}}(a_2 - a_1^*). \quad \text{(B2)}$$

In order to be consistent with the requirement that a_α^* is the Hermitian conjugate of a_α we need to assume that

$$z^* = \bar{z}, \quad \left(\frac{\partial}{\partial z}\right)^* = -\frac{\partial}{\partial \bar{z}} \quad \text{(B3)}$$

This means that the scalar product in the space X of functions of z (and $\bar{z}$) in which the Majorana representations act is given (within the accuracy of a positive constant factor) by

$$(f, g) = \frac{1}{\pi}\int \overline{f(z, \bar{z})} g(z, \bar{z}) \left(\frac{i\, dz\, d\bar{z}}{2}\right). \quad \text{(B4)}$$

(The real and imaginary parts of z play the role of the coordinates in the Schrödinger picture, $i/2\, dz\, d\bar{z} = d\,\mathrm{Re}\, z\, d\,\mathrm{Im}\, z = d^2z$.)

Canonical basis: The canonical basis can be defined as the set of normalized eigenvectors of L^0 and M^3 [see (2.22) and (2.23)]. These generators assume the following form in terms of z:

$$2L^0 = \bar{z}z - \frac{\partial^2}{\partial z \partial \bar{z}}, \quad M^3 = \frac{1}{2}\left(z\frac{\partial}{\partial z} - \bar{z}\frac{\partial}{\partial \bar{z}}\right). \quad \text{(B5)}$$

With the change of variables

$$z = \left(\frac{\rho}{2}\right)^{\frac{1}{2}} e^{i(\varphi/2)}, \quad \bar{z} = \left(\frac{\rho}{2}\right)^{\frac{1}{2}} e^{-i(\varphi/2)} \quad (\rho > 0), \quad \text{(B6)}$$

we reduce the eigenvalue problem to the set of equations

$$-i\frac{\partial}{\partial \varphi} u_{s\zeta}(\rho, \varphi) = \zeta u_{s\zeta}(\rho, \varphi),$$

$$\left[\frac{\partial}{\partial \rho}\left(\rho \frac{\partial}{\partial \rho}\right) + \frac{1}{\rho}\frac{\partial^2}{\partial \varphi^2} - \tfrac{1}{4}\rho + s + \tfrac{1}{2}\right] u_{s\zeta}(\rho, \varphi) = 0. \quad \text{(B7)}$$

The normalized solution of (B7) is given by

$$|s\zeta) \equiv u_{s\zeta}(\rho, \varphi) = \left[\frac{(s - |\zeta|)!}{(s + |\zeta|)!}\right]^{\frac{1}{2}} e^{i\zeta\varphi} e^{-(\rho/2)} \rho^{|\zeta|} L^{2|\zeta|}_{s-|\zeta|}(\rho), \quad \text{(B8)}$$

where $L_n^\alpha(\rho)$ are the Laguerre polynomials

$$L_n^\alpha(\rho) = \sum_{k=0}^{n} \binom{n + \alpha}{n - k} \frac{(-\rho)^k}{k!}, \quad \text{(B9)}$$

with normalization

$$\int_0^\infty L_n^\alpha(\rho) L_m^\alpha(\rho) \rho^\alpha e^{-\rho}\, d\rho = \frac{(\alpha + n)!}{n!}\, \delta_{mn}. \quad \text{(B10)}$$

In the original variables z and $\bar{z}$ we have

$$|s\zeta) = \left[\frac{(s - |\zeta|)!}{(s + |\zeta|)!}\right]^{\frac{1}{2}} 2^{|\zeta|} e^{-z\bar{z}} z^{|\zeta| + \zeta} \bar{z}^{|\zeta| - \zeta} L^{2|\zeta|}_{s-|\zeta|}(2z\bar{z}). \quad \text{(B11)}$$

Complete set of solutions of the Majorana equation for $\kappa > 0$: Putting

$$u_{\beta\zeta}\left(0, 0, 2\frac{\kappa}{\lambda}; z\right) = R_{\beta|\zeta|}(\rho) e^{i\zeta\varphi} \quad \text{(B12)}$$

and taking into account the equality

$$2L^3 = -z\bar{z} - \frac{\partial^2}{\partial z \partial \bar{z}} = -\frac{\rho}{2} - 2\left[\frac{\partial}{\partial \rho}\left(\rho \frac{\partial}{\partial \rho}\right) + \frac{1}{\rho}\frac{\partial^2}{\partial \varphi^2}\right], \quad \text{(B13)}$$

we obtain the following "radial" Majorana equation:

$$\left\{(1 - \beta)\left[\frac{d}{d\rho}\left(\rho \frac{d}{d\rho}\right) - \frac{\zeta^2}{\rho}\right] + \frac{1 + \beta}{4}\rho - \frac{\lambda}{2}\right\} R_{\beta|\zeta|}(\rho) = 0. \quad \text{(B14)}$$

The solution of this equation is given by

$$R_{\beta|\zeta|}(\rho) = C_{\lambda\beta|\zeta|} \Phi\left(\tfrac{1}{2} + |\zeta| - \frac{\lambda}{2(\beta^2 - 1)^{\frac{1}{2}}},\right.$$

$$\left. 1 + 2|\zeta|, \left(\frac{\beta + 1}{\beta - 1}\right)^{\frac{1}{2}} \rho\right) \rho^{|\zeta|} e^{-\frac{1}{2}(\beta + 1/\beta - 1)^{\frac{1}{2}}\rho}. \quad \text{(B15)}$$

We again find the eigenvectors of the discrete and the continuous spectra by substituting for β (4.21) and (4.28), respectively.

Solutions for the case $\kappa = 0$: For $\kappa = 0$ Eqs. (4.8) and (4.9) give

$$(2s + 1)^2 p^2 u(\mathbf{p}) = 0, \tag{B16}$$

which implies that either $s = -\frac{1}{2}$ or $p^2 = 0$. No discrete spectrum appears in either of these cases.

In the first one ($s = -\frac{1}{2}$) we again put $\mathbf{p} = (0, 0, p^3)$ and

$$\frac{p^0}{p^3} = \frac{\chi^2 - 1}{1 + \chi^2}. \tag{B17}$$

Substituting

$$u_{\chi\zeta}(0, 0, p^3; z) = R_{\chi|\zeta|}(p) e^{i\zeta\varphi} \tag{B18}$$

in the Majorana equation we find

$$\left[\frac{d}{d\rho}\left(\rho \frac{d}{d\rho}\right) - \frac{\zeta^2}{\rho} + \left(\frac{\chi}{2}\right)^2 \rho\right] R_{|\zeta|}(\rho) = 0. \tag{B19}$$

The solution of (B19) is

$$R_{|\zeta|}(\rho) = \frac{C_{\chi|\zeta|} \exp\{\frac{1}{2} i\chi\rho\}}{\Gamma(|\zeta| + 1)} \left(\frac{-\chi\rho}{4}\right)^{|\zeta|} \Phi(|\zeta| + \tfrac{1}{2}, 2|\zeta| + 1, -i\chi\rho)$$
$$= C_{\chi|\zeta|} J_{|\zeta|}\{-\tfrac{1}{2}\chi\rho\}. \tag{B20}$$

For $p^2 = (p^0)^2 - (p^3)^2 = 0$ the little group is the Euclidean group in two dimensions generated by

$$M_3 = \frac{1}{2}\left(z \frac{\partial}{\partial z} - \bar{z} \frac{\partial}{\partial \bar{z}}\right),$$
$$M_1 + N_2 = \tfrac{1}{2}(z^2 + \bar{z}^2), \quad M_2 - N_1 = \tfrac{1}{2} i(\bar{z}^2 - z^2). \tag{B21}$$

Its Casimir operator is again expressed by the square of the operator in the left-hand side of the Majorana equation

$$(M_1 + N_2)^2 + (M_2 - N_1)^2 = (z\bar{z})^2 = \tfrac{1}{4}\rho^2 = \tfrac{1}{4}(L^0 - L^3)^2. \tag{B22}$$

The Majorana equation itself reduces to

$$\rho u_\zeta(\rho, \varphi) = 0, \tag{B23}$$

its solution being

$$u_\zeta(\rho, \varphi) = \theta^{i\zeta\varphi}\delta(\rho). \tag{B24}$$

It can again be obtained as a limiting case (for $\chi \to 0$ or ∞) from the spacelike solutions and need not be taken into account separately in the completeness relation

$$\sum_\zeta \int N_\chi u^{l\eta}_{\chi\zeta}(\mathbf{p})(\bar{u}_{\chi\zeta}(\mathbf{p}) L^0)_{l'\eta'} \, d\chi = \delta^l_{l'} \delta^\eta_{\eta'}. \tag{B25}$$

(Here ζ varies over the range of all integers when we deal with the representation $[0, \frac{1}{2}]$ and it takes all possible half-odd-integer values for the representation $[\frac{1}{2}, 0]$, while N_χ is a normalization factor.)

APPENDIX C: THE LADDER REPRESENTATION OF $U(2, 2)$ AND REDUCIBLE INFINITE-COMPONENT FIELDS

It seems physically interesting to consider also fields transforming under an infinite-dimensional representation of a larger group containing $SL(2, C)$ as a subgroup. The "conformal" group $U(2, 2)$ generated by the set of all products of γ matrices satisfying

$$\gamma_R^* \gamma^0 = \gamma^0 \gamma_R \tag{C1}$$

gives a simple example of such a higher group. We first recall the definition and the main properties of the ladder representation of $U(2, 2)$ which is closely connected with the description of the representations of the Lorentz group given in Appendix A as well as with that of Sec. 2B (for the case of the self-coupled Majorana representations).

We start with the special realization (2.9) of the γ-matrices and introduce the two Dirac-conjugate 4-component operator-valued spinors

$$\varphi = \begin{pmatrix} a_1 \\ a_2 \\ b_1^* \\ b_2^* \end{pmatrix}, \quad \tilde{\varphi} = \varphi^* \gamma^0 = \begin{pmatrix} a_1^* \\ a_2^* \\ -b_1 \\ -b_2 \end{pmatrix}, \tag{C2}$$

where $a^{(*)}$, $b^{(*)}$ satisfy the commutation rules (2.11) for Bose creation and annihilation operators. As far as $[a, b^*] = 0$ we have, instead of (2.13),

$$[\varphi^A, \tilde{\varphi}_B] = \delta^A_B, \quad [\varphi^A, \varphi^B] = [\tilde{\varphi}_A, \tilde{\varphi}_B] = 0. \tag{C3}$$

Using (C3) it is easily verified that

$$\Gamma_R = \tilde{\varphi} \gamma_R \varphi \tag{C4}$$

satisfy the same commutation rules as γ_R and, because of (C1), form a Hermitian representation of the Lie algebra of $U(2, 2)$.

This is the ladder representation of $U(2, 2)$ which contains a denumerable infinity of irreducible representations all labeled by the values of the first-order Casimir operator

$$C_1 = \tilde{\varphi}\varphi = a^* a - b^* b - 2. \tag{C5}$$

This is the most degenerate discrete series of irreducible representations of $U(2, 2)$ (see Ref. 43).

Besides the well-known Fock realization of this representation, one can also introduce the "Schrödinger picture" in analogy with the case of the Majorana representations of $Sp(4, R)$ treated in Appendix

[43] I. T. Todorov ICTP Trieste Preprint IC/66/71, 1966.

B. For this purpose we put [in analogy with (B1)][44]:

$$a_\alpha = \frac{1}{\sqrt{2}}\left(\bar{z}_\alpha + \frac{\partial}{\partial z^\alpha}\right), \quad b^\alpha = \frac{1}{\sqrt{2}}\left(z^\alpha + \frac{\partial}{\partial \bar{z}_\alpha}\right),$$
$$a^{*\alpha} = \frac{1}{\sqrt{2}}\left(z^\alpha - \frac{\partial}{\partial \bar{z}_\alpha}\right), \quad b^*_\alpha = \frac{1}{\sqrt{2}}\left(\bar{z}_\alpha - \frac{\partial}{\partial z^\alpha}\right). \quad \text{(C6)}$$

The substitutions (C6) are naturally obtained if one goes to a basis with γ^5 diagonal and puts in it $\varphi_{1,2} = \partial/\partial z$,[1,2] $\varphi_{3,4} = \bar{z}_{3,4}$ and uses (B3). The z variables are suited to the study of the decomposition of the ladder representation of $U(2, 2)$ with respect to $SL(2, C)$. In terms of these variables the Lorentz generators determined from (C4) and (C6) are identical with (A12). In the z variables the first-order Casimir operator is given by

$$C_1 = z^\alpha \frac{\partial}{\partial z^\alpha} - \bar{z}_\alpha \frac{\partial}{\partial \bar{z}_\alpha} - 2, \quad \text{(C7)}$$

while Γ^μ coincide with $2A_I^\mu$ of (A13):

$$\Gamma^\mu = g^{\mu\mu} z \sigma_\mu \bar{z} - \frac{\partial}{\partial \bar{z}} \sigma_\mu \frac{\partial}{\partial z}. \quad \text{(C8)}$$

The $U(2) \times U(2)$ singlet which is contained in the $[C_1 = -2]$ representation is given by

$$|0) = (2/\pi) e^{-z\bar{z}}. \quad \text{(C9)}$$

This vector is normalized with respect to the scalar product

$$(f, g) = \iint \overline{f(z_1, z_2)} g(z_1, z_2) d^2 z_1 d^2 z_2. \quad \text{(C10)}$$

The canonical basis in the ladder representation is labeled with four numbers (including the Casimir operator C_1):

$$|C_1 n s \zeta) \quad \text{(C11)}$$

defined as eigenvalues of a complete set of commuting operators of the maximal compact subgroup $U(2) \times U(2)$:

$$(\Gamma^0 - n)\,|C_1 n s \zeta\rangle = 0,$$
$$[\mathbf{M}^2 - s(s+1)]\,|C_1 n s \zeta\rangle = 0, \quad (M^3 - \zeta)\,|C_1 n s \zeta\rangle = 0,$$
$$n = |C_1 + 2| + 2, \quad |C_1 + 2| + 4, \cdots; \quad \text{(C12)}$$
$$s = \tfrac{1}{2}n - 1, \quad \tfrac{1}{2}n - 2, \cdots \geq 0;$$
$$\zeta = -s, \quad -s + 1, \cdots, s.$$

[44] Analogous formulas, using a higher number of complex variables, have been introduced in R. L. Anderson, J. Fisher, and R. Raczka, ICTP Trieste Preprint IC/66/102, 1966.

A field transforming under an irreducible ladder representation of $U(2, 2)$ can be decomposed in terms of ordinary finite-dimensional fields (cf. Ref. 14) which depend not only on the spin s, but also on the additional quantum number n, reflecting a (possibly broken) dynamical symmetry. In the case when $C_1 = -2$ (i.e., for $a^*a - b^*b = 0$) we have a representation with the $SO(4)$ content of the possible states of the nonrelativistic hydrogen atom. $n/2$ is in this case the principal quantum number which determines the energy levels.

Finally, we shall evaluate the "boost" $V(B_\alpha)$ of the $U(2) \times U(2)$-invariant vector $|0) = |-2, 0, 0, 0)$ and show that it has a pole-type singularity when analytically continued in α at the same point as the matrix elements of the Majorana representation evaluated in Sec. 3D.

Let $p^2 = m^2 > 0$, $p^0 > 0$, and

$$N_p = \mathbf{N}\mathbf{n}_p, \quad \mathbf{n}_p = \frac{\mathbf{p}}{|\mathbf{p}|}, \quad \tanh \alpha_p = \frac{|\mathbf{p}|}{p^0}; \quad \text{(C13)}$$

then

$$V(B_p)\,|0\rangle = e^{-iN_p\alpha_p}\,|0)$$
$$= \exp\left\{\tanh\frac{\alpha}{2}\, a^*\boldsymbol{\sigma} b^* \mathbf{n}_p\right\}|0)$$
$$= \frac{2m}{p^0 + m} \exp\left\{\frac{\mathbf{p} a^* \boldsymbol{\sigma} b^*}{p^0 + m}\right\}|0). \quad \text{(C14)}$$

In terms of the variables (C6) the same result can be expressed as follows:

$$V(B_p)\,|0) = e^{iN_p\alpha_p} e^{-z\bar{z}}$$
$$= \exp\{-(\cosh\alpha_p z\bar{z} + \sinh\alpha_p z\boldsymbol{\sigma}\bar{z}\mathbf{n}_p)\}$$
$$= \exp\left\{-\frac{1}{m} p^\mu z \sigma_\mu \bar{z}\right\}. \quad \text{(C15)}$$

It is easily seen [especially if one uses (C14)] that

$$(0|\,V(B_p)\,|0) = \frac{1}{\cosh^2 \tfrac{1}{2}\alpha_p} \quad \text{(C16)}$$

which again has a pole for $\alpha_p = i\pi(2n + 1)$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Analyticity of the Envelope Diagrams*

CLIFFORD RISK†
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received 9 November 1967)

Analyticity of the envelope and open envelope diagrams is investigated, and solutions to the Landau equations are obtained which do not satisfy the usual symmetry property. The Landau curves which these solutions generate are found to have the following characteristics when the anomalous thresholds are on the unphysical sheet: (1) they are also on the unphysical sheet; (2) they have asymptotes that depend on the external masses of the diagrams; (3) the Landau curve for the envelope diagram has as an asymptote a line u = const, although the diagram has no u channel; (4) a new type of effective intersection is found.

I. INTRODUCTION

In investigating the analyticity of the envelope[1,2] and open envelope[2–5] diagrams, previous papers have been restricted to solutions of the Landau equations that satisfy the following symmetry property: Let q_i be the internal momenta of the diagram and α_i the Feynman parameters. If a combination of rotations and reflections carries the diagram into itself with $q_i \to q_{i'}$, then the solutions to the Landau equations giving physical singularities obey the relations

$$\alpha_i = \alpha_{i'}, \quad q_i \cdot q_j = \pm q_{i'} \cdot q_{j'}, \quad q_i \cdot q_{j'} = \pm q_{i'} \cdot q_j,$$

the signs of the scalar products being determined by the direction of the vectors in the diagram. The validity of this symmetry property was proven[6] under the assumption that Landau curves lying on the physical sheet are associated with positive Feynman parameters. However, Eden[1] has shown that isolated acnodes on the physical sheet are given by complex parameters, while the work of Cunningham[7] suggests that real Landau curves with complex Feynman parameters might occur quite generally.

In view of these results, the aforementioned proof cannot be regarded as valid, and the solutions to the Landau equations that do not satisfy the symmetry property must be investigated individually to determine whether or not they can give rise to physical singularities.

* Supported in part through funds provided by the National Science Foundation and the Atomic Energy Commission.

† Present address: Department of Physics, University of Michigan, Ann Arbor, Michigan.

[1] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. **2**, 656 (1961).
[2] V. A. Kolkunov, L. B. Okun, and A. P. Rudik, Zh. Eksp. Teor. Fiz. **38**, 877 (1960) [Sov. Phys.—JETP **11**, 634 (1960)]; V. A. Kolkunov, Zh. Eksp. Teor. Fiz. **40**, 678 (1961) [Sov. Phys.—JETP **13**, 474 (1961)]; D. Ya. Petrina, Zh. Eksp. Teor. Fiz. **46**, 544 (1964) [Sov. Phys.—JETP **19**, 370 (1964)].
[3] J. N. Islam, P. V. Landshoff, and J. C. Taylor, Phys. Rev. **130**, 2560 (1963).
[4] D. I. Olive and J. C. Taylor, Nuovo Cimento **24**, 814 (1962).
[5] J. N. Islam, J. Math. Phys. **4**, 872 (1963).
[6] A. Z. Patashinshii, Zh. Eksp. Teor. Fiz. **39**, 1744 (1960) [Sov. Phys.—JETP **12**, 1217 (1960)].
[7] J. Cunningham, Rev. Mod. Phys. **36**, 833 (1964).

II. THE ENVELOPE DIAGRAM

In the notation of Fig. 1,

$$s = 3 + 2(y_{23} + y_{34} + y_{24}), \tag{1a}$$
$$t = 3 + 2(y_{13} + y_{35} + y_{,5}), \tag{1b}$$

where $y_{ij} = q_i \cdot q_j$. Conservation of momentum at the vertices gives

$$y_{25} = y_{14} = 1 - m^2/2, \tag{2a}$$
$$y_{13} + y_{23} + y_{12} + (3 - M^2)/2 = 0, \tag{2b}$$
$$y_{34} + y_{35} + y_{45} + (3 - M^2)/2 = 0. \tag{2c}$$

y_{34} will be taken as the independent parameter; the other y_{ij}'s will be solved for in terms of y_{34}, and this will give the Landau curve by the parametric equations (1a, b).

The Landau equations are

$$\alpha_2 q_2 + \alpha_5 q_5 - \alpha_3 q_3 = 0, \tag{3a}$$
$$\alpha_4 q_4 + \alpha_1 q_1 - \alpha_3 q_3 = 0. \tag{3b}$$

Dotting (3a) with q_2, q_5, and q_3 gives

$$y_{35} = y_{23}y_{25} \pm [1 - (y_{23})^2]^{\frac{1}{2}}[1 - (y_{25})^2]^{\frac{1}{2}}. \tag{4a}$$

Choose new variables

$$\begin{aligned} y_{34} &= \cos(\phi + \theta/2), \quad y_{35} = \cos\beta, \\ y_{23} &= \cos(\alpha + \theta/2), \quad y_{13} = \cos\phi, \\ y_{14} &= y_{25} = \cos\theta. \end{aligned} \tag{5}$$

In terms of them, (4a) becomes

$$\beta = \theta \pm (\tfrac{1}{2}\theta + \alpha). \tag{4a'}$$

Similarly, dotting (3b) with q_4, q_1, and q_3 gives

$$\phi = \theta \pm (\tfrac{1}{2}\theta + \phi). \tag{4b}$$

In Eqs. (4a'), (4b) the minus sign will be picked; the other choices of signs give Landau curves generally on the unphysical sheet. Equations (2) and (4'), then,

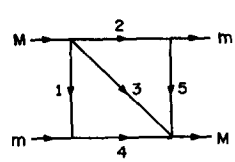


FIG. 1. Envelope diagram. Internal masses are unity, external masses are m and M.

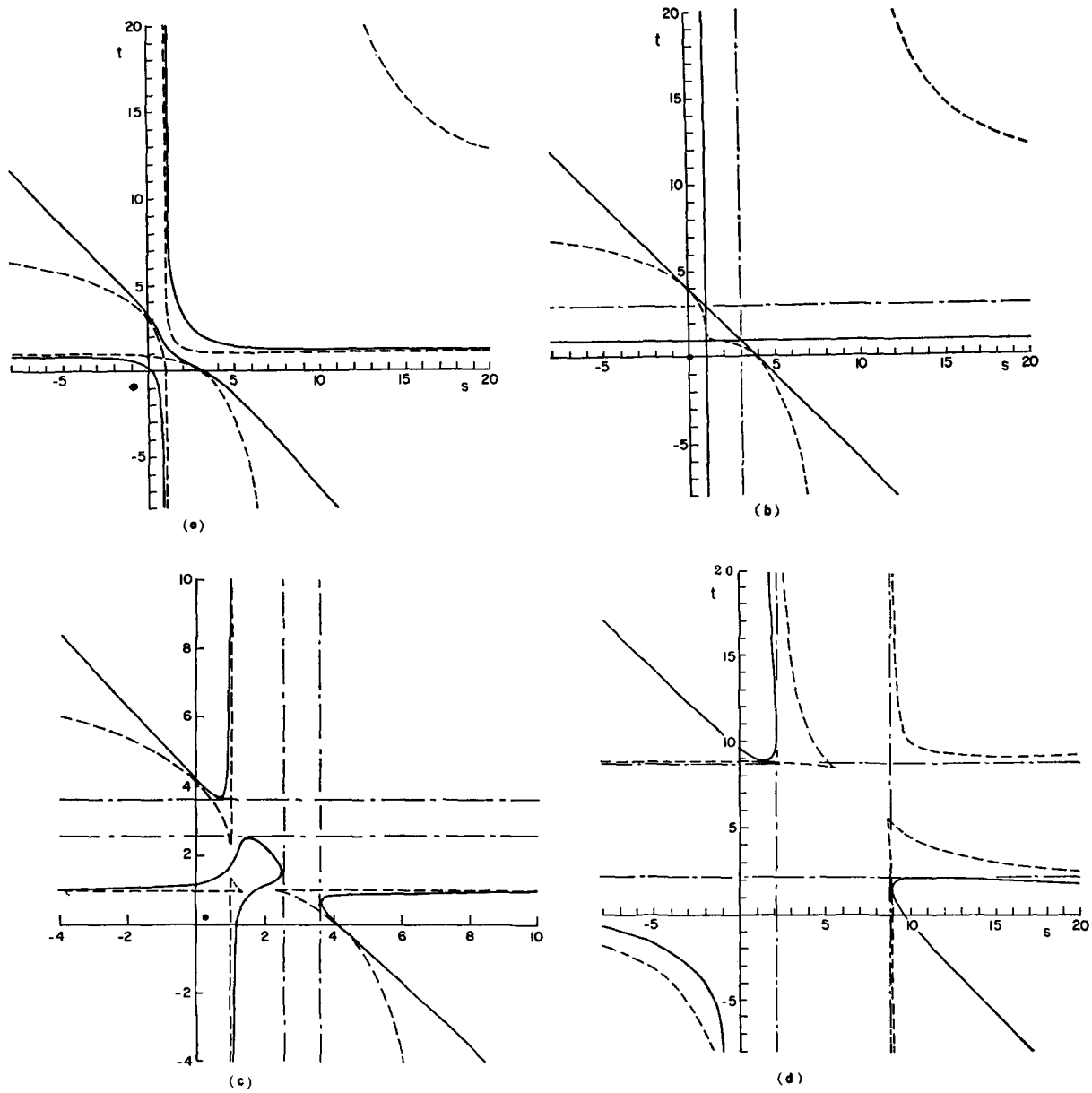


FIG. 2. Landau curves of the envelope diagram for $m = 1$. Dashed lines are I, solid lines are II, and dot-dash lines are anomalous thresholds (when real): (a) $M^2 = 0.5$; (b) $M^2 = 1.0$; (c) $M^2 = 1.05$; (d) $M^2 = 6.0$.

together give s and t in terms of ϕ and α:

$$s = s(\phi, \alpha) = 1 + [\sin(\tfrac{1}{2}\theta - \phi)]^{-1} \times \{4 \sin \tfrac{1}{2}\theta(\cos \tfrac{1}{2}\theta + \cos \phi)(1 + \cos(\tfrac{1}{2}\theta + \alpha)) + (1 - M^2) \sin(\tfrac{1}{2}\theta + \phi)\}, \quad (6a)$$

$$t = s(-\phi, -\alpha). \quad (6b)$$

It remains to determine α in terms of ϕ. This is done by dotting (3a) with q_2, q_5, q_1 to obtain

$$y_{15} = [\sin(\tfrac{1}{2}\theta + \alpha)]^{-1}\{y_{13} \sin\theta - y_{12} \sin(\tfrac{1}{2}\theta - \alpha)\}, \quad (7a)$$

and dotting (3b) with q_4, q_5, and q_1 to obtain

$$y_{15} = [\sin(\tfrac{1}{2}\theta + \phi)]^{-1}\{y_{35} \sin\theta - y_{45} \sin(\tfrac{1}{2}\theta - \phi)\}. \quad (7b)$$

Equating (7a) to (7b) then gives the relation between α and ϕ, called the tautening equation:

$$0 = 2(\cos \tfrac{1}{2}\theta + \cos\phi)(\cos \tfrac{1}{2}\theta + \cos\alpha)(\sin\alpha - \sin\phi) + (1 - M^2) \cos \tfrac{1}{2}\theta \sin(\alpha - \phi). \quad (8)$$

Equation (8) is a fourth-order equation for α. One of its solutions is $\alpha(\phi) \equiv \phi$, which was studied[8] in Refs. 1 and 2. Typical Landau curves[9] are drawn in Fig. 2.

[8] In addition to the nodes on the physical sheet found in Ref. 1, there are nodes on the unphysical sheet that were not fully accounted for.

[9] Hereafter referred to as the curve I.

Only the curve inside the crossed cuts lies on the physical sheet.

The remaining three solutions to (8) cannot be expressed so easily. However, after some straightforward algebra it can be shown that these three solutions, when substituted into (6a), (6b), together parametrize a single Landau curve[10] that is given by the following equation:

$$0 = [u - m^2(M^2 - 1)][(s - 1)(t - 1) + (M^2 - 1)(1 + 2m^2 - M^2)] + m^4(m^2 - 4) \times (M^2 - 1). \quad (9)$$

Equation (9) can be rewritten as

$$(s - t)^2 = (u - u_1)(u - u_2)\left[\frac{u - 4(M^2 - 1)}{u - m^2(M^2 - 1)}\right], \quad (10)$$

where

$$u_{1,2} = \tfrac{1}{2}m^2[M^2 + 3 \pm (M^2 - 1)^{\frac{1}{2}}(M^2 - 9)^{\frac{1}{2}}].$$

For $1 < M^2 < 9$, $u_{1,2}$ are complex, and the curve (10) must have the general form exemplified in Figs. 2(c), 2(d). For $0 < M^2 < 1$, $u_{1,2}$ are real with $4(M^2 - 1) < m^2(M^2 - 1) < u_2 < u_1$, and the curve (10) has the form of Fig. 2(a). At $M^2 = 1$, the Landau curve degenerates into the three straight lines $s = 1$, $t = 1$, and $u = 0$ [Fig. 2(b)].

From Eq. (10) and Fig. 2, the Landau curve can be seen to have the following properties:

(1) It is a cubic curve in the s–t plane, and for $M^2 \neq 1$ it has no nodes or cusps in the finite plane.

(2) It is on the unphysical sheet. This can be seen by examining the Feynman discriminant:

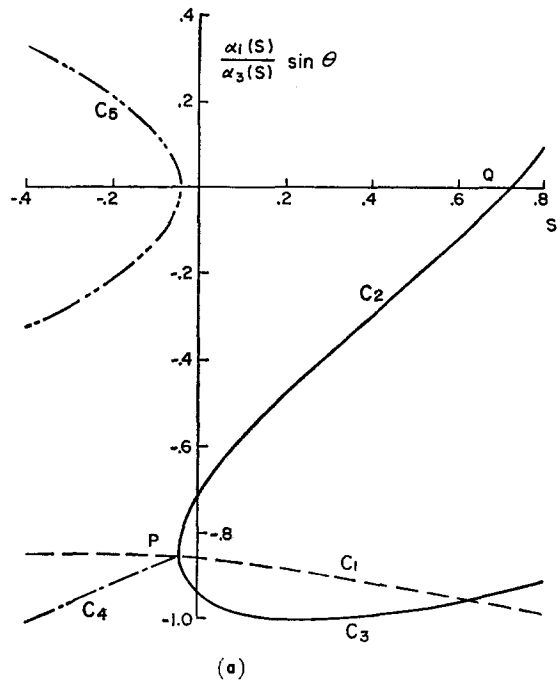
$$D(\alpha, s, t) = \alpha_2\alpha_3\alpha_4 s + \alpha_1\alpha_3\alpha_5 t + \alpha_3(\alpha_1\alpha_2 + \alpha_4\alpha_5)M^2 + [\alpha_1\alpha_4(\alpha_2 + \alpha_3 + \alpha_5) + \alpha_2\alpha_3(\alpha_1 + \alpha_3 + \alpha_5)]m^2 - [(\alpha_1 + \alpha_4)(\alpha_2 + \alpha_5) + \alpha_3(\alpha_1 + \alpha_4 + \alpha_2 + \alpha_3)]\sum_{i=1}^{5}\alpha_i. \quad (11)$$

$D(\alpha, s, t)$ is negative for positive α when $s, t < 3$, $M^2 < 3$, $m^2 < 2$; therefore, the scattering amplitude cannot be singular there. But in Fig. 2(a), each of the three branches of the Landau curve has some segment in the region $s, t < 3$. Since none of the three branches has any tangency that could bring it onto the physical sheet, the entire curve must be on the unphysical sheet.

As m^2 and M^2 vary, no developments occur which could bring any branch on the unphysical sheet, and hence they remain always nonsingular.

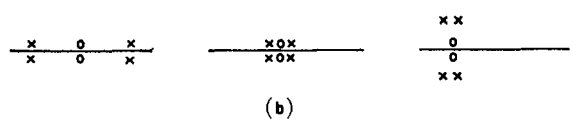
(3) It has the line $u = m^2(M^2 - 1)$ as an asymptote, even though the diagram itself has no u channel.

[10] Hereafter referred to as the curve II.



(a)

FIG. 3a. Feynman parameters for the curves I and II near their effective intersection for $m^2 = 1.0$, $M^2 = 1.05$.



(b)

FIG. 3b. Model for dissolution of a pinching singularity. Pinching α_2 values are shown by dots for I and crosses for II.

Further, the value of the asymptote depends on both the internal and external masses.

(4) It has points of tangency with the Landau curve I. These tangencies are effective intersections, but they have a new type of tangency, hitherto unobserved, in which the point of tangency P divides the Feynman parameters into real and complex values. This is illustrated in Fig. 3(a), where the quantity $r_i(s) = \sin\theta\alpha_i(s)/\alpha_3(s)$ has been plotted. For the curve I, C_1 represents the values $r_2(s)$, which equals $r_4(s)$. For the curve II, C_2 represents $r_2(s)$ for $s > s_P$, and C_3 gives $r_4(s)$ for $s > s_P$; both are real. For $s < s_P$, C_4, and C_5 give the real and imaginary parts of $r_2(s)$, $r_4(s)$; they are complex conjugates.[11]

(5) The slope of the Landau curve is given by

$$\frac{ds}{dt} = -\frac{\alpha_1\alpha_5}{\alpha_2\alpha_4} = -\frac{\sin(\tfrac{1}{2}\theta + \phi)\sin(\tfrac{1}{2}\theta + \alpha)}{\sin(\tfrac{1}{2}\theta - \phi)\sin(\tfrac{1}{2}\theta - \alpha)}. \quad (12)$$

[11] As it turns out, the curve II is doubly parametrized with respect to ϕ. For each point on the curve, one set of pinching Feynman parameters is given by Fig. 3; a second set is given by interchanging the values of α_2 and α_4. As pointed out in Ref. 7, it is this nonuniqueness of Feynman parameters that allows them to take complex values.

Hence, the Landau curve has a point of tangency P with an s threshold ($s = s_1$) if $\alpha_1 = 0$ or $\alpha_5 = 0$ at P. But for the curve I, $\alpha_1 = \alpha_5$, and therefore the line $s = s_1$ cannot be an anomalous threshold but must be a normal threshold. It can then be shown that each anomalous threshold intersects the curve I in four finite points, and none of these intersections are tangential.[12]

For the curve II, α_2 can be zero with α_4 nonzero (the point Q in Fig. 3), and it can then be concluded that each anomalous threshold has two effective tangential intersections with the curve II.

Typical examples of I and II are drawn in Fig. 2 for $m^2 = 1$. For $M^2 < 1$ [Fig. 2(a)] the anomalous thresholds are complex and on unphysical sheets. At $M^2 = 1$ [Fig. 2(b)], they become real at $s, t = 3$. For $M^2 > 1$ [Fig. 2(c)], they separate. At $M^2 = 3$, the two lower ones slip through the unphysical normal threshold at $s, t = 1$ and onto a different unphysical sheet, where they become tangent to different branches of II [Fig. 2(d)].

Finally, it should be noted that paragraph (4) could give rise to a new mechanism for the dissolution of singularities. In Fig. 3(b), the two sets[11] of pinching α_2 values collide in such a way that the contour of integration becomes free of pinching and the singularity is dissolved.

III. THE OPEN ENVELOPE DIAGRAM

For the diagram of Fig. 4, it will be assumed that solutions to the Landau equations satisfy a "half symmetry" property—namely, that $q_i \cdot q_j$'s carried into each other under a reflection about the vertical are equal. For example, $q_3 \cdot q_6 = -q_1 \cdot q_5$ and $q_1 \cdot q_6 = -q_3 \cdot q_5$. Under this assumption, the Landau curve can be expressed in terms of the parametric equations[13]:

$$s = 4 + 4y + 4v + 2y_{24} + 2y_{56},$$
$$t = 4 + 4x + 4u + 2y_{13} - 2y_{56}, \tag{13}$$

where

$$y_{56} = \frac{1}{1 - z^2}\{y(y + uz) - x(u + yz)\},$$
$$y_{24} = \frac{1}{1 - u^2}\{v(y + uz) - w(z + uy)\},$$
$$y_{13} = \frac{1}{1 - y^2}\{x(u + yz) - z(z + uy)\}; \tag{14}$$

[12] A similar result has been shown for the wigwam diagram [J. Cunningham and M. Rafique, Nucl. Phys. **B1**, 21 (1967)].

[13] u_M denotes the Mandelstam variable; u denotes the parametric variable $u(x)$.

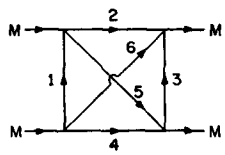


FIG. 4. Open envelope diagram. Internal masses are unity, external masses are M.

$$y = A(x) \pm B(x, u), \quad z = A(x) \mp B(x, u), \tag{15}$$
$$v = A(u) \pm B(u, x), \quad w = A(u) \mp B(u, x); \tag{16}$$
$$A(x) = \tfrac{1}{2}(b - x),$$
$$B(x, u) = \tfrac{1}{2}\left\{(1 + u)\left[2 - \frac{(b - x)^2}{1 - u}\right]\right\}^{\frac{1}{2}}, \tag{17}$$
$$b = \tfrac{1}{2}(M^2 - 3).$$

The signs in the expressions for z and w are determined by the signs in the expressions for y and v, respectively. Picking a $+$ sign for y then gives the tautening equation for u and the sign of v in terms of x:

$$\pm A(x)B(u, x)(1 - x)(1 + u)(x + u) = A(u)B(x, u)(1 - u)(1 + x)(x + u). \tag{18}$$

Eq. (18) has five different solutions for u in terms of x. One solution [$u(x) \equiv x$] is the solution studied in Refs. 2–5, and it has the "full symmetry" property that y_{ij}'s carried into each other under a horizontal or vertical reflection are equal.

Two other solutions of (18) are $u(x) = -x$ and $u(x) = -x + a + 1$. They lead to Landau curves which are the straight lines $s = 0$, $t = 0$, or $u_M = 0$.

The remaining two solutions satisfy a quadratic equation. Upon solution it leads to the Landau curves of Fig. 5. These curves have the following properties:

(1) By examining the limit $x \to 1 + \epsilon$, it can be shown that $s \to \infty$ and $t \to 16 - (5 - M^2)^2$. This asymptote is the anomalous threshold for the diagram.

Similarly, when $x \to -1 + \epsilon$, then $s \to \infty$ and $t \to -(M^2 - 1)^2$, a spurious asymptote not related to any threshold. Further, unlike the nonthreshold asymptotes of Ref. 3, the value of the asymptote depends on both the internal and external masses.

(2) Three of the Landau curves have branches lying in the same region of the s–t plane as the s reaction. Hence, by a Coleman–Norton argument,[14] they must lie on an unphysical sheet.

By examination of the Feynman discriminant,[15] it can be seen that $D(\alpha, s, t) < 0$ whenever $\alpha > 0$ and $s, t, u_M < 4$. This can be used to show that the remaining Landau curves also must lie on the unphysical sheet.

[14] S. Coleman and R. E. Norton, Nuovo Cimento **38**, 438 (1965).

[15] R. J. Eden, Phys. Rev. **119**, 1763 (1960), Eq. (8.18).

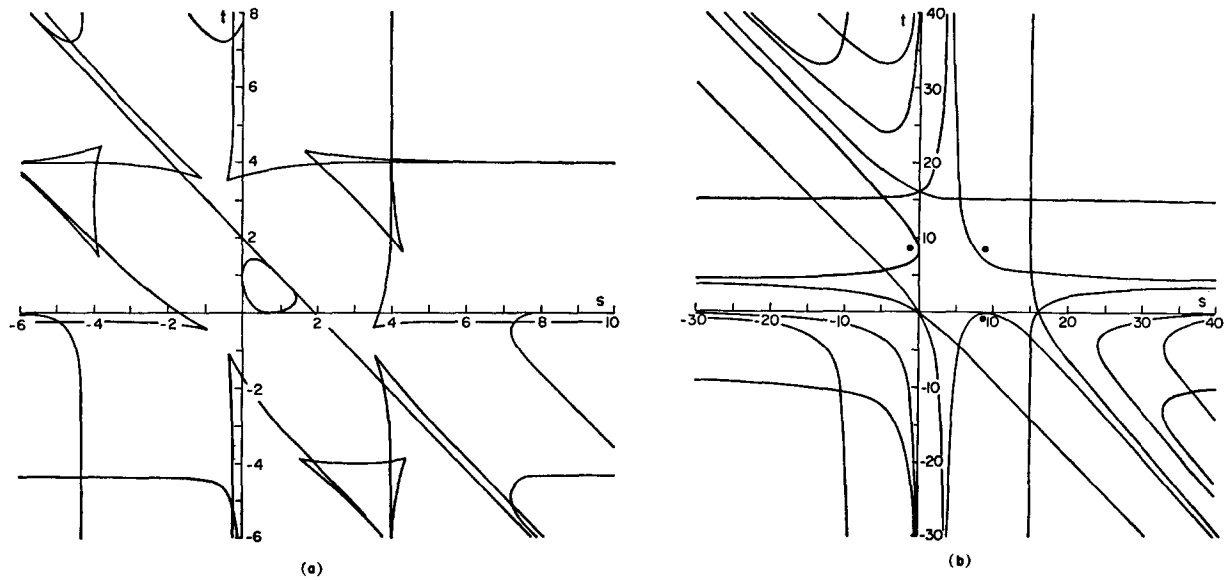


FIG. 5. Landau curves II of the envelope diagram: (a) $M^2 = 0.5$; (b) $M^2 = 4.0$.

(3) Unlike the Landau curves found in part II, the curves for this diagram have several node–cusp developments.

(4) The Landau curves found are parametrized by values of x that satisfy one of the two conditions:

(a) either x, y, z, u, v or w is real;

(b) or else all are complex, with $u = x^*$, $v = y^*$, and $w = z^*$.

By a numerical searchline technique it could be shown that the curves of Fig. 5 have no undiscovered real extensions. However, there could be entire real Landau curves, yet undiscovered, which are parametrized by complex x values other than those above.

(5) It can again be shown that the curves I and II have effective tangential intersections with nonzero Feynman parameters, and that anomalous thresholds intersect I nontangentially and intersect II both (effective) tangentially and also nontangentially.

ACKNOWLEDGMENTS

This problem was done as part of a thesis under the direction of Professor P. Federbush. I am grateful to Professor H. Cheng and Professor M. Artin for suggestions that allowed the results to be expressed in a simple and concise form. I wish to thank Professor R. J. Eden, Professor D. I. Olive, Professor G. R. Screaton, and Professor V. Teplitz for helpful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

On the Ising Model with Long-Range Interaction

A. J. F. SIEGERT
*Physics Department, Northwestern University, Evanston, Illinois**
AND
DAVID J. VEZZETTI
Physics Department, University of Illinois at Chicago Circle, Chicago, Illinois

(Received 21 June 1967)

We have obtained an expansion of the free energy per spin of an Ising model with long-range interaction in the absence of an external field, for temperatures above the Curie temperature of the Weiss–Bragg–Williams approximation (BWCP). We use as expansion parameter the reciprocal (γ) of an effective number of neighbors. Terms through order γ^2 are obtained by extracting factors from a representation of the partition function as an average over random fields. For terms of higher order, we give a diagrammatic series in which all terms through order γ^n are contained in the diagrams with not more than $2(n-1)$ bonds. The terms of order γ^3 are given explicitly. For temperatures below the BWCP we have calculated terms through order γ. Since after a few finite terms the coefficients in this γ expansion become infinite at the BWCP, we exhibit a modification of the random field representation which avoids this difficulty. We have compared our results with those of previous authors wherever available—that is, through order γ for the general case and through order γ^2 for the one-dimensional model with exponential interactions above the BWCP. The results of the modified random field representation are in formal agreement with those of previous authors. In this previous work, an equation is given, whose solution is identified as the approximate Curie point. We give arguments to show that this interpretation is not justified.

SECTION 1

The investigation of the Ising model with long-range interaction was proposed by Brout[1] in the anticipation that in the limit of infinite range, the Weiss theory of ferromagnetism is correct, and becomes a starting point for an expansion of the free energy in powers of γ, the reciprocal effective number of neighbors.[2] Originally, it was hoped that information about the phase transition would be obtained from such an expansion. However, by pointing out a discrepancy in his first-order results, Brout showed that this hope is not realized.[3] By abandoning the straightforward z^{-1} expansion in favor of a sphericalization method, Brout obtained results which do not have this discrepancy. Similar results were obtained by Horwitz and Callen.[4] Brout, as well as Horwitz and Callen, obtained his results by selective summation of diagrams. Mühlschlegel and Zittartz[5] obtained essentially the same results by the method of random fields, using a variational procedure to approximate the average over fields. These results were thought to be good approximations in the neighborhood of the true Curie point of the model, but further investigations have shown that this is not the case.[6]

In view of these difficulties, it appears useful to present the results of the straightforward γ expansion for the free energy in the absence of an external magnetic field. For temperatures above the Curie point of the Weiss–Bragg–Williams approximation (BWCP), we obtain the terms through order γ^2 by extracting factors from a variant of the Kac integral representation[7,8] of the partition function. We present a diagram series for higher-order terms in which the free energy through order γ^n is obtained from the class of diagrams with no more than $2(n-1)$ lines. We have calculated the γ^3 term explicitly. Below the BWCP, we obtain, by a slightly different method, the terms through order γ. We have included this calculation for completeness, although we cannot prove that the result includes all terms of order γ. However, the result, when specialized to the one-dimensional Kac model, agrees with the exact result obtained by Kac[8] and Baker.[9] It also agrees with the results of Refs. 1 and 5.

* Program at Northwestern University supported by the National Science Foundation and the Office of Naval Research.

[1] R. Brout, Phys. Rev. **118**, 1009 (1960).

[2] Our γ is essentially equivalent to the parameter z^{-1} used by Brout, z being the effective number of neighbors. In some of the literature γ is used to denote the reciprocal range of interaction and corresponds to our $\gamma^{1/D}$, where D is the dimensionality of the model.

[3] See Ref. 1, the discussion of Eqs. (4.2) and (4.17).

[4] G. Horwitz and H. B. Callen, Phys. Rev. **124**, 1757 (1961).

[5] B. Mühlschlegel and H. Zittartz, Z. Physik **175**, 553 (1963).

[6] M. Coopersmith and R. Brout, Phys. Rev. **130**, 2539 (1963), the last paragraph in Sec. II; R. Brout, Phys. Rev. **122**, 469 (1961), Eq. (2.1)ff.

[7] M. Kac, in *Applied Probability*, L. A. MacColl, Ed. (McGraw-Hill Book Co., New York, 1957), Vol. VIII, pp. 73–85.

[8] M. Kac, Phys. Fluids **2**, 8 (1959); and "Statistical Mechanics of Some One-dimensional Systems" in *Studies in Mathematical Analysis and Related Topics*, Gilbarg *et al.*, Eds. (Stanford University Press, Stanford, Calif., 1962).

[9] G. A. Baker, Phys. Rev. **122**, 1477 (1961).

The method of random fields[10] which we employ has the advantage of giving these results rather easily and of making obvious the reason for the breakdown of the γ expansion at the BWCP. We have removed the cause of this breakdown by employing another variant of the method of random fields. This method is similar to that of Mühlschlegel and Zittartz.[5] Our version has the advantage of showing clearly that the singularity, which under certain circumstances may occur, cannot safely be ascribed to a phase transition of the model, as was done by some of the previous authors. In the cases where these singularities occur, they occur precisely at the temperature for which the method cannot be justified. One might think that this method would give at least terms of dominant order in γ at the BWCP, since it avoids the obvious shortcomings of the straightforward γ expansion. We show, however, by comparison with the results of Kac and Helfand[11] for the two-spin correlation function of the one-dimensional Kac model, that not all terms of dominant order ($\gamma^{\frac{2}{3}}$) are given correctly, although a very good approximation is obtained for the term of largest range.

Section 2 states the problem and introduces our notation. Section 3 describes the method of random fields which converts the partition function from a sum to an integral. We describe precisely the change in the integrand which occurs at the BWCP. The physical interpretation has not yet been made rigorous, but we believe that this change in the system means physically the appearance of metastable states in the finite model.[12] It is not a phase transition, and the occurrence of metastable states in the finite model is in general neither a sufficient nor a necessary condition for a phase transition of the infinite model.[13] The maximum of the integrand occurring at temperatures above the BWCP and the largest pair of maxima occurring below the BWCP give the Weiss theory. The terms of order γ are obtained in this section by expanding the exponent in the integrand in the neighborhood of this maximum.

In Sec. 4 we change to a different integral representation, applicable only above the BWCP, and extract from the partition function factors which contribute the terms of first and second order to the free energy. The remaining factor is investigated in Sec. 5. It has the form of a grand partition function of a system of interacting particles in an external field. The relation is purely formal, but we can use this formal similarity to obtain an expansion for the terms of order γ^3 and higher. We obtain the terms of order γ^3 explicitly as an integral involving the Fourier transform of the given interaction between a pair of spins. In this section, we also obtain the above mentioned estimate of the order in γ of the terms in this expansion. Details of the proof are given in Appendix B.

In Sec. 6 we use a representation of the partition function which is valid below, as well as above, the BWCP, and obtain from it the free energy to order γ below the BWCP. In Sec. 7 another variant of the integral representation is used, which enables us to remove the singularities at the BWCP. This representation is closely related to Brout's[1] sphericalized model and to the work of Mühlschlegel and Zittartz.[5] In Sec. 8 we discuss the singularities occurring in the expansions of Sec. 7.

Finally, Sec. 9 is a summary of our results and a comparison with the work of previous authors. Appendices A and B contain details of calculations which are omitted in the text, while Appendices C and D contain the specialization of our results to the one-dimensional model with exponential interaction which we compare with the calculations of Kac[8] and Kac and Helfand,[11] based on Kac's integral equation method.

SECTION 2

We consider an Ising model of n spins located at the points of a lattice in D dimensions ($D = 1, 2, 3$). For the physical problem, we are interested in an infinite lattice. However, we perform our calculations using a lattice of finite extent and then allow the number of lattice sites to become infinite (thermodynamic limit). We give below our notation and conventions for the finite and infinite lattices. We also show how, given the interaction for the infinite lattice, one may construct a periodic interaction to be used with the finite lattice.

For the finite lattice, the number of sites is $n = m^D$, where m is an integer. Subscripts $k, l, \cdots$, denote lattice sites and are understood as D-dimensional vectors with integer-valued components. A prime is used to indicate that the term $k = l$ is excluded from the double sum $\sum'_{k,l}$. The symbols μ_k, μ_l, etc., denote the spin variables, which assume the values

[10] A short survey of the method of random fields in equilibrium statistical mechanics, in which some of the present calculations are sketched as examples, was given by one of us (A. J. F. S.) in *Analysis in Function Space*, W. T. Martin and I. Segal, Eds. (MIT Press, Cambridge, Mass., 1964), Chap. 9; also in *Statistical Physics*, Vol. III of *Brandeis Summer Institute Lectures in Theoretical Physics* (W. A. Benjamin, Inc., New York, 1963).

[11] M. Kac and E. Helfand, J. Math. Phys. **4**, 1078 (1963).

[12] The existence of metastable states for the finite Ising model with nearest neighbor interaction was shown by A. J. F. Siegert, Phys. Rev. **97**, 1456 (1955).

[13] L. van Hove (personal communication). However, the existence of metastable states, together with certain other conditions, can be shown to be a sufficient condition for condensation. See A. J. F. Siegert, Phys. Rev. **96**, 243 (1954).

± 1. The energy of a pair of spins located at sites k and l is $-J\gamma\rho_{kl}(\gamma, n)$. (The arguments γ, n will generally be omitted in the calculation.) The variable γ is the reciprocal of an effective number of neighbors[2] with $\gamma \to 0$ for infinite range, and J is a coupling parameter. Summation over the values of all the spin variables is indicated by $\sum_{\{\mu\}}$.

The partition function of this system is then given by

$$Q_n = \sum_{\{\mu\}} \exp\left(\frac{\nu\gamma}{2} \sum_{k,l}{}' \rho_{kl}\mu_k\mu_l\right), \tag{2.1}$$

where

$$\nu = J/kT. \tag{2.2}$$

The following assumptions are made about the interaction: The matrix ρ_{kl} is symmetric and depends on k and l only through the vectorial distance $k - l$, and ρ_{k-l} will be used interchangeably with ρ_{kl}. Only the pure ferromagnetic case $\rho_{kl} > 0$ is considered in the present paper, although most of the formal results are independent of this assumption.[14] Although the diagonal elements $\rho_{kk} = \rho_0$ do not appear in the physical problem, it is convenient to assign the value

$$\rho_0(\gamma, n) = 1 \tag{2.3}$$

in the calculations. Born–von Kármán boundary conditions are assumed, and, accordingly, $\rho_k(\gamma, n)$ is assumed to be a periodic function of the components of k with period $n^{1/D} = m$. The eigenvalues of ρ_{kl} are then simply the Fourier components g_τ of ρ_k. We have

$$\rho_k = \frac{1}{n} \sum_\tau g_\tau e^{2\pi i k \cdot \tau/m} \tag{2.4}$$

and

$$g_\tau = \sum_k \rho_k e^{-2\pi i k \cdot \tau/m}, \tag{2.5}$$

where τ is a vector with integer-valued components and $\sum_k$ or $\sum_\tau$ extends over the n points of the lattice.

In order to use the method of random variables in its simplest form—with real-valued random variables—we assume that the matrix ρ is positive-definite; i.e.,

$$g_\tau > 0. \tag{2.6}$$

The function $C_n(\gamma)$ is defined by

$$C_n(\gamma) = \gamma \sum_k \rho_k(\gamma, n) \tag{2.7}$$

and we assume that

$$\lim_{n\to\infty} C_n(\gamma) = C(\gamma) \tag{2.8}$$

exists. We also assume that

$$\lim_{\gamma\to 0} C(\gamma) = \lim_{\gamma\to 0} \lim_{n\to\infty} C_n(\gamma) = C \tag{2.9}$$

exists, and is different from zero, so that the second virial coefficient exists even in the limit of infinite range of interaction ($\gamma \to 0$). Note that the order of the two limits in (2.9) must not be interchanged.

The lattice of physical interest is the infinite lattice. In that case, we define an interaction matrix $\rho_k(\gamma)$. Without restricting generality, we can assume symmetry

$$\rho_k(\gamma) = \rho_{-k}(\gamma) \tag{2.10}$$

and specify

$$\rho_0(\gamma) = 1. \tag{2.11}$$

We assume purely ferromagnetic interaction

$$\rho_k(\gamma) > 0 \tag{2.12}$$

and the existence of

$$\sum_k^{(\infty)} \rho_k(\gamma) \equiv \gamma^{-1}C(\gamma); \quad \gamma > 0, \tag{2.13}$$

where the sum extends over the infinite lattice. We also assume that

$$\lim_{\gamma\to 0} C(\gamma) = C \tag{2.14}$$

exists, assuring the existence of a finite second virial coefficient for the model. We define the function $g(\omega)$, where ω is a vector in D dimensions with components in the interval $(0, 2\pi)$, by

$$g(\omega) = \sum_k^{(\infty)} e^{ik\cdot\omega}\rho_k(\gamma) \tag{2.15}$$

and restrict our calculations to the case

$$g(\omega) > 0 \tag{2.16}$$

for ω in the region specified above. (This restriction could be avoided by the use of complex-valued local fields.) The BWCP is then determined by the equation

$$\nu C(\gamma) = 1. \tag{2.17}$$

The connection between the infinite and finite lattice is made as follows. Assuming the properties (2.10)–(2.17), we may construct a sequence of finite lattices which have the proper thermodynamic limit, by defining a sequence $\rho_k(\gamma, n)$ in terms of the interaction $\rho_k(\gamma)$, by

$$\rho_k(\gamma, n) \equiv \sum_s^{(\infty)} \rho_{k+sm}(\gamma) \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma), \tag{2.18}$$

where $\sum_s^{(\infty)}$ is the sum over all vectors with integer components in the infinite lattice. The symmetry

[14] Actually, $\rho_{kl} \geq 0$ is sufficient for all the proofs, provided that $(\rho^s)_{kl} > 0$ for some positive integer $s^{(k,l)}$. This means that for any k and l there is a chain $\rho_{kr_1}, \rho_{r_1r_2}, \cdots, \rho_{r_{s-1}l}$ of positive elements connecting k and l. This excludes the case of a lattice composed of two noninteracting sublattices.

condition $\rho_k(\gamma, n) = \rho_{-k}(\gamma, n)$ and the condition $\rho_k(\gamma, n) > 0$ are satisfied since $\rho_k(\gamma)$ was assumed to have these properties. Equation (2.3) then follows from (2.18). We also have

$$\lim_{n\to\infty} \rho_k(\gamma, n) = \rho_k(\gamma) \tag{2.19}$$

and

$$\begin{aligned}\sum_k \rho_k(\gamma, n)e^{2\pi ik\cdot r/m} &= \sum_k \sum_s^{(\infty)} \rho_{k+sm}(\gamma)e^{-2\pi ik\cdot r/m} \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma) \\ &= \sum_k \sum_s^{(\infty)} \rho_{k+sm}(\gamma)e^{-2\pi i(k+sm)\cdot r/m} \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma) \\ &= \sum_{k'} \rho_{k'}(\gamma)e^{-2\pi ik'\cdot r/m} \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma), \end{aligned} \tag{2.20}$$

or

$$g_r \equiv \sum_k \rho_k(\gamma, n)e^{-2\pi ik\cdot r/m} = g\left(\frac{2\pi\tau}{m}\right) \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma). \tag{2.21}$$

Condition (2.6) is satisfied in view of (2.16). Finally,

$$C_n(\gamma) \equiv \gamma g_0 = \gamma g(0) \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma) = C(\gamma) \Big/ \sum_s^{(\infty)} \rho_{sm}(\gamma) \tag{2.22}$$

so that the conditions (2.8) and (2.9) are fulfilled.

SECTION 3

The calculations are based on the identity[15]

$$\begin{aligned}(2\pi)^{-n/2}(\det A)^{\frac{1}{2}} &\times \int_{-\infty}^{\infty}\cdots\int d^n x \exp\left(i\sum_k \xi_k x_k - \tfrac{1}{2}\sum_{k,l} x_k A_{kl} x_l\right) \\ &= \exp\left[-\tfrac{1}{2}\sum_{k,l} \xi_k (A^{-1})_{kl}\xi_l\right], \end{aligned} \tag{3.1}$$

valid for any positive-definite symmetric matrix A and any set of complex variables ξ_k. It is convenient to consider the integral as the average of $\exp(i\sum_k \xi_k x_k)$ with respect to the probability density

$$W_n(\mathbf{x}) = (2\pi)^{-n/2}(\det A)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(\mathbf{x}\cdot A\cdot \mathbf{x})\}, \tag{3.2}$$

(written in an obvious vector notation) and to use the abbreviation

$$\left\langle \exp\left\{i\sum_k \xi_k x_k\right\}\right\rangle_{\text{av x}} \tag{3.3}$$

for the left-hand side of Eq. (3.1). It is customary to call the variables x_k normally distributed or Gaussian random variables.

[15] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, N.J., 1951), p. 118.

Substituting $-i\mu_k(\nu\gamma)^{\frac{1}{2}}$ for ξ_k and ρ for A^{-1}, one then has for Q_n, defined by Eq. (2.1), the expression

$$Q_n = \exp\left(-\frac{n\nu\gamma}{2}\right)\sum_{\{\mu\}} \exp\left\langle\left((\nu\gamma)^{\frac{1}{2}}\sum_k \mu_k x_k\right)\right\rangle_{\text{av x}}. \tag{3.4}$$

The first factor compensates for the diagonal elements in $\sum_{kl}$, which are excluded in Eq. (2.1), but not in Eq. (3.1). The summation over the spin-variables can then be carried out and yields

$$Q_n = 2^n e^{-n\nu\gamma/2}\left\langle \prod_k \cosh\left(x_k(\nu\gamma)^{\frac{1}{2}}\right)\right\rangle_{\text{av x}}. \tag{3.5}$$

Equations (3.4) and (3.5) can be interpreted by the statement that the partition function of the system of interacting spins is equal to the average of the partition function of noninteracting spins in a (temperature-dependent) random magnetic field. These equations thus represent a rigorous formulation of the idea of the Weiss field.

In trying to evaluate Eq. (3.5) for $\gamma \to 0$, one is obviously led to try the approximation

$$\cosh\left(x_k(\nu\gamma)^{\frac{1}{2}}\right) \cong e^{\frac{1}{2}\nu\gamma x_k^2}. \tag{3.6}$$

The resulting average exists only if the smallest eigenvalue of ρ^{-1} is larger than $\nu\gamma$, that is, if

$$\nu\gamma g_0 = \nu\gamma \sum \rho_k < 1, \tag{3.7}$$

where g_0 is the largest eigenvalue of ρ. In the limit $n \to \infty$ this means

$$\nu C(\gamma) < 1. \tag{3.8}$$

The average of the approximate expression is easily evaluated and one obtains

$$Q_n \cong Q_n^0 \equiv 2^n e^{-n\nu\gamma/2}\{\det(I - \nu\gamma\rho)\}^{-\frac{1}{2}}. \tag{3.9}$$

This yields the free energy to order γ in agreement with Brout[1] above the BWCP. Since

$$\cosh y \le e^{\frac{1}{2}y^2}, \tag{3.10}$$

we have

$$Q_n \le Q_n^0. \tag{3.10'}$$

Before we continue with the detailed calculation, it is of interest to discuss the integral

$$\mathfrak{I} \equiv \int_{-\infty}^{\infty}\cdots\int \left\{\prod_k \cosh\left[x_k(\nu\gamma)^{\frac{1}{2}}\right]\right\} \times \exp\left[-\tfrac{1}{2}\sum_{k,l} x_k(\rho^{-1})_{kl}x_l\right] d^n x, \tag{3.11}$$

which is the important factor in Eq. (3.5).

Extrema of the integrand occur for

$$(\nu\gamma)^{\frac{1}{2}} \tanh\left[x_k(\nu\gamma)^{\frac{1}{2}}\right] - \sum_l (\rho^{-1})_{kl}x_l = 0 \tag{3.12}$$

or

$$x_k = (\nu\gamma)^{\frac{1}{2}} \sum_l \rho_{kl} \tanh [x_l(\nu\gamma)^{\frac{1}{2}}]. \tag{3.13}$$

The point $x_k = 0$ (for all k) is always a root of this set of equations. Above the BWCP it is the only root, since

$$\nu\gamma \sum_k x_k^2 = (\nu\gamma)^2 \sum_{l,j} (\rho^2)_{lj} \tanh [x_l(\nu\gamma)^{\frac{1}{2}}] \tanh [x_j(\nu\gamma)^{\frac{1}{2}}]$$
$$\leq (\nu\gamma g_0)^2 \sum_l \{\tanh [x_l(\nu\gamma)^{\frac{1}{2}}]\}^2,$$

where g_0 is the largest eigenvalue of ρ. Obviously this inequality is compatible with (3.7) only if all x_k are equal to zero. This point is then a maximum of the integrand.

At temperatures below the BWCP, the point $\mathbf{x} = 0$ is still a root of Eqs. (3.13), but is not a maximum of the integrand. There are, however, two trivial roots:

$$x_k = x \quad \text{for all } k \tag{3.14a}$$

and

$$x_k = -x \quad \text{for all } k, \tag{3.14b}$$

where x is the positive root of the equation

$$x = (\nu\gamma)^{\frac{1}{2}} \sum_l \rho_l \tanh [x(\nu\gamma)^{\frac{1}{2}}] \tag{3.15}$$

or

$$(\nu\gamma)^{\frac{1}{2}} x = \nu C_n(\gamma) \tanh [x(\nu\gamma)^{\frac{1}{2}}]. \tag{3.16}$$

The integrand has maxima of equal size at these two points. In the following discussion, we will refer to these two points as the Weiss field.

Other maxima can exist,[16] but they are smaller than the maxima at the points given by Eqs. (3.14). One sees this by writing the integrand in Eq. (3.11) in the form $e^{U(\mathbf{x})}$ with

$$U(\mathbf{x}) = -\tfrac{1}{2} \sum_{k,l} x_k(\rho^{-1})_{kl} x_l + \sum_k \ln \cosh [x_k(\nu\gamma)^{\frac{1}{2}}]. \tag{3.17}$$

At the extremum

$$U(\mathbf{x}) = \sum_k u(x_k) \tag{3.18}$$

with

$$u(x_k) = \ln \cosh [x_k(\nu\gamma)^{\frac{1}{2}}] - \tfrac{1}{2}(\nu\gamma)^{\frac{1}{2}} x_k \tanh [x_k(\nu\gamma)^{\frac{1}{2}}]. \tag{3.19}$$

The function $u(x)$ is even, and $du/dx > 0$ for $x > 0$. Furthermore, if x_{k_0} is the largest component of a solution of Eq. (3.13), and if $x_{k_0} > 0$,[17] then

$$(\nu\gamma)^{\frac{1}{2}} x_{k_0} \leq \nu\gamma \sum_l \rho_l \tanh [x_{k_0}(\nu\gamma)^{\frac{1}{2}}] \tag{3.20}$$

or

$$\frac{\tanh [x_{k_0}(\nu\gamma)^{\frac{1}{2}}]}{x_{k_0}(\nu\gamma)^{\frac{1}{2}}} \geq \left(\nu\gamma \sum_l \rho_l\right)^{-1} = \frac{\tanh [x(\nu\gamma)^{\frac{1}{2}}]}{x(\nu\gamma)^{\frac{1}{2}}} \tag{3.21}$$

from Eq. (3.16) and $x_{k_0} \leq x$. This argument can easily be generalized to show that all components of all solutions of Eq. (3.13) lie between x and $-x$, defined by Eq. (3.16). Since the equal sign in the inequality (3.20) occurs only when all components are equal to x_{k_0}, we have

$$|x_m| < x \tag{3.22}$$

for the largest components x_m of any solution other than the Weiss field and, therefore,

$$U(\mathbf{x}^0) - U(\mathbf{x}) > n(u(x) - u(x_m)), \tag{3.23}$$

where $\mathbf{x}^0$ is the Weiss field and $\mathbf{x}$ any other solution of Eq. (3.13). The inequality (3.23) does not imply that $U(\mathbf{x}^0) - U(\mathbf{x})$ is of order n, since we have not proved that $u(x) - u(x_m)$ has a lower bound when n becomes infinite. We cannot disprove the existence of solutions of Eq. (3.13), whose components have values arbitrarily close to x and $-x$ over large domains of lattice points, when n becomes very large. We can prove that no solution of Eq. (3.13), other than the Weiss field, has components which are all of the same sign (Appendix A). Integration over the neighborhood of the Weiss field only can still yield a good approximation if the contribution of the domain type maxima is essentially the same as that of the Weiss field maxima and their number sufficiently small compared to $(\exp n\gamma)$.[18]

SECTION 4

In order to carry out the calculation of the free energy to order γ^2 above the BWCP, we apply a simple transformation which results in the extraction of the factor Q_n^0 from the partition function. Q_n can be expressed in the form

$$Q_n = \sum_{\{\mu\}} \exp\left[-\tfrac{1}{2} \sum_{k,l} (\delta_{kl} - \nu\gamma\rho_{kl})\mu_k\mu_l\right] \times \exp\left[\frac{n}{2}(1 - \nu\gamma)\right]. \tag{4.1}$$

Since the matrix $I - \nu\gamma\rho$ is positive-definite above the BWCP, this can be expressed as

$$Q_n = \sum_{\{\mu\}} \left\langle \exp\left(i \sum_k \xi_k\mu_k\right)\right\rangle_{\text{av }\xi} \exp\left[\frac{n}{2}(1 - \nu\gamma)\right]$$
$$= 2^n \exp\left[\frac{n}{2}(1 - \nu\gamma)\right] \left\langle \prod_k \cos \xi_k \right\rangle_{\text{av }\xi}, \tag{4.1'}$$

[16] Some examples were pointed out to us by Dr. George Moore. These known examples result in maxima small compared with the two largest maxima.

[17] If $x_{k_0} < 0$, we consider instead the inverted solution $x'_k = -x_k$.

[18] This does not mean that the contribution of the other maxima to the integral is negligible, but only that their contribution to the free energy per particle is irrelevant. See the argument preceding Eq. (2.4) in G. F. Newell and E. Montroll, Rev. Mod. Phys. **25**, 353 (1953).

where the new random variables ξ_k are defined by the probability density

$$w_n(\xi) = (2\pi)^{-n/2}\{\det(I - \nu\gamma\rho)\}^{-\frac{1}{2}} e^{-\frac{1}{2}(\xi\cdot(I-\nu\gamma\rho)^{-1}\cdot\xi)}. \tag{4.2}$$

This result follows from Eqs. (3.1)–(3.3). With Q_n^0 defined by Eq. (3.9) we then have

$$Q_n = Q_n^0 q_n, \tag{4.3}$$

where q_n is defined by

$$q_n = e^{n/2}(2\pi)^{-n/2}\int_{-\infty}^{\infty}\cdots\int d^n\xi \times e^{-\frac{1}{2}(\xi\cdot(I-\nu\gamma\rho)^{-1}\cdot\xi)}\prod_k \cos\xi_k. \tag{4.4}$$

The following calculations are motivated by the observation that the random variables ξ_k are anticorrelated, i.e.,

$$\langle\xi_k\xi_l\rangle_{\text{av }\xi} = -\nu\gamma\rho_{kl} < 0, \quad \text{for} \quad k \neq l. \tag{4.5}$$

It seemed possible that an expansion, starting in zeroth approximation with independent random variables, could yield results, even though the range of the anticorrelation increases with γ^{-1}. This expectation, though admittedly weak, is borne out by the calculation.

For the purpose of this expansion we define the matrix $\tilde{\rho}$ as a function of ν, γ, and n, by

$$I + \nu\gamma\tilde{\rho} = (I - \nu\gamma\rho)^{-1}. \tag{4.6}$$

This equation can also be written in the form

$$\tilde{\rho} = \rho + \nu\gamma\tilde{\rho}\rho, \tag{4.7}$$

if $\nu\gamma \neq 0$. The homogeneity of ρ carries over to $\tilde{\rho}$ so that we can write $\tilde{\rho}_{kl}$ or $\tilde{\rho}_{k-l}$ for the matrix elements. We then divide the quadratic form in Eq. (4.4) into diagonal and off-diagonal terms and introduce new variables:

$$\begin{aligned} q_n &= e^{n/2}(2\pi)^{-n/2}\int_{-\infty}^{\infty}\cdots\int d^n\xi \\ &\quad\times\exp\left[-\tfrac{1}{2}\sum_k \xi_k^2(I+\nu\gamma\tilde{\rho}_0) - \frac{\nu\gamma}{2}\sum_{k,l}{}' \xi_k\xi_l\tilde{\rho}_{kl}\right] \\ &\quad\times\prod_k\cos\xi_k \\ &= e^{n/2}(2\pi)^{-n/2}u^n\int_{-\infty}^{\infty}\cdots\int d^n\eta \\ &\quad\times\exp\left[-\tfrac{1}{2}\sum_k\eta_k^2 - \frac{\nu\gamma u^2}{2}\sum_{k,l}{}'\eta_k\eta_l\tilde{\rho}_{kl}\right] \\ &\quad\times\prod_k\cos(u\eta_k), \end{aligned} \tag{4.8}$$

with η_k defined by

$$\xi_k = u\eta_k \tag{4.9}$$

and

$$u = (1 + \nu\gamma\tilde{\rho}_0)^{-\frac{1}{2}}. \tag{4.10}$$

The factor $(2\pi)^{-n/2}\exp\{-\frac{1}{2}\sum_k\eta_k^2\}$ can be considered as the probability density for independent Gaussian random variables η_k with mean zero and standard deviation unity, and we can write q_n in the form

$$q_n = e^{n/2}u^n \times\left\langle\exp\left(-\frac{\nu\gamma u^2}{2}\sum_{k,l}{}'\eta_k\eta_l\tilde{\rho}_{kl}\right)\prod_k\cos(u\eta_k)\right\rangle_{\text{av }\eta}. \tag{4.11}$$

It is convenient to write this in the form

$$q_n = q_n^{(1)}q_n^{(2)}, \tag{4.12}$$

with

$$q_n^{(1)} \equiv e^{n/2}u^n\langle\cos(u\eta)\rangle_{\text{av }\eta}^n, \tag{4.13}$$

$$q_n^{(2)} \equiv \left\langle\exp\left(-\frac{\nu\gamma u^2}{2}\sum_{k,l}{}'\eta_k\eta_l\tilde{\rho}_{kl}\right)\times\prod_k\frac{\cos(u\eta_k)}{\langle\cos(u\eta)\rangle_{\text{av }\eta}}\right\rangle_{\text{av }\eta} \tag{4.14}$$

where

$$\langle\cos(u\eta)\rangle_{\text{av }\eta} \equiv (2\pi)^{-\frac{1}{2}}\int e^{-\frac{1}{2}\eta^2}\cos(u\eta)\,d\eta = e^{-\frac{1}{2}u^2}. \tag{4.15}$$

From Eqs. (4.3) and (4.12) we then have

$$Q_n = Q_n^0 q_n^{(1)} q_n^{(2)}. \tag{4.16}$$

Substituting from Eqs. (3.9), (4.10), and (4.13), we obtain

$$\begin{aligned} \frac{1}{n}\ln Q_n &= \ln 2 - \frac{\nu\gamma}{2} - \frac{1}{2n}\ln\det(I-\nu\gamma\rho) + \tfrac{1}{2} \\ &\quad - \tfrac{1}{2}\ln(1+\nu\gamma\tilde{\rho}_0) - \tfrac{1}{2}(1+\nu\gamma\tilde{\rho}_0)^{-1} \\ &\quad + \frac{1}{n}\ln q_n^{(2)}. \end{aligned} \tag{4.17}$$

We will show later that the correction term $1/n \ln q_n^{(2)}$ is of the order γ^3, if the expansion of this term in powers of γ is semiconvergent.

The terms given in closed form can also be expressed in terms of one function $\tilde{\rho}_0(\xi)$, which is the diagonal element of a matrix $\tilde{\rho}(\xi)$ defined by a slight generalization of Eq. (4.7). With $\tilde{\rho}(\xi)$ defined by

$$\tilde{\rho}(\xi) = \rho + \xi\tilde{\rho}(\xi)\rho, \tag{4.18}$$

we have

$$\tilde{\rho}(\xi) = \rho(I - \xi\rho)^{-1} \tag{4.19}$$

and

$$\ln \det (I - \nu\gamma\rho) = \mathrm{Tr} \ln (I - \nu\gamma\rho)$$
$$= -\mathrm{Tr}\int_0^{\nu\gamma} \frac{\rho}{1 - \xi\rho}\, d\xi$$
$$= -\mathrm{Tr}\int_0^{\nu\gamma} \tilde{\rho}(\xi)\, d\xi = -n\int_0 \tilde{\rho}_0(\xi)\, d\xi. \tag{4.20}$$

Equation (4.17) can then be written in the form

$$\frac{1}{n} \ln Q_n = \ln 2 + \tfrac{1}{2}(1 - \nu\gamma)$$
$$+ \tfrac{1}{2}\int_0^{\nu\gamma} \tilde{\rho}_0(\xi)\, d\xi - \tfrac{1}{2} \ln (1 + \nu\gamma\tilde{\rho}_0(\nu\gamma))$$
$$- \tfrac{1}{2}(1 + \nu\gamma\tilde{\rho}_0(\nu\gamma))^{-1} + \frac{1}{n} \ln q_n^{(2)}. \tag{4.21}$$

The function $\tilde{\rho}_0(\xi)$ can be expressed in terms of the Fourier coefficients g_τ defined by (2.5). We then have

$$\rho_k = \frac{1}{n}\sum_\tau g_\tau e^{2\pi i k\cdot\tau/m} \tag{4.22}$$

and

$$\tilde{\rho}_k(\xi) = \frac{1}{n}\sum_\tau \frac{g_\tau}{1 - \xi g_\tau} e^{2\pi i k\cdot\tau/m}, \tag{4.23}$$

where the summation extends over all vectors τ with components $0, 1, \cdots, m-1$.

The free energy per particle, ψ, is given by

$$-\psi/kT = \lim_{n\to\infty} \frac{1}{n} \ln Q_n.$$

The limit $R(\xi)$ of $\tilde{\rho}_0(\xi)$ required for the evaluation of ψ becomes the integral

$$R(\xi) \equiv \lim_{n\to\infty} \tilde{\rho}_0(\xi) = \frac{1}{(2\pi)^D}\int_0^{2\pi}\cdots\int d^D\omega \frac{g(\omega)}{1 - \xi g(\omega)}, \tag{4.24}$$

where ω is a vector in D dimensions and

$$g(\omega) = \lim_{n\to\infty}\sum_k \rho_k(\gamma, n) e^{-ik\cdot\omega}. \tag{4.25}$$

[The sequence $\rho_k(\gamma, n)$ was introduced to have the convenience of Born–von Kármán conditions; only $\rho_k(\gamma, \infty)$ is of interest in the physical problem. We can choose a sequence such that

$$g(\omega) = \sum_k^{(\infty)} \rho_k(\gamma, \infty) e^{-ik\cdot\omega}, \tag{4.26}$$

where $\sum_k^{(\infty)}$ is the sum over all vectors k with integer components.]

We thus have for the free energy per particle, ψ, the result

$$-\psi/kT = \ln 2 + \tfrac{1}{2}(1 - \nu\gamma) + \tfrac{1}{2}\int_0^{\nu\gamma} R(\xi)\, d\xi$$
$$- \tfrac{1}{2} \ln (1 + \nu\gamma R(\nu\gamma)) - \tfrac{1}{2}(1 + \nu\gamma R(\nu\gamma))^{-1}$$
$$+ \lim_{n\to\infty} \frac{1}{n} \ln q_n^{(2)}, \tag{4.27}$$

where $R(\xi)$ is defined by Eqs. (4.24) and (4.25). A series for $1/n \ln q_n^{(2)}$ will be developed in Sec. 5, and it will be shown there that the limit $n \to \infty$ of each term in the series is of order γ^3 or higher order. The order of each term in the series will be obtained in Sec. 5 and the terms of order γ^3 will be calculated there.

It is interesting to observe the singularities of the terms of order γ^2 at the BWCP. (The terms of order γ in the free energy remain finite for the one-, two-, and three-dimensional model.) The behavior of the terms of order γ^2 is determined by the number of dimensions and the form of $g(\omega)$ for small ω, which in turn depends on the existence of the second moments of $\rho_k(\gamma)$. The integral representing $R(\nu\gamma)$ exists for any fixed temperature above the BWCP, since

$$|g(\omega)| \leq g(0) = \gamma^{-1}C(\gamma) \tag{4.28}$$

and the integrand is finite if

$$1 - \nu C(\gamma) > 0. \tag{4.29}$$

If in the neighborhood of $\omega = 0$, the function $g(\omega)$ has the form

$$g(\omega) \cong g(0) - \alpha_2\omega^2, \tag{4.30}$$

then the integral representing $R(\nu\gamma)$ becomes infinite at the BWCP for the one- and two-dimensional model, but remains finite for the three-dimensional model.

If, however,

$$g(\omega) \cong g(0) - \alpha_1 |\omega| \quad \text{for} \quad \omega \cong 0, \tag{4.31}$$

the function $R(\nu\gamma)$ becomes infinite at the BWCP for the one-dimensional model only. Since the free energy per particle must remain finite, the singularities of $R(\nu\gamma)$ indicate that the expansion in γ breaks down at the BWCP. (We have left the fourth and fifth term in the form in which they were obtained, rather than expanding them in order γ^2, since the order of the singularity is smaller in the original form.)

The form of $g(\omega)$ for small ω is determined by the moments of $\rho_k(\gamma, \infty)$. If, e.g., the moment tensor $\sum_k kk\rho_k$ exists and is diagonal with equal diagonal elements, the function $g(\omega)$ has the form (30) for sufficiently small ω. If the second moments do not exist, $g(\omega)$ can have the form (4.31) for small ω.

SECTION 5

In this section we derive a series expansion for the last term in Eq. (4.17),

$$\lim_{n\to\infty}\frac{1}{n}\ln q_n^{(2)},$$

in terms of coefficients similar to Mayer's reducible cluster coefficients, and prove that each term of the series is of order γ^3 or higher order. In order to obtain this series we first expand the exponential function in Eq. (4.14) and obtain

$$q_n^{(2)}=\sum_{N=0}^{\infty}\frac{1}{N!}(-\tfrac{1}{2}\nu\gamma u^2)^N \times\left\langle\left(\sum_{k,l}{}'\eta_k\eta_l\tilde{\rho}_{kl}\right)^N\prod_j\varphi(\eta_j)\right\rangle_{\text{av }\eta}, \quad (5.1)$$

where

$$\varphi(\eta_j)\equiv\cos(u\eta_j)/\langle\cos(u\eta_j)\rangle_{\text{av }\eta}=e^{\frac{1}{2}u^2}\cos(u\eta_j) \quad (5.2)$$

and $\prod_j$ extends over all lattice points j. We now change the notation in the sum $\sum'_{k,l}$ from points to pairs of points, so that $p_1, p_2, \cdots, p_N$ denote pairs of points, and we use p'_α, p''_α to indicate the members of the pair p_α. With this notation we have

$$\left(\sum_{k,l}{}'\eta_k\eta_l\tilde{\rho}_{kl}\right)^N=2^N\sum_{p_1}{}'\sum_{p_2}{}'\cdots\sum_{p_N}{}'\prod_{\alpha=1}^{N}(\eta_{p_\alpha'}\eta_{p_\alpha''}\tilde{\rho}_{p_\alpha}), \quad (5.3)$$

where

$$\tilde{\rho}_{p\alpha}\equiv\tilde{\rho}_{p_\alpha' p_\alpha''} \quad (5.4)$$

and the sums $\sum'_{p_\alpha}$ extend over all pairs of two different lattice points, counting each pair only once. Eq. (5.1) then becomes

$$q_n^{(2)}=\sum_{N=0}^{\infty}\frac{1}{N!}(-\nu\gamma u^2)^N \times\sum_{p_1,p_2,\cdots,p_N}{}'\prod_{\alpha=1}^{N}\tilde{\rho}_{p_\alpha}F_N(p_1,p_2,\cdots,p_N), \quad (5.5)$$

with

$$F_N(p_1,p_2,\cdots,p_N)\equiv\left\langle\prod_{\alpha=1}^{N}\eta_{p_\alpha'}\eta_{p_\alpha''}\prod_j\varphi(\eta_j)\right\rangle_{\text{av }\eta}. \quad (5.6)$$

The form of (5.5) is now analogous to that of the grand canonical partition function of a gas; the pairs correspond to the position vectors of the gas molecules, and the sums over pairs to the integrals over the coordinates. We can thus apply the Ursell development to $F_N(p_1,p_2,\cdots,p_N)$ by defining cluster functions $S_\lambda(p_1,p_2,\cdots,p_\lambda)$ in the usual way:

$$F_1(p_1)=S_1(p_1), \quad (5.7)$$

$$F_2(p_1,p_2)=S_1(p_1)S_1(p_2)+S_2(p_1,p_2), \quad (5.8)$$

$$F_3(p_1,p_2,p_3)=S_1(p_1)S_1(p_2)S_1(p_3)+S_2(p_1,p_2)S_1(p_3)+\cdots+S_3(p_1,p_2,p_3), \quad (5.9)$$

etc. Because of the independence of the variables η_k, the cluster functions S_λ vanish when the set of pairs $p_1,\cdots,p_\lambda$ is not connected, that is, when the set of pairs can be divided into two or more subsets, such that pairs in different subsets have no point in common. With $\varphi(\eta)$ defined by (5.2), S_1 vanishes, since

$$S_1(p)=\langle\eta_{p'}\eta_{p''}\varphi(\eta_{p'})\varphi(\eta_{p''})\rangle_{\text{av }\eta}=0 \quad \text{for} \quad p'\neq p'', \quad (5.10)$$

and therefore

$$S_1(p)=\langle\eta_{p'}\varphi(\eta_{p'})\rangle^2_{\text{av }\eta}, \quad (5.11)$$

and

$$\langle\eta_k\cos(u\eta_k)\rangle_{\text{av }\eta}=0. \quad (5.12)$$

With cluster coefficients b_λ defined by

$$n\lambda!\,b_\lambda=\sum_{p_1,\cdots,p_\lambda}S_\lambda(p_1,p_2,\cdots,p_\lambda)\prod_{\alpha=1}^{\lambda}\tilde{\rho}_{p_\alpha}, \quad (5.13)$$

one obtains, by the procedure of Kahn and Uhlenbeck[19]

$$\frac{1}{n}\ln q_n^{(2)}=\sum_{\lambda=2}^{\infty}(-\nu\gamma u^2)^\lambda b_\lambda. \quad (5.14)$$

It should be noted that this is not a power series in γ, since b_λ depends on γ.

The set of pairs $\{p_1,\cdots,p_\lambda\}$ in Eq. (5.13) is represented by a graph, if each pair p_α is represented by a line connecting the two lattice points p'_α and p''_α. The cluster property of S_λ restricts the sum to connected graphs. Furthermore, the number of lines entering any point of the graph must be even, since

$$\langle\eta_k^s\cos(u\eta_k)\rangle_{\text{av }\eta}=0 \quad \text{for odd } s, \quad (5.15)$$

and, therefore, $F_\lambda(p_1,\cdots,p_\lambda)$, as well as any product of functions F_λ occurring in the expression for S_λ vanishes, if this restriction is violated. We indicate these two restrictions by the subscript "ev. con." and write the definition of b_λ in the form

$$n\lambda!\,b_\lambda=\sum_{\substack{p_1,\cdots,p_\lambda\\ \text{ev. con.}}}S_\lambda(p_1,\cdots,p_2)\prod_{\alpha=1}^{\lambda}\tilde{\rho}_{p_\alpha}. \quad (5.16)$$

The coefficients b_λ in Eq. (5.14) depend on the number n of lattice sites. We will assume that for the

[19] B. Kahn and G. E. Uhlenbeck, Physica **5**, 399 (1938). See also B. Kahn in *Studies in Statistical Mechanics*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publishing Co., Amsterdam, 1965), Vol. III.

purpose of estimating the order in γ it is permissible to perform the limit $n \to \infty$ term by term. In order to show that $\lim_{n\to\infty} (1/n) \ln q_n^{(2)}$ is of order γ^3, we prove first that

$$S_\lambda(p_1, \cdots, p_\lambda) = \gamma^{m_2} \bar{S}_\lambda, \tag{5.17}$$

where $|S_\lambda|$ remains bounded for $\gamma \to 0$ and m_2 is the number of points of degree two. (The "degree" of a point is the number of lines entering it.) In order to show this, we calculate $F_\lambda(p_1, \cdots, p_\lambda)$ explicitly. If the graph $(p_1, \cdots, p_\lambda)$ has m_f points of degree f, we obtain from (5.2) and (5.6)

$$F_\lambda(p_1, \cdots, p_\lambda) = \prod_f [\langle \eta^f \cos(u\eta)\rangle_{\mathrm{av}\,\eta} / \langle \cos(u\eta)\rangle_{\mathrm{av}\,\eta}]^{m_f}, \tag{5.18}$$

where the product extends over even numbers f. Using Eq. (4.15), we obtain

$$\langle \eta^f \cos(u\eta)\rangle_{\mathrm{av}\,\eta} / \langle \cos(u\eta)\rangle_{\mathrm{av}\,\eta} = e^{u^2/2}(-1)^{f/2}\left(\frac{\partial}{\partial u}\right)^f e^{-u^2/2} = (-1)^{f/2} H_f(u), \tag{5.19}$$

where $H_f(u)$ is the Hermite polynomial of degree f in the notation of Ref. 15. We have in particular

$$H_2(u) = u^2 - 1 = (1 + \nu\gamma\tilde{\rho}_0)^{-1} - 1 = -\nu\gamma\tilde{\rho}_0/(1 + \nu\gamma\tilde{\rho}_0). \tag{5.20}$$

We will show that $\tilde{\rho}_0$ is finite for $\nu C(\gamma) < 1$, so that

$$F_\lambda(p_1, \cdots, p_\lambda) = \gamma^{m_2} \bar{F}_\lambda, \tag{5.21}$$

where $\bar{F}_\lambda$ remains bounded for $\gamma \to 0$. Equation (5.17) then follows from the inversion formula for S_λ.

To estimate the diagram sums, we note first that the matrix elements of $\tilde{\rho}$ are nonnegative since

$$\tilde{\rho} = \rho \sum_{s=0}^{\infty} (\nu\gamma\rho)^s, \tag{5.22}$$

and the matrix elements of ρ are nonnegative by assumption. The series converges when $\nu\gamma g_0 < 1$. We then obtain an estimate for the diagonal element of powers of $\tilde{\rho}$. We have

$$(\tilde{\rho}^\lambda)_0 = \frac{1}{n} \sum_\tau \left(\frac{g_\tau}{1 - \nu\gamma g_\tau}\right)^\lambda, \tag{5.23}$$

where g_τ is defined by Eq. (2.5). Since

$$0 < g_\tau \leq g_0, \tag{5.24}$$

we have

$$(\tilde{\rho}^\lambda)_0 \leq \left(\frac{1}{1 - \nu\gamma g_0}\right)^\lambda \frac{1}{n} \sum_\tau g_\tau^\lambda \leq \frac{g_0^{\lambda-1}}{(1 - \nu\gamma g_0)^\lambda} \frac{1}{n} \sum_\tau g_\tau = \frac{g_0^{\lambda-1}}{(1 - \nu\gamma g_0)^\lambda} \tag{5.25}$$

or

$$\gamma^\lambda (\tilde{\rho}^\lambda)_0 \leq \gamma M_\lambda, \tag{5.26}$$

where M_λ remains bounded for $\gamma \to 0$ for fixed $\nu\gamma g_0 = \nu C(\gamma) < 1$. We note further that

$$\frac{g_\tau}{1 - \nu\gamma g_\tau} > 0, \tag{5.27}$$

since

$$0 < g_\tau \leq g_0. \tag{5.28}$$

Therefore

$$(\tilde{\rho}^\lambda)_{kl} \leq (\tilde{\rho}^\lambda)_0. \tag{5.29}$$

We can now estimate the sum over all graphs of the type under consideration, and obtain

$$\sum_{p_1, \cdots, p_\lambda} \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq 2^{-\lambda} \sum_{k, l_1, \cdots, l_{\lambda-2}} \tilde{\rho}_{kl_1}\tilde{\rho}_{l_1 l_2} \cdots \tilde{\rho}_{l_{\lambda-1}k} = 2^{-\lambda} \sum_k (\tilde{\rho}^\lambda)_{kk} = 2^{-\lambda} n (\tilde{\rho}^\lambda)_0, \tag{5.30}$$

since every graph on the left appears in the sum on the right, and all $\tilde{\rho}_{kl}$ are nonnegative. This upper bound is still too high, but serves together with the inequality (5.26) to confirm the expectation that one obtains a factor γ^{-1} from the summation over every point except one, which one can consider fixed to start with. Summation over that point then yields the factor n. An estimate for the sum over graphs with m_f points of degree f ($f = 2, 4, \cdots$) is then obtained by observing that this condition constitutes a constraint on the sum over points, which keeps $(f/2) - 1$ points tied to one of the points and thus eliminates $(f/2) - 1$ factors γ^{-1}. Therefore we expect to find

$$\lim_{n\to\infty} \frac{1}{n} \gamma^\lambda \sum{}' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \sim \gamma^{\left[1 + \sum_{f\geq 4} m_f (f/2 - 1)\right]}, \tag{5.31}$$

where $\sum'$ means the sum over all graphs of the type considered, which have m_f points of degree $f > 2$.

In Appendix B we have made this argument rigorous for two special cases, which are sufficient for the calculation to order γ^3. We have shown there that the sum over all graphs with at least one point of degree f satisfies the inequality

$$\lim_{n\to\infty} \frac{\gamma^\lambda}{n} \sum{}' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \gamma^{f/2} M'_\lambda, \tag{5.32}$$

and the sum over graphs with at least one point of degree f and at least one other point of degree h satisfies the inequality

$$\lim_{n\to\infty} \frac{\gamma^\lambda}{n} \sum{}'' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \gamma^{f/2 + h/2 - 1} M''_\lambda, \tag{5.33}$$

where M'_λ and M''_λ remain bounded for $\gamma \to 0$.

Using these inequalities and Eq. (5.17), we have

$$\gamma^\lambda |b_\lambda| \leq \gamma^{m_2+f/2} C'_\lambda \tag{5.34}$$

with the first condition and

$$\gamma^\lambda |b_\lambda| \leq \gamma^{m_2+f/2+h/2-1} C''_\lambda \tag{5.35}$$

with the second condition, where C'_λ and C''_λ remain bounded when $\gamma \to 0$, and m_2 is the number of points of degree 2. If there is no point of degree four or higher, we have

$$\gamma^\lambda |b_\lambda| \leq \gamma^{m_2+1} C'_\lambda = \gamma^{1+\lambda} C'_\lambda . \tag{5.36}$$

If there is at least one point of degree four, there must be either two or more points of degree two, or at least one more point of degree four or higher. If the first alternative is true, one has from (34)

$$\gamma^\lambda |b_\lambda| \leq \gamma^4 C'_\lambda . \tag{5.37}$$

If the second alternative is true, one has from (35)

$$\gamma^\lambda |b_\lambda| \leq \gamma^3 C'_\lambda . \tag{5.38}$$

Every term in the series for $\lim_{n\to\infty} (1/n) \ln q_n^{(2)}$ is, therefore, of order γ^3 or higher order, since $\lambda \geq 2$.

If one accepts the heuristic argument leading to Eq. (5.31), one obtains

$$\gamma^\lambda b_\lambda \sim \gamma^{\left[1+m_2+\sum_{f\geq 4} m_f[(f/2)-1]\right]} . \tag{5.39}$$

Since

$$\lambda = \tfrac{1}{2} \sum_{f\geq 2} f m_f , \tag{5.40}$$

one has

$$\sum_{f\geq 4} m_f \frac{f}{4} = \tfrac{1}{2}(\lambda - m_2), \tag{5.41}$$

and

$$\sum_{f\geq 4} m_f \left(\frac{f}{2} - 1\right) \geq \sum_{f\geq 4} m_f \left(\frac{f}{4}\right) = \tfrac{1}{2}(\lambda - m_2), \tag{5.42}$$

so that

$$1 + m_2 + \sum_{f\geq 4} m_f \left(\frac{f}{2} - 1\right) \geq 1 + \tfrac{1}{2}(\lambda + m_2). \tag{5.43}$$

Taking $m_2 = 0$ in the inequality (5.43) shows that all terms through order γ^n are obtained by taking terms through $\lambda = 2(n-1)$ in (5.14). Some of the terms in (5.14) can still be ruled out by using the full inequality (5.43) (including m_2). For instance, terms of order γ^3 in $(1/n) \ln q_n^{(2)}$ can arise only from $\lambda = 2$ and $m_2 = 2$ (2 lines and 2 points of degree 2), and from $\lambda = 4$ and $m_2 = 0$ (4 lines and 2 points of degree 4).

These two terms are easily evaluated. From Eqs. (5.13), (5.8), (5.18), (5.19), and (5.20), one has

$$\begin{aligned} b_2 &= \frac{1}{2}\frac{1}{n} \sum_p S_2(p, p)(\tilde{\rho}_p)^2 = \frac{1}{2}\frac{1}{n} \sum_p F_2(p, p)(\tilde{\rho}_p)^2 \\ &= \frac{1}{2}\frac{1}{n} H_2^2(u) \sum_p (\tilde{\rho}_p)^2 = \frac{1}{4}\left(\frac{\nu\gamma\tilde{\rho}_0}{1+\nu\gamma\tilde{\rho}_0}\right)^2 \frac{1}{n} \sum_{k,l}' (\tilde{\rho}_{kl})^2 \\ &= \frac{1}{4}\left(\frac{\nu\gamma\tilde{\rho}_0}{1+\nu\gamma\tilde{\rho}_0}\right)^2 \frac{1}{n} \sum_k [(\tilde{\rho}^2)_{kk} - (\tilde{\rho}_{kk})^2] \\ &= \frac{1}{4}\left(\frac{\nu\gamma\tilde{\rho}_0}{1+\nu\gamma\tilde{\rho}_0}\right)^2 [(\tilde{\rho}^2)_0 - (\tilde{\rho}_0)^2]. \end{aligned} \tag{5.44}$$

In the limit $n \to \infty$, $\tilde{\rho}_0$ becomes $R(\nu\gamma)$, defined by Eq. (4.24); using (5.23), $(\tilde{\rho}^2)_0$ becomes

$$\begin{aligned} R_{22}(\nu\gamma) &= \lim_{n\to\infty} (\tilde{\rho}^2)_0 \\ &= \frac{1}{(2\pi)^D} \int_0^{2\pi} \cdots \int d^D\omega \left\{\frac{g(\omega)}{1-\nu\gamma g(\omega)}\right\}^2, \end{aligned} \tag{5.45}$$

with $g(\omega)$ given by Eq. (4.26). Only the term $(\tilde{\rho}^2)_0$ contributes a term of order γ^3 to $\gamma^2 b_2$, since $\tilde{\rho}_0$ remains bounded for $\gamma \to 0$ according to (5.26).

The only term which can contribute a γ^3 term to $\gamma^4 b_4$ is

$$b_4^0 \equiv \frac{1}{4!}\frac{1}{n} \sum_p S_4(p, p, p, p)(\tilde{\rho}_p)^4. \tag{5.46}$$

Since

$$\begin{aligned} S_4(p_1, p_2, p_3, p_4) = {} & F_4(p_1, p_2, p_3, p_4) \\ & - [F_2(p_1, p_2)F_2(p_3, p_4) \\ & + F_2(p_1, p_3)F_2(p_2, p_4) \\ & + F_2(p_1, p_4)F_2(p_2, p_3)], \end{aligned} \tag{5.47}$$

from Eqs. (5.18) and (5.19) we obtain

$$S_4(p, p, p, p) = H_4^2(u) - 3H_2^4(u), \tag{5.48}$$

with

$$u \equiv (1 + \nu\gamma\tilde{\rho}_0)^{-\frac{1}{2}},$$

and we have

$$\begin{aligned} b_4^0 &= \frac{1}{4!}(H_4^2(u) - 3H_2^4(u)) \frac{1}{2n} \sum_{k,l}' (\tilde{\rho}_{kl})^4 \\ &= \frac{1}{4!}(H_4^2(u) - 3H_2^4(u)) \tfrac{1}{2}\left(\sum_l (\tilde{\rho}_l)^4 - (\tilde{\rho}_0)^4\right), \end{aligned} \tag{5.49}$$

where $\sum_l$ extends over all points l of the lattice. From Eq. (4.23) we then obtain

$$\begin{aligned} &\sum_l (\tilde{\rho}_l)^4 \\ &= \frac{1}{n^4} \sum_l \sum_{r_1,r_2,r_3,r_4} \prod_{j=1}^4 \left(\frac{g_{r_j}}{1-\nu\gamma g_{r_j}}\right) e^{(2\pi i l/m)\cdot(r_1+r_2+r_3+r_4)} \\ &= \frac{1}{n^3} \sum_{r_1,r_2,r_3} \prod_{j=1}^3 \left(\frac{g_{r_j}}{1-\nu\gamma g_r}\right) \frac{g_{r_1+r_2+r_3}}{1-\nu\gamma g_{r_1+r_2+r_3}} . \end{aligned} \tag{5.50}$$

For $n \to \infty$, this becomes

$$R_{42}(\nu\gamma) \equiv \lim_{n\to\infty} \sum_l (\tilde{\rho}_l)^4 = (2\pi)^{-3D} \int_0^{2\pi} \cdots \int d^D\omega_1\, d^D\omega_2\, d^D\omega_3 \times \prod_{j=1}^{3} \left(\frac{g(\omega_j)}{1 - \nu\gamma g(\omega_j)} \right) \times \frac{g(\omega_1 + \omega_2 + \omega_3)}{1 - \nu\gamma g(\omega_1 + \omega_2 + \omega_3)}. \quad (5.51)$$

The terms $H_4^2(u)$ and $(\tilde{\rho}_0)^4$ contribute terms of higher order in γ. Substituting these results in Eq. (5.14), we obtain to order γ^3

$$\lim_{n\to\infty} \frac{1}{n} \ln q_n^{(2)} = (\nu\gamma)^2 \tfrac{1}{4} [\nu\gamma R(\nu\gamma)]^2 R_{22}(\nu\gamma) + (\nu\gamma)^4 \frac{1}{4!} H_4^2(1) \tfrac{1}{2} R_{42}(\nu\gamma) + ((\gamma^4)) = (\nu\gamma)^4 \{ \tfrac{1}{4} R^2(\nu\gamma) R_{22}(\nu\gamma) + \tfrac{1}{12} R_{42}(\nu\gamma) \} + ((\gamma^4)), \quad (5.52)$$

with R, R_{22}, and R_{42} defined by Eqs. (4.24), (5.45), and (5.51), respectively. Since γR_{22} and γR_{42} remain bounded for $\gamma \to 0$, the terms of (5.52) are of order γ^3 and higher order.

We note that the functions R_{22} and R_{42} become infinite at the BWCP, even in the three-dimensional case, if $g(\omega)$ is of the form

$$g(\omega) \cong g(0) - \alpha_2 \omega^2 \quad (5.53)$$

for small ω.

In order to check the numerical factors in (5.52), we can perform a Mayer expansion of Eq. (4.14). With

$$\exp\left(-\frac{\nu\gamma u^2}{2} \sum_{k,l}{}' \eta_k \eta_l \tilde{\rho}_{kl} \right) = \prod_{kl}{}'' (1 + f_{kl}), \quad (5.54)$$

where

$$f_{kl} \equiv \exp(-\nu\gamma u^2 \eta_k \eta_l \tilde{\rho}_{kl}) - 1, \quad (5.55)$$

and the double prime indicates that each pair (kl) is to be counted only once, and k is different from l, we have

$$\exp\left(-\frac{\nu\gamma u^2}{2} \sum_{k,l}{}' \eta_k \eta_l \tilde{\rho}_{kl} \right) = 1 + \tfrac{1}{2} \sum_{kl}{}' f_{kl} + \tfrac{1}{4} \sum_{k,l}{}' \sum_{i,j}{}'' f_{kl} f_{ij} + \cdots, \quad (5.56)$$

where $\sum''_{i,j}$ indicates that $i \neq j$ and the pair (ij) is different from (kl). Then we obtain

$$\left\langle \exp\left(-\frac{\nu\gamma u^2}{2} \sum_{k,l}{}' \eta_k \eta_l \tilde{\rho}_{kl} \right) \prod_j \varphi(\eta_j) \right\rangle_{\text{av}\,\eta} = 1 + \tfrac{1}{2} \sum_{k,l}{}' \left\langle \left(\frac{(\nu\gamma u^2)^2}{2!} \eta_k^2 \eta_l^2 (\tilde{\rho}_{kl})^2 + \frac{(\nu\gamma u^2)^4}{4!} \eta_k^4 \eta_l^4 (\tilde{\rho}_{kl})^4 + \cdots \right) \varphi(\eta_k) \varphi(\eta_l) \right\rangle_{\text{av}\,\eta} + \cdots = 1 + \frac{n}{2} \left[\frac{(\nu\gamma u^2)^2}{2!} H_2^2(u) \sum_{l\neq 0} (\tilde{\rho}_l)^2 + \frac{(\nu\gamma u^2)^4}{4!} H_4^2(u) \sum_{l\neq 0} (\tilde{\rho}_l)^4 \right] + \cdots. \quad (5.57)$$

SECTION 6

In order to obtain an approximation which is valid at any fixed temperature below as well as above the BWCP, we make use of the results obtained in Sec. 3 by expanding in terms of the deviations of the random fields from the Weiss field.

It is convenient to change from the random variables x_k defined by Eq. (3.2) (with $A \equiv \rho^{-1}$) to independent Gaussian random variables c_r specified by

$$\langle c_r \rangle_{\text{av c}} = 0 \quad (6.1)$$

and

$$\langle c_r c_{r'} \rangle_{\text{av c}} = \delta_{rr'}. \quad (6.2)$$

The variables x_k can then be expressed by

$$x_k = \sum_r c_r g_r^{\frac{1}{2}} u_{rk}, \quad (6.3)$$

where g_r are the eigenvalues and u_{rk} the normalized eigenvectors of ρ. The eigenvalues are given by Eq. (2.5); the eigenvectors are trigonometric functions. Note that u_{0k} is independent of k,

$$u_{0k} = n^{-\frac{1}{2}}, \quad (6.4)$$

and g_0 is given by

$$g_0 = \gamma^{-1} C_n(\gamma) \quad (6.5)$$

[Eqs. (2.5) and (2.7)].

The partition function Q_n, given by Eq. (3.5), then becomes

$$Q_n = 2^n e^{-n\nu\gamma/2} P_n, \quad (6.6)$$

with P_n defined by

$$P_n \equiv \left\langle \prod_k \cosh\left[x_k (\nu\gamma)^{\frac{1}{2}} \right] \right\rangle_{\text{av x}} = \left\langle \prod_k \cosh\left[\sum_r (\nu\gamma g_r)^{\frac{1}{2}} c_r u_{rk} \right] \right\rangle_{\text{av c}} \equiv \int \cdots \int \exp\left\{ -\tfrac{1}{2} \sum_r c_r^2 + \sum_k \ln \cosh\left[\sum_r (\nu\gamma g_r)^{\frac{1}{2}} c_r u_{rk} \right] \right\} \prod_r \frac{dc_r}{(2\pi)^{\frac{1}{2}}}. \quad (6.7)$$

We proved in Sec. 3 that the integrand assumes its largest value at two points, defined by Eqs. (3.14a), (3.14b), and (3.16), and only at these points, which in the variables c_τ are given by

$$c_\tau = 0, \quad \tau \neq 0, \tag{6.8a}$$

$$c_0 = \pm(n/\nu\gamma g_0)^{\frac{1}{2}} y, \tag{6.8b}$$

where y is the positive root of the equation

$$y = \nu C_n(\gamma) \tanh y. \tag{6.9}$$

The variable c_0 is thereby singled out. In order to carry out the expansion, we rewrite (6.7) in the form

$$\begin{aligned} P_n = &\int_{c_0>0} \exp\left\{-\tfrac{1}{2}\sum_\tau c_\tau^2 \right. \\ &+ \sum_k \ln\cosh\left(y + \sum_\tau [(\nu\gamma g_\tau)^{\frac{1}{2}} c_\tau \right. \\ &\left.\left. - n^{\frac{1}{2}} y\delta_{\tau 0}]u_{\tau k}\right)\right\} \prod_\tau \frac{dc_\tau}{(2\pi)^{\frac{1}{2}}} \\ &+ \int_{c_0<0} \exp\left\{-\tfrac{1}{2}\sum_\tau c_\tau^2 \right. \\ &+ \sum_k \ln\cosh\left(-y + \sum_\tau [(\nu\gamma g_\tau)^{\frac{1}{2}} c_\tau \right. \\ &\left.\left. + n^{\frac{1}{2}} y\delta_{\tau 0}]u_{\tau k}\right)\right\} \prod_\tau \frac{dc_\tau}{(2\pi)^{\frac{1}{2}}}. \end{aligned} \tag{6.10}$$

Changing the sign of all variables in the second integral, one obtains

$$\begin{aligned} \int_{c_0<0} = &\int_{c_0>0} \exp\left\{-\tfrac{1}{2}\sum_\tau c_\tau^2 \right. \\ &+ \sum_k \ln\cosh\left(-y - \sum_\tau [(\nu\gamma g_\tau)^{\frac{1}{2}} c_\tau \right. \\ &\left.\left. - n^{\frac{1}{2}} y\delta_{\tau 0}]u_{\tau k}\right)\right\} \prod_\tau \frac{dc_\tau}{(2\pi)^{\frac{1}{2}}}, \end{aligned} \tag{6.11}$$

which is equal to the first integral. We then change to new variables c'_τ by

$$c'_\tau = c_\tau - (n/\nu\gamma g_0)^{\frac{1}{2}} \delta_{\tau 0} y, \tag{6.12}$$

and obtain

$$\begin{aligned} P_n = 2\int_{-(n/\nu\gamma g_0)^{\frac{1}{2}} y}^{\infty} &\frac{dc'_0}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty}\cdots\int \prod_{\tau\neq 0} \frac{dc'_\tau}{(2\pi)^{\frac{1}{2}}} \\ &\times \exp\left\{-\tfrac{1}{2}[c'_0 + (n/\nu\gamma g_0)^{\frac{1}{2}} y]^2 - \tfrac{1}{2}\sum_{\tau\neq 0} c'^2_\tau \right. \\ &\left. + \sum_k \ln\cosh\left(y + \sum_\tau (\nu\gamma g_\tau)^{\frac{1}{2}} c'_\tau u_{\tau k}\right)\right\}. \end{aligned} \tag{6.13}$$

Expanding the exponent of the integrand to second order in the deviations from the Weiss field, we obtain (omitting the prime at the variables)

$$\begin{aligned} P_n \cong P_n^0 \equiv 2&\int_{c_0>-(n/\nu\gamma g_0)^{\frac{1}{2}} y} \prod_\tau \frac{dc_\tau}{(2\pi)^{\frac{1}{2}}} \\ &\times \exp\left\{-\tfrac{1}{2}(c_0 + (n/\nu\gamma g_0)^{\frac{1}{2}} y)^2 - \tfrac{1}{2}\sum_{\tau\neq 0} c_\tau^2 \right. \\ &+ \sum_k \left[\ln\cosh y + \tanh y \sum_\tau (\nu\gamma g_\tau)^{\frac{1}{2}} c_\tau u_{\tau k} \right. \\ &\left.\left. + \tfrac{1}{2}(1 - \tanh^2 y)\left(\sum_\tau (\nu\gamma g_\tau)^{\frac{1}{2}} c_\tau u_{\tau k}\right)^2\right]\right\} \end{aligned} \tag{6.14}$$

$$\begin{aligned} = 2&\int_{c_0>-(n/\nu\gamma g_0)^{\frac{1}{2}} y} \prod_\tau \frac{dc_\tau}{(2\pi)^{\frac{1}{2}}} \\ &\times \exp\left\{-\tfrac{1}{2}[c_0 + (n/\nu\gamma g_0)^{\frac{1}{2}} y]^2 - \tfrac{1}{2}\sum_{\tau\neq 0} c_\tau^2 \right. \\ &+ n\ln\cosh y + (n\nu\gamma g_0)^{\frac{1}{2}} c_0 \tanh y \\ &\left. + \tfrac{1}{2}(1 - \tanh^2 y)\sum_\tau \nu\gamma g_\tau c_\tau^2\right\}. \end{aligned} \tag{6.15}$$

Since

$$(\nu\gamma g_0)^{-\frac{1}{2}} y = (\nu\gamma g_0)^{\frac{1}{2}} \tanh y, \tag{6.16}$$

according to Eqs. (6.9) and (2.22), the terms linear in c_0 cancel, and

$$\begin{aligned} P_n^0 = 2&\exp\left\{-n\left[\frac{y^2}{2\nu C_n(\gamma)} - \ln\cosh y\right]\right\} \\ &\times \int_{c_0>-(n/\nu\gamma g_0)^{\frac{1}{2}} y} \prod_\tau \frac{dc_\tau}{(2\pi)^{\frac{1}{2}}} \\ &\times \exp\left\{-\tfrac{1}{2}\sum_\tau c_\tau^2[1 - \nu\gamma g_\tau(1 - \tanh^2 y)]\right\}, \end{aligned} \tag{6.17}$$

and

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \ln P_n^0 &= -\frac{y^2}{2\nu C(\gamma)} + \ln\cosh y - \tfrac{1}{2}\lim_{n\to\infty}\frac{1}{n} \\ &\times \sum_{\tau\neq 0} \ln[1 - \nu\gamma g_\tau(1 - \tanh^2 y)] \\ &+ \tfrac{1}{2}\lim_{n\to\infty}\frac{1}{n}\int_{-(n/\nu\gamma g_0)^{\frac{1}{2}} y}^{\infty} \frac{dc_0}{(2\pi)^{\frac{1}{2}}} \\ &\times \exp\{-c_0^2[1 - \nu C_n(\gamma)(1 - \tanh^2 y)]\} \\ &= -\frac{y^2}{2\nu C(\gamma)} + \ln\cosh y - \tfrac{1}{2}(2\pi)^{-D} \\ &\times \int_0^{2\pi}\cdots\int d^D\omega \ln[1 - \nu\gamma g(\omega)(1 - \tanh^2 y)], \end{aligned} \tag{6.18}$$

since the last term vanishes for $y > 0$.

For the free energy at temperatures below the BWCP we thus obtain

$$\psi \cong -kT\Big\{\ln 2 - \frac{y^2}{2\nu C(\gamma)} + \ln\cosh y - \tfrac{1}{2}\nu\gamma - \tfrac{1}{2}(2\pi)^{-D} \times \int_0^{2\pi}\cdots\int d^D\omega \ln\left[1 - \nu\gamma g(\omega)(1 - \tanh^2 y)\right]\Big\}. \tag{6.19}$$

Comparison of this result with Eq. (4.27) shows that in the limit $y \to 0$, i.e., $\nu C(\gamma) \to 1$, the free energy approaches the same value from both sides of the BWCP, since

$$-(2\pi)^{-D}\int_0^{2\pi}\cdots\int d^D\omega \ln\left[1 - \nu\gamma g(\omega)(1 - \tanh^2 y)\right] \equiv \int_0^{\nu\gamma(1-\tanh^2 y)} R(\xi)\, d\xi, \tag{6.20}$$

with $R(\xi)$ defined by Eq. (4.25).

The energy per particle, $\bar{\epsilon}$, is, in this approximation,

$$\bar{\epsilon} = J\frac{d}{d\nu}\left(\frac{\psi}{kT}\right) = -J\frac{d}{d\nu}\Big\{-\frac{y^2}{2\nu C(\gamma)} + \ln\cosh y - \tfrac{1}{2}\nu\gamma + \tfrac{1}{2}\int_0^{\nu\gamma(1-\tanh^2 y)} R(\xi)\, d\xi\Big\}, \tag{6.21}$$

where y is to be considered a function of ν. Since $dy/d\nu$ becomes singular for $y \to 0^+$ but $d(y^2)/d\nu$ remains finite, we use the latter and obtain

$$\bar{\epsilon} = -J\Big\{\frac{y^2}{2\nu^2 C(\gamma)} + \frac{dy^2}{d\nu}\frac{1}{2y}\left(\tanh y - \frac{y}{\nu C(\gamma)}\right) - \frac{\gamma}{2} + \frac{\gamma}{2}\Big[1 - \tanh^2 y + \nu\frac{dy^2}{d\nu}\frac{1}{2y} \times \left(-\frac{2\tanh y}{\cosh^2 y}\right)\Big] R\left(\frac{\nu\gamma}{\cosh^2 y}\right)\Big\}. \tag{6.22}$$

The second term vanishes for $y \neq 0$, and we obtain

$$\bar{\epsilon} = -J\Big\{\frac{y^2}{2\nu^2 C(\gamma)} - \frac{\gamma}{2} + \frac{\gamma}{2}\Big[1 - \tanh^2 y - \nu\frac{dy^2}{d\nu}\frac{\tanh y}{y\cosh^2 y}\Big] R\left(\frac{\nu\gamma}{\cosh^2 y}\right)\Big\}. \tag{6.23}$$

We had found that $R(\nu\gamma)$ becomes infinite at the BW Curie point for the one- and two-dimensional models if $g(\omega)$ is given by Eq. (4.30) for small ω, but remains finite for the three-dimensional model. Even in that case, however, the energy does not approach the same limit when the temperature approaches the BWCP from above and below. To show this, we solve Eq. (6.9) for small positive y and obtain

$$\frac{1}{\nu C(\gamma)} = \frac{\tanh y}{y} \cong 1 - \frac{y^2}{3}, \tag{6.24}$$

or

$$y^2 \cong 3\left(1 - \frac{1}{\nu C(\gamma)}\right), \tag{6.25}$$

and

$$d\frac{y^2}{d\nu} \cong \frac{3}{\nu^2 C(\gamma)}. \tag{6.26}$$

We thus have, with ν_c defined by $\nu_c \equiv C^{-1}(\gamma)$,

$$\bar{\epsilon}_{(\nu=\nu_c+0)} = -J\Big\{-\frac{\gamma}{2} - \gamma R(\nu_c\gamma)\Big\}, \tag{6.27}$$

while for temperatures approaching the BWCP from above we have, from Eq. (4.27), to first order in γ

$$\bar{\epsilon}_{(\nu=\nu_c-0)} = -J\Big\{-\frac{\gamma}{2} + \frac{\gamma}{2}R(\nu_c\gamma)\Big\}. \tag{6.28}$$

This discrepancy does not invalidate the results, but shows that the conditions stated for the validity of the expansion in γ are necessary. Under these conditions it is not permissible to approach the BWCP for any fixed γ, however small. Above the BWCP this can be seen also from the fact that, even for the three-dimensional model, the γ^3 terms in the energy derived from Eq. (4.27) became infinite at the BWCP, if $g(\omega)$ has the form (4.30).

SECTION 7

We have seen that after a few terms in the straightforward γ expansion of the free energy, we obtain terms which become infinite at the BWCP. This is clearly a failure of the approximation and does not indicate a phase transition. In this section we show how to avoid this by using a variant of the partition function

$$Q_n = \sum_{\{\mu\}} \exp\left(\frac{\nu\gamma}{2}\sum_{k,l}{}' \rho_{kl}\mu_k\mu_l\right). \tag{7.1}$$

Since $\mu_k^2 = 1$ and $\rho_{kk} = \rho_0 = 1$ by assumption, we may write (7.1) in the form

$$Q_n = \exp\left[\frac{n}{2}(\alpha - \nu\gamma)\right] \times \sum_{\{\mu\}} \exp\left[-\tfrac{1}{2}\sum_{k,l}(\alpha\delta_{kl} - \nu\gamma\rho_{kl})\mu_k\mu_l\right], \tag{7.2}$$

which is clearly independent of the choice of the parameter α. In order to apply the techniques of Sec. 3 to (7.2), it is necessary that the matrix $\alpha I - \nu\gamma\rho$

(with matrix elements $\alpha\delta_{kl} - \nu\gamma\rho_{kl}$) be positive-definite. This requires

$$\alpha > \nu\gamma g_0, \tag{7.3}$$

where g_0 is the zeroth Fourier coefficient of ρ_k. Proceeding as in Sec. 3, from (7.2) we obtain

$$Q_n = 2^n(2\pi)^{-n/2}\exp\left[\frac{n}{2}(\alpha - \nu\gamma)\right]\{\det(\alpha I - \nu\gamma\rho)\}^{-\frac{1}{2}} \times \int_{-\infty}^{\infty}\cdots\int \exp\left\{-\tfrac{1}{2}\sum_{k,l}[(\alpha I - \nu\gamma\rho)^{-1}]_{kl}x_k x_l\right\} \times \prod_k \cos x_k\, dx_k. \tag{7.4}$$

Defining a matrix $\tilde{\rho}(t)$ by the relation

$$(I - t\rho)^{-1} \equiv (I + t\tilde{\rho}), \tag{7.5}$$

and setting $x_k = u\xi_k(\alpha)^{\frac{1}{2}}$ with $u \equiv [1 + (\nu\gamma/\alpha)\tilde{\rho}_0]^{\frac{1}{2}}$, we obtain

$$Q_n = 2^n(2\pi)^{-n/2}u^n \times \exp\left[\frac{n}{2}(\alpha - \nu\gamma)\right]\left\{\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right)\right\}^{-\frac{1}{2}} \times \int_{-\infty}^{\infty}\cdots\int \exp\left(-\tfrac{1}{2}\sum_k \xi_k^2\right) \times \exp\left(-\frac{1}{2}\frac{\nu\gamma u^2}{\alpha}\sum_{k,l}{}' \tilde{\rho}_{kl}\xi_k\xi_l\right) \times \prod_k \cos\left[u\xi_k(\alpha)^{\frac{1}{2}}\right] d\xi_k. \tag{7.6}$$

Here $\tilde{\rho}_0$ stands for $\tilde{\rho}_{kk}(\nu\gamma/\alpha)$ which is independent of k. Finally, defining an average $\langle\ \rangle_{\text{av }\xi}$ over independent Gaussian random variables ξ_k with mean zero and mean square unity, i.e.,

$$\langle f\rangle_{\text{av }\xi} \equiv (2\pi)^{-n/2}\int_{-\infty}^{\infty}\cdots\int \exp\left(-\tfrac{1}{2}\sum_k \xi_k^2\right) f \prod_k d\xi_k, \tag{7.7}$$

we obtain

$$Q_n = (2u)^n e^{n(\alpha - \nu\gamma - u^2\alpha)/2}\left\{\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right)\right\}^{-\frac{1}{2}} h_n, \tag{7.8}$$

where

$$h_n = \left\langle \exp\left(-\frac{\nu\gamma u^2}{2\alpha}\sum_{k,l}{}' \tilde{\rho}_{kl}\xi_k\xi_l\right)\prod_k \cos\left[u\xi_k\alpha^{\frac{1}{2}}\right]\right\rangle_{\text{av }\xi} \times \langle\cos\left[u\xi_k\alpha^{\frac{1}{2}}\right]\rangle_{\text{av }\xi}^{-n}. \tag{7.9}$$

The free energy per spin, ψ, is now given by

$$-\beta\psi = \mathcal{F}_1 + \frac{1}{n}\ln h_n, \tag{7.10}$$

with

$$\mathcal{F}_1 = \ln 2 + \ln u + \tfrac{1}{2}(\alpha - \nu\gamma - u^2\alpha) - \frac{1}{2n}\ln\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right). \tag{7.11}$$

The right-hand side of (7.10) must be independent of the choice of α. However, it is possible that a propitious choice for α may result in the first few terms in the expansion (7.10) being a good approximation to the true free energy. To that end, we expand the exponential in h_n and obtain

$$h_n = 1 + \frac{1}{2}\left(\frac{\nu\gamma u^2}{2\alpha}\right)^2\sum_{i,j}{}'\sum_{k,l}{}' \tilde{\rho}_{ij}\tilde{\rho}_{kl} \times \left\langle \xi_i\xi_j\xi_k\xi_l\prod_s \cos\left[u\xi_s\alpha^{\frac{1}{2}}\right]\right\rangle_{\text{av }\xi} \times \langle\cos\left[u\xi_k\alpha^{\frac{1}{2}}\right]\rangle_{\text{av }\xi}^{-1} + \cdots, \tag{7.12}$$

or

$$h_n = 1 + \left(\frac{\nu\gamma u^2}{2\alpha}\right)^2(1 - u^2\alpha)\sum_{k,l}{}'(\tilde{\rho}_{kl})^2 + \cdots. \tag{7.13}$$

In obtaining these results, we have used the fact that

$$\langle\xi^f\cos\left[u\xi\alpha^{\frac{1}{2}}\right]\rangle_{\text{av }\xi} = 0, \quad \text{for odd } f. \tag{7.14}$$

The second term in (7.13) may be eliminated by choosing α to satisfy the equation

$$1 - u^2\alpha = 0. \tag{7.15}$$

Rewriting this equation, we obtain

$$\alpha = \left[1 + \frac{\nu\gamma}{\alpha}\tilde{\rho}_0\left(\frac{\nu\gamma}{\alpha}\right)\right] = \left[\left(I - \frac{\nu\gamma}{\alpha}\rho\right)^{-1}\right]_0. \tag{7.16}$$

With this choice of α, the next term in the expansion (7.13) also vanishes. In fact, all diagrams—with $\tilde{\rho}_{kl}$ considered as a line between vertices labeled k and l—which have at least one vertex of order less than four, vanish for this choice of α.

An alternative approach which suggests that (7.15) is a reasonable criterion is the following. The free energy ψ given by (7.10) is independent of α. We can choose α so that the approximation given by (7.11) is locally independent of α, i.e., set $d\mathcal{F}_1/d\alpha = 0$. This yields

$$\frac{d\mathcal{F}_1}{d\alpha} = (1 - u^2\alpha)\left\{\frac{1}{u}\frac{\partial u}{\partial\alpha} + \frac{1}{2\alpha} - \frac{1}{2\alpha u^2}\right\} = 0, \tag{7.17}$$

which is consistent with (7.15). In obtaining (7.17) we have used the relation

$$\ln\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right) = -n\int_0^{\nu\gamma/\alpha}\tilde{\rho}_0(t)\,dt, \tag{7.18}$$

according to (4.20) with ν replaced by ν/α.

Still a third criterion which also suggests (7.15) as the appropriate choice for α is the following. In an external magnetic field $H(r)$ the partition function

(7.1) becomes

$$Q_n = \sum_{\{\mu\}} \exp\left(\frac{\nu\gamma}{2}\sum_{k,l}{}' \rho_{kl}\mu_k\mu_l + \beta\sum_k H_k\mu_k\right), \quad (7.19)$$

where H_k is the value of the external field at site k (in appropriate units). The two-spin correlation function $\langle\mu_r\mu_s\rangle$ may be obtained from (7.19) in two distinct ways: either by differentiating twice with respect to external fields or once with respect to the interaction matrix element ρ_{rs}; that is,

$$\langle\mu_r\mu_s\rangle = \frac{1}{\beta}\frac{\partial\langle\mu_r\rangle}{\partial H_s} + \langle\mu_r\rangle\langle\mu_s\rangle$$
$$= \frac{1}{\beta^2}\left\{\frac{\partial^2 \ln Q_n}{\partial H_r \partial H_s} + \frac{\partial \ln Q_n}{\partial H_r}\frac{\partial \ln Q_n}{\partial H_s}\right\}, \quad (7.20)$$

or

$$\langle\mu_r\mu_s\rangle = \frac{1}{\nu\gamma}\frac{\partial \ln Q_n}{\partial \rho_{rs}}. \quad (7.21)$$

Therefore, in any exact calculation, the thermodynamic identity

$$\frac{1}{\nu\gamma}\frac{\partial \ln Q_n}{\partial \rho_{rs}} = \frac{1}{\beta^2}\left\{\frac{\partial^2 \ln Q_n}{\partial H_r \partial H_s} + \frac{\partial \ln Q_n}{\partial H_r}\frac{\partial \ln Q_n}{\partial H_s}\right\} \quad (7.22)$$

must obtain. Transforming (7.19) into an integral and summing over all $\mu_k = \pm 1$ results in

$$Q_n = (2u)^n \exp\left(\frac{n}{2}(\alpha - \nu\gamma)\right)$$
$$\times \left\{\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right)\right\}^{-\frac{1}{2}}(2\pi)^{-n/2}$$
$$\times \int_{-\infty}^{\infty}\cdots\int \exp\left(-\tfrac{1}{2}\sum_k \xi_k^2 - \frac{u^2\nu\gamma}{2\alpha}\sum_{k,l}{}' \tilde{\rho}_{kl}\xi_k\xi_l\right)$$
$$\times \prod_k \cos\left[u\xi_k\alpha^{\frac{1}{2}} - i\beta H_k\right] d\xi_k. \quad (7.23)$$

Defining $\langle\ \rangle_{\mathrm{av}\,\xi}$ by (7.7), (7.23) becomes

$$Q_n = (2u)^n e^{n(\alpha-\nu\gamma-u^2\alpha)/2}\left\{\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right)\right\}^{-\frac{1}{2}}$$
$$\times \left[\prod_k \cosh(\beta H_k)\right]\tilde{h}_n, \quad (7.24)$$

where

$$\tilde{h}_n \equiv \left\langle \exp\left(-\frac{u^2\nu\gamma}{2\alpha}\sum_{k,l}{}' \tilde{\rho}_{kl}\xi_k\xi_l\right)\prod_k \varphi(\xi_k, H_k)\right\rangle_{\mathrm{av}\,\xi} \quad (7.25)$$

and

$$\varphi(\xi_k, H_k) = \frac{\cos\left[u\xi_k\alpha^{\frac{1}{2}} - i\beta H_k\right]}{\langle\cos\left[u\xi_k\alpha^{\frac{1}{2}} - i\beta H_k\right]\rangle_{\mathrm{av}\,\xi}}$$
$$= \frac{e^{u^2\alpha/2}\cos\left[u\xi_k\alpha^{\frac{1}{2}} - i\beta H_k\right]}{\cosh(\beta H_k)}. \quad (7.26)$$

The free energy per spin is now given by

$$-\beta\psi = \frac{1}{n}\ln Q_n = -\beta\psi^{(0)} + \frac{1}{n}\ln\tilde{h}_n, \quad (7.27)$$

with

$$-\beta\psi^{(0)} = \ln(2u) + \tfrac{1}{2}(\alpha - \nu\gamma - u^2\alpha)$$
$$- \frac{1}{2n}\ln\det\left(I - \frac{\nu\gamma}{\alpha}\rho\right)$$
$$+ \frac{1}{n}\sum_k \ln\cosh(\beta H_k). \quad (7.28)$$

We may expand the last term in (7.27) by the technique of Sec. 5; we obtain

$$\frac{1}{n}\ln\tilde{h}_n = \sum_{\lambda=1}^{\infty}\left(-\frac{\nu\gamma u^2}{\alpha}\right)^{\lambda} b_\lambda \equiv -\beta\sum_{\lambda=1}^{\infty}\psi^{(\lambda)}, \quad (7.29)$$

where the b_λ are given by (5.13) except that they now contain an additional dependence on the external fields H_k. Explicitly, $\psi^{(0)}$ is given by (7.28) and

$$-\beta\psi^{(1)} = \frac{\nu\gamma u^4}{2n}\sum_{i,j}{}' \tilde{\rho}_{ij}\tanh(\beta H_i)\tanh(\beta H_j), \quad (7.30)$$

and

$$-\beta\psi^{(2)} = \left(\frac{\nu\gamma u^2}{\alpha}\right)^2\frac{1}{2n}\left\{\tfrac{1}{2}\sum_{i,j}{}' [(u^2\alpha - 1)^2\right.$$
$$- u^4\alpha^2\tanh^2(\beta H_i)\tanh^2(\beta H_j)](\tilde{\rho}_{ij})^2$$
$$+ u^2\alpha\sum_{\substack{i\neq j,k\\ j\neq k}}\tanh(\beta H_i)\tanh(\beta H_k)$$
$$\left.\times [(u^2\alpha - 1) - u^2\alpha\tanh^2(\beta H_j)]\tilde{\rho}_{ij}\tilde{\rho}_{jk}\right\}. \quad (7.31)$$

To get an approximate expression for $\langle\mu_r\mu_s\rangle$, we consider terms up to and including $\psi^{(2)}$ in (7.27). After taking the derivatives indicated in (7.20) and (7.21), we set all $H_k = 0$, so that our results are for the case of zero external field. In this manner, we obtain from (7.20)

$$\langle\mu_r\mu_s\rangle = (\nu\gamma u^4)\{\tilde{\rho}_{rs} + 2(u^2\alpha - 1)[(\tilde{\rho}^2)_{rs} - 2\tilde{\rho}_0\tilde{\rho}_{rs}]\}, \quad (7.32)$$

and from (7.21)

$$\langle\mu_r\mu_s\rangle = \frac{\nu\gamma}{\alpha^2}\tilde{\rho}_{rs} + (1 - u^2\alpha)\frac{u^2}{\alpha}$$
$$\times \left\{\left[I + \frac{\nu\gamma}{\alpha}\tilde{\rho}\right]^2\left[I - \frac{\nu\gamma u^2}{\alpha}(\tilde{\rho} - \tilde{\rho}_0 I)\right.\right.$$
$$\left.\left. + \left(\frac{\nu\gamma u^2}{\alpha}\right)^2(1 - 2u^2\alpha)\sum_{l\neq 0}(\tilde{\rho}_l)^2 I\right]\right\}_{rs}. \quad (7.33)$$

We note that these two expressions are not in general

equal, so that the identity (7.22) is violated. However, the identity can be preserved (at least to this order) if we choose α to satisfy (7.15), since both (7.32) and (7.33) then become

$$\langle \mu_r \mu_s \rangle = (\nu\gamma u^4)\tilde{\rho}_{rs}. \tag{7.34}$$

Using this approximation, we obtain for the energy per spin

$$\bar{\epsilon} \cong -\left(\frac{\nu\gamma}{\alpha}\right)^2 \frac{1}{2n\beta} \sum_{r,s}{}' \rho_{rs}\tilde{\rho}_{rs} = -\frac{1}{2\beta}(\alpha - 1 - \nu\gamma), \tag{7.35}$$

and for the free energy per spin

$$\begin{aligned} -\beta\psi &\cong \ln 2u + \tfrac{1}{2}(\alpha - 1 - \nu\gamma) \\ &\quad - \frac{1}{2n} \ln \det \left(I - \frac{\nu\gamma}{\alpha}\rho\right) \\ &= \ln 2u + \tfrac{1}{2}(\alpha - 1 - \nu\gamma) + \tfrac{1}{2}\int_0^{\nu\gamma/\alpha} \tilde{\rho}_0(t)\, dt. \end{aligned} \tag{7.36}$$

These expressions are well behaved at the BWCP, so that our α formalism has removed the singularities which arise there in the straight γ expansion.

SECTION 8

The results of the previous section are closely related to previous work of Brout,[1] and Mühlschlegel and Zittartz.[5] Since our notation differs from that of these authors, we give here a comparison. Brout uses the interaction Hamiltonian

$$-\mathcal{H} = \tfrac{1}{2} \sum_{i,j}{}' v_{ij}\mu_i\mu_j \tag{8.1}$$

and Fourier transform

$$v(\mathbf{q}) = \sum_j v_{ij} e^{i\mathbf{q}\cdot(\mathbf{i}-\mathbf{j})}, \tag{8.2}$$

where $\mathbf{i}$ and $\mathbf{j}$ are vectors with integer components, $\mathbf{q}$ is a reciprocal lattice vector, and v_{ii} is chosen to be zero. He then introduces a parameter δ determined by

$$1 = \frac{1}{n} \sum_{\mathbf{q}} \frac{1}{1 + \beta\delta - \beta v(\mathbf{q})}. \tag{8.3}$$

Mühlschlegel and Zittartz use

$$-\mathcal{H} = \tfrac{1}{2} \sum_{i,j}{}' V_{ij}\sigma_i\sigma_j, \tag{8.4}$$

$$v(\mathbf{p}) = \sum_j V_{ij} e^{-i\mathbf{p}\cdot(\mathbf{r}_i - \mathbf{r}_j)}, \tag{8.5}$$

and a parameter μ determined by

$$1 = \frac{1}{n} \sum_{\mathbf{p}} \frac{1}{1 - 2\beta\mu - \beta v(\mathbf{p})}. \tag{8.6}$$

The corresponding formulas in the present paper are

$$-\mathcal{H} = \frac{J\gamma}{2} \sum_{i,j}{}' \rho_{ij}\mu_i\mu_j, \tag{8.7}$$

$$g_r = \sum_j \rho_{ij} e^{-2\pi i r\cdot(i-j)/m}, \tag{8.8}$$

and

$$1 = \frac{1}{n} \sum_r \frac{1}{\alpha - \nu\gamma g_r}, \tag{8.9}$$

which may be obtained from (7.16) and (2.4).

Our formulas may thus be transferred into Brout's notation by the replacements

$$J\gamma\rho_{ij} \to v_{ij} \quad (i \neq j), \quad J\gamma(g_r - 1) \to v(\mathbf{q}),$$

and $(\alpha - 1 - \nu\gamma)/\beta \to \delta$; and into the notation of Mühlschlegel and Zittartz by $\mu_i \to \sigma_i$, $J\gamma\rho_{ij} \to v_{ij}$, $J\gamma g_r \to v(\mathbf{p})$, and $\alpha \to 1 - 2\beta\mu$. In these papers, the Curie point is determined from the equation

$$1 + \beta_c\delta - \beta_c v(0) = 0 \tag{8.10}$$

(in Ref. 1) and from

$$1 - 2\beta_c\mu - \beta_c v(0) = 0 \tag{8.11}$$

(in Ref. 5). In our notation, this would correspond to

$$\alpha - \nu_c\gamma g_0 = 0. \tag{8.12}$$

Now, we recall from Eq. (7.3) that the representation we have employed requires $\alpha > \nu\gamma g_0$, and is therefore invalid at the temperature required by (8.12). The anomalies which occur at T_c as determined from (8.10)–(8.12) are therefore spurious in the sense that they arise from the representation employed and the choice of the parameter α. It could, of course, happen that the anomalies which result from our representation still reflect a phase transition of the model, in which case we may hope that the T_c determined from (8.10)–(8.12) is close to the true T_c for the model. This hope is certainly unjustified for the case of the two-dimensional model, since in that case (8.10)–(8.12) show that the right-hand sides of (8.3), (8.6), and (8.9) become infinite while the left-hand sides of those equations remain finite, so that there is no solution T_c. To see this, writing (8.9) in the limit as $n \to \infty$, we obtain

$$1 = (2\pi)^{-D} \int_0^{2\pi} \cdots \int \frac{d^D\omega}{\alpha - \nu\gamma g(\omega)}, \tag{8.13}$$

where D is the dimensionality of the model. Assuming, as usual, that $g(\omega) \cong g(0) - A\omega^2$ for small ω, and inserting the value of α given by (8.12), we obtain

$$1 \cong (2\pi)^{-D} \int_0^{2\pi} \cdots \int \frac{d^D\omega}{\nu\gamma A\omega^2}. \tag{8.14}$$

For $D = 1, 2$, this integral diverges logarithmically or worse at the lower limit, while for $D = 3$ it remains finite. We conclude, then, that in one or two dimensions the equation determining the "Curie point" has no solution, so that the singularities in the approximation are not related to a phase transition of the model. In the one-dimensional case, where there should not be a solution T_c, we show the failure of the approximation by the other means (Appendix D). Thus, while we cannot exclude the possibility that in three dimensions the failure of the approximation is linked to the phase transition, the fact that this same approximation fails in the case of one and two dimensions makes this very doubtful.

SECTION 9

This section is a summary of our results and a comparison with previous work.

The following results have been obtained for the free energy per spin, ψ, for the model described in Sec. 2. For any fixed temperature above the BWCP, we have found (in Secs. 4 and 5) that

$$\psi = -kT\{\ln 2 + f_1 + f_2 + f_3 + ((\gamma^4))\}, \quad (9.1)$$

where $((\gamma^4))$ denotes an infinite series of terms of order γ^4 and higher order, and f_1, f_2, and f_3 are of order γ, γ^2, and γ^3, respectively, and are given by

$$f_1 = \frac{1}{2}\left[\int_0^{\nu\gamma} R(\xi)\, d\xi - \nu\gamma\right], \quad (9.2)$$

$$f_2 = -\frac{1}{2}\left[\ln(1 + \nu\gamma R(\nu\gamma)) - \frac{\nu\gamma R(\nu\gamma)}{1 + \nu\gamma R(\nu\gamma)}\right], \quad (9.3)$$

and

$$f_3 = (\nu\gamma)^4[\tfrac{1}{4}R^2(\nu\gamma)R_{22}(\nu\gamma) + \tfrac{1}{12}R_{42}(\nu\gamma)]. \quad (9.4)$$

Here

$$R(\xi) \equiv \lim_{n\to\infty} \tilde{\rho}_0(\xi) = \frac{1}{(2\pi)^D}\int_0^{2\pi}\cdots\int d^D\omega\, \frac{g(\omega)}{1 - \xi g(\omega)}, \quad (9.5)$$

$$R_{22}(\xi) \equiv \lim_{n\to\infty} (\tilde{\rho}(\xi)^2)_0 = \frac{1}{(2\pi)^D}\int_0^{2\pi}\cdots\int d^D\omega\left[\frac{g(\omega)}{1 - \xi g(\omega)}\right]^2, \quad (9.6)$$

and

$$R_{42}(\xi) \equiv \lim_{n\to\infty}\sum_l [\tilde{\rho}_l(\xi)]^4 = \frac{1}{(2\pi)^{3D}}\int_0^{2\pi}\cdots\int d^D\omega_1\, d^D\omega_2\, d^D\omega_3 \times \prod_{j=1}^{3}\frac{g(\omega_j)}{1 - \xi g(\omega_j)}\,\frac{g(\omega_1 + \omega_2 + \omega_3)}{1 - \xi g(\omega_1 + \omega_2 + \omega_3)}, \quad (9.7)$$

where

$$g(\omega) = \sum_k^{(\infty)} e^{ik\cdot\omega}\rho_k, \quad (9.8)$$

and the matrix $\tilde{\rho}(\xi)$ is obtained from the matrix of the interaction, ρ, by

$$\tilde{\rho} \equiv \rho(I - \xi\rho)^{-1}, \quad (9.9)$$

and $\tilde{\rho}_0$ is $\tilde{\rho}_{kk}(\xi)$. The term f_3 is of order γ^3, not γ^4, since $\gamma R_{22}(\nu\gamma)$ and $\gamma R_{42}(\nu\gamma)$ are bounded for $\gamma \to 0$ for any fixed temperature above the BWCP. The terms f_1 and f_2 have been obtained without diagram summation by extracting factors from a representation of the partition function as an average over Gaussian random variables. The remaining terms $f_3 + ((\gamma^4))$ are denoted by $\lim_{n\to\infty}(1/n)\ln q_n^{(2)}$, and we have obtained a series for them [see Eq. (5.14) and the definitions (4.10), (5.7)–(5.9), (5.16), (5.18), (5.19)]. We have proved that the limit as $n \to \infty$ of each term in this series is of order γ^3 or higher, and we have a heuristic argument for the order of each term in this series which shows that all terms of order γ^n are contained in the first $(2n - 3)$ terms of the series. The term f_3 given above is obtained from the first and third term of the series; the second term is of higher order.

The first-order term of (9.1) agrees with that derived in Ref. 1, which gives explicitly only the terms of order γ. Stell *et al*[20] have checked our Eq. (9.1) through order γ^2 by a different method and report agreement. Baker[9] and Kac[8] have given the exact free energy for the one-dimensional model with exponential interaction which has a phase transition in the limit $\gamma \to 0$. Kac obtained the free energy per spin as the largest eigenvalue of a certain integral equation and gives explicitly the terms through order γ^2 above the BWCP. We have specialized our results to this case in Appendix C and show agreement.

If the second moments of ρ_k exist and ρ_k satisfies certain symmetry conditions, $g(\omega)$ has the form

$$g(\omega) \cong g(0) - A\omega^2, \quad (9.10)$$

for small ω. The analytic form of $R(\nu\gamma)$ near the BWCP is then the same as the analytic form of similar integrals occurring in the theory of the spherical model.[21] Specifically, $R(\nu\gamma)$ becomes infinite for the one- and two-dimensional models, but remains finite for the three-dimensional model.

If $g(\omega)$ has the form (9.10), the function $R_{22}(\nu\gamma)$ becomes infinite at the BWCP even for the three-dimensional model. Since both R_{22} and R_{42} are positive

[20] G. Stell *et al.*, J. Math. Phys. **7**, 1532 (1966).

[21] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952); see also Newell and Montroll, Ref. 18.

above the BWCP, there can be no cancellation, and f_3 becomes infinite at the BWCP. However, our assumptions do not exclude interactions which do not have a second moment. If ρ_k does not have a second moment, then $g(\omega)$ can be of the form

$$g(\omega) \cong g(0) - B\,|\omega| \tag{9.11}$$

for small ω. In this case, $R(\nu\gamma)$ remains finite for the two-dimensional model.

For temperatures below the BWCP, we obtained, in Sec. 6, the free energy per spin in the form

$$\psi \cong -kT\Big\{\ln 2 - \frac{y}{2\nu C(\gamma)} + \ln\cosh y - \tfrac{1}{2}\nu\gamma - \frac{1}{2}\frac{1}{(2\pi)^D}\int_0^{2\pi}\cdots\int d^D\omega \times \ln\left[1 - \nu\gamma g(\omega)(1 - \tanh^2 y)\right]\Big\}, \tag{9.12}$$

with y the positive solution of the Weiss field equation

$$y = \nu C(\gamma)\tanh y, \tag{9.13}$$

and

$$C(\gamma) = \gamma g(0). \tag{9.14}$$

Equation (9.12) gives ψ through order γ below the BWCP and is in agreement with the calculation of Brout.[1] It also agrees with the result of Kac[8] and Baker[9] when it is specialized to a one-dimensional system with exponential interaction. For temperatures above the BWCP, Eq. (9.13) has only the root $y = 0$ and then (9.12) agrees with (9.1) through order γ.

In Secs. 7 and 8 we have introduced another variant of the partition function characterized by a parameter α which is chosen to satisfy the equation

$$1 = \frac{1}{n}\sum_r \frac{1}{\alpha - \nu\gamma g_r} \to \frac{1}{(2\pi)^D}\int_0^{2\pi}\cdots\int \frac{d^D\omega}{\alpha - \nu\gamma g(\omega)}. \tag{9.15}$$

This allows us to make contact with previous work of Brout[1] and Mühlschlegel and Zittartz,[5] and our results agree with those given by them. We are able to eliminate the singularities at the BWCP by this technique. We have also shown that the singularities which now occur are not related to a phase transition of the one- or two-dimensional model and are probably also spurious for the three-dimensional case. The main calculation of these sections yields the two-spin correlation function in the form

$$\langle\mu_r\mu_s\rangle \cong \frac{\nu\gamma}{\alpha^2}\,\tilde\rho_{rs}\left(\frac{\nu\gamma}{\alpha}\right), \tag{9.16}$$

with $\tilde\rho_{rs}$ defined by (9.9). In Appendix D we use (9.16) to obtain an explicit form for $\langle\mu_r\mu_s\rangle$ for the one-dimensional case with $\rho_{rs} \sim e^{-\gamma|r-s|}$, and compare our result with that of Kac and Helfand.[11] The result of their calculation is

$$\langle\mu_r\mu_s\rangle = \gamma^{\frac{2}{3}}\{a_1 e^{-b_1\gamma^{\frac{4}{3}}|r-s|} + a_2 e^{-b_2\gamma^{\frac{4}{3}}|r-s|} + \cdots\} + O(\gamma^{\frac{4}{3}}). \tag{9.17}$$

Kac and Helfand find $a_1 = 0.656$, $b_1 = 0.754$, $a_2 = 0.0019$, $b_2 = 2.91$. We have obtained $a_1 = 0.630$, $b_1 = 0.794$, $b_2 = 2.382$. We have not calculated a_2 exactly, but our estimates indicate that it is considerably smaller than the corresponding value of Kac and Helfand. Thus we see that this improvement of the method removes the most serious drawback of the straight γ expansion in which, after a few terms, the coefficients of the series become infinite at the BWCP. It does not, however, correctly yield the dominant terms ($\gamma^{\frac{2}{3}}$) in the spin correlation function, although the numerical values of the parameters in the term of longest range are very well approximated. Since the choice of α or the corresponding parameters of Brout and Mühlschlegel and Zittartz seems to be forced by the additional arguments in Sec. 7, it appears unlikely that a different choice of α will result in a substantially improved approximation at the BWCP.

ACKNOWLEDGMENTS

We would like to thank Dr. A. J. Fedro and Dr. K. W. Wong for checking some of the calculations. One of us (A. J. F. S.) would like to thank Professor Detlev Bronck and Professor M. Kac for the hospitality of the Rockefeller University, where this work was begun.

APPENDIX A

We prove here that no solution of (3.13), other than the Weiss field solution [(3.14a), (3.14b)], has components, all of which are of the same sign.

Introducing a more convenient notation by

$$y_k \equiv (\nu\gamma)^{\frac{1}{2}}x_k, \tag{A1}$$

we write Eq. (3.13) in the form

$$y_k = \nu\gamma\sum_l \rho_{kl}\tanh y_l. \tag{A2}$$

Summation over k yields

$$\sum_k y_k = \nu C_n(\gamma)\sum_l \tanh y_l, \tag{A3}$$

or

$$\sum_k (y_k - \nu C_n(\gamma)\tanh y_k) = 0. \tag{A4}$$

With

$$y \equiv (\nu\gamma)^{\frac{1}{2}}x, \tag{A5}$$

where x is defined by Eq. (3.16), we have

$$|y_k| < y \tag{A6}$$

for all solutions other than the Weiss field solution. With y_k in this interval and $\nu C_n(\gamma) > 1$,

$$y_k - \nu C_n(\gamma) \tanh y_k > 0, \quad \text{if} \quad y_k < 0, \tag{A7}$$

and vice versa. The components y_k of any solution. other than the Weiss field solution cannot, therefore, be all of the same sign.

APPENDIX B

In order to prove the inequality (5.32), we note that the matrix elements of $\tilde{\rho}$ are nonnegative according to Eq. (5.22). Furthermore, every even-connected graph with at least one point k of degree f, with $f > 2$, is contained in the class of graphs consisting of $f/2$ chains leaving the point k and returning to it, if neither the points of one chain nor the points of different chains are required to be distinct. The chains will be labeled by σ $(= 1, 2, \cdots, f/2)$ and the number of $\tilde{\rho}$ bonds in the chain σ by λ_σ. The total number of bonds in the chains must be the same as that in the graphs under consideration. We thus have

$$\sum_{p_1 \cdots p_\lambda}{}' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \sum_{\lambda_1, \cdots, \lambda_{f/2}}{}' \sum_{k} \prod_{\sigma=1}^{f/2} \left(\sum_{\{l\}} \tilde{\rho}_{kl_1} \tilde{\rho}_{l_1 l_2} \cdots \tilde{\rho}_{l'_\sigma k} \right), \tag{B1}$$

where $\sum'_{\lambda_1, \cdots, \lambda_{f/2}}$ is restricted by

$$\sum_{\sigma=1}^{f/2} \lambda_\sigma = \lambda, \tag{B2}$$

and $\sum_{\{l\}}$ means the sum over $l_1, l_2, \cdots, l'_\sigma \equiv l_{\lambda_\sigma - 1}$. The right-hand side of the inequality (B1) can be immediately evaluated in the form

$$\sum_{p_1 \cdots p_\lambda}{}' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \sum_{\lambda_1, \cdots, \lambda_{f/2}}{}' \sum_{k} \prod_{\sigma=1}^{f/2} (\tilde{\rho}^{\lambda_\sigma})_{kk} = n \sum_{\lambda_1, \cdots, \lambda_{f/2}}{}' \prod_{\sigma=1}^{f/2} (\tilde{\rho}^{\lambda_\sigma})_0, \tag{B3}$$

so that, according to (5.26),

$$\lim_{n \to \infty} \frac{\gamma^\lambda}{n} \sum_{p_1, \cdots, p_\lambda}{}' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \sum_{\lambda_1, \cdots, \lambda_{f/2}}{}' \gamma^\lambda \prod_{\sigma=1}^{f/2} \gamma^{1-\lambda_\sigma} M_{\lambda_\sigma} = \gamma^{f/2} M'_\lambda, \tag{B4}$$

where the quantities M_{λ_σ} and therefore M'_λ, also remain bounded for $\gamma \to 0$, since the sum over $\lambda_1, \lambda_2, \cdots, \lambda_{f/2}$ is a finite sum.

The sum $\sum''_{p_1, \cdots, p_\lambda}$ over graphs with at least one point of degree f and at least one point of degree h can be estimated by a similar procedure. There are at least two paths between the two points, and we have

$$\sum_{p_1, \cdots, p_\lambda}{}'' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \sum_{\lambda_1 \cdots \lambda_{f'}} \sum_{\lambda_1' \cdots \lambda_{h''}} \sum_{\lambda_a, \lambda_b} \sum_{k,l} \prod_{\sigma=1}^{f'} \left(\sum_{\{l\}} \tilde{\rho}_{kl_1} \cdots \tilde{\rho}_{l'_\sigma k} \right) \times \left(\sum_{\{j\}} \tilde{\rho}_{kj_1} \cdots \tilde{\rho}_{j_{a'}l} \right) \left(\sum_{\{i\}} \tilde{\rho}_{ki_1} \cdots \tilde{\rho}_{i_{b'}l} \right) \times \prod_{\tau=1}^{h'} \left(\sum_{\{t\}} \tilde{\rho}_{ll_1} \cdots \tilde{\rho}_{l_{\tau'}l} \right), \tag{B5}$$

where $f' = \frac{1}{2}(f-2)$, $h' = \frac{1}{2}(h-2)$, and l'_σ, means $l_{\lambda_\sigma - 1}$, j'_a means $j_{\lambda_a - 1}$, etc., and the sums over the chain lengths $\lambda_1 \cdots \lambda_{f'}$, $\lambda'_1 \cdots \lambda'_{f'}$, λ_a, λ_b, are restricted by

$$\sum_{\sigma=1}^{\frac{1}{2}(f-2)} \lambda_\sigma + \sum_{\tau=1}^{\frac{1}{2}(h-2)} \lambda'_\tau + \lambda_a + \lambda_b = \lambda. \tag{B6}$$

We obtain

$$\sum_{p_1 \cdots p_\lambda}{}'' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq n \sum_{\lambda_1 \cdots \lambda_{f'}} \sum_{\lambda_1' \cdots \lambda'_{f'}} \sum_{\lambda_a, \lambda_b} \prod_{\sigma=1}^{f'} \prod_{\tau=1}^{h'} (\tilde{\rho}^{\lambda_\sigma})_0 (\tilde{\rho}^{\lambda_a + \lambda_b})_0 (\tilde{\rho}^{\lambda'_\tau})_0, \tag{B7}$$

or

$$\lim_{n \to \infty} \frac{\gamma^\lambda}{n} \sum_{p_1 \cdots p_\lambda}{}'' \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \leq \sum_{\lambda_1 \cdots \lambda_{f'}} \sum_{\lambda_1' \cdots \lambda'_{f'}} \sum_{\lambda_a, \lambda_b} \gamma^\lambda \prod_{\sigma=1}^{f'} \gamma^{1-\lambda_\sigma} M_{\lambda_\sigma} \times \prod_{\tau=1}^{h'} \gamma^{1-\lambda'_\tau} M_{\lambda_\tau} \gamma^{1-\lambda_a-\lambda_b} M_{\lambda_a+\lambda_b} \leq \gamma^{-1+(f/2)+(h/2)} M''_\lambda, \tag{B8}$$

where M''_λ remains bounded for $\gamma \to 0$.

APPENDIX C

In this appendix we specialize the results of Secs. 4 and 6 to the case of a one-dimensional model with exponential interaction defined by

$$\rho_k(\gamma) = e^{-\gamma|k|}, \tag{C1}$$

and compare the results with exact calculations of Baker[9] and Kac.[8] We limit our computations here to terms through order γ^2 above the BWCP and order γ below the BWCP since these are the only results explicitly given in this prior work.

For the interaction given in (C1), we have

$$g(\omega) \equiv \sum_{k=-\infty}^{\infty} e^{ik\omega - \gamma|k|} = \frac{\sinh \gamma}{\cosh \gamma - \cos \omega}. \tag{C2}$$

The function $R(\xi)$ is obtained most conveniently from the Fourier coefficients $R_k(\xi)$ of $g(\omega)/(1 - \xi g(\omega))$:

$$R_k(\xi) \equiv \frac{1}{2\pi} \int_0^{2\pi} \frac{g(\omega)}{1 - \xi g(\omega)} e^{-ik\omega} \, d\omega, \tag{C3}$$

so that

$$R(\xi) = R_0(\xi). \tag{C4}$$

Explicitly, with $g(\omega)$ given by Eq. (C2), we have

$$R_k(\xi) = \frac{1}{2\pi}\int_0^{2\pi} \frac{\sinh\gamma e^{-ik\omega}}{\cosh\gamma - \xi\sinh\gamma - \cos\omega}\, d\omega. \tag{C5}$$

We write this in the form

$$R_k(\xi) = \frac{\sinh\gamma}{\sinh\tilde{\gamma}}\frac{1}{2\pi}\int_0^{2\pi} \frac{\sinh\tilde{\gamma}e^{-ik\omega}}{\cosh\tilde{\gamma} - \cos\omega}\, d\omega, \tag{C6}$$

with $\tilde{\gamma}$ defined by

$$\cosh\tilde{\gamma} \equiv \cosh\gamma - \xi\sinh\gamma, \tag{C7}$$

and

$$\tilde{\gamma} \geq 0. \tag{C8}$$

Since Eqs. (C1) and (C2) imply that

$$e^{-\gamma|k|} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\sinh\gamma}{\cosh\gamma - \cos\omega} e^{-ik\omega}\, d\omega, \tag{C9}$$

we have

$$R_k(\xi) = \frac{\sinh\gamma}{\sinh\tilde{\gamma}} e^{-\tilde{\gamma}|k|}, \tag{C10}$$

and in particular

$$R(\xi) = R_0(\xi) = \frac{\sinh\gamma}{\sinh\tilde{\gamma}} = \frac{\sinh\gamma}{[(\cosh\gamma - \xi\sinh\gamma)^2 - 1]^{\frac{1}{2}}}. \tag{C11}$$

Considering $\tilde{\gamma}$ as a function of ξ with γ as a fixed parameter, we have

$$\sinh\tilde{\gamma}\frac{d\tilde{\gamma}}{d\xi} = \frac{d}{d\xi}\cosh\tilde{\gamma} = -\sinh\gamma, \tag{C12}$$

or

$$R(\xi) = -\frac{d\tilde{\gamma}(\xi)}{d\xi}, \tag{C13}$$

and we obtain

$$\int_0^{\nu\gamma} R(\xi)\, d\xi = -\tilde{\gamma}(\nu\gamma) + \gamma = -\text{arc cosh}\,(\cosh\gamma - \nu\gamma\sinh\gamma) + \gamma. \tag{C14}$$

The BWCP defined by

$$\nu_c\gamma g(0) \equiv \nu_c C(\gamma) = 1 \tag{C15}$$

is given by

$$\nu_c\gamma\sinh\gamma/(\cosh\gamma - 1) = \nu_c\gamma\coth\gamma/2 = 1, \tag{C16}$$

or

$$\nu_c = \gamma^{-1}\tanh\gamma/2 = \tfrac{1}{2} + O(\gamma^2). \tag{C17}$$

The BW Curie temperature $J/k\nu_c$ is an increasing function of the range γ^{-1} for $\gamma > 0$, as one would expect. For any $\gamma > 0$, the function $R(\nu\gamma)$ becomes infinite at the BWCP.

Expanding in Eqs. (C11) and (C12) in powers of γ, we obtain for fixed $\nu < \nu_c$

$$R(\nu\gamma) = 1/(1 - 2\nu)^{\frac{1}{2}} + O(\gamma^2), \tag{C18}$$

and

$$\int_0^{\nu\gamma} R(\xi)\, d\xi = -\gamma[(1 - 2\nu)^{\frac{1}{2}} - 1] + O(\gamma^3). \tag{C19}$$

The free energy per particle is then obtained from Eqs. (9.1)–(9.3):

$$\begin{aligned}\psi &= -kT\Big\{\ln 2 + \tfrac{1}{2}\Big[\int_0^{\nu\gamma} R(\xi)\, d\xi - \nu\gamma\Big] \\ &\quad - \tfrac{1}{4}(\nu\gamma R(\nu\gamma))^2 + ((\gamma^3))\Big\} \\ &= -kT\Big\{\ln 2 + \frac{\gamma}{2}[1 - \nu - (1 - 2\nu)^{\frac{1}{2}}] \\ &\quad - \frac{\gamma^2}{4}\frac{\nu^2}{1 - 2\nu} + ((\gamma^3))\Big\},\end{aligned} \tag{C20}$$

for any fixed temperature above the BWCP. Writing this in the form

$$\begin{aligned}\lim_{n\to\infty} Q_n^{1/n} = e^{-\psi/kT} = 2e^{-\nu\gamma/2}\Big\{1 + \frac{\gamma}{2}[1 - (1 - 2\nu)^{\frac{1}{2}}] \\ + \frac{\gamma^2}{8}[1 - (1 - 2\nu)^{\frac{1}{2}}]^2 - \frac{\gamma^2}{4}\frac{\nu^2}{1 - 2\nu} + ((\gamma^3))\Big\},\end{aligned} \tag{C21}$$

we establish complete agreement with the result obtained by Kac's integral equation method.[8] With the aid of Eq. (C10), it would not be difficult to evaluate the functions R_{22} and R_{24} in our result, and obtain the free energy to third order in γ.

For any fixed temperature above and below the BWCP, we obtain the free energy per particle through $O(\gamma)$ from Eqs. (6.19) and (6.20) or (9.12). Observing that $C(\gamma) = 2 + O(\gamma^2)$, we obtain

$$\begin{aligned}\psi \cong -kT\Big\{\ln 2 - \frac{y^2}{4\nu} + \ln\cosh y \\ + \frac{\gamma}{2}\{1 - \nu - [1 - 2\nu(1 - \tanh^2 y)]^{\frac{1}{2}}\}\Big\},\end{aligned} \tag{C22}$$

where y is equal to zero, or to the positive root of the Weiss field equation (6.9)

$$y = 2\nu\tanh y, \tag{C23}$$

for temperatures above or below the BWCP, respectively. Equation (C22) is in agreement through order γ with the results of Kac.[8]

APPENDIX D

In this appendix we test the approximation developed in Sec. 7 by applying it to the two-spin correlation function $\langle \mu_r \mu_s \rangle$ for a one-dimensional model with interaction (C1). In the limit of infinite size of the model, the matrix elements $\tilde{\rho}_{rs}$ are given by R_k [Eq. (C10)] with $k = |r - s|$, and $\tilde{\gamma}$ given by (C7). At the BWCP ($\nu = \frac{1}{2}$), we solve (7.15) approximately for small γ and obtain

$$\alpha \cong 1 + (\nu\gamma)^{\frac{2}{3}}. \tag{D1}$$

Then (C7) and (C8) give

$$\tilde{\gamma} \cong (2)^{-\frac{1}{3}}\gamma^{\frac{4}{3}}. \tag{D2}$$

Collecting these results, we find for the leading term of longest range

$$\langle \mu_r \mu_s \rangle \cong 0.630\gamma^{\frac{2}{3}} e^{-0.794\gamma^{\frac{4}{3}}|r-s|}. \tag{D3}$$

This result is discussed and compared with the exact result of Kac and Helfand[11] in Sec. 9, following (9.16).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Hyperbolic Differential Equations in Two Dimensions*

W. KUNDT
University of Hamburg
AND
E. T. NEWMAN
University of Pittsburgh

(Received 21 June 1967)

We analyze the general two-dimensional hyperbolic differential equation of second order by means of a *substitution* method. Our main interest lies in the *support* of the solutions, i.e., in an answer to the question: under what circumstances can a signal be transmitted along *null rays*? It turns out that, in general, a signal *spreads*, i.e., fills the entire future of an event. However, reasonably large classes of differential equations do permit nonspreading (*characteristic propagation*) solutions. As examples it is shown that multipole solutions of the *flat* space–time *scalar wave equation* and *Maxwell equations* fall into the nonspreading class, whereas multipole solutions of the corresponding equations in a *curved* Schwarzschild background always show spreading (or continuous reflection).

1. INTRODUCTION

The propagation of disturbances of most physical systems can be mathematically described by *second-order normal hyperbolic* differential equations. In the simplest cases, these equations admit of wave solutions which propagate at some characteristic velocity, i.e., at some velocity which only depends on the medium (but not on frequencies, intensities, pulse shapes). We say that such solutions have the *characteristic propagation* property. In general media, such simple solutions are absent. The front of any disturbance will still propagate at some (maximum) characteristic velocity, but there will be a *tail*, or wake, to the wave which travels at all smaller speeds. This tail can be considered as a result of continuous *backward scattering*, or reflection, of the wave; we say in this case that the solution *spreads*. Our main effort will be directed towards finding criteria which tell whether a given equation has characteristic propagation solutions or not.

In this article we restrict ourselves to problems of *one space dimension* only (which form an exceptional case in the general theory of hyperbolic differential equations). Note that this restriction nevertheless allows the treatment of *separable* problems in which the solutions can be expanded in terms of complete orthonormal functions with respect to two space variables. Examples of this kind are media with plane symmetry, spherical symmetry, or cylindrical symmetry.[1]

In the literature[2] we could not find results relating to *spreading*, except in very special cases in which explicit

* This work has been supported by the Office of Aerospace Research, United States Air Force.

[1] *Note Added in Proof:* The tail problem in *three space dimensions* has recently been solved by McLenaghan for the wave equation in Einstein spaces: [R. G. McLenaghan, Proc. Cambridge Phil. Soc. (1968) (to be published)].

[2] In historical order, we mention the following books on hyperbolic differential equations: (a) J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Dover Publications, Inc., New York, 1952); (b) J. Leray, "Hyperbolic Differential Equations," Lecture notes, Princeton University, 1952; (c) R. Courant and D. Hilbert, *Methods of Mathematical Physics Vol. II* (Interscience Publishers, Inc., New York, 1962); (d) L. Hörmander, *Linear Partial Differential Operators* (Springer-Verlag, New York, 1963); (e) I. M. Gel'fand and G. E. Schilov, *Verallgemeinerte Funktionen (Distributionen) III* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1964).

APPENDIX D

In this appendix we test the approximation developed in Sec. 7 by applying it to the two-spin correlation function $\langle \mu_r \mu_s \rangle$ for a one-dimensional model with interaction (C1). In the limit of infinite size of the model, the matrix elements $\tilde{\rho}_{rs}$ are given by R_k [Eq. (C10)] with $k = |r - s|$, and $\tilde{\gamma}$ given by (C7). At the BWCP ($\nu = \frac{1}{2}$), we solve (7.15) approximately for small γ and obtain

$$\alpha \cong 1 + (\nu\gamma)^{\frac{2}{3}}. \tag{D1}$$

Then (C7) and (C8) give

$$\tilde{\gamma} \cong (2)^{-\frac{1}{3}}\gamma^{\frac{4}{3}}. \tag{D2}$$

Collecting these results, we find for the leading term of longest range

$$\langle \mu_r \mu_s \rangle \cong 0.630\gamma^{\frac{2}{3}} e^{-0.794\gamma^{\frac{4}{3}}|r-s|}. \tag{D3}$$

This result is discussed and compared with the exact result of Kac and Helfand[11] in Sec. 9, following (9.16).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Hyperbolic Differential Equations in Two Dimensions*

W. KUNDT
University of Hamburg
AND
E. T. NEWMAN
University of Pittsburgh

(Received 21 June 1967)

We analyze the general two-dimensional hyperbolic differential equation of second order by means of a *substitution* method. Our main interest lies in the *support* of the solutions, i.e., in an answer to the question: under what circumstances can a signal be transmitted along *null rays*? It turns out that, in general, a signal *spreads*, i.e., fills the entire future of an event. However, reasonably large classes of differential equations do permit nonspreading (*characteristic propagation*) solutions. As examples it is shown that multipole solutions of the *flat* space–time *scalar wave equation* and *Maxwell equations* fall into the nonspreading class, whereas multipole solutions of the corresponding equations in a *curved* Schwarzschild background always show spreading (or continuous reflection).

1. INTRODUCTION

The propagation of disturbances of most physical systems can be mathematically described by *second-order normal hyperbolic* differential equations. In the simplest cases, these equations admit of wave solutions which propagate at some characteristic velocity, i.e., at some velocity which only depends on the medium (but not on frequencies, intensities, pulse shapes). We say that such solutions have the *characteristic propagation* property. In general media, such simple solutions are absent. The front of any disturbance will still propagate at some (maximum) characteristic velocity, but there will be a *tail*, or wake, to the wave which travels at all smaller speeds. This tail can be considered as a result of continuous *backward scattering*, or reflection, of the wave; we say in this case that the solution *spreads*. Our main effort will be directed towards finding criteria which tell whether a given equation has characteristic propagation solutions or not.

In this article we restrict ourselves to problems of *one space dimension* only (which form an exceptional case in the general theory of hyperbolic differential equations). Note that this restriction nevertheless allows the treatment of *separable* problems in which the solutions can be expanded in terms of complete orthonormal functions with respect to two space variables. Examples of this kind are media with plane symmetry, spherical symmetry, or cylindrical symmetry.[1]

In the literature[2] we could not find results relating to *spreading*, except in very special cases in which explicit

* This work has been supported by the Office of Aerospace Research, United States Air Force.

[1] *Note Added in Proof:* The tail problem in *three space dimensions* has recently been solved by McLenaghan for the wave equation in Einstein spaces: [R. G. McLenaghan, Proc. Cambridge Phil. Soc. (1968) (to be published)].

[2] In historical order, we mention the following books on hyperbolic differential equations: (a) J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Dover Publications, Inc., New York, 1952); (b) J. Leray, "Hyperbolic Differential Equations," Lecture notes, Princeton University, 1952; (c) R. Courant and D. Hilbert, *Methods of Mathematical Physics Vol. II* (Interscience Publishers, Inc., New York, 1962); (d) L. Hörmander, *Linear Partial Differential Operators* (Springer-Verlag, New York, 1963); (e) I. M. Gel'fand and G. E. Schilov, *Verallgemeinerte Funktionen (Distributionen) III* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1964).

solutions were obtained. The characteristic propagation property is related to *Huygens' principle*,[3] whose precise scope of validity is not known (compare Hadamard's conjecture and its counterexamples[4]). In one space dimension, Huygens' principle is known not to be valid[5]; but characteristic propagation can be considered a *modified Huygens principle* for one space dimension.

In trying to find criteria for the modified Huygens principle, we develop a (seemingly) new approach towards solving, or analyzing, the general second-order hyperbolic differential equation in two dimensions. The approach consists of repeatedly replacing the searched-for solution by suitably defined potentials. This *substitution method* gives closed-form solutions in terms of finite-order *potentials* for large classes of equations and gives integral representations in terms of "*negative-order potentials*" for other classes of equations; these two classes have nonvoid intersection. If applied to the *characteristic initial-value problem*, the substitution method gives partial answers to the characteristic propagation problem.[6]

In Sec. 4 the substitution method is presented and reduced to the study of a sequence of functions in two variables. This *substitution sequence* is crucial to our analysis. Some of its properties are surveyed in Sec. 5 and proven in Secs. 8–12. The main results concerning spreading are derived in Secs. 6 and 7. In a sense to be specified below, we find that, among all solutions, the *spreading* ones form an *open set* whose *boundary points* are those with *characteristic propagation*.

In this article we restrict ourselves to the *homogeneous* (rather than inhomogeneous) equation. As shown in the literature,[7] the general solution of the homogeneous problem implies the general solution of the inhomogeneous problem via Green's celebrated integral theorem for which one needs the so-called *Riemann–Green function*, viz., the solution of the (homogeneous) adjoint equation for special boundary conditions.[8]

The present analysis arose from the question whether electromagnetic waves in curved empty space-time suffer backward scattering—the answer to which appears to be positive. More precisely, (even) the simplest case of a Maxwell field on a Schwarzschild background gives rise to spreading. Further *physical examples* are deformations of a (nonhomogeneous) rope (which may even be immersed in damping surroundings), electromagnetic currents in a cable, or nonrelativistic quantum-mechanical probability amplitudes. These examples will be briefly treated in Sec. 8.

2. REDUCTION TO RIGHT AND LEFT NORMAL FORM

The general *homogeneous second-order hyperbolic* differential equation can be written as

$$L\phi = 0, \tag{2.1}$$

with the linear operator

$$L := g^{ab}\nabla_a\partial_b + h^a\partial_a + i. \tag{2.2}$$

Here $g^{ab} = g^{(ab)}$, h^a, and i are

$$\binom{n+1}{2} + n + 1 = \binom{n+2}{2}$$

arbitrary functions of the n coordinates x^a; ∂_a denotes (ordinary) partial differentiation, and ∇_a covariant differentiation.

The equation $L\phi = 0$ can be simplified (i) by a *conformal transformation* of the metric g^{ab}, (ii) by suitable *coordinate choice*, and (iii) by a *factor transformation* on the unknown function ϕ, by means of which $n + 2$ coefficients can be eliminated.

In what follows we restrict ourselves to $n = 2$ dimensions. Every 2-dimensional metric is *conformally flat*:

$$g^{ab} = |g|^{-\frac{1}{2}}\eta^{ab}, \quad g := \det(g_{ab}), \quad g_{ab}g^{bc} := \delta_a^c, \tag{2.3}$$

where η^{ab} is a flat-space metric tensor of signature $(+\ -)$. By adapting coordinates, we can choose η^{ab} in one of the two normal forms

$$(\eta^{ab}) \overset{*}{=} \left\{ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ \text{(a)}; \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \ \text{(b)} \right\}. \tag{2.4}$$

Choice (a) means that we have introduced two *null coordinates* u and v which are related to the Minkowski coordinates x and t of choice (b) through

$$\begin{aligned} u &= (x - t)/2, \\ v &= (x + t)/2, \end{aligned} \quad \text{whence} \quad \partial_x^2 - \partial_t^2 = \partial_v\partial_u. \tag{2.5}$$

---

[3] Huygens' principle is formulated, and discussed in Ref. 2(c), and in the following: (a) L. Asgeirsson, Commun. Pure Appl. Math. **9**, 307 (1956); (b) S. Helgason, Acta Math. **102**, 239 (1959).

[4] Hadamard's conjecture, and some of its counterexamples, are presented in Ref. 3(a).

[5] See, e.g., Ref. 2(c), p. 765.

[6] *Note Added in Proof:* F. G. Friedlander has pointed out to us that "our" *substitution method* was already known to Laplace in 1773, and used at length by Darboux; *see* G. Darboux, *Leçons sur la théorie des surfaces* (Gauthier-Villars, Paris, 1915), 2me éd., vol. II, book 4, Secs. 2–9. Darboux presents a more powerful (and sophisticated) approach to one of the problems treated by us in Sec. 11 (symmetrical double-terminating sequences).

[7] See, e.g., Ref. 2(c), pp. 449–461.

[8] As a review article on the Riemann–Green function, we mention E. T. Copson, Arch. Ratl. Mech. Anal. **1**, 324 (1962).

For any choice of the coordinates, the covariant d'Alembertian takes the simplified form

$$g^{ab}\nabla_a\partial_b = |g|^{-\frac{1}{2}}\,\partial_a|g|^{\frac{1}{2}}\,g^{ab}\partial_b = |g|^{-\frac{1}{2}}\,\partial_a\eta^{ab}\partial_b. \quad (2.6)$$

Combining the above simplifications with the factor transformation

$$\phi =: f\psi, \quad (2.7)$$

from (2.2) and (2.6) we get

$$L\phi = f|g|^{-\frac{1}{2}}\{\partial_a\eta^{ab}\partial_b + |g|^{\frac{1}{2}}[h^a + (\ln f^2)^{,a}]\partial_a - j\}\psi \quad (2.8)$$

with

$$j := -|g|^{\frac{1}{2}}\{h^a\partial_a \ln|f| + f^{-1}\nabla_a^a f + i\}, \qquad f^{,a} := g^{ab}\partial_b f. \quad (2.9)$$

The first-order differential term in L can obviously be annihilated iff h_a is a gradient. In general, we use f to annihilate at least one of the two first-order components, and for the coordinate choice (2.4a) we obtain

$$L\phi = f|g|^{-\frac{1}{2}}\{\partial_v\partial_u + h\partial_u - j\}\psi, \quad (2.10)$$

where

$$\partial_u \ln f^2 = -h_u, \quad h := h_v + \partial_v \ln f^2. \quad (2.11)$$

Finally, with

$$k := \exp\int^v dv' h(u, v'), \quad j := jk, \quad (2.12)$$

Eq. (2.1) becomes equivalent to

$$\boxed{\{\partial_v k\partial_u - j\}\psi = 0,}, \quad (2.13)$$

where ∂_v acts on everything to the right of it.

We call (2.13) the *right normal form* of Eq. (2.1). It contains the two coefficients j and k; u and v are characteristic or null coordinates. The function k can be eliminated, i.e., set equal to one through a suitable gauge iff it is a product of a function of u and a function of v, which takes place iff h_a is a gradient. Physically, a nontrivial k corresponds to the presence of *dissipation*, i.e., to nonconservation of the energy of the system.

Before we ask for the degree of uniqueness of the right normal form, we are interested in a transition to the corresponding *left normal form* (the roles of u and v interchanged):

$$\{\partial_u\tilde{k}\partial_v - \tilde{j}\}\chi = 0, \quad (2.14)$$

which is obtained from (2.13) via the factor transformation

$$\psi =: k^{-1}\chi. \quad (2.15)$$

By straightforward calculation one finds

$$\tilde{k} = k^{-1}, \quad \frac{\tilde{j}}{\tilde{k}} = \frac{j}{k} + \partial_v\partial_u \ln|k|. \quad (2.16)$$

Let us remark that the *adjoints* of Eqs. (2.13) and (2.14) are obtained by interchanging ∂_u and ∂_v; consequently, right and left normal form are adjoints of each other iff $k = \text{const}$, in which case they are equal and *self-adjoint*.

3. GAUGE GROUP; INVARIANTS

In Eqs. (2.13) and (2.14) we obtained the right and left normal form of the homogeneous second-order hyperbolic differential equation in two dimensions. It contains the two coefficients j and k. We inquire into their degree of uniqueness.

The first step in our reduction to normal form was a conformal simplification of the metric g^{ab}. We introduced two null coordinates u and v defined by $g^{ab}\nabla_a\partial_b = |g|^{-\frac{1}{2}}\,\partial_v\partial_u$. The corresponding *null rays* $v = \text{const}$, $u = \text{const}$, are the two null eigendirections of g^{ab}, and are thereby unique. Consequently, u and v are unique up to arbitrary monotonic transformations

$$\tilde{u} = \tilde{u}(u), \qquad \tilde{v} = \tilde{v}(v), \quad (3.1)$$

which form the only possible *coordinate gauges* left.

Once our equation is in the normal form (2.13)

$$\{\partial_v k\partial_u - j\}\psi = 0,$$

the only *factor transformation* $\psi =: f\chi$ which preserves this shape obeys $f_{,u} = 0$, and the only possible *conformal factor* to be absorbed into this shape is a function of u. Considering the combined effect of a *coordinate gauge*, a *factor transformation*, and a *conformal gauge*, we find that the coefficient k is invariant up to a u-dependent factor and a v-dependent factor, and the same applies to j. $k^{-1}j$ is *invariant* under combined factor and conformal gauges but behaves under coordinate gauges (3.1) as

$$k^{-1}j = \tilde{u}_{,u}\tilde{v}_{,v}\tilde{k}^{-1}\tilde{j}, \quad (3.2)$$

i.e., like a *scalar density*. The same applies to the normal form of the d'Alembertian

$$\partial_v\partial_u = \tilde{u}_{,u}\tilde{v}_{,v}\partial_{\tilde{v}}\partial_{\tilde{u}}. \quad (3.3)$$

From these results we infer that the operator

$$\boxed{D := \partial_v\partial_u \ln} \quad (3.4)$$

maps *scalars* or *scalar densities* into *scalar densities*, and that the operator

$$E := kj^{-1}D \quad (3.5)$$

maps scalars or scalar densities into invariants. Equally, one finds that Ej and Ek are *invariants* under arbitrary gauges. (N.B.: Ej means "E applied to j!")

Of outstanding physical interest are *static media* for which both j and k can be chosen as functions of $x = u + v$ only [cf. (2.5)]. A function $j(u + v)$ satisfies

$$0 = D(j_{,u}j_{,v}^{-1}), \tag{3.6}$$

and this equation is gauge independent, i.e., *characterizes* static coefficients in any normal form.

4. SUBSTITUTION METHOD; POTENTIALS

In this section, we consider the *characteristic initial-value* problem for "*incoming*" ("left traveling") *radiation*. By "incoming (left traveling) radiation" we mean solutions of (2.13) in the future light cone of some event (chosen as the origin of spacetime) which vanish along the left boundary (ray). (See Fig. 1. Notice that with our coordinate choice u decreases towards the future!) An analogous treatment can, of course, be given under the *outgoing* radiation condition, and the solution of the *general* characteristic initial-value problem can be obtained as a superposition of an incoming and an outgoing solution.

Our *substitution* will be based on the once-integrated right normal form (2.13) for zero outgoing data ($\partial_u\psi = 0$ at $v = 0$), which reads

$$j_0\partial_u\psi = \int_0^v dv' j\psi, \quad j_0 := k. \tag{4.1}$$

This equation turns out to be form-invariant if we express the function ψ in terms of a (generalized) *potential* ψ_2 defined by

$$j\psi_2 := \int_0^v dv' j\psi. \tag{4.2}$$

Indeed we obtain [cf. (3.4)]

$$\begin{aligned} j_0\partial_u\psi &= \frac{j_0}{j}\left(\partial_u\partial_v j\psi_2 - \frac{j_{,u}}{j}\partial_v j\psi_2\right) \\ &= \frac{j_0}{j}(\partial_v[\partial_u j\psi_2 - j_{,u}\psi_2] + j\psi_2 Dj). \end{aligned} \tag{4.3}$$

But the quantity in square brackets is equal to $j\partial_u\psi_2$. On the other hand, (4.1) and (4.2) yield

$$j_0\partial_u\psi = j\psi_2; \tag{4.4}$$

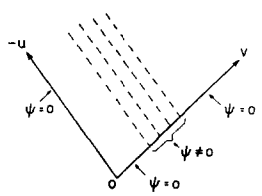


FIG. 1. Incoming (or left traveling) radiation.

comparison with (4.3) gives

$$\partial_v j\partial_u\psi_2 = j\psi_2(jj_0^{-1} - Dj), \tag{4.5}$$

or, for zero outgoing data ($\partial_u\psi_2 = 0$ at $v = 0$),

$$j\partial_u\psi_2 = \int_0^v dv' j_2\psi_2, \tag{4.6}$$

with

$$j_2 := j(jj_0^{-1} - Dj). \tag{4.7}$$

All steps of the foregoing transformation can be followed backwards to derive Eq. (4.1) from Eq. (4.6), and we have shown that the *integro-differential equations* (4.1) and (4.6) are *equivalent* (under the incoming radiation condition) if ψ and its *potential* ψ_2 are related by definition (4.2). This remarkable result allows one to replace a differential equation by another one of the same shape but with (in general) different coefficients.

Moreover, a potential substitution analogous to (4.2) can be performed any number of times, both "forward" and "backward," resulting in

$$\boxed{j_{k-1}\partial_u\psi_k = \int_0^v dv' j_k\psi_k}, \quad -\infty < k < \infty, \tag{4.8}$$

$$j_k\psi_{k+1} = \int_0^v dv' j_k\psi_k, \tag{4.9}$$

$$\boxed{\frac{j_{k+1}}{j_k} = \frac{j_k}{j_{k-1}} - Dj_k}. \tag{4.10}$$

Iteration of Eq. (4.9) gives a closed form expression for ψ in terms of its kth-order potential ψ_{k+1}:

$$\boxed{\psi = j^{-1}\partial_v j j_2^{-1}\partial_v j_2 j_3^{-1}\cdots\partial_v j_{k-1}j_k^{-1}\partial_v j_k\psi_{k+1}}, \quad \text{for } k > 0, \tag{4.11}$$

and

$$\boxed{\begin{aligned}\psi = j_0^{-1}\int_0^v dv_1 j_0 j_{-1}^{-1}\int_0^{v_1} dv_2 j_{-1}j_{-2}^{-1}\cdots \\ \int_0^{v_{k-1}} dv_k j_{-(k-1)}j_{-k}^{-1}\int_0^{v_k} dv_{k+1} j_{-k}\psi_{-k}\end{aligned}}, \tag{4.12}$$

for potentials ψ_{-k} of *negative order* $-(k + 1)$, $k \geq 0$.

The importance of this result can be gathered from the following remark: Suppose the sequence j_k *terminates* to the *right*, i.e., we have $j_{n+1} \equiv 0$, $j_n \neq 0$ for some $n > 0$. In this case, Eq. (4.8) for $k = n + 1$ reads $\partial_u\psi_{n+1} = 0$, so that ψ_{n+1} is an arbitrary function of v; and formula (4.11) gives the *general incoming solution* as a modified *n*th-*order derivative* [of an arbitrary *n*th order potential $\psi_{n+1}(v)$].

Correspondingly, for *left termination*

$$j_{-(n+1)}^{-1} \equiv 0, \quad j_{-n}^{-1} \neq 0,$$

the *general incoming solution* is obtained from formula (4.12) as a modified *n*th-*order integral* [of an arbitrary $(n+1)$th-order potential $\psi_{-n}(v)$].

We have seen that the whole sequence of equations (4.8) is determined by the two coefficients $j_0, j_1 := j$ in the right normal form (4.1). According to our invariance considerations in Sec. 3, the sequence of functions $j_k(u, v)$ defined recursively in (4.10) is *invariant* under combined factor and conformal gauges. We call it the *substitution sequence*. Under coordinate gauges, all *ratios* $j_k^{-1} j_{k+1}$ transform like *scalar densities* (which implies that their vanishing is an invariant property). Sequences related in this way form equivalence classes.

In Eq. (2.16) we found that the left normal form is related to the right normal form according to

$$\tilde{j}_0^{-1}\tilde{j} = j_0^{-1} j + D j_0, \quad \tilde{j}_0 = j_0^{-1}, \tag{4.13}$$

which, via Eq. (4.10), implies that

$$\tilde{j}_0^{-1}\tilde{j} = j_{-1}^{-1} j_0, \quad \text{whence} \quad \tilde{j} = j_{-1}^{-1}, \tag{4.14}$$

and recursively

$$\tilde{j}_k = j_{-k}^{-1}. \tag{4.15}$$

We call j_{-k}^{-1} the *mirrored* sequence. Our last equation tells that the *incoming* and the *outgoing* problem are described by mirrored sequences.

If a sequence terminates to the right (left), its mirrored sequence terminates to the left (right). In either case, formulas (4.11), (4.12) provide *closed form solutions*. We can therefore solve the general characteristic initial-value problem (and especially determine the Riemann–Green function) whenever the sequence *terminates* to (at least) one side.

For further analysis it is important to know that there is a mutual linear dependence between the *k*th-order *transverse derivatives* $\partial_u^k \psi$ and the *k*th-order potentials ψ_{k+1}. We claim

$$\boxed{\partial_u^k \psi = \sum_{l=1}^{k} a_l^k j_l \psi_{l+1}}, \quad \text{for} \quad k \geq 1, \tag{4.16}$$

with coefficient functions a_l^k which are recursively determined by

$$a_l^k = 0, \quad \text{for} \quad l = 0, \quad l > k,$$
$$a_k^k = j_0^{-1}, \quad a_l^{k+1} = a_{l-1}^k + j_l^{-1} \partial_u j_l a_l^k. \tag{4.17}$$

The simple proof by induction is omitted; it makes use of Eqs. (4.8) and (4.9) which imply that

$$\partial_u \psi_l = j_{l-1}^{-1} j_l \psi_{l+1}. \tag{4.18}$$

We now show how formula (4.12) may be simplified. To this end we prove

$$\boxed{\int_{x_0}^{x} dx_1 g_1(x_1) \int_{x_0}^{x_1} dx_2 g_2(x_2) \int_{x_0}^{x_2} \cdots \int_{x_0}^{x_{n-1}} dx_n g_n(x_n) f(x_n) = \int_{x_0}^{x} dx_n G_n(x, x_n) f(x_n)}, \tag{4.19}$$

where

$$\boxed{G_n(x, x_n) := g_n(x_n) \int_{x_n}^{x} dx_{n-1} g_{n-1}(x_{n-1}) \int_{x_{n-1}}^{x} dx_{n-2} g_{n-2}(x_{n-2}) \int_{x_{n-2}}^{x} \cdots \int_{x_2}^{x} dx_1 g_1(x_1)}. \tag{4.20}$$

The proof is done by *n*-fold integration by parts, starting at the left end. We only demonstrate the first step, writing "lhs" for the left-hand side of Eq. (4.19):

$$\text{lhs} = \int_{x_0}^{x} dx_1 g_1(x_1) \cdot \int_{x_0}^{x} dx_2 g_2(x_2) \int_{x_0}^{x_2} \cdots \int_{x_0}^{x_{n-1}} dx_n g_n f - \int_{x_0}^{x} dx_2 \int_{x_0}^{x_2} dx_1 g_1(x_1) \cdot g_2(x_2) \int_{x_0}^{x_2} \cdots \int_{x_0}^{x_{n-1}} dx_n g_n f$$
$$= \int_{x_0}^{x} dx_2 \left[g_2(x_2) \int_{x_2}^{x} dx_1 g_1(x_1) \right] \int_{x_0}^{x_2} dx_3 g_3(x_3) \int_{x_0}^{x_3} \cdots \int_{x_0}^{x_{n-1}} dx_n g_n f.$$

In the second step, the square bracket takes the role of $g_1(x_1)$ in the first step, and so on.

For later use let us mention the following properties of the *integrating kernel* $G_n(x, x_n)$ in (4.19):

$$\partial_x^k G_n(x, x) = 0 \quad \text{for} \quad k \leq n-1,$$
$$\partial_x^n G_n(x, x) = \prod_{k=1}^{n} g_k(x). \tag{4.21}$$

They follow straightforwardly from the definition (4.20) once we note that $G_n(x, x_n)$ is stationary in all upper limits x, except the one at the right end. For instance,

$$\partial_x G_n(x, x_n) = g_1(x) g_n(x_n) \int_{x_n}^{x} dx_{n-1} g_{n-1} \int_{x_{n-1}}^{x} \cdots \int_{x_3}^{x} dx_2 g_2(x_2). \tag{4.22}$$

Application of (4.19), (4.20) to (4.12) yields

$$\boxed{\psi = \jmath_0^{-1}(u, v)\int_0^v dv' H_{-k}(u, v, v_1)\psi_{-k}(u, v')}, \tag{4.23}$$

where H_{-k} is constructed from the $\jmath_{-(l-1)}\jmath_{-l}^{-1}$, $\jmath_{-k}$ in the same way as G_k is constructed from the g_l:

$$\boxed{H_{-k}(v, v_k) := \jmath_{-k}(v_k)\int_{v_k}^v dv_{k-1}\jmath_{-k}^{-1}\jmath_{-(k-1)}\int_{v_{k-1}}^v dv_{k-2}\jmath_{-(k-1)}^{-1}\jmath_{-(k-2)}\int_{v_{k-2}}^v, \cdots, \int_{v_1}^v dv_0\jmath_{-1}^{-1}\jmath_0}, \tag{4.24}$$

for $k > 0$ (we have suppressed the additional u dependence), and where

$$H_0(v, v_0) = \jmath_0(v_0). \tag{4.25}$$

5. PROPERTIES OF SUBSTITUTION SEQUENCES

In the last section we defined the substitution sequence $\jmath_k(u, v)$ and hinted at its importance in the analysis of solutions. Several of our future conclusions will depend on nontrivial properties of this sequence which we shall prove in Sec. 8–12. For the benefit of the reader, the present section gives a summary of some of the results obtained.

(A) We speak of *right-terminating* sequences if $\jmath_k$ vanishes identically for some $k \geq 1$, and of *left-terminating* sequences if $\jmath_{-k}^{-1}$ vanishes identically for some $k \geq 1$. If both conditions are satisfied at the same time, we speak of *double termination*. The number of finite, nonvanishing elements in a sequence will be called its *length*. Lemma 1 states that all these possibilities are realizable.

Lemma 1: There exist sequences of any length $L \geq 1$. The general sequence of length L depends on $2L$ arbitrary functions of one variable. The general right- (or left-) terminating sequence depends on one arbitrary function of two variables. Almost all sequences do not terminate to either side.

Proof (sketch): The proof is now sketched. A left- (right-) terminating sequence is determined by its left (right) end element, i.e., by the last finite, nonvanishing element at that end. For double termination, the left-end element has to satisfy an ordinary linear homogeneous differential equation with arbitrary coefficients.

(B) There are sequences which can be expressed in *closed form*. They are the ones which satisfy

$$D\jmath = c(\jmath_0^{-1}\jmath), \quad D\jmath_0 = c_0(\jmath_0^{-1}\jmath) \tag{5.1}$$

with arbitrary constants c and c_0. One finds

$$\jmath_k = \jmath_0^{-k+1}\jmath^k(1 - c) \times (1 - 3c + c_0)\cdots\left(1 - \binom{k}{2}c + \binom{k-1}{2}c_0\right) \tag{5.2}$$

for $k \geq 2$, and

$$\jmath_{-k+1}^{-1} = \jmath^{k-1}\jmath_0^{-k}(1 + c_0) \times (1 + 3c_0 - c)\cdots\left(1 + \binom{k}{2}c_0 - \binom{k-1}{2}c\right) \tag{5.3}$$

for negative indices ($k \geq 2$). In the simplest special case, both $\jmath$ and $\jmath_0$ are constant.

(C) Of special interest to physics are *static sequences* whose elements can and will be chosen as functions of $x = u + v$ alone [cf. (3.6)]. We will show that all members of *double-terminating static* sequences are *rational functions* in x and (a finite number of) $e^{\omega_k x}$ with possibly complex frequencies ω_k. Special cases are rational functions in x alone. However, this structure is not sufficient for double termination.

(D) *Self-adjoint* equations give rise to *symmetrical* sequences (in the gauge $\jmath_0 = 1$), which are defined as those equalling their mirrored ones:

$$\jmath_{-k} = \jmath_k^{-1}. \tag{5.4}$$

This is a direct consequence of Eq. (4.15) relating the right and left normal form. As a consequence, terminating symmetrical sequences are double-terminating. We have a conjecture concerning the analytic shape of their members (see Sec. 11).

(E) We *summarize*: There exist *nonterminating*, *right-terminating*, *left-terminating*, and *double-terminating* sequences. *Symmetrical* sequences are either nonterminating or double-terminating. None of these possibilities is lost if one restricts the sequences to *static* ones. Analytic expressions for all these classes are either known or conjectured, and criteria for their occurrence available. Several of the known solutions in the *literature* belong to nonterminating sequences.[6]

(F) Let us finally relate some properties of the sequences to well-known *physical* problems. The simplest physical media are static and nondissipative. Their corresponding substitution sequence is static and symmetrical. For instance, the radial behavior of

Maxwell multipole waves in flat space-time gives rise to static, symmetric, terminating sequences whereby the order of termination is related to the order of the pole. The corresponding problem on a Schwarzschild background, however, gives rise to nonterminating (static, symmetric) sequences.

6. SPREADING; CHARACTERISTIC PROPAGATION

We are now ready to discuss the support of solutions, i.e., give partial answers to the question under what circumstances a wave propagates with or without a *tail.*

The *support* of a function is the closure of the set on which the function does not vanish. The question for the existence of a wave tail is a question for the support of the solution (within the domain of dependence). For instance, let us consider kth-order right-terminating sequences. Formula (4.11) shows that in this case an incoming solution can only be different from zero along null rays $v = \text{const}$ at which ψ_{k+1} does not vanish. We call this phenomenon *characteristic propagation.* In a slightly more sophisticated language this means that the v support of the solution is contained in the support of ψ_{k+1}. [The v support of ψ, for $0 \leq -u \leq u_0$, is the closure of the set of v- values at which ψ does not vanish for some $-u \in (0, u_0)$.]

We just saw that for right-terminating sequences the support of ψ_{k+1} contains the support of the solution and (clearly) also the support of the (characteristic) *initial state* $\psi(0, v)$. In general, if the initial state is of compact (or bounded) support, the kth-order potential need not have compact support: the solution *spreads.* However, all ψ_{k+1} of compact support give rise to characteristic propagation.

One might conjecture that characteristic propagation was a privilege of right-terminating sequences. The conjecture is false; but we will show below that characteristic propagation is a measure-zero phenomenon within the set of all equations and initial states. We are going to develop sufficient conditions for spreading.

In what follows we assume the coefficients f_0, f, and initial states $\psi(0, v)$ to be *infinitely differentiable.* Well-known existence theorems[9] then guarantee the same differentiability structure for the solution $\psi(u, v)$; and we know that the expansion

$$\psi(u, v) \cong \sum_{k=0}^{\infty} \frac{u^k}{k!} \partial_u^k \psi(0, v) \tag{6.1}$$

is (at least) an *asymptotic expansion.*[10] That is, if one extends the summation on the right-hand side up to the Kth term only, Eq. (6.1) is valid with a possible deviation of the order of the $(K + 1)$th term.

Our further discussion will be based upon Eqs. (4.16) and (4.17), which express the kth transverse derivative $\partial_u^k \psi$ as a linear combination of the first k potentials ψ_l with nonvanishing highest coefficient $a_k^k = f_0^{-1}$. We conclude:

Lemma 2: The union of the supports of all the potentials at $u = 0$ is contained in the v support of the solution for $0 \leq -u \leq u_0$, $u_0 > 0$.

Proof: One reads off from (6.1) for $u = 0$ that the support of the initial state (= 0th potential) is contained in the v support of the solution. Inductively, suppose that v_0 is contained in the support of ψ_{k+1} but not of ψ_l for $l \leq k$. Equations (4.16) and (4.17) imply that v_0 must be contained in the support of $\partial_u^k \psi$, and hence in the v support of ψ, because different terms in a power-series expansion cannot cancel each other identically in some interval.

Lemma 2 has important consequences. Suppose the *initial state* $\psi(0, v)$ has *compact support* contained in the interval $(0, v_0)$, say. According to Eq. (4.11), the lth potential is a modified lth-order integral of the initial state and is of the form (4.12). This integral may vanish for all $v \geq v_0$, at least for some small values of l. Suppose it vanishes for $l \leq k$, but not for $l = k + 1$. In this case, we have $\psi_{k+1} = \text{const} \neq 0$ for $v \geq v_0$; i.e., the support of ψ_{k+1} extends to infinity. In other words, either all potentials are supported by $(0, v_0)$ or the v support of the solution includes all points from v_0 to infinity.

We can give sufficient conditions for the latter case to happen, conditions which are based upon the fact that repeated integration of a function with compact support leads eventually to a function with noncompact support. This fact is a generalization of a converse of the classical theorem of Rolle, which says that the derivative of a differentiable function on the real line has a zero between each two zeros of the function. We prove:

Lemma 3: Let $f(x)$ be an infinitely differentiable function whose support is an interval. Let $Z[f]$ be the number of zeros of $f(x)$ whereby nth-order zeros are not counted if n is even, and whereby the two end points are counted onefold each. Then $Z[f'] \geq Z[f] + 1$.

[9] See Ref. 2(c), p. 470, or Refs. 2(d) and 2(c).

[10] A. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956).

Lemma 3 says that, on differentiation, one wins zeros, or conversely, that on integration one loses zeros. Its proof follows from the fact that a function has a relative extremum between each two of its zeros, so that f' has an odd-order zero between each two zeros of f. The number of intervals between zeros equals the number of zeros minus one, and the proposed inequality results when one takes account of the two end points.

If one applies Lemma 3 to a function f with $Z[f] = N + 2$, then, after at most N-fold integration, one arrives at a function g with $Z[g] \leq 2$, which means that g has no zeros in the interior of the interval. A further integration necessarily results in a monotonic function. Consequently, the *n*th-*order integral* of a function f supported by an interval must have a *noncompact support* if

$$n > Z[f] - 2. \tag{6.2}$$

We want to extend Lemma 3 to the potentials ψ_k which are obtained from ψ by repeated integration analogous to Eq. (4.12). Here we have to multiply each integral by a ratio of $\mathcal{J}_l$'s before the next integration. If all these ratios are free of zeros (and infinities) or have only a limited number of them, a generalized Lemma 3 still applies, because the proof takes only account of the number of zeros in the interval. That is, if the (repeated) zeros of all the elements of a nonright-terminating sequence do not lie dense everywhere on the v axis, there will be potentials of noncompact support for all initial states with a finite number of (odd-order) zeros whose support is contained in an interval between accumulation points of zeros. The sequences (5.1) and (5.2) form examples of this kind for almost all values of c and c_0.

Let us make precise what we mean by "spreading." We say that a solution *spreads* (strongly) if its support is equal to the domain of dependence (or future) of the nonzero initial data. (The future of a set is the union of the future light cones through all the points of the set.)

With this definition we can infer from Lemma 2 that an incoming solution *spreads* whenever there are potentials of *noncompact support*. Actually, the above proof only applies to a retarded time slice ($0 \leq -u \leq u_0$) of infinitesimal thickness. But it is not difficult to prove spreading globally under this condition by means of an argument by contradiction.

Further insight can be gained by considering *left-terminating* sequences. They will provide examples of *characteristic propagation* for nonright-terminating sequences (and certain initial states). In (4.23) and (4.24) we obtained the general incoming solution for $(n + 1)$th-order left termination:

$$\psi = \mathcal{J}_0^{-1} \int_0^v dv' H_{-n}(u, v, v') \psi_{-n}(v').$$

In order to check on characteristic propagation, we have to assume $\psi(0, v)$ of compact support, contained in the interval $(0, v_0)$, say, and ask whether the integral can vanish identically for $u \leq 0$, $v \geq v_0$. If it does vanish for $v \geq v_0$, then so does the integral

$$I(u, v, v_1) := \int_0^v dv' H_{-n}(u, v_1, v') \psi_{-n}(v') \quad \text{for all} \quad u \leq 0, v \geq v_0, v_1 \geq v_0, \tag{6.3}$$

because $\psi_{-n}(v)$ vanishes for $v \geq v_0$. But $I(u, v, v_1)$ is an analytic function in u and v_1 if the functions $\mathcal{J}_k(u, v)$ are analytic in u and v, and $\psi_{-n}(v)$ is continuous, so that (at least) under this assumption I has to vanish identically in u and v_1 for all $v \geq v_0$. Conversely, the condition

$$I(u, v, v_1) = 0, \quad \text{for} \quad v \geq v_0, \tag{6.4}$$

is clearly sufficient for characteristic propagation of all initial states supported by $(0, v_0)$.

In order to evaluate the criterion (6.4), let us restrict ourselves to *static* sequences in which case v can be replaced by $v + u = x$ in all the relevant equations. The condition for characteristic propagation now reads

$$0 = \int_u^x dx' H_{-n}(x_1, x') \psi_{-n}(x' - u), \quad \text{for} \quad x \geq v_0 + u, \tag{6.5}$$

identically in x_1. Or, using the fact that $\psi_{-n}(v)$ vanishes outside the interval $(0, v_0)$, we get

$$0 = \int_{-\infty}^{\infty} dx' H_{-n}(x_1, u + x') \psi_{-n}(x') = (H_{-n} \star \check{\psi}_{-n})(x_1, u), \tag{6.6}$$

where $\check{\psi}_{-n}(v) := \psi_{-n}(-v)$, and $\star$ denotes *convolution* with respect to the second argument of H_{-n}.

Under *Fourier* or *Laplace* transformation, convolution maps into ordinary (pointwise) multiplication. Hence we get the equivalent criterion

$$0 = \hat{H}_{-n}(x_1, k) \hat{\psi}_{-n}(-k) \tag{6.7}$$

for the Fourier or Laplace transforms $\hat{H}_{-n}$ and $\hat{\psi}_{-n}$ of H_{-n} and ψ_{-n}.

Now ψ_{-n} is by assumption a function of compact support, which implies that its transform $\hat{\psi}_{-n}$ is *entire analytic*. Consequently, $\hat{\psi}_{-n}$ has at most discrete zeros which cannot accumulate at finite points.

Criterion (6.7) says that the k support of $\hat{H}_{-n}(x_1, k)$ must be contained in the set of zeros of $\hat{\psi}_{-n}(-k)$. This restricts H_{-n} to (at most) a series in exponentials with polynomial factors (depending on the initial state). Almost periodic functions are special cases of this kind.

So far we have ignored the fact that criterion (6.7) has to be satisfied *identically* in x_1, which imposes a severe extra restriction upon the (left-terminating) sequence considered. This extra condition drops out for $n = 0$ [cf. (4.25)], in which case we have

$$H_0(x_1, x) = \mathcal{j}_0(x), \tag{6.8}$$

and (6.7) simplifies to

$$0 = \mathcal{j}_0(k)\hat{\psi}_0(-k). \tag{6.9}$$

Any function $\mathcal{j}_0$ can "generate" a first-order left-terminating sequence. The criterion for characteristic propagation can therefore certainly be satisfied in this case, namely, by choosing $\mathcal{j}_0$ as a series in exponentials with polynomial factors and choosing the initial state such that Eq. (6.9) is satisfied.

Another important special case deserves mentioning: If the substitution sequence is *left-terminating*, *static*, and *periodic* in space, the integral in condition (6.5) can be understood as being extended over a (possibly multiple) period of H_{-n} containing the support of ψ_{-n}. By continuing ψ_{-n} periodically, one can expand H_{-n} and ψ_{-n} in *Fourier series* (rather than Fourier integrals), and criterion (6.7) becomes a countable number of conditions on the corresponding (discrete) Fourier coefficients.

Let us *summarize* this section. We have seen that for *right-terminating* sequences there exists a functional class of initial states with compact support which propagate along characteristics (namely, those derivable from corresponding potentials). For *left-terminating* sequences, characteristic propagation can also take place, but only under highly restricting conditions on the sequence and the corresponding initial states. *Nonterminating* sequences give rise to spreading for almost all initial states, but we have only been able to present a proof for restricted classes of sequences and initial states, and there are indications that exceptional cases exist. In the following section we will derive further criteria for spreading.

7. FURTHER CRITERIA FOR SPREADING (BY EXPANSIONS)

In this section we will prove spreading under certain assumptions by means of several different expansions.

(A) Whenever the coefficients, and solutions, possess *asymptotic expansions* in x^{-1} or v^{-1} *around infinity*, one can apply Lemma 3 (of the last section) to prove that, in general, all initial states of compact support must spread. In order to avoid lengthy formulas we restrict our calculations to *symmetric* sequences. We assume $\mathcal{j}(u, v)$ of the form

$$\mathcal{j} = \sum_{l=2}^{\infty} J_l(u)x^{-l} \tag{7.1}$$

and try to solve Eq. (2.13), or rather its integrated equivalent

$$\partial_u\psi = \int_{\infty}^{v} dv' \mathcal{j}\psi + \partial_u\phi(u), \tag{7.2}$$

with the ansatz

$$\psi = \sum_{l=0}^{\infty} \Psi_l(u)x^{-l}. \tag{7.3}$$

That is, we try to solve the *general characteristic initial-value* problem based upon the two null rays $u = 0$ and $v = \infty$; [$\phi(u)$ is the initial state at $v = \infty$, recently called the *news function*].

Note that in Eq. (7.2), v and v' can be replaced by x and x' ($x = u + v$). Insertion of (7.1) and (7.3) into (7.2) leads to ($' := \partial_u$)

$$\sum_{k=0}^{\infty} x^{-k}[\Psi_k' - (k-1)\Psi_{k-1}] = -\sum_{k=1}^{\infty} x^{-k}k^{-1}\sum_{l=0}^{k-1} J_{k-l+1}\Psi_l + \phi'(u), \tag{7.4}$$

whence

$$\Psi_0' = \phi',$$

$$\Psi_k' = k^{-1}\sum_{l=0}^{k-1} \Psi_l[\delta_l^{k-1}k(k-1) - J_{k+1-l}] \quad \text{for} \quad k \geq 1. \tag{7.5}$$

From this infinite system of ordinary differential equations, the functions $\Psi_k(u)$ can be successively determined by quadratures. Assume $\phi(u)$ of compact support contained in the interval $(0, u_0)$. Under what circumstances can we prove spreading, i.e., nonvanishing of some $\Psi_k(u)$ for $u \geq u_0$ if $\Psi_k(0) = 0$? Notice that the roles of u and v are interchanged as compared to the discussion of spreading in the rest of this article, and that we have considered a time reversed problem.

A simple answer can be given if we assume $\mathcal{j}$ to be *static* so that the expansion coefficients J_l are all constant. In this case, Ψ_k becomes a constant linear combination of iterated integrals of $\phi(u)$ up to order k:

$$\Psi_k(u) = \sum_{l=1}^{k} c_l^k \int_0^u du_1 \int_0^{u_1} du_2 \cdots \int_0^{u_{l-1}} du_l \phi(u_l), \quad k \geq 1, \tag{7.6}$$

with highest coefficient

$$k!\, c_k^k = -J_2(2 - J_2)(6 - J_2)\cdots(k(k-1) - J_2) \tag{7.7}$$

for $J_2 \neq 0$. If $J_l = 0$ for $2 \leq l \leq p$, the sum in (7.6) extends only up to $l = k - p + 1$, with highest coefficient

$$c_{k-p+1}^k = -\frac{(k-1)!}{p!} J_{p+1}, \quad k \geq p, \tag{7.8}$$

and all the coefficients Ψ_k with $1 \leq k \leq p - 1$ vanish identically. Combining the two cases, we see that Ψ_k is a linear combination of iterated integrals of ϕ whose order tends to infinity for $k \to \infty$ (at least) unless $J_2 = n(n-1)$ for some positive integer n. In all cases except possibly this latter one, therefore, Lemma 3 implies *spreading*.

In (7.5) we obtained $\Psi_l' = 0$ for $J_2 = 0$, $1 \leq l \leq p - 1$, [p as in (7.8)]. The corresponding Ψ_l are the *conserved quantities* of Newman and Penrose.[11]

(B) We are now going to prove that "almost all" solutions spread. To this end we observe that an incoming solution (as defined in Sec. 4) is fully described by the two coefficients j_0, $j(u, v)$ in the normal form, and by its initial state $\Psi := \psi(0, v)$. Consequently, the triplets of functions $\psi := \{j_0, j, \Psi\}$ characterize solutions. They form in a natural way an infinite-dimensional vector space. Each finite-dimensional linear subspace carries a natural topology induced by its Euclidean metric. We can therefore legitimately talk of *neighborhoods* of *solutions* whenever we deal with finite-dimensional linear subspaces.[12] With these definitions we prove:

Lemma 4: In every finite-dimensional linear subspace of solutions which contains a spreading one, the *spreading* solutions form an *open set* whose boundary points are those with characteristic propagation.

Proof: We first show that the spreading solutions form an open set. To this end, consider a one-dimensional neighborhood of a given solution:

$$\psi = \psi^{(0)} + \epsilon\psi^{(1)}, \quad \psi^{(k)} := \{j_0^{(k)}, j^{(k)}, \Psi^{(k)}\}. \tag{7.9}$$

ψ is linear, hence analytic in ϵ. A well-known existence theorem[13] guarantees that the corresponding solutions must likewise be *analytic* in ϵ so that for sufficiently small $|\epsilon|$,

$$\psi = \sum_{k=0}^{\infty} \epsilon^k \psi^{(k)} \tag{7.10}$$

with well determined functions $\psi^{(k)}(u, v)$. By assumption, the zero-order term $\psi^{(0)}$ spreads. Consequently, ψ must spread for sufficiently small $|\epsilon|$. The generalization from one dimension to a finite number of them is obvious.

In order to show that the solutions with characteristic propagation form boundary points, we consider a "straight line" connecting one of them with a spreading solution. On this line, the set of characteristic propagation solutions cannot have *interior* points. For suppose there was an interior point, and consider a nearest boundary point $\psi^{(0)}$. By assumption, in any neighborhood of $\psi^{(0)}$ there are spreading solutions, which implies that at least one of the "coefficients" $\psi^{(k)}$ in the expansion (7.10) of ψ spreads. But then ψ must spread in a full ϵ neighborhood except possibly at $\epsilon = 0$, and one arrives at a contradiction. Consequently, the characteristic propagation solutions can be nowhere open in a finite-dimensional subspace which contains a spreading solution.

(C) A third type of expansion, called expansion in *progressing waves*, is used in the literature for the construction of solutions.[14] It can likewise be used for a discussion of spreading. One makes the ansatz

$$\psi = \sum_{k=0}^{\infty} g_k(u, v)\chi_k(v) \tag{7.11}$$

with

$$\chi_k := \int_0^v dv' \chi_{k-1}, \quad \chi_0 \text{ arbitrary}, \tag{7.12}$$

and finds that the right normal form (4.1) is solved identically in the χ_k by

$$\begin{aligned} \partial_u g_0 &= 0, \\ j_0 \partial_u g_{k+1} &= (j - \partial_v j_0 \partial_u) g_k. \end{aligned} \tag{7.13}$$

One has a freedom in the choice of the *initial data* $g_k(0, v)$ of which one can dispose by setting

$$g_k(0, v) = \delta_k^0. \tag{7.14}$$

This choice implies that $\chi_0(v)$ is the initial state. An application to problems with a right-terminating sequence teaches, however, that there is another natural choice which in some cases can be defined by

$$g_k(-\infty, v) = \delta_k^0. \tag{7.15}$$

In general, the initial state is given by (7.11) with

[11] The "conserved quantities" of an asymptotically flat space-time were discovered, and described in the following: (a) E. T. Newman and R. Penrose, Phys. Rev. Letters **15**, 231 (1965); (b) E. T. Newman and R. Penrose, Proc. Roy. Soc. (London) Ser. A (1968) (to be published).

[12] The topology induced in this way on the full space is the *inductive limit* topology.

[13] See Ref. 9.

[14] See, e.g., Ref. 2(c), pp. 620–636, 760–764.

$u = 0$, and its convergence must be postulated. If (7.11) converges for $u = 0$, it converges (at least) for some neighborhood of $u = 0$.

Can it happen that the series in (7.11) is finite? As a consequence of (7.12), a *termination* implies that g_k vanishes identically beyond some k. We will show that this can only happen for *right-terminating* sequences. If the sequence does not terminate, its individual terms will eventually have noncompact support according to Lemma 3. In general, therefore, the solution will *spread* for nonright-terminating sequences. But there are exceptional cases in which the infinite series (7.11) has a compact support.

Let us derive a *closed-form expression* for the functions $g_k(u, v)$ satisfying (7.13) with $g_0 = 1$. To this end we define the modified derivatives of g_k:

$$g_k^{[l]} := \partial_u \frac{j_{l-2}}{j_{l-1}} \partial_u \frac{j_{l-3}}{j_{l-2}} \partial_u \cdots \frac{j_1}{j_2} \partial_u \frac{j_0}{j_1} \partial_u g_k, \quad l \geq 2,$$
$$g_k^{[1]} := \partial_u g_k, \tag{7.16}$$

and claim

$$g_{k+1}^{[l]} = j_{l-1}^{-2} j_{l-2} j_l g_k^{[l-1]} - j_{l-1}^{-1} \partial_v j_{l-1} g_k^{[l]}, \quad l \geq 2,$$
$$g_{k+1}^{[1]} = j_0^{-1} j_1 g_k - j_0^{-1} \partial_v j_0 g_k^{[1]}. \tag{7.17}$$

The bottom line is equivalent to the defining equation (7.13); it starts a *proof* by *induction*. We skip the step $l = 1 \to l = 2$ as it is perfectly analogous to the general step $l \to l + 1$. Assume (7.17) for l, multiply by $j_{l-1} j_l^{-1}$, and integrate by parts:

$$\frac{j_{l-1}}{j_l} g_{k+1}^{[l]} = \frac{j_{l-2}}{j_{l-1}} g_k^{[l-1]} - \partial_v \frac{j_{l-1}}{j_l} g_k^{[l]} - \frac{j_{l,v}}{j_l} \frac{j_{l-1}}{j_l} g_k^{[l]}. \tag{7.18}$$

Application of ∂_u yields, with (7.16),

$$g_{k+1}^{[l+1]} = g_k^{[l]} - \partial_v g_k^{[l+1]} - D j_l \cdot \frac{j_{l-1}}{j_l} g_k^{[l]} - \frac{j_{l,v}}{j_l} g_k^{[l+1]}, \tag{7.19}$$

and $D j_l (j_l^{-1} j_{l-1}) = 1 - j^{l-1} (j_l^2)^{-1} j_{l+1}$ leads to

$$g_{k+1}^{[l+1]} = j_l^{-2} j_{l-1} j_{l+1} g_k^{[l]} - j_l^{-1} \partial_v j_l g_k^{[l+1]}. \tag{7.20}$$

Q.E.D.

Next we have $g_0 = 1$, whence $g_1^{[1]} = j_0^{-1} j$. We claim

$$\boxed{g_k^{[k]} = \frac{j_k}{j_{k-1}}}. \tag{7.21}$$

For the proof by induction from k to $k + 1$, insert (7.21) into (7.17) for $l = k + 1$, and observe that $g_k^{[l]} = 0$ for $l \geq k + 1$.

Formula (7.21) shows that $g_k^{[k]}$ vanishes for a $(k - 1)$th-order *right-terminating* sequence, so that $g_k = 0$ can be achieved by a suitable choice of the initial data (constants of integration, which are functions of v). We omit a further analysis along these lines.

(D) Let us mention that there are yet *other expansions* which lend themselves to an analysis of the solutions. For instance, the asymptotic expansion in terms of inverse powers can be paralleled by an expansion in terms of exponentials e^{-kx}. However, with none of these expansions were we able to give *exhaustive* criteria for spreading.

8. CLOSED-FORM EVALUABLE SEQUENCES; EXAMPLES

In the following sections we will prove a number of astonishing properties of substitution sequences, some of which have already been listed in Sec. 5. Our main concern is the finding of *closed-form expressions* for the elements of certain classes of sequences, or at least of *criteria* for special cases to occur, in order to be able to tell the properties of a given differential equation. In this section we present some formulas and simple examples.

(A) The substitution sequence was defined in Eq. (4.10) for any given pair of functions j_0, j_1. One finds recursively [compare (3.5)]

$$\frac{j_2}{j_1} = \frac{j_1}{j_0} - D j_1 =: \frac{j_1}{j_0} I_1, \quad I_1 = 1 - E j_1, \tag{8.1}$$

where I_1 is an invariant [compare (3.5)] and

$$\frac{j_3}{j_2} = \frac{j_1}{j_0} - D j_1 j_2 =: \frac{j_1}{j_0} I_2, \quad I_2 = 1 - E^2 \frac{j_1}{j_0} I_1, \tag{8.2}$$

$$\boxed{\frac{j_{k+1}}{j_k} = \frac{j_1}{j_0} - D j_1 j_2 \cdots j_k =: \frac{j_1}{j_0} I_k}. \tag{8.3}$$

From the identity

$$j_{k+1} = \frac{j_{k+1}}{j_k} \frac{j_k}{j_{k-1}} \cdots \frac{j_2}{j_1} j_1$$

one obtains

$$\boxed{j_{k+1} = j_1^{k+1} j_0^{-k} I_k I_{k-1} \cdots I_2 I_1}, \tag{8.4}$$

and from (8.3)

$$I_k = 1 - E j_1 j_2 \cdots j_k. \tag{8.5}$$

Insertion of (8.4) into (8.5) yields

$$\boxed{I_k = 1 - E j_1^{\binom{k+1}{2}} j_0^{-\binom{k}{2}} I_1^{k-1} I_2^{k-2} \cdots I_{k-2}^2 I_{k-1}}. \tag{8.6}$$

This equation is a recursive formula for the *invariants* I_k defined in (8.3). Once the I_k are known, the elements $\mathcal{J}_k$ of the sequence are given by Eq. (8.4). We shall need these formulas in the sequel.

The invariants I_k are *constant* if and only if $E\mathcal{J}_1$ and $E\mathcal{J}_0$ are constant. In this case, all members of the sequence can be expressed in closed form, which we have done in Eqs. (5.2) and (5.3). Let us find the corresponding functions $\mathcal{J}_1, \mathcal{J}_0$! We distinguish several cases:

Case 1:

$$E\mathcal{J}_0 = 0$$

means

$$\partial_v \partial_u \ln|\mathcal{J}_0| = 0, \tag{8.7}$$

whose general solution is

$$\mathcal{J}_0 = g(u)h(v). \tag{8.8}$$

In this case, a conformal plus, factor gauge can be applied to achieve $\mathcal{J}_0 = 1$; and the remaining equation $E\mathcal{J}_1 = c_1$ simplifies to ($\mathcal{J}_1 =: \mathcal{J}$)

$$D\mathcal{J} = c\mathcal{J}. \tag{8.9}$$

We derive the complete solution of this equation in Sec. 11.

At present, let us content ourselves with the special case of a *static* equation for which (8.9) becomes

$$\ln''|\mathcal{J}| = c\mathcal{J}, \quad \left(' := \frac{d}{dx}\right). \tag{8.10}$$

Multiplication with $2\ln'|\mathcal{J}|$ yields

$$(\ln'|\mathcal{J}|)^{2'} = 2c\mathcal{J}'. \tag{8.11}$$

Again we have to distinguish different cases.

(a) $c = 0$ implies $\ln'|\mathcal{J}| = a$ ($a, b =$ const), whence

$$\boxed{\mathcal{J} = be^{ax}}. \tag{8.12}$$

For $c \neq 0$, Eq. (8.11) integrates to

$$(\ln'|\mathcal{J}|)^2 = 2c(\mathcal{J} + a). \tag{8.13}$$

We distinguish further:

(b) for $a = 0$ one gets $\mathcal{J}' = (2c\mathcal{J}^3)^{\frac{1}{2}}$, so that

$$\boxed{\mathcal{J} = \frac{2}{c(x - x_0)^2}}. \tag{8.14}$$

(c) For $a \neq 0$, (8.13) leads to

$$(2c)^{\frac{1}{2}}(x - x_0) = \int \frac{d\mathcal{J}}{\mathcal{J}(\mathcal{J} + a)^{\frac{1}{2}}} = \frac{2}{a^{\frac{1}{2}}} \begin{Bmatrix} \operatorname{arc\,tanh} f \\ \operatorname{arc\,coth} f \end{Bmatrix},$$

$$f := \left(1 + \frac{\mathcal{J}}{a}\right)^{\frac{1}{2}}, \tag{8.15}$$

whence

$$\boxed{\mathcal{J} = -a \cosh^{-2}\left[\left(\frac{ac}{2}\right)^{\frac{1}{2}}(x - x_0)\right]} \tag{8.16}$$

or

$$\boxed{\mathcal{J} = a \sinh^{-2}\left[\left(\frac{ac}{2}\right)^{\frac{1}{2}}(x - x_0)\right]}. \tag{8.17}$$

In these last two solutions of (8.10), the product ac can be positive or negative (without violating the reality of $\mathcal{J}$). If it is negative, "cosh" or "sinh" change into "sin" of some real argument. The four classes of solutions (8.12), (8.14), (8.16), and (8.17) exhaust all solutions of Eq. (8.10). Note that (8.14) can be obtained from (8.17) as a limiting case as $a \to 0$. This completes our discussion of Case 1.

Case 2: $E\mathcal{J}_0 = c_0$, $E\mathcal{J} = c$, $c_0 c \neq 0$. We conclude $\mathcal{J} = \mathcal{J}_0^{c/c_0}$, whence

$$\mathcal{J}\mathcal{J}_0^{-1} = \mathcal{J}^\alpha, \quad \alpha := 1 - c_0 c^{-1} \tag{8.18}$$

and

$$D\mathcal{J} = c\mathcal{J}^\alpha. \tag{8.19}$$

Again, two cases have to be distinguished:

(a) $\alpha = 0$. The general solution of (8.19) is obtained from (8.8) by multiplication with a special solution; it reads

$$\boxed{\mathcal{J} = e^{c\alpha^2/2} g(u)h(v)}; \tag{8.20}$$

(b) $\alpha \neq 0$. In this case we set

$$f := \mathcal{J}^\alpha, \tag{8.21}$$

and obtain from Eq. (8.19)

$$Df = \alpha c f, \tag{8.22}$$

which is identical in shape with Eq. (8.9). We therefore know all sequences with *constant invariants.*

(B) A fairly general shape of a wave equation encountered in physics is

$$\{\partial_y \lambda(y)\partial_y - \mu(y)\partial_t^2 - \nu(y)\}\phi = 0. \tag{8.23}$$

For instance, a vibrating rope is described by (8.23) with $\lambda(y)$ being the *tension coefficient*, $\mu(y)$ the *mass density* per unit length, and $\nu(y) = 0$. The reduction to normal form is achieved by setting

$$\frac{dx}{dy} := \left(\frac{\mu}{\lambda}\right)^{\frac{1}{2}}, \quad \begin{Bmatrix} u := (x - t)/2 \\ v := (x + t)/2 \end{Bmatrix}, \tag{8.24}$$

and

$$\phi =: (\lambda\mu)^{-\frac{1}{4}}\psi. \tag{8.25}$$

One finds

$$\mathcal{J} = (\lambda\mu)^{-\frac{1}{4}}\partial_x^2(\lambda\mu)^{\frac{1}{4}} + \mu^{-1}\nu, \quad \mathcal{J}_0 = 1. \tag{8.26}$$

The corresponding sequence is *static* and *symmetric* as it has to be for a time-independent medium without dissipation. Nothing can be said about termination before one has specified the two functions $\lambda\mu(x)$ and $\mu^{-1}\nu(x)$.

The propagation of an *electromagnetic test field* in an *empty spherically symmetric* space-time can be essentially described by an equation of the shape (8.23). (For a "test field" one ignores its reaction upon the space-time geometry, i.e., replaces the right-hand side of Einstein's equation by zero.) The electric and magnetic field form a bivector $F_{\alpha\beta} = F_{[\alpha\beta]}$ (in 4-dimensional language) which satisfies Maxwell's equations

$$\Phi_{[\alpha\beta,\lambda]} = 0, \quad \Phi_{\alpha\beta} := F_{\alpha\beta} + i\overset{*}{F}_{\alpha\beta}, \qquad (8.27)$$

in the absence of electric sources. (A comma denotes ordinary differentiation; an asterisk denotes passing to the dual.) The *complex bivector* $\Phi_{\alpha\beta}$ has only three linearly independent (complex) components each of which obeys a generalized covariant wave equation[15] as a consequence of (8.27). For spherical symmetry, it is convenient to introduce a *tetrad* which is covariantly constant along the (outgoing) *radial null geodesics* and to expand the tetrad components of $\Phi_{\alpha\beta}$ in terms of *spherical harmonics*. Call "Φ" the contraction of $\Phi_{\alpha\beta}$ with the timelike tetrad bivector containing the (outgoing) radial null direction, and $r^{-1}\phi_l := \Phi_l(r, t)$ its lth expansion coefficient (l describing the representation of the rotation group). One then finds[16] for ϕ_l an equation of the shape (8.23) with

$$\lambda(r) = r^2\left(1 - \frac{2m}{r}\right),$$
$$\mu(r) = r^2\left(1 - \frac{2m}{r}\right)^{-1}, \qquad (8.28)$$
$$\nu(r) = l(l+1) - \frac{2m}{r},$$

where r stands for y and m is the mass parameter of the spherically symmetric empty space-time (which is known to be a *Schwarzschild* space-time). The corresponding normal form can be easily gathered from Eqs. (8.24), (8.25), and (8.26):

$$\frac{dx}{dr} = \frac{r}{r - 2m}, \qquad \psi = r^2\Phi_l, \qquad (8.29)$$

and

$$\boxed{j = \left(1 - \frac{2m}{r}\right)\frac{l(l+1)}{r^2}}, \quad j_0 = 1. \qquad (8.30)$$

Note that for $m \neq 0$, r is different from the normalized radial coordinate x defined in (8.29).

Let us mention that the *scalar wave equation* in a Schwarzschild space-time can be analogously reduced and leads to

$$\boxed{j = \left(1 - \frac{2m}{r}\right)\left[\frac{l(l+1)}{r^2} + \frac{2m}{r^3}\right]}, \quad j_0 = 1. \qquad (8.31)$$

Formulas (8.30) and (8.31) deserve a discussion. In *flat space-time*, i.e., for $m = 0$, we have $x = r$, and j is of the form (8.14) (with constant invariants). Equation (5.2) shows that the sequence is (double-) *terminating* of order l. All potentials of order l give rise to characteristic propagation. Physically, l is the order of the *multipole moment* considered.

In *curved space-time*, i.e., for $m \neq 0$, the sequences determined by (8.30) or (8.31) *cannot terminate* because j is not a rational function in x and $e^{\omega x}$ (see Sec. 11). In order to show that all the corresponding solutions ψ with an asymptotic expansion around infinity must *spread*, we have to slightly modify the approach taken in Sec. 7 under (A): We expand ψ in terms of negative powers of r (rather than x), and enter into Eq. (7.2). As a result, we get a three-term recursion relation for the expansion coefficients $\Psi_k(u)$ of Ψ, which reads

$$k\Psi_k' = [k(k-1) - l(l+1)]\Psi_{k-1} - 2m\begin{Bmatrix} k(k-2) \\ (k-1)^2 \end{Bmatrix}\Psi_{k-2} \qquad (8.32)$$

for $k \geq 1$. Here the upper line refers to (8.30), the lower line to (8.31). It is now easily seen that no Ψ_k' can vanish identically (for $m \neq 0$), so that Lemma 3 implies spreading.

The *scalar monopole* wave in a Schwarzschild space-time is by itself a nontrivial example. Here, from Eqs. (8.1) and (8.2), one finds

$$\frac{j_2}{j} = -\frac{3}{r^2}\left(1 - \frac{2m}{r}\right)\left(1 - \frac{6m}{r}\right),$$
$$\frac{j_3}{j_2} = -\frac{2}{r^2}\left(1 - \frac{2m}{r}\right)\left(1 - \frac{6m}{r}\right)^{-2} \times \left(4 - 77\frac{m}{r} + 24 \cdot 19\frac{m^2}{r^2} - 36 \cdot 25\frac{m^3}{r^3}\right). \qquad (8.33)$$

[15] R. C. Tolman, *Relativity Thermodynamics and Cosmology* (Oxford University Press, London, 1934), Eq. (108.1).

[16] A discussion of Maxwell's equations formulated for complex tetrad components of the field bivector can be found in the following: (a) E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962); (b) A. I. Janis and E. T. Newman, J. Math. Phys. **6**, 902 (1965).

No unexpected simple structure law has been recognized by us.

9. CRITERIA FOR NONTERMINATION

How can one know whether a given substitution sequence terminates or not? We are going to give partial answers to this important question for *static* sequences. For simplicity's sake let us assume all sequences considered to be analytic. Their elements can be locally represented by convergent power series in x or x^{-1}. Let us write

$$f_K(x) := \sum_{k=K}^{\infty} a_k x^k, \quad a_K \neq 0, \tag{9.1}$$

for any power series which starts with a term of order K, where K stands for an arbitrary integer. The two expansions mentioned above need separate treatment.

(A) Assume j_0 and j to be of *finite order* at *infinity*, i.e.,

$$j_0^{-1} j = f_K(x^{-1}), \quad j = f_L(x^{-1}). \tag{9.2}$$

We claim

Lemma 5: The sequences defined by Eq. (9.2) cannot terminate to the right (left) if

$$\begin{aligned} &\text{(a) } K \leq 1, \\ &\text{(b) } K = 2 \text{ and } a_2 \neq c_n, \\ &\text{(c) } K \geq 3 \text{ and } \begin{Bmatrix} L \geq 0 \\ K \geq L \end{Bmatrix} \text{ for } \begin{Bmatrix} \text{right} \\ \text{left} \end{Bmatrix} \text{ termination.} \end{aligned} \tag{9.3}$$

Here, a_2 is the coefficient of x^{-2} in the expansion of $j_0^{-1} j$, and c_n runs through the sequence ($n \geq 1$)

$$c_n = \begin{Bmatrix} n(n+L-1) \\ n(n+K-L-1) \end{Bmatrix} \text{ for } \begin{Bmatrix} \text{right} \\ \text{left} \end{Bmatrix} \text{ termination.} \tag{9.4}$$

Proof: Equation (8.3) implies

$$\frac{j_{n+1}}{j_n} = \frac{j}{j_0} - D j^n \left(\frac{j_2}{j}\right)^{n-1} \left(\frac{j_3}{j_2}\right)^{n-2} \cdots \left(\frac{j_n}{j_{n-1}}\right). \tag{9.5}$$

A trivial calculation shows

$$D f_N(x^{-1}) = N x^{-2} + f_3(x^{-1}), \tag{9.6}$$

so that under the conditions of the Lemma

$$j_n^{-1} j_{n+1} = f_K(x^{-1}) + f_2(x^{-1}). \tag{9.7}$$

Obviously, the right-hand side cannot vanish if $K \leq 1$, which implies that the sequence cannot terminate (to the right) in this case. If $K = 2$, it may happen that the term of order -2 on the right-hand side cancels for some n. In this situation, one finds $f_2(x^{-1}) = -c_n x^{-2} + f_3(x^{-1})$ with c_n given in (9.4) so that cancellation (to the right) cannot happen if $a_2 \neq c_n$; [see the calculation following Eq. (12.1)]. Finally, if $K \geq 3$ and $L \geq 0$, the term of order -2 in (9.7) is strictly negative. In order to study termination to the left, one need only consider the behavior to the right of the mirrored sequence $\tilde{j}_k = j_{-k}^{-1}$, in which $j_0^{-1} j$ and j_0^{-1} play the roles of $j_0^{-1} j$ and j before. One has

$$j_0^{-1} = f_{K-L}(x^{-1}), \tag{9.8}$$

so that in (9.3) and (9.4), L has to be replaced by $K - L$ for left-termination. Lemma 5 is thereby proven.

(B) Assume j_0 and j to be of *finite order* at some *point* x_0 which we choose as the origin of a Taylor expansion, i.e.,

$$j_0^{-1} j = f_K(x), \quad j = f_L(x). \tag{9.9}$$

We claim

Lemma 6: The sequences defined by Eq. (9.9) cannot terminate to the right (left) if

$$\begin{aligned} &\text{(a) } K \leq -3, \\ &\text{(b) } K = -2 \text{ and } \begin{Bmatrix} \text{either } a_{-1} \neq 0 \\ \text{or } \quad a_{-2} \neq d_n \end{Bmatrix}, \\ &\text{(c) } K = -1, \\ &\text{(d) } K \geq 0 \text{ and } \begin{Bmatrix} L \leq -1 \\ K \leq L - 1 \end{Bmatrix} \text{ for } \begin{Bmatrix} \text{right} \\ \text{left} \end{Bmatrix} \\ &\qquad\qquad \text{termination.} \end{aligned} \tag{9.10}$$

Here a_k are the coefficients of x^k in the expansion of $j_0^{-1} j$, and d_n runs through the sequence ($n \geq 1$)

$$d_n = \begin{Bmatrix} n(n-L-1) \\ n(n-K+L-1) \end{Bmatrix} \text{ for } \begin{Bmatrix} \text{right} \\ \text{left} \end{Bmatrix} \text{ termination.} \tag{9.11}$$

Proof: In the present case, formula (9.5) takes the form

$$j_n^{-1} j_{n+1} = f_K(x) + N x^{-2} + f_0(x), \tag{9.12}$$

whose right-hand side can clearly not vanish (identically) for $K \leq -3$ or $a_{-1} \neq 0$. This proves (a), (c), and the first line of (b). For $K = -2$, the term of order -2 on the right-hand side can only vanish for some n if a_{-2} assumes one of the integers given in (9.11); this finishes the proof of (b). Finally, if $K \geq 0$, and $L \leq -1$, the pole term on the right-hand side is strictly negative. Again, as in the proof of Lemma 5, the discussion of non-left-termination is reduced to that of non-right-termination by considering the mirrored sequence.

10. IDENTITIES FOR LEFT-TERMINATING SEQUENCES

We are going to derive an infinite set of identities among the elements of left-terminating sequences. More precisely, we assume $j_{-1}^{-1} \equiv 0$. In this case, the general incoming solution has the shape (4.12) for $k = 0$:

$$j_0\psi = \int_0^v dv' j_0(u, v')\psi_0(v'). \tag{10.1}$$

Assume j_0 to be analytic, and substitute it by its Taylor series expansion in u around $u = u_0$. As a result

$$\partial_u^k j_0\psi(u_0, v) = \int_0^v dv' \partial_u^k j_0(u_0, v') \cdot \psi_0(v'). \tag{10.2}$$

Now express ψ_0 as a modified derivative of its first potential ψ [according to (4.9)], perform an integration by parts, express ψ as a modified derivative of ψ_2, integrate by parts, and so on. After $k + 1$ steps, one gets (with j_{-1} replaced by "one" in the sum)

$$\partial_u^k j_0\psi = \sum_{l=0}^{k} j_0^{[k,l]} j_{l-1}\psi_{l+1} + \int_0^v dv' j_0^{[k,k+1]} j_k \psi_{k+1}, \tag{10.3}$$

where

$$\boxed{j_0^{[k,l]} := (-1)^l \partial_v \frac{j_{l-2}}{j_{l-1}} \partial_v \frac{j_{l-3}}{j_{l-2}} \partial_v \cdots \frac{j_0}{j_1} \partial_v \frac{\partial_u^k j_0}{j_0}} \tag{10.4}$$

for $l \geq 2$, and

$$j_0^{[k,0]} := \partial_u^k j_0 \quad \text{for} \quad l = 0,$$
$$j_0^{[k,1]} := -\partial_v j_0^{-1} \partial_u^k j_0 \quad \text{for} \quad l = 1.$$

$j_0^{[k,l]}$ is a modified lth-order v derivative of an ordinary kth-order u derivative of j_0.

On the other hand, one can give a proof by *induction* of the following linear expansion:

$$\partial_u^k j_0\psi = \sum_{l=0}^{k} b_l^k j_l \psi_{l+1} \quad \text{for} \quad k \geq 1, \tag{10.5}$$

with

$$b_l^k = 0 \quad \text{for} \quad k < l, \quad b_k^k = 1,$$
$$b_l^{k+1} = b_{l-1}^k + j_l^{-1} \partial_u j_l b_l^k, \tag{10.6}$$

which is perfectly analogous to the proof of formula (4.16).

We compare Eqs. (10.3) and (10.5), which have to be equal for $j_{-1}^{-1} = 0$ identically in ψ. Let us show that $j_0^{[k,k+1]}$ must vanish by setting $\psi_{k+1} = j_0^{[k,k+1]} j_k$ for $0 \leq v \leq v_0$, and $\psi_{k+1} = 0$ for $v > v_0$. In this case, all the terms in the two sums vanish for $v > v_0$, whereas the integral term in (10.3) is positive definite. This implies the vanishing of $j_0^{[k,k+1]}$ if $j_k \not\equiv 0$. (Otherwise, for $j_k \equiv 0$, $j_0^{[k,k+1]}$ would not be defined.) Next we claim that the two sums are identical term by term. For a proof, choose $j_l\psi_{l+1} = \text{const}$ recursively for $l = k, k - 1, \cdots, 1$, in which case all the lower-order terms vanish identically. We can now compare the respective coefficients of ψ_{l+1}, and obtain

$$\boxed{j_0^{[k,l]} = b_l^k \frac{j_l}{j_{l-1}}} \quad \text{for} \quad j_{-1}^{-1} = 0. \tag{10.7}$$

Note especially that $b_k^k = 1$, so that

$$\frac{j_k}{j_{k-1}} = j_0^{[k,k]}. \tag{10.8}$$

A comparison of these formulas with Eqs. (8.3) and (8.6) reveals remarkable *identities*.

11. DOUBLE-TERMINATING SEQUENCES

With the knowledge of formula (10.8), we are ready to derive closed-form expressions for the *left-end elements* of all double-terminating sequences.

To this end let us number the elements of the sequence such that j_0 is the left-end element. Under these circumstances, *double termination* is equivalent with

$$j_0^{[n,n]} = 0 \tag{11.1}$$

identically for some n. This differential equation can be easily integrated with respect to v if repeated use is made of Eq. (10.8) and of the definition (10.4). One finds

$$\sum_{l=0}^{n} c_l \partial_u^l j_0 = 0, \quad c_l = c_l(u), \quad c_n \equiv 1. \tag{11.2}$$

That is, j_0 satisfies an ordinary nth-order differential equation in u with arbitrary u-dependent coefficients. Its general solution is necessarily of the form

$$\boxed{j_0 = \sum_{k=1}^{n} g_k(u) h_k(v)}. \tag{11.3}$$

Here g_k and h_k are arbitrary functions. In order to see this let us assume $f(u, v)$ of the shape (11.3), and derive for it an equation of the form (11.2). We u-differentiate f up to $n - 1$ times:

$$\partial_u^l f(u, v) = \sum_{k=1}^{n} \partial_u^l g_k(u) h_k(v), \quad 0 \leq l \leq n - 1, \tag{11.4}$$

and conclude: Either the n derivatives $\partial_u^l g_k$ are linearly dependent with respect to l

$$\sum_{l=0}^{n-1} d_l(u) \partial_u^l g_k = 0, \tag{11.5}$$

in which case one gets

$$\sum_{l=0}^{n-1} d_l(u) \partial_u^l f = 0, \tag{11.6}$$

from which (11.2) is obtained by differentiation and multiplication; or there exists an inverse matrix g_l^{*m} with

$$\sum_{l=0}^{n-1} g_l^{*m}\partial_u^l g_k = \delta_k^m, \tag{11.7}$$

in which case one concludes

$$\partial_u \sum_{l=0}^{n-1} g_l^{*m}\partial_u^l f = 0 \quad \text{for} \quad 0 \leq m \leq n-1. \tag{11.8}$$

Each of these n equations takes the form (11.2) when divided by g_{n-1}^{*m}; the latter matrix elements cannot all vanish.

What is the general form of $\jmath_0$ for *double-terminating static* sequences? In this case, $\partial_u = \partial_v = \partial_x$; and from (11.1) one derives Eq. (11.2) but with constant coefficients c_l. As is well known, the general integral of an nth-order ordinary differential equation with constant coefficients is a linear combination of n terms, each being a nonnegative power of x times an exponential function in x:

$$\boxed{\jmath_0 = \sum_{k,l} c_{k,l} x^k e^{\omega_l x}}. \tag{11.9}$$

And conversely, every such sum satisfies an ordinary differential equation with constant coefficients.

It may be interesting to observe that both the functions of type (11.3) and those of type (11.9) are closed under *ring* operations (linear combination and multiplication), and differentiation.

Our next problem is the determination of all *symmetrical double-terminating* sequences. These sequences are of particular interest because they correspond to nondissipative media. Their determination turns out to be much harder than any of the earlier calculations, and we can only solve some lowest-order cases. However, from these lowest-order results one can guess at the general structure of the higher-order cases. Again, we restrict our calculations to the left-end element $\jmath_0$. But once $\jmath_0$ is determined, with $\jmath_n \equiv 1$, there are closed form expressions for its successor $\jmath_{n+1}$:

$$\begin{aligned}\jmath_{n+1} &= -D\jmath_0\jmath_1\cdots\jmath_{n-1} \\ &= (-1)^{n-1}[\jmath_0\cdot D\jmath_0\cdot D\jmath_0\jmath_1\cdot D\cdots D\jmath_0\jmath_1\cdots\jmath_{n-2}]^{-1}.\end{aligned} \tag{11.10}$$

They can be obtained from (8.3) for $\jmath_0^{-1}\jmath_1 = 0$ by shifting the sequence index by one, and observing that $\jmath_{n-1} = \jmath_{n+1}^{-1}$ for $\jmath_n \equiv 1$.

We treat the lowest nontrivial case: a *symmetrical double-terminating* sequence of *length three*. Its defining equations can be written as follows:

$$\jmath_{-1}^{-1} \equiv 0, \quad \jmath_1 = 1, \quad \jmath_3 \equiv 0. \tag{11.11}$$

Notice that they are equivalent to Eq. (8.9) for $c = 1$; but c can be absorbed into the definition of u, say.

The equations $\jmath_{-1}^{-1} \equiv 0 \equiv \jmath_3$ were solved in (11.3), viz.,

$$\jmath_0 = \sum_{k=1}^{3} g_k(u)h_k(v). \tag{11.12}$$

If neither $\jmath_1$ nor $\jmath_2$ is to vanish identically, the sum in (11.12) must not be reducible to a shorter sum; which implies that the g_k's and h_k's have *nonconstant ratios*. The remaining equation $\jmath_1 = 1$ reads [compare (8.1)]

$$1 = -\jmath_0 D\jmath_0. \tag{11.13}$$

We spell it out, and insert the expression (11.12):

$$-\sum_k g_k h_k = \sum_{k<l}(g_k'g_l - g_k g_l')(h_k'h_l - h_k h_l'). \tag{11.14}$$

(A prime denotes differentiation with respect to the argument.)

Now we observe that, by means of suitable gauges, $\jmath_0$ can be multiplied by a product of an arbitrary function of u and an arbitrary function of v. We dispose of this gauge by setting

$$g_3 = 1, \quad h_3 = 1. \tag{11.15}$$

Equation (11.14) thus simplifies to

$$\begin{aligned}&-(g_1h_1 + g_2h_2 + 1) \\ &\quad = g_1'h_1' + g_2'h_2' + (g_1'g_2 - g_1g_2')(h_1'h_2 - h_1h_2'),\end{aligned} \tag{11.16}$$

with nonconstant g_k, h_k.

Equation (11.16) implies that there exist (at least) two linear relations with constant coefficients

$$\operatorname{lin}(g_k', g_k'') = 0,$$

so that

$$\left.\begin{aligned} g_1' &= c_{11}g_1 + c_{12}g_2 + c_1u + d_1 \\ g_2' &= c_{21}g_1 + c_{22}g_2 + c_2u + d_2 \end{aligned}\right\}. \tag{11.17}$$

This step, from (11.16) to (11.17), is lengthy and is achieved by a discussion of several possible cases. We express the result in vector notation:

$$\begin{aligned} g' &= Cg + l, \quad g := \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}, \\ C &:= \begin{pmatrix} c_{11}, c_{12} \\ c_{21}, c_{22} \end{pmatrix}, \quad l := \begin{pmatrix} c_1u + d_1 \\ c_2u + d_2 \end{pmatrix}. \end{aligned} \tag{11.18}$$

The general solution of (11.18) reads

$$g = e^{Cu}\left\{a + \int^u du' e^{-Cu'} l(u')\right\}, \quad a := \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \text{const.} \tag{11.19}$$

For symmetry reasons, h must be given by an analogous expression

$$h = e^{Dv}\left\{b + \int^v dv' e^{-Dv'} m(v')\right\}. \tag{11.20}$$

In vector notation Eq. (11.16) reads

$$-1 = {}^t gh + {}^t g'h' + {}^t gTg' \cdot {}^t hTh', \quad T := \begin{bmatrix} 0, 1 \\ -1, 0 \end{bmatrix}, \tag{11.21}$$

with "t" standing for "transposed." After (11.19) and (11.20) have been inserted, there exist two cases:

(1) If $\det(C) \neq 0$, one concludes successively that $l = 0$, $TCT = {}^tC$, ${}^tCD = -1$, and $m = 0$. The corresponding solution reads

$$\boxed{\begin{aligned} &g = e^{Cu}\, a, \quad h = e^{-{}^tC^{-1}v}\, b, \\ &\operatorname{tr}(C) = 0, \quad {}^taTCa \cdot {}^tbT\,{}^tC^{-1}b = 1 \end{aligned}}. \tag{11.22}$$

It is *six-parametric*, and determines j_0 via Eq. (11.12):

$$\boxed{j_0 = 1 + {}^t gh}. \tag{11.23}$$

(2) If $\det(C) = 0$, after some discussion one finds that g_k and h_k are *polynomials* of maximal *degree* two, and (11.21) results in nine equations for twelve unknowns. Presumably, these solutions are limiting cases of (11.22).

As a last problem, we turn our attention towards *static symmetrical double-terminating* sequences. Formulas (8.16) and (8.17) give two-parametric classes of such solutions for every *length* $2l + 1$ if $c = 2/l(l + 1)$. But the defining equations

$$j_{-1}^{-1} \equiv 0, \quad j_l = 1, \quad j_{2l+1} \equiv 0 \tag{11.24}$$

result in a $2l$th-order equation for j_0, so that the general solution of order l must be *$2l$-parametric.* For $l \geq 2$, therefore, (8.16) and (8.17) do not represent the general solution.

Assume $l = 2$. The corresponding static sequence must be of the form (11.9) with precisely five terms, and the equation $j_2 = 1$ can be reduced to

$$(j_0 j_0''' - j_0'^2)(j_0'''' - 1) = (j_0 j_0'''' - 2j_0' j_0''') j_0'''' + j_0'''^3. \tag{11.25}$$

By (11.9) and the special known solutions, one is tempted to try the ansatz

$$\boxed{j_0(x) = \sum_{k=1}^{2} a_k \cos \omega_k (x - x_k) + \prod_{k=1}^{2} \omega_k^{-2}}, \tag{11.26}$$

which actually is a solution for

$$a_k = \frac{\epsilon_k}{\omega_k^2 \prod_{l<k} (\omega_k^2 - \omega_l^2)}, \quad \epsilon_k = \pm 1. \tag{11.27}$$

Moreover, each cosine function in (11.26) can be replaced by either cosh or sinh. Note that (11.26) with (11.27) is *four-parametric* as wanted.

There is good reason to conjecture that the ansatz (11.26) holds for any positive order l if the upper summation and multiplication limits are changed from 2 to l. Likewise, formula (11.22) gives rise to generalizing guesses. We leave the proofs of such generalizations as an unsolved problem.

12. CANCELLATION OF POLES

In Sec. 6 we found that accumulating zeros, or poles of $j_k^{-1} j_{k+1}$, play an important role in the discussion of spreading. Can one make general statements about the *zeros*, or *poles*, of the elements of a sequence? In what follows we restrict ourselves to *static* sequences for which we prove a surprising lemma relating the zeros of successive elements.

Suppose some member of a (static) sequence has a zero of order n_0 at x_0 whereas its predecessor is nonzero at that point. For later convenience, let us choose this member as j_0 (by shifting the index if necessary). From (8.3) we get

$$j_k^{-1} j_{k+1} = j_{-1}^{-1} j_0 - D j_0 j_1 \cdots j_k, \tag{12.1}$$

which shows that $j_0^{-1} j$ has a second-order pole (at x_0), whence j has a zero of order $n - 2$. Recursively, j_k will have a *zero* of *order* $n - 2k$ unless the product $j_0 j_1 \cdots j_k$ happens to be nonvanishing and regular. This product has a zero of order $n + (n - 2) + \cdots + (n - 2k) = (n - k)(k + 1)$; it is therefore *finite* for $k = n$. Consequently, $j_n^{-1} j_{n+1}$ is necessarily regular (at x_0), and j_{n+1} may be expected to have an nth-order pole (like j_n). If this were so, all the succeeding ratios $j_l^{-1} j_{l+1}$ would have second-order poles. But this expectation is false: j_{n+1} is necessarily regular! A direct proof of this "cancellation of poles" turned out to be hopeless except for lowest orders. We are going to sketch an indirect proof which is based upon formula (10.7). Let us collect all statements into the following:

Lemma 7: If at some point x_0 some member of a substitution sequence has an nth-order zero while its predecessor does not vanish, then its kth successor has a zero of order $n - 2k$ for $k \leq n$, and its $(n + 1)$th successor is regular.

Proof: We have already verified all statements of the Lemma except the last one: j_{n+1} is regular if j_0 has an nth-order zero. Its verification will be the concern of the rest of this section.

Let us first of all observe that under the above circumstances, j_{n+1} is regular (at x_0) if and only if $j_n^{-1} j_{n+1}$ has an nth-order zero. Secondly, we have the identity

$$j_0 j_1 \cdots j_k = j_0^{k+1} \left(\frac{j_1}{j_0}\right)^k \left(\frac{j_2}{j_1}\right)^{k-1} \cdots \left(\frac{j_k}{j_{k-1}}\right) \tag{12.2}$$

and insert it into (12.1) thus:

$$\frac{\jmath_{k+1}}{\jmath_k} = \frac{\jmath_0}{\jmath_{-1}} - \Big[(k+1)D\jmath_0 + \sum_{l=1}^{k}(k-l+1)D\frac{\jmath_l}{\jmath_{l-1}}\Big]. \quad (12.3)$$

For $k < n$, the left-hand side and (hence) the square bracket must have a second-order pole whereas $\jmath_{-1}^{-1}\jmath_0$ is of order n. This implies that

$$D\frac{\jmath_{k+1}}{\jmath_k} = D(x-x_0)^{-2}\Big[1 + O(x-x_0) + c(x-x_0)^2\frac{\jmath_0}{\jmath_{-1}}\Big] = 2(x-x_0)^{-2} + D\Big[1 + O(x-x_0) + c(x-x_0)^2\frac{\jmath_0}{\jmath_{-1}}\Big], \quad (12.4)$$

where $O(x-x_0)$ vanishes at x_0; and it transpires that the contribution of $\jmath_{-1}^{-1}\jmath_0$ to the left-hand side is (at least) of order n. But we want to verify that $\jmath_n^{-1}\jmath_{n+1}$ is of order n, which means that nth-order contributions like $\jmath_{-1}^{-1}\jmath_0$ can be omitted. Omission of $\jmath_{-1}^{-1}\jmath_0$ means dealing with a sequence that *terminates* at $k = -1$, and we can apply formula (10.7):

$$\frac{\jmath_{n+1}}{\jmath_n} \equiv \jmath_0^{[k,n+1]}(b_{n+1}^k)^{-1} \text{ modulo } (x-x_0)^n, \quad \text{for } k \geq n+1. \quad (12.5)$$

We now infer from the definition (10.4) that $\jmath_0^{[k,n+1]}$ is *regular* at x_0 for all $k \geq n+1$. Our proof would therefore be completed as soon as we could show that b_{n+1}^k had a pole of order $\geq n$ (at x_0) for some $k \geq n+1$. Our conclusion will be by contradiction, and will only be sketched.[17]

From the recursion relations (10.6) one easily derives

$$b_k^{k+1} = \partial_u \ln|\jmath_0\jmath_1\cdots\jmath_k|. \quad (12.6)$$

We conclude exactly as below Eq. (12.1) that if $\jmath_0$ has an nth-order zero (at x_0), then b_k^{k+1} has a first-order pole for $0 \leq k \leq n-1$, is regular for $k = n$, and has a first-order pole for $k = n+1$ iff $\jmath_{n+1}$ has a pole or a zero. Correspondingly, b_k^{k+l} will have an lth-*order pole* unless all the terms contributing to this pole cancel. Let us show that such a cancellation actually takes place for $n-l+1 \leq k \leq n$.

To this end we calculate the coefficient $c(k,l)$ of the lth-order pole of b_k^{k+l}:

$$c(k,l) := (x-x_0)^l b_k^{k+l}|_{x=x_0}. \quad (12.7)$$

From (12.6) for $0 \leq k \leq n$ one finds

$$c(k,1) = (k+1)(n-k) = \Big(\frac{n+1}{2}\Big)^2 - \kappa^2, \qquad \kappa := k - \frac{n-1}{2}. \quad (12.8)$$

Note that κ runs through the half odds (rather than integers) for n even. Next, from Eqs. (10.6) one gathers

$$c(k,2) = \sum_{\kappa=-(n-1)/2}^{k-(n-1)/2}\Big[\Big(\frac{n+1}{2}\Big)^2 - \kappa^2\Big](-2\kappa), \quad \text{for } 0 \leq k \leq n. \quad (12.9)$$

The latter sum is an *even function* of $k-(n-2)/2$, which therefore vanishes for $k = n-1$, and trivially for $k = n$. This is a special case of the general formula

$$\sum_{\kappa=-l}^{m} \text{odd}\,(\kappa) = \text{even}\,(m+\tfrac{1}{2}), \quad \text{for } |m| \leq l \quad (12.10)$$

whenever κ increases in steps of one; which is verified by observing that

$$\sum_{\kappa=-l}^{m-\frac{1}{2}} - \sum_{\kappa=-l}^{-m-\frac{1}{2}} = \sum_{\kappa=-(m-\frac{1}{2})}^{m-\frac{1}{2}}. \quad (12.11)$$

Along the same lines one gets for the $(l+1)$th-order pole-term coefficient

$$c(k,l+1) = -2\sum_{\kappa=-(n-1)/2}^{k-(n-1)/2} c(\kappa,l)\Big(\kappa + \frac{l-1}{2}\Big), \quad \text{for } 0 \leq k \leq n, \quad (12.12)$$

and a proof by induction shows that $c(k,l)$ is an even function of $k-(n-l)/2$ which *vanishes* for $n-l+1 \leq k \leq n$.

What is the leading pole-term of b_{n+1}^k? From our above result, and the recursive relation (10.6) one learns that the only contribution to the lth-order pole of b_{n+1}^{n+1+l} comes from $\jmath_{n+1}$, and is given by

$$c(n+1,l) = m(m-1)\cdots(m-l+1), \quad l > 0, \quad (12.13)$$

when $\jmath_{n+1}$ has an mth-order zero. If $\jmath_{n+1}$ had a pole, m would be negative, and $c(n+1,l)$ could never vanish. Which would lead to a contradiction for $l \geq n$. This proves that $\jmath_{n+1}$ must be regular.

ACKNOWLEDGMENTS

The authors would like to thank Dr. J. Ehlers, Dr. A. Exton, Dr. A. Janis, Dr. A. Trautman, and D. Pajorski for many valuable discussions. A particular thanks is extended to Dr. R. Kreps.

One of the authors (W. K.) wishes to express his gratitude to the University of Pittsburgh and to NASA for supporting this work.

[17] The following formulas have been developed in collaboration with Dr. R. Kreps.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Class of Solutions of the Physical 3-Particle Unitarity Equations in Static S-Matrix Theory

JEAN-PIERRE LEBRUN*†
Institute for Theoretical Physics, McGill University, Montreal, Canada

(Received 9 January 1968)

A class of formal exact solutions of the 3-particle unitarity equations are obtained under the restriction that the S matrix be invariant under a contraction of the Poincaré group known as the static limit. The main tool is the study of a homogeneous, multidimensional, Riemann–Hilbert boundary-value problem in the complex plane of a subenergy variable. It allows one to construct an ansatz for the S matrix that satisfies the four sets of 3-particle unitarity equations recurrently; i.e., the sets of Eqs. (II), (III), and (IV) are automatically satisfied, provided that (I) is satisfied. The set (I) is then reduced to a nonlinear boundary-value problem on the inelastic cut in the complex energy plane of the scattering amplitudes. The most general solution is obtained in terms of a particular solution; an algorithm is then described that produces such a particular solution. Finally, the relation of crossing to our exact 3-particle unitary S matrix is discussed.

1. INTRODUCTION

It has been frequently suggested that an approximate scheme for the S matrix involving only normal threshold singularities can serve as a basis for a complete S-matrix theory encompassing all Landau–Cutkosky singularities.[1]

The multiparticle physical unitarity equations are interpreted as giving the compound discontinuities of the connected amplitudes—considered as the real boundary values of analytic functions in the product of the independent variable's complex planes—across the energetically permissible normal threshold cuts.[2]

The higher Landau singularities are then generated by iteration of the unitarity equations and the theory is built in successive stages: general properties such as Hermitian analyticity, the existence of particle poles in unphysical regions, the TCP theorem, and crossing symmetry are first made plausible within the normal threshold singularity scheme and then reconsidered at each successive stage of sophistication.

Little has been achieved so far in terms of actual solution of the multiparticle physical unitarity equation beyond the elastic approximation, which is easily enforced on the partial-wave amplitudes.[3]

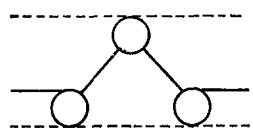


FIG. 1. Triangle singularity absent in static limit.

However, besides the interest of getting insight into the highly nonlinear structure of the unitarity equations—for instance, with a view toward continuation in the angular-momentum variables—there is at least one physical case for which the next dominant singularity is the second normal threshold: the lowest inelastic channel coupled to the $N\pi$ system is the $N\pi\pi$ system, and it is the only one of interest up to 1 GeV (approximately); it influences even the low-energy region by making certain partial waves $(P_{11}^{+}, D_{13}^{-}, \cdots)$ strongly inelastic.

In this paper, a class of formal, exact solutions of the 3-particle physical unitarity equations[4] will be described under the restriction that the S matrix be invariant under a "contraction" of the Poincaré group known as the "static limit."

This limiting situation would be obtained if the particles making the asymptotic states could be classified in two kinds: infinite mass "baryons" interacting with relativistic spinless "mesons." The ensuing considerable technical simplification is the existence of a common partial-wave expansion for all "baryon"–"meson" pairs. The number of independent kinematic variables is reduced from 2, 5, and 8 for the 4-, 5-, and 6-point connected amplitudes to 1, 2, and 3, respectively. An extra bonus for our purpose is the absence of triangle singularities of the type represented on Fig. 1: the Landau curve degenerates in the static limit.

* Research supported by the National Research Council of Canada under tenure of a postdoctoral fellowship.

† Certain results described here were derived in a different context in the author's Ph.D. thesis, M.I.T. (May, 1967, unpublished).

[1] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S Matrix* (Cambridge Univ. Press, London, 1966), and original literature cited therein; in particular, see the following: D. I. Olive, Phys. Rev. **135B**, 745 (1964); Nuovo Cimento **37**, 1422 (1965); J. Gunson, J. Math. Phys. **6**, 827, 852 (1965); H. P. Stapp, *High Energy Physics and Elementary Particles*, A. Salam, Ed. (I.A.E.A., Vienna, 1965).

[2] See, for instance, J. B. Boyling, Nuovo Cimento **33**, 1356 (1964b).

[3] Particular 3-particle unitary solutions have been obtained in some field theory models: P. R. Amado, Phys. Rev. **122**, 696 (1961) for the second sector of the Lee model; and J. B. Bronzan, J. Math. Phys. **7**, 1351 (1966) and J. Math. Phys. **8**, 6 (1967) for the charged scalar-field theory.

[4] Three-particle unitarity has been discussed in different contexts; for instance, J. Aitchison and R. Pasquier [Phys. Rev. **152**, 1274 (1967)] on the Khuri–Treiman amplitudes and L. Fleming [Phys. Rev. **135B**, 551 (1964)] on the isobar model.

We now summarize briefly the salient features of our solution of the 3-particle unitarity equations.

The production amplitudes coupled to a given scattering channel, when continued analytically around the total energy normal thresholds, suggest that one consider a homogeneous multidimensional Riemann–Hilbert boundary-value problem in the complex plane of a subenergy variable.

An ansatz for the 4-, 5-, and 6-point S-matrix elements is then developed in terms of the fundamental matrix solution of this Riemann–Hilbert boundary-value problem; it incorporates the following physically motivated features:

1. the connectedness structure of the S matrix;
2. the factorization of residues of unphysical region (baryon) poles;
3. a "protection" conjecture specifying that all 4-point scattering amplitudes appearing in disconnected diagrams or "subdiagrams" (see Fig. 2) ought to be replaced by their elastic approximations (Note: the crossing symmetric elastic scattering amplitudes are an input in the present problem, since the purpose is to incorporate the second-normal threshold);
4. Extended factorization conjecture. This generalization of 2 for the complete, connected, multiparticle amplitudes is suggested by the very structure of the unitarity discontinuity formulas. It implies a factorization of the connected part of the 6-point amplitude—after the physical region cross-energy poles have been taken care of—into a production amplitude in the final state, times a prescattering correction, or, equivalently, into an "absorption" amplitude in the initial state, times a rescattering correction. For the 5-point amplitudes (production or "absorption"), a similar ansatz (product of an inelastic 4-point amplitude times a rescattering or prescattering correction) requires a compensating structure that can only be understood within the framework of a discussion of crossing symmetry.

When the above conjectures are implemented, the 3-particle unitarity equations are shown to be satisfied recurrently; i.e., the sets of Eqs. (II), (III), and (IV) of (3.6)–(3.9) are identically satisfied, provided (I) is. [The set (I) gives the inelastic, unitarity total discontinuity of the scattering amplitudes.]

The inelastic unitarity equation for each scattering amplitude (I) is then reduced to a nonlinear boundary-value problem on the inelastic cut in the energy plane. The most general solution is obtained in terms of a particular solution and an algorithm is described that constructs a particular solution.

Finally, the last section discusses the consistency of crossing symmetry with the 3-particle unitary approximation, i.e., the appproximation to the S matrix obtained by arbitrarily breaking the normal-threshold singularity structure after the first inelastic normal threshold, in the total energy variable.

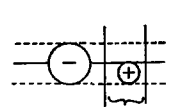


FIG. 2. Six-point disconnected subdiagram (between vertical lines) and protected 4-point amplitude.

2. THE STATIC LIMIT[5]

To show briefly how the static limit makes certain aspects of the 3-body problem more tractable and allows one to concentrate on the structure of the 3-particle unitarity equation, the transformation property of a multiparticle S-matrix element under an element of the Poincaré group (a, A) will be recalled (see Barut[5]):

$$S(K_i, K_f) = \exp\left[i\left(\sum_f k_f - \sum_i k_i\right)\cdot a\right]\left[\prod_f D^{(S_f)}(A'_f)\right] \times \left[\prod_i D^{(S_i)}(A'_i)^*\right] S(\Lambda^{-1}K_i, \Lambda^{-1}K_f). \quad (2.1)$$

$A' = B^{-1}_{k\leftarrow p} A B_{q\leftarrow p}$ is an element of the "little" group of $p = (m, 0, 0, 0)$ (case of massive particles); $B = HU$, where U is a unitary matrix (taken equal to unity if all spins are measured with respect to the direction O_z); H, the Hermitian factor is uniquely determined to be

$$H = \left(\frac{k_i^\mu \sigma_\mu}{m_i}\right)^{\frac{1}{2}} = \cosh(\chi_i/2) + \hat{\mathbf{k}}_i \cdot \boldsymbol{\sigma} \sinh(\chi_i/2),$$
$$\cosh(\chi_i/2) = (1/\sqrt{2})[(k_i^0/m_i) + 1]^{\frac{1}{2}}, \quad (2.2)$$
$$\sinh(\chi_i/2) = (1/\sqrt{2})[(k_i^0/m_i) - 1]^{\frac{1}{2}}.$$

This complicated transformation property, where spins and linear momenta are coupled, is to be contrasted with the simple transformation property of S-matrix elements under isospin rotations:

$$S(K_i, K_f) = \left[\prod_f D^{(I_f)}(u)\right]\left[\prod_i D^{(I_i)}(u)\right]^* S(K_i, K_f), \quad (2.3)$$

[5] The static limit is believed to be qualitatively exact in describing low-energy baryon–meson scattering. Since the classical work of G. F. Chew and F. E. Low [Phys. Rev. **101**, 1570 (1956)], there has been a revival of interest stimulated by the reciprocal bootstrap. See the following recent references: T. Cook, C. J. Goebel, and B. Sakita, Phys. Rev. Letters **15**, 35 (1965); V. Singh and B. Udgaonkar, Phys. Rev. **149**, 1164 (1966); M. L. Whippman, Phys. Rev. **152**, 1269 (1966). About formulas (2.1), (2.2), and (2.3), see A. O. Barut, *The Theory of the Scattering Matrix* (The Macmillan Co., New York, 1967), p. 27.

where u is an arbitrary rotation in 3-dimensional isospin space. From (2.3), it is possible to define S-matrix elements that are eigenstates of the total isospin, and independent of its third component (isotropic S-matrix elements).

A correspondingly simple situation can be achieved for angular momenta by imposing the following simplifications:

1. All particles endowed with spin ("Baryons") have an infinite mass. The corresponding

$$\cosh(\chi_i/2) \to 1, \quad \sinh(\chi_i/2) \to 0,$$

and the $D^{(S)}(A')$ become independent of linear momentum.

2. The other relativistic particles are spinless, giving factors $D^{(0)}(A')$ that are also obviously independent of momentum.

3. Only baryon–meson interactions are retained. As a result of these restrictions, it is possible to carry out a partial-wave expansion in the common center-of-mass of each meson–baryon pair and introduce amplitudes of total angular momentum J. A complete set of 3-particle states is labelled by: (a) the absolute value of each meson momentum and its orbital angular momentum; (b) the total angular momentum and isospin of a baryon–meson pair $\dot{\alpha} = (\bar{J}, \bar{I})$; and (c) the total angular momentum and isospin of the baryon–meson–meson system $\alpha = (J, I)$. A concise notation is:

$$|[[N\pi_{k_1 l_1}]_{\dot{\alpha}}\pi_{k_2 l_2}]_{\alpha}; \overset{\text{in}}{\text{out}}\rangle.$$

(Practically, we shall take $l_1 = l_2 = 1$, since the Nucleon pole occurs in p waves only, at least in the static limit).[6] A recoupling transformation between two possible coupling schemes taking into account the Bose character of the mesons is denoted by

$$|[[N\pi_{k_1 l_1}]_{\dot{\alpha}}\pi_{k_2 l_2}]_{\alpha}; \overset{\text{in}}{\text{out}}\rangle = \sum_{\beta} \lambda^{\alpha}{}_{\alpha\beta}\, |[[N\pi_{k_2 l_2}]_{\dot{\beta}}\pi_{k_1 l_1}]_{\alpha}; \overset{\text{in}}{\text{out}}\rangle, \tag{2.4}$$

$$\lambda^{\alpha}{}_{\alpha\beta} = \lambda^{J}{}_{\bar{J}\bar{J}'}\lambda^{I}{}_{\bar{I}\bar{I}'},$$

$$\lambda^{J}{}_{\bar{J}\bar{J}'} = -(-)^{\bar{J}+\bar{J}'}[(2\bar{J}+1)(2\bar{J}'+1)]^{\frac{1}{2}}\begin{Bmatrix} l_1 & \frac{1}{2} & \bar{J} \\ l_2 & J & \bar{J}' \end{Bmatrix},$$

$$\sum_{\mu} \lambda^{\alpha}{}_{\alpha\mu}\lambda^{\alpha}{}_{\mu\beta} = \delta_{\alpha\beta}. \tag{2.5}$$

[6] For p waves, we take $\rho^2(\omega) = (k/R\omega)\gamma(\omega)$,

$$\gamma(\omega) = [12\pi]^{-1}k^3F^2(\omega),$$

where $F(\omega)$ is the phenomenological static form factor. With this convention for $\rho(\omega)$, an $NN\pi$ vertex gives a factor $3g$. [$g^2/4\pi \simeq 0.8$; see E. Henley and W. Thirring, *Elementary Quantum Field Theory*, (McGraw-Hill Book Co., New York, 1962), Chap. 18.] The analytic structure of $F(\omega)$ never appears and its presence in the phase space factors of unitarity integrals ensures convergence. Note the conversion of summation over $k = |\mathbf{k}|$ into energy integrals:

$$\sum_{k} \to R\pi^{-1}\int_{\mu}^{\infty} \omega k^{-1}\, d\omega.$$

(R is the normalization radius).

3. CLUSTER DECOMPOSITION OF THE S MATRIX AND 3-PARTICLE UNITARITY

In accordance with general S-matrix theory principles, we exhibit the connectedness structure of the S matrix by removing all disconnected contributions and factoring the energy-conserving δ functions (Fig. 3). We also factor out a centrifugal barrier factor $k^{(2l+1)/2}$ per meson line to insure the boundedness at threshold of the A amplitudes and a static phenomenological form factor $F(\omega)$ per external meson line[6]:

$$S^{\alpha}{}_{k'k} = \delta_{k'k} + 2\pi i\delta(\omega' - \omega)\rho^2(\omega)A^{\alpha}(\omega), \tag{3.1}$$

$$S^{\alpha';\alpha}_{k_1'k_2';k} = 2\pi i\delta(\omega_1' + \omega_2' - \omega) \times \rho(\omega_1')\rho(\omega_2')\rho(\omega)A^{\alpha}{}_{\dot{\alpha}'}(\omega_1';\omega), \tag{3.2}$$

$$S^{\dot{\alpha};\alpha}_{k';k_1k_2} = 2\pi i\delta(\omega' - \omega_1 - \omega_2) \times \rho(\omega')\rho(\omega_1)\rho(\omega_2)\bar{A}^{\alpha}{}_{\dot{\alpha}}(\omega_1;\omega'), \tag{3.3}$$

$$S^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2} = S^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2}(D) + 2\pi i\delta(\omega_1' + \omega_2' - \omega_1 - \omega_2) \times \rho(\omega_2')\rho(\omega_1')\rho(\omega_2)\rho(\omega_1)A^{\alpha}{}_{\dot{\alpha}'}(\omega_1';\omega;\omega_1). \tag{3.4}$$

The disconnected part of the 6-point S-matrix element, according to Fig. 3 is

$$S^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2}(D) = \delta^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2} + \tfrac{1}{2}\{\delta_{\dot{\alpha}'\dot{\alpha}}\delta_{k_2'k_2}2\pi i\delta(\omega_1' - \omega_1)\rho^2(\omega_1)A^{\alpha}(\omega_1) + \cdots +\}. \tag{3.5}$$

The dots stand for three other terms required by the Bose character of the ingoing and outgoing meson lines. The 3-particle Kronecker symbol is

$$\delta^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2} = \tfrac{1}{2}\{\delta_{\dot{\alpha}'\dot{\alpha}}\delta_{k_1'k_1}\delta_{k_2'k_2} + \lambda^{\alpha}_{\dot{\alpha}'\dot{\alpha}}\delta_{k_1'k_2}\delta_{k_2'k_1}\}. \tag{3.6}$$

Using (3.1), (3.5) can be reexpressed:

$$S^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2}(D) = \tfrac{1}{2}\{\delta_{\dot{\alpha}'\dot{\alpha}}S^{\dot{\alpha}}_{k_1'k_1}\delta_{k_2'k_2} + \lambda^{\alpha}_{\dot{\alpha}'\dot{\alpha}}S^{\dot{\alpha}}_{k_1'k_2}\delta_{k_2'k_1} + \lambda^{\alpha}_{\dot{\alpha}'\dot{\alpha}}S^{\dot{\alpha}}_{k_2'k_1}\delta_{k_1'k_2} + \sum_{\dot{\mu}}\lambda^{\alpha}_{\dot{\alpha}'\dot{\mu}}S^{\dot{\mu}}_{k_2'k_2}\lambda^{\alpha}{}_{\dot{\mu}\dot{\alpha}}\delta_{k_1'k_1}\} - \delta^{\dot{\alpha}'\dot{\alpha};\alpha}_{k_1'k_2';k_1k_2}. \tag{3.5'}$$

The Bose character of the mesons implies certain invariance properties of the 5- and 6-point amplitudes. For the production amplitudes,

$$A^{\alpha}_{\dot{\alpha}}(\omega_1';\omega) = \sum_{\beta}\lambda^{\alpha}{}_{\dot{\alpha}\beta}A^{\alpha}_{\beta}(\omega - \omega_1';\omega). \tag{3.7}$$

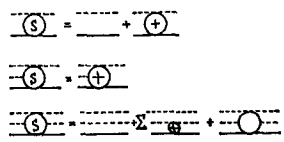


FIG. 3. Connectedness structure of S-matrix elements. *Note:* the δ function and kinematical factors in front of the connected parts are omitted.

Let us now define what is meant by a 3-particle unitary approximation to the S matrix. The "elastic" unitarity relations $\sum_{k''} S^{\alpha'*}_{k''k} S^{\alpha}_{k''k} = \delta_{k'k}$ are exact between 2- and 3-particle normal thresholds. A set of scattering amplitudes[7] satisfying these relations in the whole physical range of energy and crossing symmetric is what we call an "elastic" approximation to the S matrix; it is an input for our purpose.

In view of the importance of the second-normal threshold in the low-energy πN system, it is natural to attempt breaking the S-matrix normal-threshold singularity structure after this extra singularity has been included. This leads one to consider the following four sets of coupled equations relating the 4-, 5-, and 6-point S-matrix elements.

$$\text{(I)} \sum_{k''} S^{\alpha *}_{k''k'} S^{\alpha}_{k''k} + \sum_{\alpha''} \sum_{k_1'' k_2''} S^{\alpha'';\alpha *}_{k_1'' k_2'';k'} S^{\alpha'';\alpha}_{k_1'' k_2'';k} = \delta_{k'k}, \tag{3.8}$$

$$\text{(II)} \sum_{k''} S^{\alpha';\alpha *}_{k'';k_1' k_2'} S^{\alpha}_{k''k} + \sum_{\alpha''} \sum_{k_1'' k_2''} S^{\alpha''\alpha';\alpha *}_{k_1'' k_2'';k_1' k_2'} S^{\alpha'';\alpha}_{k_1'' k_2'';k} = 0, \tag{3.9}$$

$$\text{(III)} \sum_{k''} S^{\alpha *}_{k'';k'} S^{\alpha;\alpha}_{k'';k_1 k_2} + \sum_{\alpha''} \sum_{k_1'' k_2''} S^{\alpha'';\alpha *}_{k_1'' k_2'';k'} S^{\alpha'',\alpha';\alpha}_{k_1'' k_2'';k_1 k_2} = 0, \tag{3.10}$$

$$\text{(IV)} \sum_{k''} S^{\alpha';\alpha *}_{k'';k_1' k_2'} S^{\alpha;\alpha}_{k'';k_1 k_2} + \sum_{\alpha''} \sum_{k_1'' k_2''} S^{\alpha''\alpha';\alpha *}_{k_1'' k_2'';k_1' k_2'} S^{\alpha'',\alpha;\alpha}_{k_1'' k_2'';k_1 k_2} = \delta^{\alpha'\alpha;\alpha}_{k_1' k_2';k_1 k_2}. \tag{3.11}$$

Again, the relations (I)–(IV) are exact between the 3- and 4-particle normal thresholds. Their extension in the whole energy range above the 3-particle normal threshold constitutes the 3-particle unitary approximation, a useful concept in physical situations where the 3-particle channel dominates the other inelastic channels (case of $\pi N \leftrightarrow \pi\pi N$ below 1 GeV). The difficulties encountered with respect to certain crossing symmetry properties by breaking the S-matrix singularity structure in that way, are discussed and partially resolved in the last section.

When the connectedness structure (3.1)–(3.4) is substituted into (I)–(IV), the 3-particle unitarity equations take a form represented diagrammatically on Fig. 4. In the derivation, use has actually been made of the "protection" prescription in the form represented on Fig. 5; i.e., of the fact that, between the 3- and 4-particle normal thresholds in the total energy, the "protected" 4-point amplitudes can only have the 2-particle normal threshold.

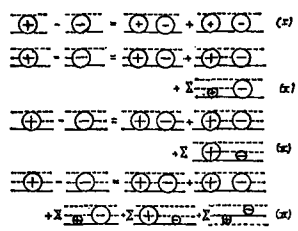


FIG. 4. The 3-particle unitarity relations in terms of A amplitudes.

In the context of the 3-particle unitary approximation defined above, the substitution in the whole energy range of the "elastic" approximation for the protected 4-point amplitudes becomes a conjecture justified only by our ability to produce in this way a class of solutions to (I)–(IV).

The interpretation of the relations (I)–(IV) as read off of their diagrammatic representation (Fig. 4) will now be given.

(I) implies for the scattering amplitudes the elastic and inelastic normal thresholds at $z = \mu - i\epsilon$, $z = 2\mu - 2i\epsilon$, respectively, in the total energy variable. (The total and subenergy variables ω and ω_1 are promoted to complex variables; the physical masses are given a negative infinitesimal imaginary part in accordance with Feynman's $i\epsilon$ prescription.)

(II), for the production amplitudes besides the normal thresholds at $z = \mu - i\epsilon$, $z = 2\mu - 2i\epsilon$, implies a two-sheeted normal threshold at $\mu - i\epsilon$ in the complex subenergy plane z_1 and an extra branch point—also two-sheeted—at $z - \mu + i\epsilon$ arising from the redundant subenergy variable z_2. The physical region for the production amplitudes is $\omega > 2\mu$, $\mu < \omega_1 < \omega - \mu$ and is represented by the dotted real axes segments of the z_1 and z planes, respectively (Fig. 6).

(III), unrepresented on Fig. 4, gives for the absorption 5-point amplitudes a singularity structure completely analogous to (II), except for the fact that the final-state subenergy variable is replaced by an initial-state subenergy variable.

(IV) exhibits for the 6-point amplitudes 2- and 3-particle normal thresholds in the total energy variable $z = \mu - i\epsilon$, $z = 2\mu - 2i\epsilon$; 2-particle normal thresholds in the final and initial subenergy variables $z_1 = \mu - i\epsilon$, $z_1 = z - \mu + i\epsilon$, $z_1' = \mu - i\epsilon$, $z_1' = z - \mu + i\epsilon$, and, finally, physical-region single-particle discontinuities at $z_1 = z_1'$ and $z_1 = z_2'$.

These single-particle discontinuities originate in the cross-energy pole terms pictured on Fig. 7. The

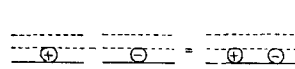


FIG. 5. Protected relation used in deriving the relations (I)–(IV) of Fig. 4.

[7] In the elastic approximation we have

$$S^{\alpha}_{k'k} = \delta_{k'k} + 2\pi i \delta(\omega' - \omega)\rho^2(\omega)\mathcal{A}_{\alpha}(\omega).$$

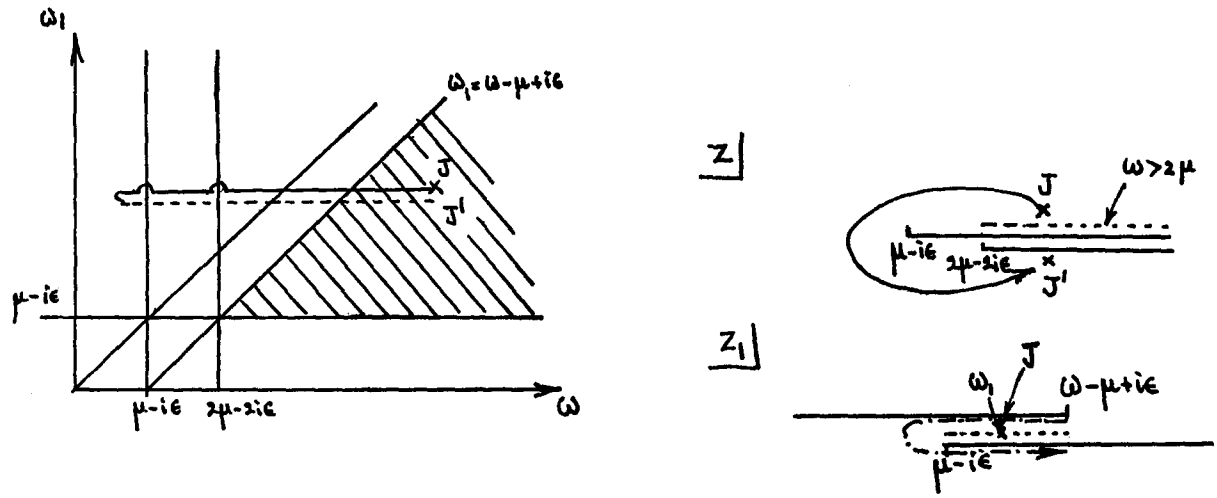


FIG. 6. The singularity structure in the complex (z_1, z) planes implied by Eq. (II) or (III) and the physical region; the analytic continuation path of Sec. (IV) is also represented.

complete contribution of these terms,

$$2^{-1}\Big\{\sum_{\dot\mu'}\lambda^{\alpha}_{\alpha'\dot\mu'}\mathcal{A}_{\dot\mu'}(\omega_2')\lambda^{\alpha}_{\dot\mu'\alpha}D_1^{(+)}\mathcal{A}_{\alpha}(\omega_1) + \sum_{\dot\mu\dot\mu'}\lambda^{\alpha}_{\alpha'\dot\mu'}\mathcal{A}_{\dot\mu'}(\omega_2')\lambda^{\alpha}_{\dot\mu'\dot\mu}D_2^{(+)}\mathcal{A}_{\dot\mu}(\omega_2)\lambda^{\alpha}_{\dot\mu\alpha} - \mathcal{A}_{\alpha'}(\omega_1')D_2^{(-)}\mathcal{A}_{\alpha}(\omega_1)\lambda^{\alpha}_{\alpha'\alpha} - \mathcal{A}_{\alpha'}(\omega_1')D_1^{(-)}\sum_{\dot\mu}\lambda^{\alpha}_{\alpha'\dot\mu}\mathcal{A}_{\dot\mu}(\omega_2)\lambda^{\alpha}_{\dot\mu\alpha}\Big\}, \quad (3.12)$$

can be split into two parts, using the discontinuity equation for the baryon propagators:

$$D_1^{(\pm)} = \mathcal{P}(\omega_1' - \omega_1)^{-1} \pm i\pi\delta(\omega_1' - \omega_1);$$
$$D_2^{(\pm)} = \mathcal{P}(\omega_1' - \omega_2)^{-1} \pm i\pi\delta(\omega_1' - \omega_2). \quad (3.13)$$

The first part, containing the pole terms proper, will be dealt with after (3.17); the second part, containing the energy-conserving δ functions, is combined with the disconnected part (3.5′) into a new "disconnected" part bilinear in "protected" scattering S-matrix elements (see Fig. 8):

$$S^{\alpha'\alpha;\alpha}_{k_1'k_2';k_1k_2}(D') = \tfrac{1}{4}\cdot\Big\{\mathbb{S}^{\alpha'}_{k_1'k_2}\lambda^{\alpha}_{\alpha'\alpha}\mathbb{S}^{\alpha}_{k_2'k_1} + \sum_{\dot\mu}\lambda^{\alpha}_{\alpha'\dot\mu}\mathbb{S}^{\dot\mu}_{k_2'k_2}\lambda^{\alpha}_{\dot\mu\alpha}\mathbb{S}^{\alpha}_{k_1'k_1} + \mathbb{S}^{\alpha'}_{k_1'k_1}\sum_{\dot\mu}\lambda^{\alpha}_{\alpha'\dot\mu}\mathbb{S}^{\dot\mu}_{k_2'k_2}\lambda^{\alpha}_{\dot\mu\alpha} + \sum_{\dot\mu\dot\mu'}\lambda^{\alpha}_{\alpha'\dot\mu'}\mathbb{S}^{\dot\mu'}_{k_2'k_1}\lambda^{\alpha}_{\dot\mu'\dot\mu}\mathbb{S}^{\dot\mu}_{k_1'k_2}\lambda^{\alpha}_{\dot\mu\alpha}\Big\}. \quad (3.14)$$

The convenience of (3.14) arises from the fact that the $\mathbb{S}$ are simply given phase factors:

$$\mathbb{S}^{\alpha}_{k_1'k_1} = \delta_{k_1'k_1}\exp\,[2i\delta^{e}_{\alpha}(\omega_1)],$$

where the δ^{e}_{α} are the elastic phase shifts.

Returning to the production amplitudes, we now describe the single-particle intermediate states in the subenergy and cross-energy variables [Figs. 9(a) and (b)] giving rise to unphysical region poles in the subenergy variable z_1 in the static limit. The particular importance of these poles is due to the observation that the axes $\omega_1 = 0$, $\omega_2 = 0$ frame the sector of the physical region in the diagram of Fig. 10; certain crossing symmetry properties must be satisfied on these lines that will be described in the last section.

The analytic expressions of the final subenergy pole terms of Fig. 9(a), since they contain "unprotected" 4-point amplitudes, are

$$(3g/\sqrt{2})\omega_1'^{-1}\delta_{\alpha,\frac{11}{22}}A^{\alpha}(\omega); \quad (3g/\sqrt{2})\omega_2'^{-1}\lambda^{\alpha}_{\alpha,\frac{11}{22}}A^{\alpha}(\omega). \quad (3.15)$$

The cross-energy pole terms of Fig. 9(b), on the contrary, contain "protected" scattering amplitudes to be replaced by their "elastic" approximations; their analytic expressions are

$$-(3g/\sqrt{2})\omega_1'^{-1}\sum_{\dot\mu}\lambda^{\alpha}_{\alpha\dot\mu}\mathcal{A}_{\dot\mu}(\omega)\lambda^{\alpha}_{\dot\mu,\frac{11}{22}};$$
$$-(3g/\sqrt{2})\omega_2'^{-1}\mathcal{A}_{\alpha}(\omega)\lambda^{\alpha}_{\alpha,\frac{11}{22}}. \quad (3.16)$$

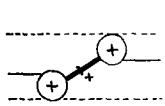


FIG. 7. Cross-energy pole present in the connected 6-point amplitude.

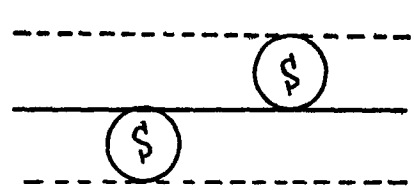


FIG. 8. New disconnected part of the 6-point S-matrix element.

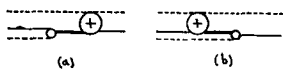


FIG. 9. (a) Subenergy pole terms of the production amplitudes. Note the "unprotected" 4-point amplitude. (b) Cross-energy pole term of the production amplitudes. Note the "protected" 4-point amplitude.

Recapitulating, the driving pole terms of the production amplitudes in the z_1 complex plane for fixed complex total energy are

$$D^{\alpha}_{\alpha}(z_1'; z) = d^{\alpha}_{\alpha}(z)[z - z_1']^{-1} + \sum_{\mu} \lambda^{\alpha}_{\alpha\mu}\, d^{\alpha}_{\mu}(z) z_1'^{-1},$$

$$d^{\alpha}_{\alpha}(z) = 3g\lambda^{\alpha}_{\alpha\,\frac{1}{2}\frac{1}{2}}[\mathcal{A}_{\alpha}(z) - A^{\alpha}(z)]. \tag{3.17}$$

Similarly, the pole terms at $z_1 = 0$ and $z_1 = z$ of the connected parts of the 6-point amplitudes—as redefined by (3.14')—contain the full production amplitudes in their residues (see Fig. 11):

$$(3g/\sqrt{2})\{\delta_{\alpha,\frac{1}{2}\frac{1}{2}} z_1^{-1} + \lambda^{\alpha}_{\alpha,\frac{1}{2}\frac{1}{2}}(z - z_1)^{-1}\} A^{\alpha}{}_{\alpha'}(z_1'; z).$$

(Here we have chosen the ingoing subenergy variable, but, of course, a similar expression can be written for the outgoing subenergy variable with the full 5-point absorption amplitudes $\tilde{A}^{\alpha}_{\alpha}$ appearing in the residues.)

Finally, the pole terms in the z_1 variable arising from (3.12) can be shown to give a vanishing total contribution to the connected part of the 6-point amplitudes. For instance, the residue of the pole at $z_1' = z_1$ is

$$2^{-1}\left\{\sum_{\mu} \lambda^{\alpha}_{\alpha'\mu}\mathcal{A}_{\mu}(z_2')\lambda^{\alpha}_{\mu\alpha}\mathcal{A}_{\alpha}(z_1) - \mathcal{A}_{\alpha'}(z_1')\sum_{\mu}\lambda^{\alpha}_{\alpha'\mu}\mathcal{A}_{\mu}(z_2)\lambda^{\alpha}_{\mu\alpha}\right\}.$$

When it becomes multiplied by the fundamental matrix solution (see next section) in the z_1 plane expressing prescattering, the contributions of the two terms cancel out. [The proof is identical to the proof of identities (5.8) and (5.9) and will not be reproduced here.]

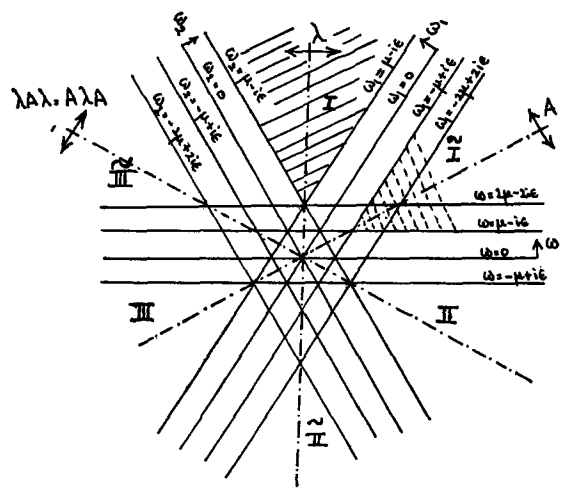


FIG. 10. Complete singularity structure and crossing symmetry of the production amplitudes.

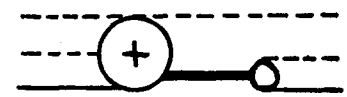


FIG. 11. Initial subenergy-variable pole of the 6-point amplitudes.

4. THE HOMOGENEOUS RIEMANN-HILBERT PROBLEM

Let us continue $A^{\alpha}_{\alpha'}(z_1'; z)$ analytically from a physical region point J at $z_1' = \omega_1'$, $z = \omega$ around the total-energy normal thresholds while keeping $z_1' = \omega_1'$ fixed in its physical region. The analytic continuation path is represented on Fig. 6. It is observed that, in the process, the redundant subenergy branch point moves around $\mu - i\epsilon$. The continued amplitude has only the last two unitarity discontinuity terms of (II) (see Fig. 12). Just such a discontinuity equation has been shown by Polkinghorne, Olive, and Landshoff[1] to hold in a certain energy interval even in the presence of triangle singularities.[8]

The discontinuity of the production amplitudes or their analytic continuations in the z plane is given by the expression (Fig. 12)

$$\mathrm{disc}_{\omega_1' \in R_1} A^{\alpha}{}_{\alpha'}(\omega_1'; \omega) = \sum_{k''} 2\pi i \delta(\omega'' - \omega_1')\rho^2(\omega'')\mathcal{A}_{\alpha'}(\omega_1')A_{\alpha'}(\omega_1' - i0; \omega).$$

This relation is true, irrespective of the relative positions of the R_1 and R_2 cuts, i.e., of whether or not the analytic continuation has been performed [a consequence of Steinmann's relations, see Ref. 9]. A similar discontinuity equation can be written for R_2.

Taking advantage of the "protection" assumption, they lead one to consider the following homogeneous, multidimensional Riemann–Hilbert boundary-value problem in the complex z_1 plane[10]:

$$\text{"}h\text{"}\begin{cases} \phi^{\alpha}_{\alpha}(\omega_1 + i0; \omega) = \exp[2i\delta^{e}_{\alpha}(\omega_1)]\phi^{\alpha}_{\alpha}(\omega_1 - i0; \omega); \\ \qquad\qquad \omega_1 \in R_1 \equiv [\mu, \infty], \quad (4.1) \\ \sum_{\mu} \lambda^{\alpha}_{\alpha\mu}\phi^{\alpha}_{\mu}(\omega_1 + i0; \omega) \\ = \exp[-2i\delta^{e}_{\alpha}(\omega - \omega_1)]\sum_{\mu}\lambda^{\alpha}_{\alpha\mu}\phi^{\alpha}_{\mu}(\omega_1 - i0; \omega); \\ \qquad\qquad \omega_1 \in R_2 \equiv [-\infty, \omega - \mu]. \quad (4.2) \end{cases}$$

The dimension [i.e., the number of equation of (4.1)

[8] P. V. Landshoff, D. I. Olive, J. Polkinghorne, J. Math. Phys. 7, 1593 (1966).

[9] D. I. Olive, Nuovo Cimento 37, 1422 (1965); see especially p. 1429.

[10] The multidimensional homogeneous Riemann–Hilbert problem is treated for the case of closed contours by N. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953). The case of discontinuous coefficients or open contours—of interest here—is treated by N. Vekua, *Systems of Singular Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1967). As far as the Hilbert problem is concerned, we confine ourselves here to the class of sectionally holomorphic (or meromorphic) functions with polynomial boundedness at infinity and whose boundary-values on the cuts satisfy a Hölder condition.

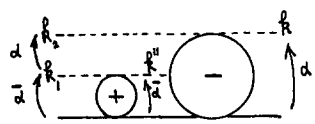


FIG. 12. Subenergy-variable discontinuity term giving rise to a Riemann–Hilbert boundary condition on $R_1 = [\mu, \infty]$.

or (4.2)] of "h" is equal to the number n of subchannels $\dot{\alpha} = (\bar{J}, \bar{I})$ coupled to the scattering channel $\alpha = (J, I)$. Again, the relations (4.1) and (4.2) are *exact* (with δ_α substituted for δ_α^e) in the physical region when $2\mu < \omega < 3\mu$ (this condition restricts the overlap of R_1 and R_2); in the context of a 3-particle unitary approximation, we extend them to the whole R_1, R_2 cuts, thus neglecting further thresholds at $z_1 = 2\mu - 2i\epsilon$, $z_1 = \omega - 2\mu + 2i\epsilon$, etc. Besides, our singularity structure is also incomplete in view of the crossing symmetry properties of the production amplitudes requiring, for instance, branch points at $z_1 = -\mu + i\epsilon$, $z_1 = z + \mu - i\epsilon$, etc. The possibility of embedding the Riemann–Hilbert problem "h" into a larger Riemann–Hilbert boundary-value problem "H", once these extra branch cuts are recognized, will be discussed in Appendix B.

We proceed now with the construction of the most general solution of "h" having the correct driving poles (3.15). Quite generally,

$$\Phi_\alpha^\alpha(z_1; z) = \Psi_\alpha^\alpha(z_1; z) F^\alpha(z_1; z), \tag{4.3}$$

where $\Psi_\alpha^\alpha(z_1; z)$ has no poles in the finite plane and satisfies (4.1) and (4.2) on R_1, R_2, whereas F^α is meromorphic in the z_1 plane with poles at $z_1 = 0$, $z_1 = z$.

By a theorem of Muskhelishvili,[9]

$$\Psi_\alpha^\alpha(z_1; z) = \sum_{(i)} \Omega_{\alpha(i)}^\alpha(z_1; z) P_i(z_1; z), \tag{4.4}$$

where the $\Omega_{\alpha(i)}^\alpha(z_1; z)$, $(i = 1, \cdots, n)$ form a fundamental matrix of vector solutions of "h." We define briefly the fundamental-matrix solution of "h": Let $\Omega_{\alpha(1)}^\alpha(z_1; z)$ be the sectionally holomorphic vector solution (n components labelled by $\dot{\alpha}-$), having lowest degree $(-\kappa_1)$ at infinity in the z_1 plane; the dependence on z is parametric through the position of the end point of R_2. The second fundamental solution $\Omega_{\alpha(2)}^\alpha(z_1; z)$ has the lowest degree at infinity $(-\kappa_2)$, compatible with orthogonality to $\Omega_{\alpha(1)}^\alpha$ with polynomial weight functions. Proceeding successively, one can define exactly n such independent fundamental solutions, where n is the dimension of the Riemann–Hilbert problem "h."

Among the other properties of the fundamental matrix used here, we note the following:

1. A necessary and sufficient condition for Ω to be a fundamental matrix of "h" is that its determinant be nonvanishing everywhere in the finite plane and the determinant of X, defined as

$$X_{\alpha(i)}^\alpha(z_1; z) = z_1^{\kappa_i} \Omega_{\alpha(i)}^\alpha(z_1; z),$$

be different from zero at infinity.

2. A sectionally holomorphic solution of "h" is a linear combination with polynomial coefficients of the fundamental solutions; this is exactly the statement (4.4).

Since the Ω matrix defined above is not unique and since it will be convenient to handle an Ω matrix with simple transformation property under the permutation of the two rescattering meson lines, we add to the properties 1 and 2 the following theorem, proved in Appendix A.

Theorem: There exists one and only one fundamental-matrix solution to "h" with the following property under exchange of the cuts R_1 and R_2:

$$\sum_{\dot{\mu}} \lambda_{\dot{\alpha}\dot{\mu}}^\alpha \Omega_{\dot{\mu}(i)}^\alpha(z - z_1; z) = \Omega_{\dot{\alpha}(i)}^\alpha(z_1; z) c_i. \tag{4.5}$$

An immediate *corollary* is: the $c_i = \pm 1$ are the eigenvalues of the λ matrix, and the coefficients of the dominant power of $\Omega_{\dot{\alpha}(i)}^\alpha$ at infinity are the corresponding eigenvectors of λ^α. Indeed,

$$\sum_{\dot{\mu}} \lambda_{\dot{\alpha}\dot{\mu}}^\alpha X_{\dot{\mu}(i)}^\alpha(\infty; z) = c_i X_{\dot{\alpha}(i)}^\alpha(\infty; z). \tag{4.6}$$

Returning to the construction of $\Phi_\alpha^\alpha(z_1; z)$, from (4.3) and (4.4), we have

$$\Phi_\alpha^\alpha(z_1; z) = \sum_{(i)} \Omega_{\alpha(i)}^\alpha(z_1; z) F_i^\alpha(z_1; z),$$
$$F_i^\alpha(z_1; z) = f_i^\alpha(z_1; z) z_1^{-1} + g_i^\alpha(z_1; z)(z_1 - z)^{-1}, \tag{4.7}$$

and the f_i^α, g_i^α are entire functions of z_1. Bose symmetry in the final state gives a simple relation between f_i^α and g_i^α. From (4.5) and

$$\Phi_\alpha^\alpha(z_1; z) = \sum_{\dot{\mu}} \lambda_{\dot{\alpha}\dot{\mu}}^\alpha \Phi_{\dot{\mu}}^\alpha(z - z_1; z), \tag{4.8}$$

we find that

$$g_i^\alpha(z_1; z) = c_i f_i^\alpha(z - z_1; z). \tag{4.9}$$

Besides, identification of the pole residues at $z_1 = 0$, $z_1 = z$ with (3.15) imposes the conditions

$$\sum_{(i)} \Omega_{\alpha(i)}^\alpha(z_1; z) g_i^\alpha(z; z) = d_\alpha^\alpha(z),$$
$$\sum_{(i)} \Omega_{\alpha(i)}^\alpha(0; z) f_i^\alpha(0; z) = \sum_{\dot{\mu}} \lambda_{\dot{\alpha}\dot{\mu}}^\alpha d_{\dot{\mu}}^\alpha(z). \tag{4.10}$$

One relation is clearly redundant in view of (4.9). Since $\det \|\Omega_{\alpha(i)}^\alpha(z_1; z)\| \neq 0$ in the finite z_1 plane, (4.10) can readily be inverted, yielding the most general solution of "h" with the correct poles at

$z_1 = 0$, $z_1 = z$ and a prescribed degree N at infinity:

$$\Phi^\alpha_\alpha(z_1; z) = \sum_{(i)\mu} \Omega^\alpha_{\mu(i)}(z_1; z)\{c_i z_1^{-1} + (z - z_1)^{-1}\}[\Omega^\alpha(z; z)]^{-1}_{(i)\mu}\, d^\alpha_\mu(z) + \sum_{(i)} \Omega^\alpha_{\alpha(i)}(z_1; z) P'_i(z_1; z). \quad (4.11)$$

The

$$P'_i(z_1; z) = c_i z_1^{-1}[f^\alpha_i(z_1; z) - f^\alpha_i(0; z)] + (z - z_1)^{-1}[f^\alpha_i(z - z_1; z) - f^\alpha_i(0; z)] \quad (4.12)$$

are polynomials in z_1 of degree $N + \kappa_i - 1$ with arbitrary functions of z as coefficients.

The representation (4.11) still involves considerable freedom, especially insofar as the z dependence is concerned. Crossing, of course, by relating the analytic structure and the asymptotic behaviors in the z_1 and z complex planes is expected to restrict or eliminate this arbitrariness. But, of course, crossing also obliges us to take into account other branch points in unphysical regions that have not been incorporated in our model. Then the $P'_i(z_1; z)$ cease to be entire functions with finite degree at infinity (i.e., polynomials) in the z_1 plane but are still regular across the cuts R_1 and R_2.

Within our 3-particle unitary model—with truncated singularity structure both in relation to higher unitary approximations and crossing symmetry—it will become apparent in the next section that it is impossible to satisfy (I)–(IV) with an ansatz based on (4.11) unless the analytic dependence in z of the $P'_i(z_1; z)$ is the same as the one exhibited by the pole terms.

The effective representation of the production amplitudes adopted here is thus

$$A^\alpha_\alpha(z_1; z) = \sum_{(i)\mu} \Omega^\alpha_{\alpha(i)}(z_1; z)\left\{\frac{c_i P_i(z_1; z)}{z_1} + \frac{P_i(z - z_1; z)}{z - z_1}\right\} \times [\Omega^\alpha(z; z)]^{-1}_{(i)\mu} d^\alpha_\mu(z). \quad (4.11')$$

In our model, the functions $P_i(z_1; z)$ are polynomials of degree $N + \kappa_i + 1$ in the z_1 plane such that $P_i(0; z) = 1$; their coefficients are regular functions of z (with possibly a pole at $z = 0$ to be discussed later); the branch point at $z = 2\mu$ comes entirely from the "unprotected" $A^\alpha(z)$. If the fundamental-matrix solution of "h" can be maintained in the presence of other singularities by enlarging the Riemann–Hilbert problem (a matter taken up in Appendix B), we expect the representation (4.11′) still to be valid; the $P_i(z_1; z)$ are then to be free of the branch points at $z_1 = \mu - i\epsilon$, $z_1 = z - \mu + i\epsilon$ in the z_1 plane and $z = \mu - i\epsilon$, $z = 2\mu - 2i\epsilon$ in the z plane. (They cease being simply polynomials.)

5. THE ANSATZ AND THE RECURRENT SOLUTION OF (I)–(IV)

Let us first define the following convenient expressions describing final-state rescattering and initial-state prescattering:

$$Y^\alpha_{\alpha\beta}(\omega_1; \omega) = 2\pi i\delta(\omega_1 + \omega_2 - \omega)\rho(\omega_1)\rho(\omega_2)\rho^{-1}(\omega)\frac{k}{R\omega} \times \sum_{(i)} \Omega^\alpha_{\alpha(i)}(\omega_1; \omega)\left\{\frac{P_i(\omega_2; \omega)}{\omega_2} + \frac{c_i P_i(\omega_1; \omega)}{\omega_1}\right\} \times [\Omega^\alpha(\omega; \omega)]^{-1}_{(i)\beta}\lambda^\alpha_{\beta,\frac{11}{22}}, \quad (5.1)$$

$$Z^\alpha_\alpha(\omega_1; \omega) = \sum_\beta Y^\alpha_{\alpha\beta}(\omega_1; \omega). \quad (5.2)$$

The basic ansatz for the S-matrix elements is

$$S^{\alpha';\alpha}_{k_1'k_2';k} = Z^\alpha_{\alpha'}(\omega_1'; \omega)S^\alpha_{k'k} - \sum_{\beta'} Y^\alpha_{\alpha'\beta'}(\omega_1'; \omega)S^{\beta'}_{k'k}, \quad (5.3)$$

$$S^{\alpha;\alpha}_{k';k_1k_2} = Z^\alpha_\alpha(\omega_1; \omega)S^\alpha_{k'k} - \sum_\beta Y^\alpha_{\alpha\beta}(\omega_1; \omega)S^\beta_{k'k}, \quad (5.4)$$

$$S^{\alpha'\alpha;\alpha}_{k_1'k_2';k_1k_2} = S^{\alpha'\alpha;\alpha}_{k_1'k_2';k_1k_2}(D') + S^{\alpha';\alpha}_{k_1'k_2';k}Z^\alpha_\alpha(\omega_1; \omega) = S^{\alpha'\alpha;\alpha}_{k_1'k_2';k_1k_2}(D') + Z^\alpha_{\alpha'}(\omega_1; \omega)S^{\alpha;\alpha}_{k';k_1k_2}. \quad (5.5)$$

Substituting (5.3), (5.4), and (5.5) into (II), (III), and (IV), for (II) and (IV) we obtain successively

$$\left\{\sum_{k''} S^{\alpha';\alpha *}_{k'';k_1'k_2'}S^\alpha_{k''k} + \sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha *}_{k_1''k_2'';k_1'k_2'}S^{\alpha'';\alpha}_{k_1''k_2'';k}\right\} = Z^\alpha_{\alpha'}(\omega_1'; \omega)^*\left\{\sum_{k''} S^{\alpha *}_{k''k'}S^\alpha_{k''k} + \sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha'';\alpha *}_{k_1''k_2'';k'}S^{\alpha'';\alpha}_{k_1''k_2'';k}\right\} + \sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha}_{k_1''k_2'';k_1'k_2'}(D')^*S^{\alpha'';\alpha}_{k_1''k_2'';k} - \sum_{\dot\mu} Y^\alpha_{\alpha'\dot\mu}(\omega_1'; \omega)^*\sum_{k''} S^{\dot\mu *}_{k''k'}S^\alpha_{k''k}, \quad (5.6)$$

$$\left\{\sum_{k''} S^{\alpha';\alpha *}_{k'';k_1'k_1'}S^{\alpha;\alpha}_{k'';k_1k_2} + \sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha *}_{k_1''k_2'';k_1'k_2'}S^{\alpha''\alpha;\alpha}_{k_1''k_2'';k_1k_2}\right\} = \delta^{\alpha'\alpha;\alpha}_{k_1'k_2';k_1k_2} + Z^\alpha_{\alpha'}(\omega_1; \omega)\left\{\sum_{k''} S^{\alpha';\alpha *}_{k'';k_1'k_2'}S^\alpha_{k''k} + \sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha *}_{k_1''k_2'';k_1'k_2'}S^{\alpha'';\alpha}_{1''k_2'';k}\right\} + \sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha}_{k_1''k_2'';k_1'k_2'}(C)^*S^{\alpha''\alpha';\alpha}_{k_1''k_1'';k_1k_2}(D') - S^{\alpha';\alpha *}_{k_1'k_2';k}\sum_{\dot\mu}[Y^\alpha_{\alpha'\dot\mu}(\omega_1'; \omega)\exp(2i\delta^e_{\dot\mu}(\omega))]. \quad (5.7)$$

All undesirable terms cancel out and (I)–(IV) are recurrently satisfied [i.e., (II) and (III) are satisfied if (I) is, and (IV) are satisfied if (II) or (III) is],

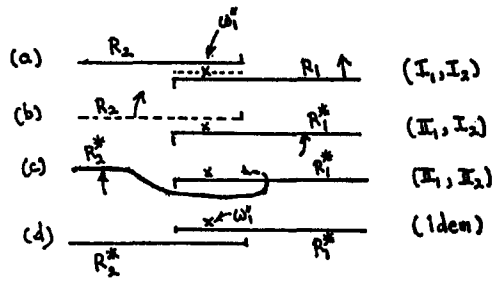


FIG. 13. Motion of the subenergy-variable cuts depicting the effect of multiplication of Ω by $S(D)^*$.

provided we can prove the two similar identities:

$$\sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha}_{k_1''k_2'';k_1'k_2'}(D')^* \sum Y^{\alpha}_{\alpha'\dot{\mu}}(\omega_1''; \omega)S^{\dot{\mu}}_{k''k} = Z^{\alpha}_{\alpha'}(\omega_1'; \omega)^*\delta_{k'k}, \quad (5.8)$$

$$\sum_{\alpha''}\sum_{k_1''k_2''} S^{\alpha''\alpha';\alpha}_{k_1''k_2'';k_1'k_2'}(D')^*Z^{\alpha}_{\alpha''}(\omega_1''; \omega) = \sum_{\dot{\mu}} [Y^{\alpha}_{\alpha'\dot{\mu}}(\omega_1'; \omega) \exp(2i\delta^e_{\dot{\mu}}(\omega))]^*. \quad (5.9)$$

Proof of (5.8), (5.9): Multiplication by the phase factors in $S(D)^*$ affects the fundamental matrix $\Omega^{\alpha}_{\alpha'(i)}(\omega_1''; \omega)$ by bringing the physical region point ω_1'' on the second sheet of both subenergy-variable cuts. This actually requires giving $\omega = \omega_1'' + \omega_2''$ also an infinitesimal, negative, imaginary part, as shown by the sequence of Fig. 13. On this figure, only the physical sheet cuts are represented; they are either visible or not, according to whether or not the sheet of the representative point is visible after application of each phase factor. The fundamental matrix of situation (d) obtained from (e) by an infinitesimal deformation—and thus topologically equivalent to situation (e)—is $\Omega^{\alpha}_{\alpha(i)}(\omega_1''; \omega)^*$.

Analogously, the factor $S^{\dot{\mu}}_{k''k}$ on the right of Y takes one on the second sheet of R_1 (only one cut is operative at $\omega_1 = \omega$!) of $[\Omega^{\alpha}(\omega; \omega)]^{-1}_{(i)\dot{\mu}}$. Finally, we obtain (since $P_i = P_i^*$) the complex conjugate of Z or Y, respectively, and (5.8) and (5.9) are established.

6. THE INELASTIC BOUNDARY-VALUE PROBLEM FOR THE SCATTERING AMPLITUDES

Having established that our basic ansatz for the 4-5-, and 6-point S-matrix elements recurrently satisfies the four sets of 3-particle unitarity equation, we proceed to substitute (5.3)–(5.5) into (I). Recalling [see Ref. 6] that $\gamma(\omega) = k^3F^2(\omega)/12\pi$, (I) is rewritten as

$$A^{\alpha}(\omega) - A^{\alpha}(\omega)^* = 2i\left\{\gamma(\omega)\,|A^{\alpha}(\omega)|^2 + \frac{1}{\pi}\cdot\int_{\mu}^{\omega-\mu} d\omega_1''\gamma(\omega_1'')\gamma(\omega - \omega_1'') \times \sum_{\alpha''} |A^{\alpha}_{\alpha''}(\omega_1''; \omega)|^2\right\}. \quad (6.1)$$

Redefining $Y^{\alpha}_{\alpha\beta}$ and Z^{α}_{α} to omit the kinematical factors kept in (5.1), we write

$$A^{\alpha}_{\alpha''}(\omega_1''; \omega) = \sum_{\beta} Y^{\alpha}_{\alpha''\beta}(\omega_1''; \omega)\mathcal{A}_{\beta}(\omega) - Z^{\alpha}_{\alpha''}(\omega_1''; \omega)A^{\alpha}(\omega) = R^{\alpha}_{\alpha''}(\omega_0''; \omega) - Z^{\alpha}_{\alpha''}(\omega_0''; \omega)A^{\alpha}(\omega). \quad (6.2)$$

From (6.1) and (6.2), the inelastic scattering amplitudes are found to satisfy the inhomogeneous nonlinear boundary-value problem in the ω plane on the right-hand cut (the index α is omitted in this section from now on):

$$(2i)^{-1}[G(\omega)A(\omega + i0) - G^*(\omega)A(\omega - i0)] = \Gamma(\omega)A(\omega + i0)A(\omega - i0) + C(\omega), \quad \omega \in R \equiv [\mu, \infty], \quad (6.3)$$

with

$$G(\omega) = 1 + 2i\pi^{-1}\cdot\int_{\mu}^{\omega-\mu} d\omega_1''\gamma(\omega_1'')\gamma(\omega - \omega_1'') \times \sum_{\alpha''} R^{\alpha}_{\alpha''}(\omega_1''; \omega)^*Z^{\alpha}_{\alpha''}(\omega_1''; \omega), \quad (6.4)$$

$$\Gamma(\omega) = \gamma(\omega) + \pi^{-1}\cdot\int_{\mu}^{\omega-\mu} d\omega_1''\gamma(\omega_1'')\gamma(\omega - \omega_1'') \times \sum_{\alpha''} |Z^{\alpha}_{\alpha''}(\omega_1''; \omega)|^2, \quad (6.5)$$

$$C(\omega) = \pi^{-1}\cdot\int_{\mu}^{\omega-\mu} d\omega_1''\gamma(\omega_1'')\gamma(\omega - \omega_1'')\sum_{\alpha''} |R^{\alpha}_{\alpha''}(\omega_1''; \omega)|^2. \quad (6.6)$$

As a first step in studying (6.3), let us assume we do have a particular solution ψ. Then let the most general solution be $A = \psi + \chi$, where A and ψ have identical discontinuities across all the—so far unspecified—left-hand cuts and the same poles with the same residues. On the right-hand cut, (6.3) splits into

$$\begin{cases} (2i)^{-1}\{G\psi_+ - G^*\psi_-\} - \Gamma\psi_+\psi_- - C = 0, & (6.7) \\ (2i)^{-1}\{\chi_+[G - 2i\Gamma\psi_-] - \chi_-[G^* + 2i\Gamma\psi_+]\} = \Gamma\psi_+\psi_-. & (6.8) \end{cases}$$

The sectionally holomorphic $\chi(z)$ has only the right-hand inelastic cut across which it satisfies the nonlinear but homogeneous boundary-value problem (6.8).

We treat first the auxiliary linear boundary-value problem:

$$(2i)^{-1}\{\Delta'_+[G - 2i\Gamma\psi_-] - \Delta'_-[G^* + 2i\Gamma\psi_+]\} = 0, \quad \mathrm{disc}_L\,\Delta' = 0. \quad (6.9)$$

This is a Riemann-type problem, i.e., a relation between the real and imaginary parts of a sectionally holomorphic function $\Delta(z)$ on an open contour. It is

observed that

$$\Delta'(z) = \exp\left\{\pi^{-1}\int_{\mu}^{\infty} d\omega[\omega - z]^{-1} \times \arg\,[G^*(\omega) + 2i\Gamma(\omega)\psi(\omega + i0)]\right\} \quad (6.10)$$

certainly satisfies (6.9) on R and is bounded at threshold.[11] The point at infinity as an end-point of the cut R is best studied by performing the inversion $u = z^{-1}$. Considering

$$\Delta(u^{-1}) = \exp\left\{\pi^{-1}\int_{\mu^{-1}}^{\infty} du'[u' - u]^{-1} \times \arg\,[G^*(u'^{-1}) + 2i\Gamma(u'^{-1})\psi_+(u'^{-1})]\right\}$$
$$= \exp\{\varphi(u)\}, \quad (6.11)$$

$$\kappa = \lim_{u\to 0} \pi^{-1}\varphi(u) = \lim_{u\to 0} \pi^{-1}[G^* + 2i\Gamma\psi_+]. \quad (6.12)$$

The index κ so defined is certainly bounded but will, in general, be different from an integer; $\varphi(u)$ is assumed to belong to the class H_0; i.e., it belongs to the class H of Hölder functions everywhere on $[0, \mu^{-1}]$ except at $u = 0$, where it undergoes a discontinuity of the first kind. Then, the limiting behavior of the Cauchy integral $\varphi(u)$ when $u = 0$ is approached is

$$\lim_{u\to 0} \varphi(u^{-1}) = -\pi^{-1} \arg\,[G^*(\infty) + 2i\Gamma(\infty)\psi_+(\infty)] \times \log u^{-1} + \varphi_0(u^{-1}), \quad (6.13)$$

where $\varphi_0(u^{-1})$ is regular when $u \to 0$. Hence

$$\lim_{u\to 0} \Delta'(u^{-1}) = u^{\kappa}\Delta_0'(u^{-1}), \quad \text{where } \Delta_0' \text{ is bounded at } u = 0. \quad (6.14)$$

To achieve a weak singularity at $u = 0$, it is necessary to multiply by a power of u; let it be λ such that $-1 < \lambda + \kappa < +1$. The classification of the possibilities concerning $u = 0$ is given by the short table:

(1) "special" end (κ integer): $\lambda + \kappa = 0$ solution bounded at origin;

(2) "nonspecial" end ($\kappa \neq$ integer)
- Class I: $\lambda + \kappa > 0$, solution vanishing at origin;
- Class II: $\lambda + \kappa < 0$, solution weakly discontinuous at origin.

$u^{\lambda}\Delta'(u^{-1})$ is the fundamental sectionally holomorphic solution in the u plane [for a given class in Case (2)].

[11] The threshold is a "special end" in Muskhelishvili's nomenclature, since $G^*(z^*)$ can be regarded as the continuation into the lower half-plane of $G(z)$ around $z = \mu$.

Reverting to the z plane, the factor $z^{-\lambda}$ would bring a pole or a zero at $z = 0$, and thus must be removed when constructing the fundamental solution in the z plane:

$$\Delta'(z) = \exp\left\{\frac{z}{\pi}\int_{\mu}^{\infty} \frac{d\omega}{\omega(\omega - z)} \times \arg\,[G^*(\omega) + 2i\Gamma(\omega)\psi_+(\omega)]. \quad (6.15)\right.$$

Let $\chi = \Delta'\chi'$ in (6.8); we get

$$(2i)^{-1}\,\mathrm{disc}_R\,[-\chi']^{-1} = \frac{\Gamma|\Delta'|}{|G - 2i\Gamma\psi_-|} \equiv \Gamma'. \quad (6.16)$$

$-[\chi'(z)]^{-1}$ is a Herglotz function determined by its discontinuity across the real axis, up to a Wigner R function. Since the elastic approximation is recovered by the limit $\Gamma \to \gamma$, $\Delta' \to 1$, $G \to 1$, we reach the conclusion that there is a one-to-one correspondence between the elastic and inelastic scattering amplitudes.[12]

We proceed now with an algorithm constructing a particular solution of (6.3). Since the purpose of a 3-particle unitary approximation, as we see it, is to incorporate the second-normal threshold singularity, we take the discontinuities across the left-hand cuts over from the input crossing-symmetric elastic approximation. We thus have a Cauchy boundary-value problem on L (given discontinuity) and (6.3) on R, and we will worry later about reestablishing crossing symmetry for the inelastic scattering amplitudes.[13]

Let us first solve the auxiliary problem of finding the effect of a small perturbation given as a density function satisfying the Hölder condition on R:

$$(2i)^{-1}[\psi_+(C + \epsilon)G - \psi_-(C + \epsilon)G^*] = \Gamma\psi_+(C + \epsilon)\psi_-(C + \epsilon) + [C + \epsilon]. \quad (6.17)$$

We write $\psi(C + \epsilon) \equiv \psi + \xi$ with $\mathrm{disc}_L\,\psi(C + \epsilon) = \mathrm{disc}_L\,\psi(C)$ or $\mathrm{disc}_L\,\xi = 0$. The perturbation ξ satisfies the boundary-value problem on R:

$$(2i)^{-1}\{\xi_+(G - 2i\Gamma\psi_-(C)) - \xi_-(G^* + 2i\Gamma\psi_+(C))\} = \Gamma\xi_+\xi_- + \epsilon. \quad (6.18)$$

In the limit $\epsilon \to 0$ uniformly, we have the trivial particular solution $\xi \equiv 0$. For this particular solution $\Gamma\xi_+\xi_- = O(\epsilon^2)$, and the problem is linearized to order

[12] This issue was obscured in the second sector of the Lee model and the charged scalar theory by the fact that two successive inversions of an analytic function were necessary to obtain a Herglotz function whose discontinuity across the cut is known [see M. Feinroth, Ph.D. thesis, The Massachusetts Institute of Technology, 1967 (unpublished)]. The procedure applicable in these two special cases does not extend to the more general situation considered here.

[13] In principle, an iteration scheme in the z plane can be set up using the inelastic amplitudes resulting from this section to compute the left-hand cut discontinuities in the first approximation.

ϵ^2. Let

$$\Delta(z; C) = \exp\left\{\frac{z}{\pi}\int_{\mu}^{\infty}\frac{d\omega}{\omega(\omega - z)} \times \arg\left[G^* + 2i\Gamma\psi_+(\omega; C)\right]\right\}, \quad (6.19)$$

$$\psi(z; C + \epsilon) = \psi(z; C) + \Delta(z; C)\cdot\frac{1}{\pi}\int_{\mu}^{\infty}\frac{d\omega}{\omega - z} \times \frac{\epsilon(\omega)}{|\Delta_-(\omega - i0; C)[G^* + 2i\Gamma\psi_+(C)]|}. \quad (6.20)$$

This defines $\psi(z; C + \epsilon) = \mathcal{F}[\psi(z; C); \epsilon]$ as a functional of $\psi(z; C)$ and $\epsilon(\omega)$.

We now partition the inhomogeneous term $C(\omega)$ into N (for instance) equal Hölder density functions on R:$\epsilon(\omega) = N^{-1}C(\omega)$, and recall that $\mathrm{disc}_L\,\psi(z; C = 0) = \mathrm{disc}_L\,\psi(z; C)$. Provided $\epsilon(\omega)$ is well behaved, we have

$$\psi(z; C) = \lim_{N\to\infty}\mathcal{F}^N[\psi(z; C = 0); C/N]. \quad (6.21)$$

The homogeneous boundary-value problem on R satisfied by $\psi(z; C = 0)$ is

$$(2i)^{-1}[\psi_+(C = 0)G - \psi_-(C = 0)G^*] = \Gamma\psi_+(C = 0)\psi_-(C = 0), \quad (6.22)$$

i.e., actually a particular case of (6.8). According to our general method, let

$$\psi(z; C = 0) = \Delta(z)\cdot\chi(C = 0),$$

$$\Delta(z) = \exp\left\{\pi^{-1}z\int_{2\mu}^{\infty}d\omega[\omega(\omega - z)]^{-1}\arg G^*(\omega)\right\},$$

$$(2i)^{-1}\,\mathrm{disc}_R\,[-\chi^{-1}(C = 0)] = \frac{\Gamma|\Delta|}{|G|} = \Gamma'(C = 0)$$

(Note: $\Gamma' = \gamma$ for $\mu < \omega < 2\mu$). (6.23)

Let $\mathrm{disc}_L\,\psi^{-1}(C = 0) = \mathrm{disc}_L\,\psi_e^{-1}$; then

$$\psi^{-1}(C = 0) = \frac{\Delta^{-1}}{2\pi i}\cdot\int_L\frac{d\omega}{\omega - z}\Delta\,\mathrm{disc}_L\,\psi_e^{-1} - \frac{\Delta^{-1}}{\pi}\cdot\int_R\frac{\Gamma'(C = 0)\,d\omega}{\omega - z} - \Delta^{-1}\times\text{Wigner } R \text{ function}. \quad (6.24)$$

Since $\Delta \to 1$, $\Gamma' \to \gamma$ in the elastic approximation, the one-to-one correspondence between elastic and inelastic scattering amplitudes is quite apparent from (6.24).

In conclusion, given a set of crossing-symmetric elastic-unitary amplitudes, we can, in principle, (a) construct a particular set of inelastic scattering amplitudes belonging to a 3-particle unitary S matrix through (6.20), (6.21), and (6.24), and (b) construct the most general of such amplitudes by adding the solution given by (6.15) and (6.16) of the homogeneous boundary-value problem (6.8).

7. CROSSING OF PRODUCTION AMPLITUDES

It is obvious that our construction of a 3-particle unitary S matrix violates many crossing-symmetry requirements. This is, of course, due to the fact that we have confined ourselves to physical region singularities and those singularities that frame the physical region, i.e., the poles at $\omega_1 = 0$ and $\omega_2 = 0$.

However, re-establishing crossing-symmetry invariance of the production amplitudes is not wholly a question of recognizing the existence of further unphysical region cuts with their boundary-value conditions. Some subtle points arise in connection with the "protection" prescription that enabled us to solve the 3-particle physical unitarity equation.

Let us first enumerate the complete "*a priori*" crossing-symmetry structure of the exact production amplitudes.

A. Substitution Law-Type Crossing Relations

A set of n-point amplitudes whose arguments are taken in the physical region is related by a linear substitution to the *same* set of n-point amplitudes in an unphysical region. The scattering amplitudes obey only crossing relations of this type:

$$A^{\alpha}(\omega) = \sum_{\beta}A^{\alpha\beta}_{(\frac{1}{2},\frac{1}{2})}A^{\beta}(-\omega). \quad (7.1)$$

The corresponding substitution laws of the production amplitudes are

$$A^{\alpha}_{\dot\alpha}(\omega_1\omega_2; \omega) = \sum_{\beta}A^{\alpha\beta}_{\dot\alpha}A^{\beta}_{\dot\alpha}(\omega_1, -\omega; -\omega_2), \quad (7.2)$$

with

$$A^{\alpha\beta}_{\dot\alpha} = (2\beta + 1)\begin{Bmatrix}1 & \frac{1}{2} & \alpha\\ 1 & \dot\alpha & \beta\end{Bmatrix} = (2I_\beta + 1)(2J_\beta + 1)\begin{Bmatrix}1 & \frac{1}{2} & I_\alpha\\ 1 & I_{\dot\alpha} & I_\beta\end{Bmatrix}\begin{Bmatrix}1 & \frac{1}{2} & J_\alpha\\ 1 & J_{\dot\alpha} & J_\beta\end{Bmatrix} \quad \text{for } l_1 = l_2 = 1. \quad (7.3)$$

Such crossing relations involve the crossing of two meson lines and preserve the "protected" character of rescattering blobs in unitarity discontinuity formulas (Fig. 14).

The substitution law (6.2) combined with Bose symmetry in the final state generates an invariance

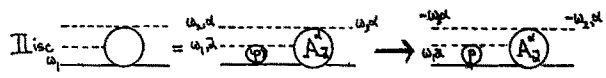


FIG. 14. Substitution law-type crossing. The protectedness of the rescattering blob is preserved.

group of the production amplitudes isomorphic to the permutation group of three objects. The elements of this invariance group in matrix notation are

Symmetry with respect to axis $\omega_1 = \omega_2$:

$$A(\omega_1\omega_2; \omega) = [\lambda]A(\omega_2\omega_1; \omega); \tag{7.4}$$

Symmetry with respect to axis $\omega = -\omega_1$:

$$A(\omega_1\omega_2; \omega) = [A]A(\omega_1, -\omega; -\omega_2); \tag{7.4'}$$

Rotation of $-2\pi/3$ in diagram of Fig. 10:

$$A(\omega_1\omega_2; \omega) = [A\lambda]A(-\omega\omega_1; -\omega_2); \tag{7.4''}$$

Rotation of $+2\pi/3$ in diagram of Fig. 10:

$$A(\omega_1\omega_2; \omega) = [\lambda A]A(\omega_2, -\omega; -\omega_1); \tag{7.4'''}$$

Symmetry with respect to axis:

$$A(\omega_1\omega_2; \omega) = [A\lambda A]A(-\omega\omega_2; -\omega_1). \tag{7.4''''}$$

Closure arises from the identity $[A\lambda A] = [\lambda A\lambda]$, itself a consequence of

$$\sum_\gamma (2\gamma+1)\begin{Bmatrix} \frac{1}{2} & 1 & \alpha \\ \dot\alpha & 1 & \gamma \end{Bmatrix}\begin{Bmatrix} 1 & \frac{1}{2} & \dot\alpha \\ 1 & \gamma & \dot\beta \end{Bmatrix}\begin{Bmatrix} \beta & \dot\beta & 1 \\ \gamma & \frac{1}{2} & 1 \end{Bmatrix} = \begin{Bmatrix} \frac{1}{2} & 1 & \beta \\ 1 & \frac{1}{2} & \dot\beta \\ \alpha & \dot\alpha & 1 \end{Bmatrix}. \tag{7.5}$$

B. Transfer-Type Crossing Relations

These are crossing relations connecting through a linear substitution set of n-point amplitudes to *another* set of n-point amplitudes whose kinematic variables are taken in an unphysical region, for instance, production five-point amplitudes to absorption five-point amplitudes through crossing of one meson line. This type of crossing does not conserve the "protected" or "unprotected" character of 4-point blobs appearing in unitarity discontinuity formulas, as exemplified by Fig. 15.

Crossing of one meson line thus transforms a subenergy-variable normal threshold into a total-energy variable normal threshold for an amplitude related to the original one by time reversal. In the process, the total and subchannel indices $(\alpha, \dot\alpha)$ exchange their roles. Crossing symmetry with respect to the axis $\omega_2 = 0$, if $\tilde A$ denote the absorption amplitudes, is written as

$$A^\alpha_{\dot\alpha}(\omega_1\omega_2; \omega) = f(\dot\alpha, \alpha)\tilde A^{\dot\alpha}_\alpha(\omega, -\omega_2; \omega_1), \tag{7.6}$$

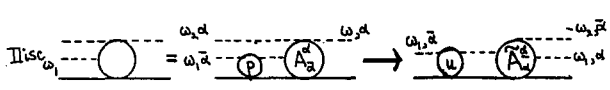


FIG. 15. Transfer-type crossing. The protectedness of the rescattering blob is altered.

where $f(\dot\alpha, \alpha)$ plays the role of a raising or lowering operator for the channel indices. [Every time it is feasible, the lower index $\dot\alpha = (\dot J, \dot I)$ denotes the subchannel, and the upperscript $\alpha = (J, I)$ denotes the total channel.]

$f(\dot\alpha, \alpha)$ can readily be determined by observing that Bose symmetry for the production amplitudes in the physical region implies

$$f(\dot\alpha, \alpha)\tilde A^\alpha_{\dot\alpha}(\omega, -\omega_2; \omega_1) = \sum_{\dot\beta} \lambda^\alpha{}_{\dot\alpha\dot\beta} f(\dot\beta, \alpha)\tilde A^{\dot\beta}_\alpha(\omega, -\omega_1; \omega_2), \tag{7.7}$$

i.e.,

$$\lambda^\alpha{}_{\dot\alpha\dot\beta} f(\dot\beta, \alpha) f^{-1}(\dot\alpha, \alpha) = A^{\alpha\dot\beta}_{\dot\alpha}.$$

Exchanging the numerical values of $(\dot\alpha, \alpha)$, we see that $f(\dot\alpha, \alpha) = f(\dot\alpha)f^{-1}(\alpha)$, i.e.,

$$f(\alpha) = (-)^{J+I}[(2J+1)(2I+1)]^{\frac{1}{2}}.$$

Hence the useful relations

$$A^{\alpha\dot\beta}_{\dot\alpha} = \lambda^\alpha{}_{\dot\alpha\dot\beta} f(\dot\beta, \alpha), \quad \lambda^\alpha{}_{\dot\alpha\dot\beta} = A^{\alpha\dot\beta}_{\dot\alpha} f(\alpha, \dot\beta). \tag{7.8}$$

The adjunction of a new symmetry with respect to the axis $\omega_2 = 0$ in the diagram of Fig. 10 generates two new symmetries in view of the invariance group of (i):

(1) Reflection with respect to axis $\omega_1 = 0$:

$$A^\alpha_{\dot\alpha}(\omega_1\omega_2; \omega) = \sum_{\dot\beta,\beta} \lambda^\alpha{}_{\dot\alpha\dot\beta} A^{\alpha\dot\beta}_\beta f(\dot\beta, \beta)\tilde A^{\dot\beta}_\beta(-\omega_1\omega; \omega_2). \tag{7.9}$$

Proof:

$$\begin{aligned} A^\alpha_{\dot\alpha}(\omega_1\omega_2; \omega) &= \sum_{\dot\beta} \lambda^\alpha{}_{\dot\alpha\dot\beta} A^\alpha_{\dot\beta}(\omega_2\omega_1; \omega) \\ &= \sum_{\dot\beta} \lambda^\alpha{}_{\dot\alpha\dot\beta} f(\dot\beta, \alpha)\tilde A^{\dot\beta}_\alpha(\omega, -\omega_1; \omega_2) \\ &= \sum_{\dot\beta\beta} \lambda^\alpha{}_{\dot\alpha\dot\beta} f(\dot\beta, \alpha)\lambda^{\dot\beta}{}_{\dot\alpha\beta}\tilde A^{\dot\beta}_\beta(-\omega_1\omega; +\omega_2). \end{aligned}$$

Finally from $f(\dot\beta, \alpha) = f(\dot\beta, \beta)f(\beta, \dot\alpha)$ and

$$f(\beta, \alpha)\lambda^{\dot\beta}{}_{\dot\alpha\beta} = A^{\alpha\dot\beta}_{\dot\alpha},$$

(7.9) follows.

(2) Reflection with respect to axis $\omega = 0$:

$$\begin{aligned} &A^\alpha_{\dot\alpha}(\omega_1\omega_2; \omega) \\ &= \sum_{\dot\beta\beta\dot\gamma\nu} A^{\alpha\dot\beta}_{\dot\alpha}\lambda^{\dot\beta}{}_{\dot\alpha\beta} A^{\beta\dot\gamma}_{\dot\beta} f(\dot\beta, \dot\gamma)\lambda^{\dot\beta}_{\dot\gamma\nu}\tilde A^{\dot\beta}_\nu(-\omega_2, -\omega_1; -\omega). \end{aligned} \tag{7.10}$$

Proof:

$$\begin{aligned} &A^\alpha_{\dot\alpha}(\omega_1\omega_2; \omega) \\ &\quad= \sum_{\dot\beta} A^{\alpha\dot\beta}_{\dot\alpha} A^{\dot\beta}_{\dot\alpha}(\omega_1, -\omega; -\omega_2) \\ &\quad= \sum_{\dot\beta\beta\dot\gamma} A^{\alpha\dot\beta}_{\dot\alpha}\lambda^\beta{}_{\dot\alpha\dot\beta} A^{\beta\dot\gamma}_{\dot\beta} f(\dot\beta, \dot\gamma)\tilde A^{\dot\beta}_{\dot\gamma}(-\omega_1, -\omega_2; -\omega). \end{aligned}$$

Hence (7.10). It is worth noting that the last relation of the proof gives the symmetry operation with respect to the origin in the diagram of Fig. 10.

If time-reversal invariance holds—and we have assumed that it does in our basic ansatz—the relations

(7.6), (7.9), and (7.10) become new symmetry properties of the production amplitudes. The complete invariance group represented on Fig. 10 is reminiscent of the crystallographic 2-dimensional point group labeled (6 mm). We have the following negative statement: (a) exact and complete crossing symmetry; (b) time-reversal invariance, and (c) 3-particle physical region unitarity—as defined earlier—are incompatible because a reflection with respect to $\omega_2 = 0$, for instance, would generate a 3-particle normal threshold in the subenergy variable ω_1. Such a discontinuity term represented on Fig. 16 is clearly relevant to the 4-particle unitary approximation. Such a difficulty in arbitrarily breaking the S-matrix normal threshold singularity structure was, of course, absent in the elastic approximation, due to the circumstance that it involved only amplitudes with an equal number of ingoing and outgoing lines.

However, a certain type of approximate crossing is implied by our basic ansatz for the production amplitudes. It is recalled that in (5.3) and (5.4) compensating structures (the second terms) were needed to satisfy 3-particle unitarity. To understand the origin of these terms, we note that the difficulty mentioned above does not appear, provided we add the prescription that for each meson line crossed, the "protection" character of the 4-point blobs be changed. Then, the discontinuity term of Fig. 16 is not generated and it is possible to reconcile 3-particle unitarity with this "approximate" crossing invariance. The physical meaning of the compensating terms in (5.3) and (5.4) is then understood: they are required by an "approximate crossing symmetry" on the demarcation lines $\omega_1 = 0$, $\omega_2 = 0$ between the physical region sector (Fig. 10) and the crossed regions sectors. Indeed, from $f(\dot{\alpha}, \alpha)\lambda^{\alpha}_{\alpha,\frac{1}{2}\frac{1}{2}} = -\lambda^{\alpha}_{\dot{\alpha},\frac{1}{2}\frac{1}{2}}$, we have

$$f(\dot{\alpha}, \alpha)d^{\alpha}_{\alpha}(\omega) = (3g/\sqrt{2})\lambda^{\alpha}_{\alpha,\frac{1}{2}\frac{1}{2}}[A^{\alpha}(\omega) - \mathcal{A}_{\alpha}(\omega)] \to d^{\alpha}_{\dot{\alpha}}(\omega), \tag{7.11}$$

by changing the protectedness of the 4-point amplitudes.

It can also be checked explicitly that all pole terms "approximately" cross into each other when the pole term at $\omega = 0$ of Fig. 17 is added:

$$3g\sqrt{2}\,\omega^{-1}\delta_{\alpha,\frac{1}{2}\frac{1}{2}} f(\dot{\alpha}, \alpha)\mathcal{A}^{\alpha}(\omega_1). \tag{7.12}$$

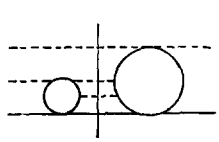


FIG. 16. This subenergy discontinuity term is obtained by crossing one meson line in a 3-particle unitarity total-energy discontinuity term; it appears in the 4-particle unitarity relations.

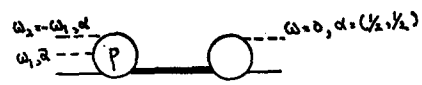


FIG. 17. The pole term at $\omega = 0$ in the z plane.

In Appendix B, an "approximately" crossing-symmetric homogeneous Riemann–Hilbert problem (of dimension n^2) is set up that includes the unphysical region cuts at $z_1 = -\mu + i\epsilon$ and $z_1 = \omega + \mu - i\epsilon$, but yields the n Riemann–Hilbert problems of Sec. IV (labeled by α) when the cuts L_1, L_2 are neglected.

However, since the boundary conditions on L_1 and L_2 make use of the inelastic phaseshifts, an "approximately" crossing symmetric representation of the production amplitudes built in this way cannot be useful except in an iteration scheme (somewhat analogous to the N/D method for the scattering amplitudes).

An approximate calculation of the P^{+}_{11} inelasticity parameter in πN scattering using the framework set up here will be described elsewhere.

ACKNOWLEDGMENTS

The author is indebted to Professor J. B. Bronzan for his support and encouragement during the completion of his Ph.D. thesis at The Massachusetts Institute of Technology and to Professor K. Huang and Professor H. Feshbach for their kind interest. Thanks are also due to Dr. Lam for valuable comments on the manuscript and, more generally, to the theoretical group at McGill for its hospitality.

APPENDIX A

Theorem: There exists one and only one fundamental matrix solution with the following property under exchange of the cuts R_1 and R_2:

$$\sum_{\beta} \lambda^{\alpha}_{\alpha\beta}\Omega^{\alpha}_{\beta(i)}(z - z_1; z) = \Omega^{\alpha}_{\alpha(i)}(z_1; z)c_i. \tag{4.5}$$

Proof: Let $\Omega^{\alpha}_{\alpha(i)}(z_1; z)$ be any fundamental matrix solution of "h." The individual vector solutions have degree $-\kappa_i$ at infinity (κ_i: "individual indices" such that $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_n$ $\kappa = \sum_i \kappa_i$, total index). Now, $\sum_{\beta} \lambda^{\alpha}_{\alpha\beta}\Omega^{\alpha}_{\beta(i)}(z - z_1; z)$ is readily seen to satisfy exactly the Riemann–Hilbert problem as $\Omega^{\alpha}_{\alpha(i)}(z_1; z)$. It also has the same degree at infinity in virtue of a theorem by Sugawara and Kanazawa[14] [indeed, $|\exp(2i\delta_{\alpha}(\omega))| = 1$, so that $\Omega^{\alpha}_{\alpha(i)}(z_1; z)$ is bounded by the same power of z_1 in the upper and lower half-planes]. By application of the fundamental theorem

[14] M. Sugawara and A. Kanazawa, Phys. Rev. 123, 1895 (1961).

(4.4), we have successively

$$\sum_\beta \lambda^\alpha{}_{\alpha\beta}\Omega^\alpha{}_{(1)}(z - z_1; z) = \Omega^\alpha_{\alpha(1)}(z_1; z)c_1, \quad c_1 \neq 0, \tag{A1}$$

$$\sum_\beta \lambda^\alpha{}_{\alpha\beta}\Omega^\alpha_{\beta(2)}(z - z_1; z) = \Omega^\alpha_{\alpha(1)}(z_1; z)P_{\kappa_1-\kappa_n}(z_1) + \Omega^\alpha_{\alpha(2)}(z_1; z)c_2, \tag{A2}$$

$$\vdots$$

$$\sum_\beta \lambda^\alpha{}_{\alpha\beta}\Omega^\alpha_{\beta(n)}(z - z_1; z) = \Omega^\alpha_{\alpha(1)}(z_1; z)P_{\kappa_1-\kappa_n}(z_1) + \cdots + \Omega^\alpha_{\alpha(n)}(z_1; z)c_n, \quad c_n \neq 0.$$

The first relation is in the desired form. Substituting $(z - z_1)$ for z_1 in (A2), we have

$$c_2^{-1}\sum_\beta \lambda^\alpha{}_{\alpha\beta}\Omega^\alpha_{\beta(2)}(z_1; z) = c_2^{-1}\Omega^\alpha_{\alpha(1)}(z - z_1; z)P_{\kappa_1-\kappa_2}(z - z_1) + \Omega^\alpha_{\alpha(2)}(z - z_1; z).$$

Combining this last relation with (A1) and (A2), we have

$$(c_2 - c_2^{-1})\Omega^\alpha_{\alpha(2)}(z_1; z) + \{P_{\kappa_1-\kappa_2}(z_1) + c_1c_2P_{\kappa_1-\kappa_2}(z - z_1)\}\Omega^\alpha_{\alpha(1)}(z_1; z) = 0.$$

Because of the linear independence—with arbitrary polynomial coefficients—of the fundamental vector solutions, no such relationship can be satisfied unless

$$c_2 = c_2^{-1}; \quad P_{\kappa_1-\kappa_2}(z_1) + c_1c_2P_{\kappa_1-\kappa_2}(z - z_1) = 0. \tag{A3}$$

This stringent restriction on the polynomial $P_{\kappa_1-\kappa_2}(z_1)$ allows us to define a new fundamental vector solution:

$$\Omega'^\alpha_{\alpha'2)}(z_1; z) = \Omega^\alpha_{\alpha(2)}(z_1; z) + 2^{-1}c_2P_{\kappa_1-\kappa_2}(z_1)\Omega^\alpha_{\alpha(1)}(z_1; z),$$

such that

$$\sum_\beta \lambda^\alpha{}_{\alpha\beta}\Omega'^\alpha_{\beta(2)}(z - z_1; z) = c_2\Omega'^\alpha_{\alpha(2)}(z_1; z)$$

[explicitly using (A3)].

One proceeds similarly to construct the other $\Omega'^\alpha_{\alpha(i)}(z_1; z)$, $i > 2$.

APPENDIX B: THE ENLARGED RIEMANN-HILBERT PROBLEM

As is apparent on Fig. 18, approximate crossing symmetry—i.e., maximal crossing symmetry compatible with the 3-particle unitarity approximation—requires two extra branch points in the z_1 plane at $z_1 = -\mu + i\epsilon$ and $z_1 = z + \mu - i\epsilon$, respectively. These two branch points are mapped by crossing—with $[A\lambda A] = [\lambda A\lambda]$ and $[A]$ as substitution matrices, respectively—on the branch point at $z = \mu - i\epsilon$ in the z plane. The phase relation on this total energy branch cut is

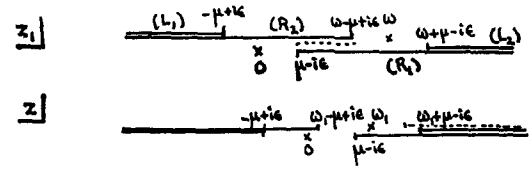


FIG. 18. The enlarged Riemann-Hilbert problem in the z_1 and z planes.

$$A^\alpha_\alpha(\omega_1; \omega + i0) = \exp[2i\delta^i_\alpha(\omega)]A^\alpha_\alpha(\omega_1; \omega - i0), \tag{B1}$$

where the scattering phaseshift belongs to the 3-particle approximation, since the total energy prescattering blob is "unprotected." The application of the crossing substitution laws (7.4′) and (7.4‴) then generates the boundary conditions on L_1 and L_2.

The complete n^2-dimensional resulting homogeneous Riemann-Hilbert problem in the z_1 plane is then

$$\text{“}H\text{”}\begin{cases} \phi^\alpha_\alpha(\omega_1 + i0; \omega) = \exp[2i\delta^e_\alpha(\omega_1)]\phi^\alpha_\alpha(\omega_1 - i0; \omega) & \text{for } (\omega_1 \in R_1), \\ \sum_\beta \lambda^\alpha{}_{\alpha\beta}\phi^\alpha_\beta(\omega_1 + i0; \omega) = \exp[-2i\delta^e_\alpha(\omega - \omega_1)]\sum_\beta \lambda^\alpha{}_{\alpha\beta}\phi^\alpha_\beta(\omega_1 - i0; \omega) & \text{for } (\omega_1 \in R_2), \\ \sum_\beta A^{\alpha\beta}_\alpha\phi^\beta_\alpha(\omega_1 + i0; \omega) = \exp[2i\delta^i_\alpha(\omega_1 - \omega)]\sum_\beta A^{\alpha\beta}_\alpha\phi^\beta_\alpha(\omega_1 - i0; \omega) & \text{for } (\omega_1 \in L_2), \\ \sum_{\beta\beta} [\lambda A\lambda]^{\alpha\beta}_{\alpha\beta}\phi^\beta_\beta(\omega_1 + i0; \omega) = \exp[-2i\delta^i_\alpha(-\omega)]\sum_{\beta\beta} [\lambda A\lambda]^{\alpha\beta}_{\alpha\beta}\phi^\beta_\beta(\omega_1 - i0; \omega) & \text{for } (\omega_1 \in L_1). \end{cases} \tag{B2, B3}$$

Since the matrix λ and the phase factor commute in the last relation, we can write (B3) with $[A\lambda]$ instead of $[\lambda A\lambda] = [A\lambda A]$; this corresponds to the fact that the thresholds $\omega_1 = -\mu$, $\omega = \mu$ are also interchanged by a rotation of $(-2\pi/3)$ in the diagram of Fig. 9. Let $\mathbf{\Omega}^\alpha_{\alpha(s)}(z_1; z)$ be a $(n^2 \times n^2)$ fundamental matrix solution of "H," where the index s runs from 1 to n^2. In the limit $\delta^i_\alpha(\omega) \to 0$, i.e., when the cuts L_1 and L_2 are neglected, $\mathbf{\Omega}^\alpha_{\alpha(s)}$ degenerates into n diagonal blocs, the n fundamental matrix solutions of the n smaller (n-dimensional) Riemann-Hilbert problems on R_1 and R_2 labeled by α.

Since the boundary conditions on L_1 and L_2 are exchanged by Bose symmetry just the way R_1 and R_2 were exchanged in the small problem, we can still apply the theorem of Appendix A to construct a fundamental matrix solution with simple Bose symmetry property. By the way this new fundamental

matrix is constructed, the index (s) splits into (i,j) with $i,j = 1,\cdots,n$:

$$\sum_{\beta} \lambda^{\alpha}{}_{\alpha\beta}\Omega^{\alpha(j)}_{\beta(i)}(z_1; z) = \Omega^{\alpha(j)}_{\alpha(i)}(z - z_1; z)c^{\alpha}_{(i)}. \quad \text{(B4)}$$

Let us now turn to the large Riemann–Hilbert problem in the z plane also represented on Fig. 16. Let $\Omega^{\alpha}_{\alpha}(z; z_1)$ be its fundamental matrix solution; it depends only parametrically on z_1 through the endpoint positions of some cuts. Since crossing maps the point at infinity in the z plane onto itself, by the theorem of Appendix A there exists a unique fundamental matrix solution satisfying

$$\Omega^{\alpha}_{\alpha}(z; z_1) = \sum_{\beta} A^{\alpha\beta}_{\alpha}\Omega^{\beta}_{\alpha}(z_1 - z; z_1). \quad \text{(B5)}$$

It is readily verified that the ansatz

$$\Omega^{\alpha}_{\alpha}(z; z_1) = \sum_{\beta\beta} [A\lambda A]^{\alpha\beta}_{\alpha\beta}\Omega^{\beta}_{\beta}(-z; -z_1)$$

satisfies (B5) and hence we reach the conclusion that the large $\Omega^{\alpha(j)}_{\alpha(i)}(z; z_1)$ gives us the solution of the homogeneous Riemann–Hilbert problems in both the z_1 and z planes (trivially, also z_2).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Intrinsic Spinor Techniques with Applications to the Lorentz Group and the Dirac Equation

C. P. LUEHR AND M. ROSENBAUM
TEMPO, Center for Advanced Studies, General Electric Company, Santa Barbara, California

(Received 11 June 1968)

An intrinsic spinor formalism is developed by extending the axiomatic approach of a previous paper [J. Math. Phys. 9, 284 (1968)] in which abstract tensor techniques in Minkowski space were discussed. Comparable notational and calculational advantages are achieved, and, in particular, the cumbersome spinor indices are not required. The advantages and practicality of the method are evidenced in a discussion of the Dirac equation and a novel derivation of the two- and four-dimensional spinor representations of the homogeneous Lorentz group.

1. INTRODUCTION

In a previous paper,[1] (hereafter referred to as I) we described an abstract formalism for tensor analysis in Minkowski space and demonstrated its notational simplicity and calculational advantages by applying it to several problems in special relativity. The purpose of the present paper is to show how a corresponding intrinsic spinor formalism entailing similar advantages can be developed. A notable feature of this formalism is that the cumbersome manipulations associated with spinor indices can be avoided, a fact which results in considerable algebraic simplification.

To begin, we give in Sec. 2 the basic definitions and axioms of the two-dimensional spinor space and its conjugate space in intrinsic notation. Dyadics attached to the spinor space and their operations are then defined, and the procedure for relating the intrinsic formalism to the conventional one in terms of components[2] is established.

In Sec. 3, unimodular linear spinor transformations are discussed and proven to be always expressible in exponential form. In Sec. 4 it is shown that by merely defining an appropriate scalar product in the space of Hermitian spinor dyadics, it becomes isomorphic to Minkowski space. Thus, world vectors and Hermitian spinor dyadics become intrinsically identical entities, and can therefore be denoted by the same symbol. It is only their alternate representations, either in terms of an orthonormal Minkowski basis or a spinor basis, that are distinct. It is also shown that in the conventional component notation, the matrices connecting world vectors with spinors are simply hybrid components of the unit dyadic for Minkowski space.

Section 5 deals with the two-dimensional spinorial representation of the restricted homogeneous Lorentz transformation, following an approach that is different from others appearing in the literature.[3] Essentially,

[1] C. P. Luehr and M. Rosenbaum, J. Math. Phys. 9, 284 (1968). We note that in this paper four-vectors in Minkowski space were denoted by underlined boldface (Roman) type.

[2] See, for example, E. M. Corson, *Tensors, Spinors and Relativistic Wave-Equations* (Blackie and Son Ltd., London, 1953); E. Cartan, *The Theory of Spinors* (The Massachusetts Institute of Technology Press, Cambridge, Mass., 1966); J. Rzewuski, *Field Theory* (Państowowe Wydawnictwo Naukowe, Warsaw, 1958), Part I.

[3] See, for example, W. L. Bade and H. Jehle, Rev. Mod. Phys. 25, 714 (1953), and references therein; P. Roman, *Theory of Elementary Particles* (North-Holland Publ. Co., Amsterdam, 1961), second ed; J. L. Synge, *Relativity: The Special Theory* (North-Holland Publ. Co., Amsterdam, 1965), second ed., p. 102; A. J. Macfarlane, J. Math. Phys. 3, 1116 (1962); I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon Press Ltd., Oxford, 1963).

matrix is constructed, the index (s) splits into (i,j) with $i,j = 1,\cdots,n$:

$$\sum_{\beta} \lambda^{\alpha}{}_{\alpha\beta}\Omega^{\alpha(j)}_{\beta(i)}(z_1; z) = \Omega^{\alpha(j)}_{\alpha(i)}(z - z_1; z)c^{\alpha}_{(i)}. \quad \text{(B4)}$$

Let us now turn to the large Riemann–Hilbert problem in the z plane also represented on Fig. 16. Let $\Omega^{\alpha}_{\alpha}(z; z_1)$ be its fundamental matrix solution; it depends only parametrically on z_1 through the endpoint positions of some cuts. Since crossing maps the point at infinity in the z plane onto itself, by the theorem of Appendix A there exists a unique fundamental matrix solution satisfying

$$\Omega^{\alpha}_{\alpha}(z; z_1) = \sum_{\beta} A^{\alpha\beta}_{\alpha}\Omega^{\beta}_{\alpha}(z_1 - z; z_1). \quad \text{(B5)}$$

It is readily verified that the ansatz

$$\Omega^{\alpha}_{\alpha}(z; z_1) = \sum_{\beta\beta} [A\lambda A]^{\alpha\beta}_{\alpha\beta}\Omega^{\beta}_{\beta}(-z; -z_1)$$

satisfies (B5) and hence we reach the conclusion that the large $\Omega^{\alpha(j)}_{\alpha(i)}(z; z_1)$ gives us the solution of the homogeneous Riemann–Hilbert problems in both the z_1 and z planes (trivially, also z_2).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Intrinsic Spinor Techniques with Applications to the Lorentz Group and the Dirac Equation

C. P. LUEHR AND M. ROSENBAUM
TEMPO, Center for Advanced Studies, General Electric Company, Santa Barbara, California

(Received 11 June 1968)

An intrinsic spinor formalism is developed by extending the axiomatic approach of a previous paper [J. Math. Phys. 9, 284 (1968)] in which abstract tensor techniques in Minkowski space were discussed. Comparable notational and calculational advantages are achieved, and, in particular, the cumbersome spinor indices are not required. The advantages and practicality of the method are evidenced in a discussion of the Dirac equation and a novel derivation of the two- and four-dimensional spinor representations of the homogeneous Lorentz group.

1. INTRODUCTION

In a previous paper,[1] (hereafter referred to as I) we described an abstract formalism for tensor analysis in Minkowski space and demonstrated its notational simplicity and calculational advantages by applying it to several problems in special relativity. The purpose of the present paper is to show how a corresponding intrinsic spinor formalism entailing similar advantages can be developed. A notable feature of this formalism is that the cumbersome manipulations associated with spinor indices can be avoided, a fact which results in considerable algebraic simplification.

To begin, we give in Sec. 2 the basic definitions and axioms of the two-dimensional spinor space and its conjugate space in intrinsic notation. Dyadics attached to the spinor space and their operations are then defined, and the procedure for relating the intrinsic formalism to the conventional one in terms of components[2] is established.

In Sec. 3, unimodular linear spinor transformations are discussed and proven to be always expressible in exponential form. In Sec. 4 it is shown that by merely defining an appropriate scalar product in the space of Hermitian spinor dyadics, it becomes isomorphic to Minkowski space. Thus, world vectors and Hermitian spinor dyadics become intrinsically identical entities, and can therefore be denoted by the same symbol. It is only their alternate representations, either in terms of an orthonormal Minkowski basis or a spinor basis, that are distinct. It is also shown that in the conventional component notation, the matrices connecting world vectors with spinors are simply hybrid components of the unit dyadic for Minkowski space.

Section 5 deals with the two-dimensional spinorial representation of the restricted homogeneous Lorentz transformation, following an approach that is different from others appearing in the literature.[3] Essentially,

[1] C. P. Luehr and M. Rosenbaum, J. Math. Phys. 9, 284 (1968). We note that in this paper four-vectors in Minkowski space were denoted by underlined boldface (Roman) type.

[2] See, for example, E. M. Corson, *Tensors, Spinors and Relativistic Wave-Equations* (Blackie and Son Ltd., London, 1953); E. Cartan, *The Theory of Spinors* (The Massachusetts Institute of Technology Press, Cambridge, Mass., 1966); J. Rzewuski, *Field Theory* (Państowowe Wydawnictwo Naukowe, Warsaw, 1958), Part I.

[3] See, for example, W. L. Bade and H. Jehle, Rev. Mod. Phys. 25, 714 (1953), and references therein; P. Roman, *Theory of Elementary Particles* (North-Holland Publ. Co., Amsterdam, 1961), second ed; J. L. Synge, *Relativity: The Special Theory* (North-Holland Publ. Co., Amsterdam, 1965), second ed., p. 102; A. J. Macfarlane, J. Math. Phys. 3, 1116 (1962); I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon Press Ltd., Oxford, 1963).

we arrive at the double-valued homomorphism straightforwardly by taking advantage of one of the basic forms of the Lorentz transformation derived in I, which, when rewritten in terms of a spinor basis, can then be factored into a tensor product of spinor transformations.

Section 6 is devoted to a discussion of the Dirac equation, making use of intrinsic spinor techniques. The Dirac $\mathbf{\Gamma}^\mu$ appear as symmetric spinor dyadics formed from the basis vectors of Minkowski space, and their anticommutation relations are easily derived. It is also shown how the different matrix representations may be obtained. In addition, we give a concise derivation of the continuity equation for the Dirac field expressed in spinor form. Section 7 contains an interesting method of extending the results of Sec. 5 to obtain explicit four-dimensional intrinsic spinor representations of the homogeneous restricted Lorentz transformations.

Finally, in Sec. 8, we extend our results to cover the full homogeneous Lorentz group by including intrinsic spinor representations for the inversions.

2. INTRINSIC SPINOR ANALYSIS

Let $\mathcal{S}_2$ denote a two-dimensional symplectic space, i.e., a linear vector space over the field $\mathcal{F}$ of complex numbers in which there is a nondegenerate skew-symmetric bilinear inner product $\boldsymbol{u}\cdot\boldsymbol{v}$. Explicitly, given any $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{S}_2$, and $\alpha \in \mathcal{F}$, then:

$$\boldsymbol{u}\cdot\boldsymbol{v} = -\boldsymbol{v}\cdot\boldsymbol{u}, \tag{1a}$$

$$(\alpha\boldsymbol{u})\cdot\boldsymbol{v} = \alpha(\boldsymbol{u}\cdot\boldsymbol{v}), \quad \boldsymbol{u}\cdot(\alpha\boldsymbol{v}) = \alpha(\boldsymbol{u}\cdot\boldsymbol{v}), \tag{1b}$$

$$(\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{w} = \boldsymbol{u}\cdot\boldsymbol{w} + \boldsymbol{v}\cdot\boldsymbol{w},$$

$$\boldsymbol{u}\cdot(\boldsymbol{v}+\boldsymbol{w}) = \boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{u}\cdot\boldsymbol{w}, \tag{1c}$$

$$\boldsymbol{u}\cdot\boldsymbol{v} = 0 \quad \text{for all} \quad \boldsymbol{v}\in\mathcal{S}_2 \quad \text{implies} \quad \boldsymbol{u} = 0. \tag{1d}$$

For any basis $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ in $\mathcal{S}_2$, the reciprocal basis $\boldsymbol{h}^1$, $\boldsymbol{h}^2$ is defined to satisfy

$$\boldsymbol{h}^a\cdot\boldsymbol{h}_b = \delta^a_b, \tag{2}$$

where δ^a_b is the ordinary Kronecker delta. If we now require that

$$\boldsymbol{h}_1\cdot\boldsymbol{h}_2 = 1, \tag{3}$$

then

$$\boldsymbol{h}^1 = -\boldsymbol{h}_2, \quad \text{and} \quad \boldsymbol{h}^2 = \boldsymbol{h}_1. \tag{4}$$

In terms of these bases, any vector in $\mathcal{S}_2$ can be written as[4]

$$\boldsymbol{u} = u^a\boldsymbol{h}_a = u_a\boldsymbol{h}^a, \tag{5}$$

where u^a and u_a are contravariant and covariant components of $\boldsymbol{u}$, respectively. From Eq. (2) it readily follows that

$$u^a = \boldsymbol{h}^a\cdot\boldsymbol{u} = -\boldsymbol{u}\cdot\boldsymbol{h}^a,$$
$$u_a = \boldsymbol{u}\cdot\boldsymbol{h}_a = -\boldsymbol{h}_a\cdot\boldsymbol{u}, \tag{6}$$

and

$$\boldsymbol{u}\cdot\boldsymbol{v} = u_a v^b \boldsymbol{h}^a\cdot\boldsymbol{h}_b = u_a v^b\delta^a_b = u_a v^a,$$
$$\boldsymbol{u}\cdot\boldsymbol{v} = u^a v_b \boldsymbol{h}_a\cdot\boldsymbol{h}^b = -u^a v_b\delta^b_a = -u^a v_a. \tag{7}$$

The antilinear operation of complex conjugation, which maps $\mathcal{S}_2$ onto the conjugate[5] space $\bar{\mathcal{S}}_2$, can now be defined axiomatically by the rules

$$\overline{(\boldsymbol{u}+\boldsymbol{v})} = \bar{\boldsymbol{u}} + \bar{\boldsymbol{v}},$$
$$\overline{(\alpha\boldsymbol{u})} = \bar{\alpha}\bar{\boldsymbol{u}},$$
$$\overline{(\boldsymbol{u}\cdot\boldsymbol{v})} = \bar{\boldsymbol{u}}\cdot\bar{\boldsymbol{v}}, \tag{8}$$
$$\bar{\bar{\boldsymbol{u}}} = \boldsymbol{u}.$$

Accordingly, the complex conjugate of Eq. (5) gives

$$\bar{\boldsymbol{u}} = \bar{u}^a\bar{\boldsymbol{h}}_a = \bar{u}_a\bar{\boldsymbol{h}}^a, \tag{9}$$

where the components $\bar{u}^a$ and $\bar{u}_a$ are usually denoted in the literature by the "dotted indices":

$$u^{\dot{a}} \equiv \bar{u}^a, \quad u_{\dot{a}} \equiv \bar{u}_a.$$

In I, we reviewed dyadic algebra in Minkowski space. Similar definitions and operations[6] can be considered in $\mathcal{S}_2$. Thus, the tensor product (written as $\boldsymbol{uv}$ or $\boldsymbol{u}\otimes\boldsymbol{v}$) of two vectors $\boldsymbol{u}$, $\boldsymbol{v}$ in $\mathcal{S}_2$ constitutes a dyad.[7] A dyadic[8] is a sum of dyads:

$$\mathbf{K} = \sum_l \boldsymbol{u}_l\boldsymbol{v}_l.$$

The two products of $\mathbf{K}$ with a vector $\boldsymbol{w}\in\mathcal{S}_2$ are

$$\mathbf{K}\cdot\boldsymbol{w} = \left(\sum_l \boldsymbol{u}_l\boldsymbol{v}_l\right)\cdot\boldsymbol{w} = \sum_l \boldsymbol{u}_l(\boldsymbol{v}_l\cdot\boldsymbol{w}),$$
$$\boldsymbol{w}\cdot\mathbf{K} = \boldsymbol{w}\cdot\left(\sum_l \boldsymbol{u}_l\boldsymbol{v}_l\right) = \sum_l (\boldsymbol{w}\cdot\boldsymbol{u}_l)\boldsymbol{v}_l. \tag{10}$$

Additional operations are:

1. The scalar of $\mathbf{K}$, denoted by $\mathbf{K}_s$:

$$\mathbf{K}_s = \left(\sum_l \boldsymbol{u}_l\boldsymbol{v}_l\right)_s = \sum_l \boldsymbol{u}_l\cdot\boldsymbol{v}_l. \tag{11}$$

2. The transpose of $\mathbf{K}$, denoted by $\tilde{\mathbf{K}}$ or $\mathbf{K}_T$:

$$\tilde{\mathbf{K}} \equiv \mathbf{K}_T = \sum_l (\boldsymbol{u}_l\boldsymbol{v}_l)_T = \sum_l \boldsymbol{v}_l\boldsymbol{u}_l. \tag{12}$$

[4] The summation convention on repeated indices is always implied unless stipulated otherwise.

[5] We will use a bar to denote complex conjugation of spinors and spinor components. A star will be reserved for complex conjugation of complex four-vectors.

[6] Care must be exercised, however, in carrying out this analogy, since there are some concepts in $\mathcal{M}_4$ (such as the vector of a dyadic and "cross products") which do not apply to $\mathcal{S}_2$.

[7] This is what is called in some of the literature an affine tensor attached to the vector space $\mathcal{S}_2$. See, e.g., A. Lichnerowicz, *Elements of Tensor Calculus* (Methuen and Co. Ltd., London, 1962).

[8] Spinor dyadics will be denoted by boldface capital Roman type and boldface Greek type.

3. The product of $\mathbf{K}$ with another dyadic $\mathbf{M} = \sum_k \boldsymbol{w}_k \boldsymbol{y}_k$, which yields a new dyadic:

$$\mathbf{N} = \mathbf{K} \cdot \mathbf{M} = \left(\sum_l \boldsymbol{u}_l \boldsymbol{v}_l\right) \cdot \left(\sum_k \boldsymbol{w}_k \boldsymbol{y}_k\right) = \sum_l \sum_k (\boldsymbol{v}_l \cdot \boldsymbol{w}_k) \boldsymbol{u}_l \boldsymbol{y}_k. \tag{13}$$

4. The double product of two dyadics, which yields a scalar:

$$\mathbf{K} : \mathbf{M} = \left(\sum_l \boldsymbol{u}_l \boldsymbol{v}_l\right) : \left(\sum_k \boldsymbol{w}_k \boldsymbol{y}_k\right) = \sum_l \sum_k (\boldsymbol{u}_l \cdot \boldsymbol{w}_k)(\boldsymbol{v}_l \cdot \boldsymbol{y}_k). \tag{14}$$

5. The exterior product of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{S}_2$:

$$\boldsymbol{u} \wedge \boldsymbol{v} = \boldsymbol{u}\boldsymbol{v} - \boldsymbol{v}\boldsymbol{u}. \tag{15}$$

6. The complex conjugate of a dyadic

$$\bar{\mathbf{K}} = \overline{\left(\sum_l \boldsymbol{u}_l \boldsymbol{v}_l\right)} = \sum_l \bar{\boldsymbol{u}}_l \bar{\boldsymbol{v}}_l. \tag{16}$$

Observe that by making use of Eqs. (1a), (10), and (12), one obtains

$$\mathbf{K} \cdot \boldsymbol{u} = -\boldsymbol{u} \cdot \tilde{\mathbf{K}}, \quad \tilde{\mathbf{K}} \cdot \boldsymbol{u} = -\boldsymbol{u} \cdot \mathbf{K}. \tag{17}$$

Also note that, from Eqs. (11) and (14), we have

$$(\tilde{\mathbf{K}} \cdot \mathbf{M})_s = (\mathbf{K} \cdot \tilde{\mathbf{M}})_s = \mathbf{K} : \mathbf{M}. \tag{18}$$

In terms of its contravariant, covariant, and mixed components, a dyadic in $\mathcal{S}_2$ can be written in the following alternate ways:

$$\mathbf{K} = K^{ab}\boldsymbol{h}_a\boldsymbol{h}_b = K_{ab}\boldsymbol{h}^a\boldsymbol{h}^b = K^a{}_b\boldsymbol{h}_a\boldsymbol{h}^b = K_a{}^b\boldsymbol{h}^a\boldsymbol{h}_b, \tag{19}$$

from which we find that

$$\begin{aligned} K^{ab} &= \boldsymbol{h}^a\boldsymbol{h}^b : \mathbf{K} = -\boldsymbol{h}^a \cdot \mathbf{K} \cdot \boldsymbol{h}^b, \\ K_{ab} &= \mathbf{K} : \boldsymbol{h}_a\boldsymbol{h}_b = -\boldsymbol{h}_a \cdot \mathbf{K} \cdot \boldsymbol{h}_b, \\ K^a{}_b &= \boldsymbol{h}^a \cdot \mathbf{K} \cdot \boldsymbol{h}_b, \\ K_a{}^b &= \boldsymbol{h}_a \cdot \mathbf{K} \cdot \boldsymbol{h}^b. \end{aligned} \tag{20}$$

The previous definitions enable us to express the product $(\mathbf{K} \cdot \boldsymbol{u})$ as

$$\begin{aligned} (\mathbf{K} \cdot \boldsymbol{u})^a &= \boldsymbol{h}^a \cdot \mathbf{K} \cdot \boldsymbol{h}_b u^b = K^a{}_b u^b = -K^{ab}u_b, \\ (\mathbf{K} \cdot \boldsymbol{u})_a &= -\boldsymbol{h}_a \cdot \mathbf{K} \cdot \boldsymbol{h}^b u_b = -K_a{}^b u_b = K_{ab}u^b. \end{aligned} \tag{21}$$

The identity, or unit dyadic $\mathbf{I}_2$, which is defined to satisfy the equation

$$\mathbf{I}_2 \cdot \boldsymbol{u} = \boldsymbol{u} \cdot \mathbf{I}_2 = \boldsymbol{u}, \tag{22}$$

for all $\boldsymbol{u} \in \mathcal{S}_2$, can be written as

$$\mathbf{I}_2 = \boldsymbol{h}_a\boldsymbol{h}^a = -\boldsymbol{h}^a\boldsymbol{h}_a. \tag{23}$$

Note that the unit dyadic is antisymmetric, i.e.,

$$\tilde{\mathbf{I}}_2 = -\mathbf{I}_2. \tag{24}$$

The metric in $\mathcal{S}_2$ is defined in terms of the components of $\mathbf{I}_2$. Thus, by Eq. (20), we have

$$\begin{aligned} \omega_{ab} &\equiv (\mathbf{I}_2)_{ab} = \mathbf{I}_2 : \boldsymbol{h}_a\boldsymbol{h}_b = -\boldsymbol{h}_a \cdot \mathbf{I}_2 \cdot \boldsymbol{h}_b = -\boldsymbol{h}_a \cdot \boldsymbol{h}_b, \\ \omega^{ab} &\equiv (\mathbf{I}_2)^{ab} = \boldsymbol{h}^a\boldsymbol{h}^b : \mathbf{I}_2 = -\boldsymbol{h}^a \cdot \mathbf{I}_2 \cdot \boldsymbol{h}^b = -\boldsymbol{h}^a \cdot \boldsymbol{h}^b, \end{aligned} \tag{25}$$

and using Eqs. (3) and (4) results in

$$\begin{aligned} &\omega_{12} = \omega^{12} = -1, \quad \omega_{21} = \omega^{21} = 1, \\ &\omega_{11} = \omega_{22} = \omega^{11} = \omega^{22} = 0. \end{aligned} \tag{26}$$

We can now derive the rules for raising and lowering indices:

$$\begin{aligned} u^a &= \boldsymbol{h}^a \cdot \boldsymbol{u} = \boldsymbol{h}^a \cdot \boldsymbol{h}^b u_b = -u_b \boldsymbol{h}^b \cdot \boldsymbol{h}^a = u_b \omega^{ba}, \\ u_a &= \boldsymbol{u} \cdot \boldsymbol{h}_a = u^b \boldsymbol{h}_b \cdot \boldsymbol{h}_a = -\boldsymbol{h}_a \cdot \boldsymbol{h}_b u^b = \omega_{ab}u^b. \end{aligned} \tag{27}$$

In the next section we will make use of the determinant $|\mathbf{K}|$ of a dyadic $\mathbf{K}$ in $\mathcal{S}_2$, which can be defined according to

$$(\mathbf{K} \cdot \boldsymbol{u}) \wedge (\mathbf{K} \cdot \boldsymbol{v}) = |\mathbf{K}|\, \boldsymbol{u} \wedge \boldsymbol{v}. \tag{28}$$

That this is indeed equivalent to the usual definition in terms of components can be seen from the following considerations.

First take the scalar of Eq. (28) and note that $(\boldsymbol{u} \wedge \boldsymbol{v})_s = 2\boldsymbol{u} \cdot \boldsymbol{v}$. Hence,

$$(\mathbf{K} \cdot \boldsymbol{u}) \cdot (\mathbf{K} \cdot \boldsymbol{v}) = |\mathbf{K}|\, \boldsymbol{u} \cdot \boldsymbol{v} \tag{29a}$$

or

$$-(\boldsymbol{u} \cdot \tilde{\mathbf{K}}) \cdot (\mathbf{K} \cdot \boldsymbol{v}) = |\mathbf{K}|\, \boldsymbol{u} \cdot \mathbf{I}_2 \cdot \boldsymbol{v}. \tag{29b}$$

Since $\boldsymbol{u}$ and $\boldsymbol{v}$ are arbitrary, we must have

$$\tilde{\mathbf{K}} \cdot \mathbf{K} = -|\mathbf{K}|\, \mathbf{I}_2. \tag{30}$$

The fact that $\tilde{\mathbf{K}} \cdot \mathbf{K}$ must be proportional to the identity $\mathbf{I}_2$ can be verified by taking the transpose of $\tilde{\mathbf{K}} \cdot \mathbf{K}$:

$$(\tilde{\mathbf{K}} \cdot \mathbf{K})_{\mathrm{T}} = -\tilde{\mathbf{K}} \cdot \tilde{\tilde{\mathbf{K}}} = -\tilde{\mathbf{K}} \cdot \mathbf{K},$$

and noting that there is only one independent antisymmetric dyadic. Taking the scalar of Eq. (30) yields

$$(\tilde{\mathbf{K}} \cdot \mathbf{K})_s = \mathbf{K} : \mathbf{K} = 2\,|\mathbf{K}|, \tag{31}$$

where we have made use of Eq. (18), and

$$(\mathbf{I}_2)_s = -2.$$

In terms of components, Eq. (31) becomes

$$|\mathbf{K}| = -\tfrac{1}{2}K^a{}_b K_a{}^b = (\tfrac{1}{2})K^a{}_b K^c{}_d \omega_{ac}\omega^{bd}. \tag{32}$$

Q.E.D.

The gradient operation $\boldsymbol{D}$ in $\mathcal{S}_2$ can be defined by the infinitesimal relation

$$dF = -d\boldsymbol{u} \cdot \boldsymbol{D}F = -du^a \boldsymbol{h}_a \cdot \boldsymbol{D}F, \tag{33}$$

where $F(\boldsymbol{u})$ is an arbitrary function of $\boldsymbol{u}$. Regarding F

as a function of the components u^a, we can alternately write

$$dF = du^a \partial_a F = -d\boldsymbol{u} \cdot \boldsymbol{h}^a \partial_a F, \tag{34}$$

where

$$\partial_a \equiv \partial/\partial u^a.$$

Since $d\boldsymbol{u}$ and F are arbitrary, we find by comparison of Eqs. (33) and (34) that

$$\boldsymbol{D} = \boldsymbol{h}^a \partial_a. \tag{35}$$

This expression leads immediately to

$$\partial_a = -\boldsymbol{h}_a \cdot \boldsymbol{D} = \boldsymbol{D} \cdot \boldsymbol{h}_a. \tag{36}$$

3. TRANSFORMATIONS OF SPINORS

In this section we shall consider those linear transformations $\mathbf{S}(\boldsymbol{u})$ of $\mathcal{S}_2$ into itself which preserve the inner product. Given any such $\mathbf{S}(\boldsymbol{u})$, let $\mathbf{S}$ be the dyadic which realizes the transformation, i.e.,

$$\mathbf{S} \cdot \boldsymbol{u} = \mathbf{S}(\boldsymbol{u}).$$

The preservation of the inner product can then be expressed as

$$(\mathbf{S} \cdot \boldsymbol{u}) \cdot (\mathbf{S} \cdot \boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}, \tag{37}$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{S}_2$. By comparison of Eq. (37) with (29a) we can immediately infer that $|\mathbf{S}| = 1$, i.e., $\mathbf{S}$ is unimodular. Thus from Eq. (30) we have

$$\tilde{\mathbf{S}} \cdot \mathbf{S} = -\mathbf{I}_2. \tag{38}$$

We will now show that $\mathbf{S}$ can be stated in the exponential form

$$\mathbf{S} = \tau e^{\mathbf{M}},$$

where τ has one of the values ± 1, and $\mathbf{M}$ is a symmetric ($\tilde{\mathbf{M}} = \mathbf{M}$) dyadic. To this end, first note that since any dyadic can be expressed as a sum of an antisymmetric and a symmetric dyadic, we can write

$$\mathbf{S} = \alpha \mathbf{I}_2 + \mathbf{T}, \tag{39}$$

where $\tilde{\mathbf{T}} = \mathbf{T}$. From Eq. (30) we have

$$\mathbf{T} \cdot \mathbf{T} = \tilde{\mathbf{T}} \cdot \mathbf{T} = -|\mathbf{T}|\, \mathbf{I}_2, \tag{40}$$

and substituting Eq. (39) into (38) together with (40) yields

$$\alpha^2 + |\mathbf{T}| = 1. \tag{41}$$

This last result suggests the identification

$$\alpha^2 = \cos^2 m, \tag{42a}$$

$$|\mathbf{T}| = \sin^2 m, \tag{42b}$$

where m is generally a complex number. A value of m satisfying Eq. (42) is

$$m = \begin{cases} \sin^{-1}(|\mathbf{T}|^{\frac{1}{2}}), & \text{if } |\mathbf{T}| \neq 0, \\ 0, & \text{if } |\mathbf{T}| = 0. \end{cases}$$

Note that in the case $|\mathbf{T}| \neq 0$ any one of the multiple values of m is permissible. Equation (42a) implies that either $\cos m = \alpha$, or $-\cos m = \alpha$. Accordingly we can choose τ to satisfy

$$\tau \cos m = \alpha. \tag{43}$$

Defining

$$\mathbf{M} = \frac{\tau m \mathbf{T}}{(|\mathbf{T}|)^{\frac{1}{2}}} = \frac{\tau m \mathbf{T}}{\sin m} \tag{44}$$

and substituting into Eq. (39) together with (43) gives

$$\mathbf{S} = \tau[(\cos m)\mathbf{I}_2 + (\sin m) m^{-1}\mathbf{M}]. \tag{45}$$

In the special case where $m = 0$, Eq. (45) is to be understood in the limiting sense as $m \to 0$, i.e., $(\sin m)m^{-1} \to 1$. Finally, by observing that

$$(m^{-1}\mathbf{M}) \cdot (m^{-1}\mathbf{M}) = |\mathbf{T}|^{-1}\, \mathbf{T} \cdot \mathbf{T} = -\mathbf{I}_2, \tag{46}$$

i.e., $m^{-1}\mathbf{M}$ behaves like the imaginary unit i, we can immediately write Eq. (45) as

$$\mathbf{S} = \tau \exp[m(m^{-1}\mathbf{M})] = \tau e^{\mathbf{M}}, \tag{47}$$

which is our desired result.

In the ensuing sections, the exponential form of $\mathbf{S}$, together with some of our results given in I, will be particularly useful for attaining a different and more direct approach than others appearing in the literature, for establishing the homomorphism of the homogeneous restricted Lorentz group $\mathcal{L}_p$ to the group $\mathcal{C}_2$ of two-dimensional unimodular transformations.

4. CONNECTION BETWEEN FOUR-VECTORS IN MINKOWSKI SPACE AND SPINOR DYADICS

Consider the tensor product space $\bar{\mathcal{S}}_2 \otimes \mathcal{S}_2$ generated from dyads of the form $\bar{\boldsymbol{u}} \otimes \boldsymbol{u}$ with real multipliers. Equivalently, $\bar{\mathcal{S}}_2 \otimes \mathcal{S}_2$ is the space of all Hermitian dyadics, i.e., dyadics satisfying

$$\mathbf{K}^\dagger \equiv \overline{\mathbf{K}}_{\mathrm{T}} = \mathbf{K}. \tag{48}$$

The double product of any two dyadics in this space is real, as can be seen from

$$\overline{\mathbf{K} : \mathbf{M}} = \overline{\mathbf{K}_{\mathrm{T}} : \mathbf{M}_{\mathrm{T}}} = \mathbf{K}^\dagger : \mathbf{M}^\dagger = \mathbf{K} : \mathbf{M}. \tag{49}$$

Furthermore, with a scalar product $\mathbf{K} \odot \mathbf{M}$ in $\bar{\mathcal{S}}_2 \otimes \mathcal{S}_2$ defined by

$$\mathbf{K} \odot \mathbf{M} = -\mathbf{K} : \mathbf{M}, \tag{50}$$

the space becomes isomorphic to $\mathcal{M}_4$ (Minkowski space). To show this, it is sufficient to find a set of dyadics $\mathbf{E}_\mu$ having the scalar products

$$\mathbf{E}_\mu \odot \mathbf{E}_\nu = g_{\mu\nu}, \tag{51}$$

where $g_{\mu\nu}$ is the Minkowski metric

$$g_{11} = g_{22} = g_{33} = -g_{00} = 1,$$
$$g_{\mu\nu} = 0 \quad (\mu \neq \nu).$$

It is readily verified that the following set satisfies the above conditions:

$$\begin{aligned}
\mathbf{E}^0 &= -\mathbf{E}_0 = -(2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}_1 \otimes \boldsymbol{h}_1 + \bar{\boldsymbol{h}}_2 \otimes \boldsymbol{h}_2) \\
&= -(2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}^1 \otimes \boldsymbol{h}^1 + \bar{\boldsymbol{h}}^2 \otimes \boldsymbol{h}^2), \\
\mathbf{E}^1 &= \mathbf{E}_1 = -(2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}_1 \otimes \boldsymbol{h}_2 + \bar{\boldsymbol{h}}_2 \otimes \boldsymbol{h}_1) \\
&= (2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}^1 \otimes \boldsymbol{h}^2 + \bar{\boldsymbol{h}}^2 \otimes \boldsymbol{h}^1), \\
\mathbf{E}^2 &= \mathbf{E}_2 = -i(2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}_1 \otimes \boldsymbol{h}_2 - \bar{\boldsymbol{h}}_2 \otimes \boldsymbol{h}_1) \\
&= i(2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}^2 \otimes \boldsymbol{h}^1 - \bar{\boldsymbol{h}}^1 \otimes \boldsymbol{h}^2), \\
\mathbf{E}^3 &= \mathbf{E}_3 = -(2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}_1 \otimes \boldsymbol{h}_1 - \bar{\boldsymbol{h}}_2 \otimes \boldsymbol{h}_2) \\
&= (2)^{-\frac{1}{2}}(\bar{\boldsymbol{h}}^1 \otimes \boldsymbol{h}^1 - \bar{\boldsymbol{h}}^2 \otimes \boldsymbol{h}^2).
\end{aligned} \tag{52}$$

Thus we have the isomorphism:

$$\mathbf{E}_\mu \leftrightarrow \underline{\mathbf{e}}_\mu$$
$$\mathbf{A} = A^\mu \mathbf{E}_\mu \leftrightarrow \underline{\mathbf{a}} = a^\mu \underline{\mathbf{e}}_\mu, \quad \text{with } a^\mu = A^\mu, \tag{53}$$

where $\underline{\mathbf{e}}_\mu$ is an orthonormal basis ($\underline{\mathbf{e}}_\mu \cdot \underline{\mathbf{e}}_\nu = g_{\mu\nu}$) in $\mathcal{M}_4$.

In order to relate these results with the component formalism conventionally used, note that a dyadic $\mathbf{A}$ in $\bar{\mathcal{S}}_2 \otimes \mathcal{S}_2$ can be expressed in terms of the Minkowski basis $\mathbf{E}_\mu$ or $\mathbf{E}^\mu$ as

$$\mathbf{A} = A^\mu \mathbf{E}_\mu = A_\mu \mathbf{E}^\mu$$

or, alternately, in terms of spinor bases as in Eq. (19):

$$\mathbf{A} = A^{\dot{a}b} \bar{\boldsymbol{h}}_a \boldsymbol{h}_b = A_{\dot{a}b} \bar{\boldsymbol{h}}^a \boldsymbol{h}^b = \cdots.$$

Consequently, making use of Eqs. (51) and (50), we find

$$A^\mu = \mathbf{E}^\mu \odot \mathbf{A} = \mathbf{E}^\mu \odot \bar{\boldsymbol{h}}^a \boldsymbol{h}^b A_{\dot{a}b} = -(\mathbf{E}^\mu : \bar{\boldsymbol{h}}^a \boldsymbol{h}^b) A_{\dot{a}b}. \tag{54}$$

Conversely,

$$A_{\dot{a}b} = \mathbf{A} : \bar{\boldsymbol{h}}_a \boldsymbol{h}_b = A^\mu (\mathbf{E}_\mu : \bar{\boldsymbol{h}}_a \boldsymbol{h}_b). \tag{55}$$

Now observe that the identity[9] in $\mathcal{M}_4$,

$$\mathsf{I}_4 = \mathbf{E}_\mu \mathbf{E}^\mu = \mathbf{E}^\mu \mathbf{E}_\mu, \tag{56}$$

can be expressed in terms of a hybrid of four-vector and spinor components

$$\mathsf{I}_4 = (\mathsf{I}_4)^{\mu\dot{a}b} \mathbf{E}_\mu \bar{\boldsymbol{h}}_a \boldsymbol{h}_b = (\mathsf{I}_4)_{\mu\dot{a}b} \mathbf{E}^\mu \bar{\boldsymbol{h}}^a \boldsymbol{h}^b, \tag{57}$$

where

$$(\mathsf{I}_4)^{\mu\dot{a}b} = \mathbf{E}^\mu \odot \mathsf{I}_4 : \bar{\boldsymbol{h}}^a \boldsymbol{h}^b = \mathbf{E}^\mu : \bar{\boldsymbol{h}}^a \boldsymbol{h}^b \tag{58a}$$

and

$$(\mathsf{I}_4)_{\mu\dot{a}b} = \mathbf{E}_\mu \odot \mathsf{I}_4 : \bar{\boldsymbol{h}}_a \boldsymbol{h}_b = \mathbf{E}_\mu : \bar{\boldsymbol{h}}_a \boldsymbol{h}_b. \tag{58b}$$

[9] Note that dyadics in $\mathcal{M}_4$ are tetradics in $\mathcal{S}_2$. They will be denoted, as in I, by sans serif type or boldface Greek capitals.

Hence, Eqs. (54) and (55) become

$$A^\mu = -(\mathsf{I}_4)^{\mu\dot{a}b} A_{\dot{a}b} \tag{59}$$

and

$$A_{\dot{a}b} = (\mathsf{I}_4)_{\mu\dot{a}b} A^\mu, \tag{60}$$

respectively. All other possible combinations with covariant, contravariant, and mixed indices can be obtained from Eqs. (59) or (60) by the usual rules for raising or lowering indices of four-vector and spinor components. In particular, it is interesting to note that the hybrid covariant components of I_4, which are explicitly given by

$$\begin{aligned}
(\mathsf{I}_4)_{0\dot{a}b} &= (2)^{-\frac{1}{2}}(\delta_{\dot{a}\dot{1}}\delta_{b1} + \delta_{\dot{a}\dot{2}}\delta_{b2}), \\
(\mathsf{I}_4)_{1\dot{a}b} &= (2)^{-\frac{1}{2}}(\delta_{\dot{a}\dot{1}}\delta_{b2} + \delta_{\dot{a}\dot{2}}\delta_{b1}), \\
(\mathsf{I}_4)_{2\dot{a}b} &= (2)^{-\frac{1}{2}}(-i\delta_{\dot{a}\dot{1}}\delta_{b2} + i\delta_{\dot{a}\dot{2}}\delta_{b1}), \\
(\mathsf{I}_4)_{3\dot{a}b} &= (2)^{-\frac{1}{2}}(\delta_{\dot{a}\dot{1}}\delta_{b1} - \delta_{\dot{a}\dot{2}}\delta_{b2}),
\end{aligned} \tag{61}$$

are proportional to the identity and the three Pauli matrices.

As an additional remark which is of interest here, we show how the equations of stereographic projection can be obtained from our spinor formalism in a rather straightforward way. To this end, consider any Hermitian dyadic

$$\mathbf{A} = A^{\dot{a}b} \bar{\boldsymbol{h}}_a \otimes \boldsymbol{h}_b, \quad A^{\dot{a}b} = (A^{\dot{b}a})^* \tag{62}$$

and let

$$\bar{\boldsymbol{v}} = A^{\dot{a}1} \bar{\boldsymbol{h}}_a, \quad \bar{\boldsymbol{w}} = A^{\dot{a}2} \bar{\boldsymbol{h}}_a.$$

Then,

$$\mathbf{A} = \bar{\boldsymbol{v}} \otimes \boldsymbol{h}_1 + \bar{\boldsymbol{w}} \otimes \boldsymbol{h}_2. \tag{63}$$

Since a stereographic projection maps a null vector into a spinor, we add the further requirement that $\mathbf{A} \odot \mathbf{A} = 0$. But

$$\mathbf{A} \odot \mathbf{A} = -2\overline{(\boldsymbol{v} \cdot \boldsymbol{w})}(\boldsymbol{h}_1 \cdot \boldsymbol{h}_2) = -2\overline{(\boldsymbol{v} \cdot \boldsymbol{w})}.$$

Therefore,

$$\boldsymbol{v} \cdot \boldsymbol{w} = 0,$$

which implies that $\boldsymbol{v}$ and $\boldsymbol{w}$ are proportional, i.e.,

$$\boldsymbol{w} = \alpha \boldsymbol{v}.$$

Thus,

$$\mathbf{A} = \bar{\boldsymbol{v}} \otimes \boldsymbol{h}_1 + \bar{\alpha}\bar{\boldsymbol{v}} \otimes \boldsymbol{h}_2 = \bar{\boldsymbol{v}} \otimes \boldsymbol{s},$$

with

$$\boldsymbol{s} = \boldsymbol{h}_1 + \bar{\alpha}\boldsymbol{h}_2. \tag{64}$$

Furthermore, from the Hermiticity requirement,

$$\mathbf{A}^\dagger = \bar{\boldsymbol{s}} \otimes \boldsymbol{v} = \mathbf{A} = \bar{\boldsymbol{v}} \otimes \boldsymbol{s},$$

it follows that $\boldsymbol{v}$ and $\boldsymbol{s}$ must be proportional:

$$\boldsymbol{v} = \beta \boldsymbol{s}, \quad \beta = \bar{\beta}.$$

Consequently,

$$\mathbf{A} = \beta \bar{\boldsymbol{s}} \otimes \boldsymbol{s}. \tag{65}$$

Note that by making use of the first equation in (52), we have

$$\mathbf{E}^0 \circ \mathbf{A} = (2)^{-\frac{1}{2}}\beta(\bar{s}_1 s_1 + \bar{s}_2 s_2),$$

which means that $\mathbf{A}$ is on the future (past) sheet of the null cone if $\beta > 0$ ($\beta < 0$). For our purposes, we shall restrict attention to the case $\beta > 0$. Then let $\boldsymbol{u} = (\beta)^{\frac{1}{2}}\boldsymbol{s}$, so we can write

$$\mathbf{A} = \bar{\boldsymbol{u}} \otimes \boldsymbol{u}. \tag{66}$$

In terms of components, Eq. (66) can be displayed as

$$A^{\mu} = -(\mathsf{I}_4)^{\mu\dot{a}b}A_{\dot{a}b} = -(\mathsf{I}_4)^{\mu\dot{a}b}u_{\dot{a}}u_b. \tag{67}$$

Explicitly, adopting the notation $\bar{\xi} = u_1$, $\bar{\eta} = u_2$, Eqs. (67) become

$$\sqrt{2}\,A^0 = \bar{\xi}\xi + \bar{\eta}\eta, \qquad \sqrt{2}\,A^1 = \bar{\xi}\eta + \bar{\eta}\xi$$
$$\sqrt{2}\,A^2 = -i(\bar{\xi}\eta - \bar{\eta}\xi), \quad \sqrt{2}\,A^3 = \bar{\xi}\xi - \bar{\eta}\eta. \tag{68}$$

Within a proportionality factor, the above are the equations of stereographic projection usually appearing in the literature.[10]

5. TWO-DIMENSIONAL SPINORIAL REPRESENTATION OF THE RESTRICTED HOMOGENEOUS LORENTZ GROUP

It was shown in I [Eqs. (106), (121), and (125)] that a restricted homogeneous Lorentz transformation can be written in the form

$$\begin{aligned}\mathsf{L} &= \exp(\mathbf{a} \times \mathsf{I}_3 + \underline{\mathbf{e}}_0 \wedge \mathbf{b}) \\ &= \tau \exp(-iq_k\boldsymbol{\Sigma}_k) \cdot \tau \exp(iq_k^*\boldsymbol{\Sigma}_k^*),\end{aligned} \tag{69}$$

where

$$\tau = \pm 1,$$
$$\mathbf{q} = \tfrac{1}{2}(\mathbf{a} - i\mathbf{b}), \tag{70}$$
$$\boldsymbol{\Sigma}_k = \underline{\mathbf{e}}_k \cdot \boldsymbol{\Sigma} = i\underline{\mathbf{e}}_k \times \mathsf{I}_3 - \underline{\mathbf{e}}_0 \wedge \underline{\mathbf{e}}_k, \tag{71}$$

and

$$(\mathbf{q} \cdot \boldsymbol{\Sigma}) \cdot (\mathbf{q}^* \cdot \boldsymbol{\Sigma}^*) = (\mathbf{q}^* \cdot \boldsymbol{\Sigma}^*) \cdot (\mathbf{q} \cdot \boldsymbol{\Sigma}). \tag{72}$$

Making use of these results, together with the intrinsic spinor formalism of the preceding sections, we will show how the transformations

$$\exp(-i\mathbf{q} \cdot \boldsymbol{\Sigma}) \quad \text{and} \quad \exp(i\mathbf{q}^* \cdot \boldsymbol{\Sigma}^*)$$

are directly related to the spinor operators (which were introduced in Sec. 3) $\mathbf{S} = \exp(\mathbf{M})$ in $\mathcal{S}_2$, and $\bar{\mathbf{S}} = \exp(\bar{\mathbf{M}})$ in $\bar{\mathcal{S}}_2$, respectively.

As a short digression intended to attain further algebraic simplification, we introduce the following additional operations on tetradics:

$$(\bar{v}w\bar{y}z)^{\ddagger} \equiv -\bar{v}\bar{y}wz, \tag{73a}$$
$$(\bar{v}\bar{y}wz) \diamond \bar{u}u \equiv [(\bar{v}\bar{y}) \cdot \bar{u}][(wz) \cdot u], \tag{73b}$$
$$(\bar{v}w) \barwedge (\bar{y}z) \equiv \bar{v}w\bar{y}z - \bar{y}z\bar{v}w. \tag{73c}$$

[10] See, for example, V. I. Smirnov, *Linear Algebra and Group Theory* (McGraw-Hill Book Co., Inc., New York, 1961); B. Kursunoğlu, *Modern Quantum Theory* (W. H. Freeman and Co., San Francisco, 1962).

From these definitions and Eq. (50), we have the useful result

$$(vw\bar{y}z) \circ \bar{u}u = -(\bar{v}wyz) : \bar{u}u = (vw\bar{y}z)^{\ddagger} \diamond \bar{u}u. \tag{74}$$

Returning to the main theme of this section, and following the formalism of I, we denote a Lorentz transformation of a four-vector $\underline{\mathbf{x}}$ in $\mathcal{M}_4$ by

$$\underline{\mathbf{x}}' = \mathsf{L} \cdot \underline{\mathbf{x}}.$$

Moreover, because of Eqs. (50) and (52), this may be written as

$$\mathbf{X}' = \mathsf{L} \circ \mathbf{X}, \tag{75}$$

where $\mathbf{X}$ is now a Hermitian dyadic in $\bar{\mathcal{S}}_2 \otimes \mathcal{S}_2$, and L is a tetradic operating by means of the product defined in Eq. (50).

It will be convenient to have Eq. (75) expressed in the form

$$\mathbf{X}' = \mathsf{L}^{\ddagger} \diamond \mathbf{X}, \tag{76}$$

where use has been made of Eq. (74). In our present notation, Eq. (69) becomes

$$\mathsf{L} = \tau \exp(-iq_k\boldsymbol{\Sigma}_k) \circ \tau \exp(iq_k^*\boldsymbol{\Sigma}_k^*), \tag{77}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_k &= i\mathbf{E}_k \times \mathsf{I}_3 - \mathbf{E}_0 \wedge \mathbf{E}_k = i\mathbf{E}_k \times \mathbf{E}_l\mathbf{E}_l - \mathbf{E}_0 \wedge \mathbf{E}_k, \\ \boldsymbol{\Sigma}_k^* &= -i\mathbf{E}_k \times \mathbf{E}_l\mathbf{E}_l - \mathbf{E}_0 \wedge \mathbf{E}_k.\end{aligned} \tag{78}$$

Performing the operation defined by Eq. (73a) on the above expression for L yields[11]

$$\mathsf{L}^{\ddagger} = \tau \exp(-iq_k\boldsymbol{\Sigma}_k^{\ddagger}) \diamond \tau \exp(iq_k^*\boldsymbol{\Sigma}_k^{*\ddagger}). \tag{79}$$

It is now a straightforward matter to write $\boldsymbol{\Sigma}_k^{\ddagger}$ explicitly in terms of the spinor bases by making use of Eqs. (52). Thus, for $\boldsymbol{\Sigma}_1^{\ddagger}$, we have

$$\begin{aligned}\boldsymbol{\Sigma}_1 &= -i\mathbf{E}_2 \wedge \mathbf{E}_3 - \mathbf{E}_0 \wedge \mathbf{E}_1 \\ &= \tfrac{1}{2}[(\bar{h}_1h_2 - \bar{h}_2h_1) \barwedge (\bar{h}_1h_1 - \bar{h}_2h_2) \\ &\quad + (\bar{h}_1h_1 + \bar{h}_2h_2) \barwedge (\bar{h}_1h_2 + \bar{h}_2h_1)] \\ &= -(\bar{h}_1h_2) \barwedge (\bar{h}_2h_2) - (\bar{h}_2h_1) \barwedge (\bar{h}_1h_1) \\ &= -\bar{h}_1h_2\bar{h}_2h_2 + \bar{h}_2h_2\bar{h}_1h_2 - \bar{h}_2h_1\bar{h}_1h_1 \\ &\quad + \bar{h}_1h_1\bar{h}_2h_1;\end{aligned} \tag{80a}$$

$$\begin{aligned}\boldsymbol{\Sigma}_1^{\ddagger} &= \bar{h}_1\bar{h}_2h_2h_2 - \bar{h}_2\bar{h}_1h_2h_2 \\ &\quad + \bar{h}_2\bar{h}_1h_1h_1 - \bar{h}_1\bar{h}_2h_1h_1 \\ &= (\bar{h}_1\bar{h}_2 - \bar{h}_2\bar{h}_1)(h_2h_2 - h_1h_1) \\ &= \bar{\mathsf{I}}_2(h_1h_1 - h_2h_2).\end{aligned}$$

[11] Note that the exponentials in Eqs. (77) and (79) are respectively defined by means of the two series

$$\exp(\mathsf{T}) = \mathsf{I}_4 + \mathsf{T} + (1/2!)(\mathsf{T} \circ \mathsf{T}) + (1/3!)(\mathsf{T} \circ \mathsf{T} \circ \mathsf{T}) + \cdots,$$

$$\exp(\mathsf{T}^{\ddagger}) = \mathsf{I}_4^{\ddagger} + \mathsf{T}^{\ddagger} + (1/2!)(\mathsf{T}^{\ddagger} \diamond \mathsf{T}^{\ddagger}) + (1/3!)(\mathsf{T}^{\ddagger} \diamond \mathsf{T}^{\ddagger} \diamond \mathsf{T}^{\ddagger}) + \cdots.$$

Similarly, one finds that

$$\mathbf{\Sigma}_2^{\ddagger} = \bar{\mathbf{I}}_2(-i\boldsymbol{h}_1\boldsymbol{h}_1 - i\boldsymbol{h}_2\boldsymbol{h}_2), \tag{80b}$$

$$\mathbf{\Sigma}_3^{\ddagger} = \bar{\mathbf{I}}_2(-\boldsymbol{h}_1\boldsymbol{h}_2 - \boldsymbol{h}_2\boldsymbol{h}_1), \tag{80c}$$

$$\mathbf{\Sigma}_1^{*\ddagger} = (\bar{\boldsymbol{h}}_1\bar{\boldsymbol{h}}_1 - \bar{\boldsymbol{h}}_2\bar{\boldsymbol{h}}_2)\mathbf{I}_2, \tag{81a}$$

$$\mathbf{\Sigma}_2^{*\ddagger} = (i\bar{\boldsymbol{h}}_1\bar{\boldsymbol{h}}_1 + i\bar{\boldsymbol{h}}_2\bar{\boldsymbol{h}}_2)\mathbf{I}_2, \tag{81b}$$

$$\mathbf{\Sigma}_3^{*\ddagger} = (-\bar{\boldsymbol{h}}_1\bar{\boldsymbol{h}}_2 - \bar{\boldsymbol{h}}_2\bar{\boldsymbol{h}}_1)\mathbf{I}_2. \tag{81c}$$

Making the identifications

$$\begin{aligned} \boldsymbol{\sigma}_1 &= \boldsymbol{h}_1\boldsymbol{h}_1 - \boldsymbol{h}_2\boldsymbol{h}_2, \\ \boldsymbol{\sigma}_2 &= -i\boldsymbol{h}_1\boldsymbol{h}_1 - i\boldsymbol{h}_2\boldsymbol{h}_2, \\ \boldsymbol{\sigma}_3 &= -\boldsymbol{h}_1\boldsymbol{h}_2 - \boldsymbol{h}_2\boldsymbol{h}_1, \end{aligned} \tag{82}$$

and

$$\begin{aligned} \boldsymbol{\sigma}_1^{\dagger} = \bar{\boldsymbol{\sigma}}_1 &= \bar{\boldsymbol{h}}_1\bar{\boldsymbol{h}}_1 - \bar{\boldsymbol{h}}_2\bar{\boldsymbol{h}}_2, \\ \boldsymbol{\sigma}_2^{\dagger} = \bar{\boldsymbol{\sigma}}_2 &= i\bar{\boldsymbol{h}}_1\bar{\boldsymbol{h}}_1 + i\bar{\boldsymbol{h}}_2\bar{\boldsymbol{h}}_2, \\ \boldsymbol{\sigma}_3^{\dagger} = \bar{\boldsymbol{\sigma}}_3 &= -\bar{\boldsymbol{h}}_1\bar{\boldsymbol{h}}_2 - \bar{\boldsymbol{h}}_2\bar{\boldsymbol{h}}_1, \end{aligned} \tag{83}$$

we obtain

$$\mathbf{\Sigma}_k^{\ddagger} = \bar{\mathbf{I}}_2\boldsymbol{\sigma}_k \equiv \bar{\mathbf{I}}_2 \otimes \boldsymbol{\sigma}_k, \quad \mathbf{\Sigma}_k^{*\ddagger} = \bar{\boldsymbol{\sigma}}_k\mathbf{I}_2 \equiv \bar{\boldsymbol{\sigma}}_k \otimes \mathbf{I}_2. \tag{84}$$

Note that the components

$$(\boldsymbol{\sigma}_k)^a{}_b = \boldsymbol{h}^a \cdot \boldsymbol{\sigma}_k \cdot \boldsymbol{h}_b \tag{85}$$

and

$$(\bar{\boldsymbol{\sigma}}_k)^{\dot{a}}{}_{\dot{b}} = \bar{\boldsymbol{h}}^a \cdot \bar{\boldsymbol{\sigma}}_k \cdot \bar{\boldsymbol{h}}_b = [(\boldsymbol{\sigma}_k)^a{}_b]^* \tag{86}$$

are the right-handed and left-handed Pauli spin matrices, respectively.

Substituting Eq. (84) into (79) results in

$$\begin{aligned} \mathsf{L}^{\ddagger} &= \tau \exp(-iq_k\bar{\mathbf{I}}_2 \otimes \boldsymbol{\sigma}_k) \diamond \tau \exp(iq_k^*\bar{\boldsymbol{\sigma}}_k \otimes \mathbf{I}_2) \\ &= \tau \exp(iq_k^*\bar{\boldsymbol{\sigma}}_k) \otimes \tau \exp(-iq_k\boldsymbol{\sigma}_k) = \bar{\mathbf{S}} \otimes \mathbf{S}, \end{aligned} \tag{87}$$

where

$$\mathbf{S} = \tau \exp(-iq_k\boldsymbol{\sigma}_k). \tag{88}$$

Furthermore, since any four-vector can be written as a sum of two null vectors, each of which in turn can be put in the form of Eq. (66), we can express any four-vector $\mathbf{X}$ as

$$\mathbf{X} = \alpha\bar{\boldsymbol{u}} \otimes \boldsymbol{u} + \beta\bar{\boldsymbol{v}} \otimes \boldsymbol{v}. \tag{89}$$

Hence, the Lorentz transformation of $\mathbf{X}$, given by Eq. (76), becomes

$$\begin{aligned} \mathsf{L}^{\ddagger} \diamond \mathbf{X} &= (\bar{\mathbf{S}} \otimes \mathbf{S}) \diamond (\alpha\bar{\boldsymbol{u}} \otimes \boldsymbol{u} + \beta\bar{\boldsymbol{v}} \otimes \boldsymbol{v}) \\ &= \alpha(\bar{\mathbf{S}} \cdot \bar{\boldsymbol{u}}) \otimes (\mathbf{S} \cdot \boldsymbol{u}) + \beta(\bar{\mathbf{S}} \cdot \bar{\boldsymbol{v}}) \otimes (\mathbf{S} \cdot \boldsymbol{v}). \end{aligned} \tag{90}$$

From Eq. (90), we see that to an arbitrary restricted homogeneous Lorentz transformation L, there corresponds a spinor transformation $\mathbf{S}$, operating in $\mathcal{S}_2$, which is defined up to a sign by Eq. (88), i.e.,

$$\mathsf{L} \leftrightarrow \tau \exp(-iq_k\mathbf{\Sigma}_k) \leftrightarrow \mathbf{S} = \tau \exp(-iq_k\boldsymbol{\sigma}_k). \tag{91}$$

Referring to Eqs. (82), we see that $\tilde{\boldsymbol{\sigma}}_k = \boldsymbol{\sigma}_k$, and that any symmetric dyadic attached to $\mathcal{S}_2$ may be written as

$$\mathbf{M} = -iq_k\boldsymbol{\sigma}_k. \tag{92}$$

Hence all elements of the set $\mathcal{C}_2$ of unimodular transformations described in Sec. 3 can be generated from Eq. (88) by varying the parameters q_k, which were defined in terms of the Lorentz transformation parameters by Eq. (70).

In addition, observe that if

$$\mathsf{L}_1 \leftrightarrow \mathbf{S}_1, \quad \mathsf{L}_2 \leftrightarrow \mathbf{S}_2,$$

then

$$\begin{aligned} \mathsf{L}_1^{\ddagger} \diamond \mathsf{L}_2^{\ddagger} &= (\bar{\mathbf{S}}_1 \otimes \mathbf{S}_1) \diamond (\bar{\mathbf{S}}_2 \otimes \mathbf{S}_2) = (\bar{\mathbf{S}}_1 \cdot \bar{\mathbf{S}}_2) \otimes (\mathbf{S}_1 \cdot \mathbf{S}_2) \\ &= \overline{(\mathbf{S}_1 \cdot \mathbf{S}_2)} \otimes (\mathbf{S}_1 \cdot \mathbf{S}_2), \end{aligned}$$

i.e.,

$$\mathsf{L}_1 \odot \mathsf{L}_2 \leftrightarrow \mathbf{S}_1 \cdot \mathbf{S}_2. \tag{93}$$

Thus, finally, we can conclude that Eq. (91) establishes a double-valued homomorphism of the restricted homogeneous Lorentz group onto the group $\mathcal{C}_2$ of two-dimensional unimodular transformations.

In closing this section, we emphasize the over-all notational simplicity and calculational advantages of our basic procedure, which was made possible by the combined use of the intrinsic tensor formalism of I together with the intrinsic spinor formalism of the preceding sections of this paper. These advantages become even more apparent when comparing our results with the usual component formalism appearing in the literature.[2,3] In particular, we have arrived at the familiar homomorphism realized by Eq. (91) in a different and direct manner involving essentially two simple steps: The application on L of the transposition operation defined in Eq. (73a), followed by a mere change of basis.

6. SPINORIAL FORM OF THE DIRAC EQUATION

In Sec. 2, we introduced the spaces $\mathcal{S}_2$ and $\bar{\mathcal{S}}_2$, in each of which an antisymmetric scalar product was defined. Up to now, however, there was no need to define a scalar product between an element in $\mathcal{S}_2$ and an element in $\bar{\mathcal{S}}_2$. In order to discuss the Dirac equation, we shall define such a product by

$$\boldsymbol{u} \cdot \bar{\boldsymbol{v}} = 0, \tag{94}$$

for all $\boldsymbol{u} \in \mathcal{S}_2$ and $\boldsymbol{v} \in \bar{\mathcal{S}}_2$.

Thus, in the direct-sum space $\mathcal{S}_2 + \bar{\mathcal{S}}_2$, the product of any two elements $(\boldsymbol{u} + \bar{\boldsymbol{v}})$ and $(\boldsymbol{w} + \bar{\boldsymbol{y}})$ is

$$(\boldsymbol{u} + \bar{\boldsymbol{v}}) \cdot (\boldsymbol{w} + \bar{\boldsymbol{y}}) = \boldsymbol{u} \cdot \boldsymbol{w} + \bar{\boldsymbol{v}} \cdot \bar{\boldsymbol{y}}. \tag{95}$$

Making use of the above definitions and our intrinsic spinor formalism, we can now easily show that the Dirac equation in its customary form is equivalent to

the set of linear equations:

$$\begin{aligned} \Box \cdot \varphi + k\bar{\chi} &= 0, \\ \tilde{\Box} \cdot \bar{\chi} + k\varphi &= 0, \end{aligned} \tag{96a}$$

where $\varphi(\mathbf{X}) \in \mathcal{S}_2$ and $\bar{\chi}(\mathbf{X}) \in \bar{\mathcal{S}}_2$ are functions of the coordinate four-vector $\mathbf{X} \in \bar{\mathcal{S}}_2 \otimes \mathcal{S}_2 (\leftrightarrow \mathcal{M}_4)$, and the four-gradient operation $\Box$ is defined by the infinitesimal equation

$$df(\mathbf{X}) = d\mathbf{X} \odot \Box f(\mathbf{X})$$

as was done in I. To start with, note that in view of the identity,

$$\Box = \mathbf{E}^\mu \mathbf{E}_\mu \odot \Box = \mathbf{E}^\mu \partial_\mu, \quad \partial_\mu \equiv \partial/\partial X^\mu,$$

Eqs. (96a) become

$$\begin{aligned} \mathbf{E}^\mu \cdot \partial_\mu \varphi + k\bar{\chi} &= 0, \\ \tilde{\mathbf{E}}^\mu \cdot \partial_\mu \bar{\chi} + k\phi &= 0. \end{aligned} \tag{96b}$$

Moreover, because of Eqs. (94) and (95), Eqs. (96a) can be combined to give

$$(\Box + \tilde{\Box}) \cdot \psi + k\psi = 0, \tag{97}$$

where $\psi = \varphi + \bar{\chi}$. Similarly, Eqs. (96b) yield

$$(\mathbf{E}^\mu + \tilde{\mathbf{E}}^\mu) \cdot \partial_\mu \psi + k\psi = 0. \tag{98}$$

If we now make the identifications

$$k = mc/((2)^{\frac{1}{2}} \hbar), \tag{99}$$

$$\mathbf{\Gamma}^\mu = i(2)^{\frac{1}{2}} (\mathbf{E}^\mu + \tilde{\mathbf{E}}^\mu), \tag{100}$$

Eq. (98) can be written as

$$\mathbf{\Gamma}^\mu \cdot \partial_\mu \psi + i(\hbar)^{-1} mc\psi = 0. \tag{101}$$

It only remains to show that the four quantities $\mathbf{\Gamma}^\mu$ satisfy the anticommutation relations

$$\{\mathbf{\Gamma}^\mu, \mathbf{\Gamma}^\nu\} \equiv \mathbf{\Gamma}^\mu \cdot \mathbf{\Gamma}^\nu + \mathbf{\Gamma}^\nu \cdot \mathbf{\Gamma}^\mu = -2g^{\mu\nu}(\mathbf{I}_2 + \bar{\mathbf{I}}_2). \tag{102}$$

A simple proof, which makes use of Theorems 1 and 2 of the Appendix, together with Eqs. (94) and (100), follows:

$$\begin{aligned} &\mathbf{\Gamma}^\mu \cdot \mathbf{\Gamma}^\nu + \mathbf{\Gamma}^\nu \cdot \mathbf{\Gamma}^\mu \\ &= -2[(\mathbf{E}^\mu + \tilde{\mathbf{E}}^\mu) \cdot (\mathbf{E}^\nu + \tilde{\mathbf{E}}^\nu) + (\mathbf{E}^\nu + \tilde{\mathbf{E}}^\nu) \cdot (\mathbf{E}^\mu + \tilde{\mathbf{E}}^\mu)] \\ &= -2(\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu + \mathbf{E}^\mu \cdot \tilde{\mathbf{E}}^\nu + \tilde{\mathbf{E}}^\nu \cdot \mathbf{E}^\mu + {}^\nu \cdot \tilde{\mathbf{E}}^\mu) \\ &= -2[\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu + \tilde{\mathbf{E}}^\nu \cdot \mathbf{E}^\mu + \overline{(\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu + \tilde{\mathbf{E}}^\nu \cdot \mathbf{E}^\mu)}]. \end{aligned} \tag{103}$$

Consider separately the following cases:

Case 1: $\mu \neq \nu$.

$$(\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu)_s = \mathbf{E}^\mu : \mathbf{E}^\nu = -\mathbf{E}^\mu \odot \mathbf{E}^\nu = 0.$$

Therefore, by Theorem 1 of the Appendix,

$$\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu = (\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu)_T = -\tilde{\mathbf{E}}^\nu \cdot \mathbf{E}^\mu;$$

i.e.,

$$\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu + \tilde{\mathbf{E}}^\nu \cdot \mathbf{E}^\mu = 0.$$

Consequently, Eq. (103) becomes

$$\{\mathbf{\Gamma}^\mu, \mathbf{\Gamma}^\nu\} = 0 \quad \text{for} \quad \mu \neq \nu. \tag{104}$$

Case 2: $\mu = \nu$.

$$(\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\mu)_T = -\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\mu.$$

By Theorem 2 of the Appendix,

$$\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\mu = -\tfrac{1}{2}(\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\mu)_s \mathbf{I}_2.$$

But

$$(\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\mu)_s = \mathbf{E}^\mu : \mathbf{E}^\mu = -\mathbf{E}^\mu \odot \mathbf{E}^\mu = -g^{\mu\mu}.$$

Hence,

$$\tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\mu = \tfrac{1}{2} g^{\mu\mu} \mathbf{I}_2,$$

and substituting into Eq. (103), with $\mu = \nu$, yields

$$\{\mathbf{\Gamma}^\mu, \mathbf{\Gamma}^\mu\} = -2g^{\mu\mu}(\mathbf{I}_2 + \bar{\mathbf{I}}_2). \tag{105}$$

Finally, combining Eqs. (104) and (105) gives the anticommutation relations of Eq. (102). Q.E.D.

Thus Eq. (101) is indeed the customary form of the Dirac equation, and the set of Eqs. (96a) is its equivalent spinorial form.

In order to relate our work with the conventional matrix formalism[12] of the Dirac γ^μ, the four quantities $\mathbf{\Gamma}^\mu$, which were defined intrinsically by Eq. (100), must be expressed with respect to a particular basis. A simple choice is

$$\boldsymbol{l}_1 = \boldsymbol{h}_1, \quad \boldsymbol{l}_2 = \boldsymbol{h}_2, \quad \boldsymbol{l}_3 = \bar{\boldsymbol{h}}^1, \quad \boldsymbol{l}_4 = \bar{\boldsymbol{h}}^2. \tag{106a}$$

The corresponding reciprocal basis, defined to satisfy

$$\boldsymbol{l}^\alpha \cdot \boldsymbol{l}_\beta = \delta^\alpha_\beta \quad (\alpha, \beta = 1, 2, 3, 4) \tag{107}$$

is given by

$$\boldsymbol{l}^1 = \boldsymbol{h}^1, \quad \boldsymbol{l}^2 = \boldsymbol{h}^2, \quad \boldsymbol{l}^3 = -\bar{\boldsymbol{h}}_1, \quad \boldsymbol{l}^4 = -\bar{\boldsymbol{h}}_2. \tag{106b}$$

In terms of Eqs. (106a) and (106b), we write

$$\mathbf{\Gamma}^\mu = (\mathbf{\Gamma}^\mu)^\alpha{}_\beta \boldsymbol{l}_\alpha \boldsymbol{l}^\beta, \tag{108}$$

and from Eqs. (100) and (52) one readily finds that:

$$\begin{aligned} &(\mathbf{\Gamma}^0)^1{}_3 = (\mathbf{\Gamma}^0)^2{}_4 = -(\mathbf{\Gamma}^0)^3{}_1 = -(\mathbf{\Gamma}^0)^4{}_2 = i, \\ &(\mathbf{\Gamma}^1)^1{}_4 = (\mathbf{\Gamma}^1)^2{}_3 = (\mathbf{\Gamma}^1)^3{}_2 = (\mathbf{\Gamma}^1)^4{}_1 = i, \\ &(\mathbf{\Gamma}^2)^1{}_4 = -(\mathbf{\Gamma}^2)^2{}_3 = (\mathbf{\Gamma}^2)^3{}_2 = -(\mathbf{\Gamma}^2)^4{}_1 = 1, \\ &(\mathbf{\Gamma}^3)^1{}_3 = -(\mathbf{\Gamma}^3)^2{}_4 = (\mathbf{\Gamma}^3)^3{}_1 = -(\mathbf{\Gamma}^3)^4{}_2 = i, \end{aligned} \tag{109}$$

and all other components are zero.

Similarly, it is a simple matter to show that the most frequently used representation of the $\mathbf{\Gamma}^\mu$ (known

[12] R. H. Good, Rev. Mod. Phys. **27**, 187 (1955); S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson and Co., Evanston, Ill., 1961).

as the Dirac representation) is obtained from the following basis:

$$\begin{aligned} \boldsymbol{m}_1 &= (2)^{-\frac{1}{2}}(\boldsymbol{h}_1 - i\bar{\boldsymbol{h}}^1), \\ \boldsymbol{m}_2 &= (2)^{-\frac{1}{2}}(\boldsymbol{h}_2 - i\bar{\boldsymbol{h}}^2), \\ \boldsymbol{m}_3 &= -(2)^{-\frac{1}{2}}(\boldsymbol{h}_1 + i\bar{\boldsymbol{h}}^1), \\ \boldsymbol{m}_4 &= -(2)^{-\frac{1}{2}}(\boldsymbol{h}_2 + i\bar{\boldsymbol{h}}^2). \end{aligned} \tag{110}$$

We next give a concise derivation of the continuity equation for the Dirac field expressed in intrinsic spinor form. We begin by multiplying Eq. (97) on the left by $\bar{\psi} = \bar{\varphi} + \chi$ to obtain

$$\bar{\psi} \cdot (\square + \tilde{\square}) \cdot \psi + k\bar{\psi} \cdot \psi = 0. \tag{111}$$

Observing that $\overline{\square} = \tilde{\square}$ (because $\mathbf{E}_\mu$ is Hermitian), and taking the complex conjugate of Eq. (111), results in

$$\psi \cdot (\square + \tilde{\square}) \cdot \bar{\psi} + k\psi \cdot \bar{\psi} = 0. \tag{112}$$

Adding the above two equations gives

$$-(\square + \tilde{\square}) : (\bar{\psi}\psi) = 0$$

or

$$\square \odot (\bar{\psi}\psi + \psi\bar{\psi}) = 0. \tag{113}$$

Furthermore, recalling Eq. (94), one sees that Eq. (113) is equivalent to

$$\square \odot \mathbf{J} = 0, \tag{114}$$

where

$$\mathbf{J} = \bar{\varphi}\varphi + \bar{\chi}\chi \tag{115}$$

is proportional to the current density.

In terms of components, Eq. (115) yields

$$\begin{aligned} J^\mu = \mathbf{E}^\mu \odot \mathbf{J} &= \mathbf{E}^\mu \odot (\bar{\varphi}\varphi + \bar{\chi}\chi) = \mathbf{E}^\mu \odot (\bar{\psi}\psi + \psi\bar{\psi}) \\ &= (\bar{\psi} \cdot \mathbf{E}^\mu \cdot \psi + \psi \cdot \mathbf{E}^\mu \cdot \bar{\psi}) \\ &= \bar{\psi} \cdot (\mathbf{E}^\mu + \tilde{\mathbf{E}}^\mu) \cdot \psi \\ &= -i(2)^{-\frac{1}{2}}(\bar{\psi} \cdot \boldsymbol{\Gamma}^\mu \cdot \psi). \end{aligned} \tag{116}$$

As a final remark, which serves to relate $\bar{\psi}$ to the conventional adjoint Dirac spinor, note that, for any basis $\boldsymbol{n}_\alpha$,

$$(\bar{\psi})_\beta = \bar{\psi} \cdot \boldsymbol{n}_\beta = \overline{-(\psi \cdot \boldsymbol{n}^\alpha \boldsymbol{n}_\alpha)} \cdot \boldsymbol{n}_\beta = (\psi^\alpha)^* \bar{\boldsymbol{n}}_\alpha \cdot \boldsymbol{n}_\beta. \tag{117}$$

Furthermore, if the basis is chosen to satisfy

$$\boldsymbol{n}_\beta \bar{\boldsymbol{n}}_\beta = i\boldsymbol{\Gamma}^0, \tag{118}$$

then

$$\begin{aligned} \bar{\boldsymbol{n}}_\alpha \cdot \boldsymbol{n}_\beta &= \delta^\lambda_\alpha \bar{\boldsymbol{n}}_\lambda \cdot \boldsymbol{n}_\beta = \boldsymbol{n}^\alpha \cdot \boldsymbol{n}_\lambda \bar{\boldsymbol{n}}_\lambda \cdot \boldsymbol{n}_\beta \\ &= \boldsymbol{n}^\alpha \cdot (i\boldsymbol{\Gamma}^0) \cdot \boldsymbol{n}_\beta = (i\boldsymbol{\Gamma}^0)^\alpha{}_\beta. \end{aligned} \tag{119}$$

In this case we can write

$$(\bar{\psi})_\beta = (\psi^\alpha)^* (i\boldsymbol{\Gamma}^0)^\alpha{}_\beta, \tag{120}$$

which is the conventional matrix definition of the adjoint Dirac spinor. Examples of bases fulfilling the above conditions are the ones given in Eqs. (106a) and (110).

7. FOUR-DIMENSIONAL SPIN REPRESENTATION OF THE RESTRICTED HOMOGENEOUS LORENTZ GROUP

In Sec. 5 we have shown how the intrinsic spinor formalism can be used to obtain a double-valued two-dimensional spin representation of the restricted homogeneous Lorentz group. We will now show how these results can be extended to the direct-sum space $\mathcal{S}_2 + \bar{\mathcal{S}}_2$ in order to obtain a four-dimensional homomorphism. The calculational advantages and aesthetic appeal of the method become even more evident here, and the expressions for the transformations in terms of the Dirac $\boldsymbol{\Gamma}^\mu$ emerge in a most natural and direct way.

To begin with, for every restricted homogeneous Lorentz transformation, we define a corresponding spinor transformation on $\psi = \varphi + \bar{\chi}$ by

$$\psi' = \mathbf{S} \cdot \varphi + \bar{\mathbf{S}} \cdot \bar{\chi}, \tag{121}$$

where $\mathbf{S}$ is given by Eq. (88). Moreover, since $\mathbf{S} \cdot \bar{\chi} = 0$ and $\bar{\mathbf{S}} \cdot \varphi = 0$, Eq. (121) becomes

$$\psi' = \boldsymbol{\Lambda} \cdot \psi, \tag{122}$$

where

$$\boldsymbol{\Lambda} = \mathbf{S} + \bar{\mathbf{S}} = \tau \exp(-iq_k\tilde{\boldsymbol{\sigma}}_k) + \tau \exp(iq_k^*\tilde{\boldsymbol{\sigma}}_k). \tag{123}$$

Now, noting that

$$\tilde{\boldsymbol{\sigma}}_k \cdot \boldsymbol{\sigma}_l = \boldsymbol{\sigma}_l \cdot \tilde{\boldsymbol{\sigma}}_k = 0,$$

and recalling Eq. (70), we can write

$$\begin{aligned} \boldsymbol{\Lambda} &= \tau \exp(-iq_k\boldsymbol{\sigma}_k + iq_k^*\tilde{\boldsymbol{\sigma}}_k) \\ &= \tau \exp[-\tfrac{1}{2}ia_k(\boldsymbol{\sigma}_k - \tilde{\boldsymbol{\sigma}}_k) - \tfrac{1}{2}b_k(\boldsymbol{\sigma}_k + \tilde{\boldsymbol{\sigma}}_k)]. \end{aligned} \tag{124}$$

To express this result in terms of the Dirac $\boldsymbol{\Gamma}^\mu$, we first make use of the property

$$\mathbf{I}_2 : \mathbf{I}_2 = (\tilde{\mathbf{I}}_2 \cdot \mathbf{I}_2)_s = -(\mathbf{I}_2)_s = 2$$

to write

$$\tilde{\boldsymbol{\sigma}}_k = \tfrac{1}{2}(\tilde{\boldsymbol{\sigma}}_k \otimes \mathbf{I}_2) : \mathbf{I}_2 = \tfrac{1}{2}\boldsymbol{\Sigma}_k^{*\ddagger} : \mathbf{I}_2 = -\tfrac{1}{2}\boldsymbol{\Sigma}_k^* \diamond \mathbf{I}_2, \tag{125}$$

where Eq. (84) has been used. Substituting from Eq. (78), this last expression takes the form

$$\begin{aligned} \tilde{\boldsymbol{\sigma}}_k &= -\tfrac{1}{2}(i\epsilon_{klm}\mathbf{E}_l\mathbf{E}_m - \mathbf{E}_0\mathbf{E}_k + \mathbf{E}_k\mathbf{E}_0) \diamond \mathbf{I}_2 \\ &= \tfrac{1}{2}(i\epsilon_{klm}\mathbf{E}_l \cdot \mathbf{I}_2 \cdot \tilde{\mathbf{E}}_m - \mathbf{E}_0 \cdot \mathbf{I}_2 \cdot \tilde{\mathbf{E}}_k + \mathbf{E}_k \cdot \mathbf{I}_2 \cdot \tilde{\mathbf{E}}_0) \\ &= \tfrac{1}{2}(i\epsilon_{klm}\mathbf{E}^l \cdot \tilde{\mathbf{E}}^m + \mathbf{E}^0 \cdot \tilde{\mathbf{E}}^k - \mathbf{E}^k \cdot \tilde{\mathbf{E}}^0). \end{aligned} \tag{126}$$

Combining the above equation with its conjugate,

given by

$$\boldsymbol{\sigma}_k = \tfrac{1}{2}(-i\epsilon_{klm}\tilde{\mathbf{E}}^l \cdot \mathbf{E}^m + \tilde{\mathbf{E}}^0 \cdot \mathbf{E}^k - \tilde{\mathbf{E}}^k \cdot \mathbf{E}^0), \quad (127)$$

we now obtain

$$\begin{aligned}\boldsymbol{\sigma}_k - \bar{\boldsymbol{\sigma}}_k &= \tfrac{1}{2}[-i\epsilon_{klm}(\mathbf{E}^l \cdot \tilde{\mathbf{E}}^m + \tilde{\mathbf{E}}^l \cdot \mathbf{E}^m) \\ &\quad - (\mathbf{E}^0 \cdot \tilde{\mathbf{E}}^k - \tilde{\mathbf{E}}^0 \cdot \mathbf{E}^k) \\ &\quad + (\mathbf{E}^k \cdot \tilde{\mathbf{E}}^0 - \tilde{\mathbf{E}}^k \cdot \mathbf{E}^0)] \\ &= \tfrac{1}{2}\{\tfrac{1}{2}i\epsilon_{klm}\boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m + (\mathbf{I}_2 - \bar{\mathbf{I}}_2) \\ &\quad \cdot [(\mathbf{E}^0 \cdot \tilde{\mathbf{E}}^k + \tilde{\mathbf{E}}^0 \cdot \mathbf{E}^k) \\ &\quad - (\mathbf{E}^k \cdot \tilde{\mathbf{E}}^0 + \tilde{\mathbf{E}}^k \cdot \mathbf{E}^0)]\} \\ &= \tfrac{1}{2}[\tfrac{1}{2}i\epsilon_{klm}\boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m - \tfrac{1}{2}(\mathbf{I}_2 - \bar{\mathbf{I}}_2) \\ &\quad \cdot (\boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k - \boldsymbol{\Gamma}^k \cdot \boldsymbol{\Gamma}^0)] \\ &= \tfrac{1}{2}[\tfrac{1}{2}i\epsilon_{klm}\boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m - (\mathbf{I}_2 - \bar{\mathbf{I}}_2) \cdot \boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k]. \end{aligned} \quad (128)$$

Similarly, adding Eqs. (126) and (127) gives

$$\begin{aligned}\boldsymbol{\sigma}_k + \bar{\boldsymbol{\sigma}}_k &= \tfrac{1}{2}[i\epsilon_{klm}(\mathbf{E}^l \cdot \tilde{\mathbf{E}}^m - \tilde{\mathbf{E}}^l \cdot \mathbf{E}^m) \\ &\quad + (\mathbf{E}^0 \cdot \tilde{\mathbf{E}}^k + \tilde{\mathbf{E}}^0 \cdot \mathbf{E}^k) \\ &\quad - (\mathbf{E}^k \cdot \tilde{\mathbf{E}}^0 + \tilde{\mathbf{E}}^k \cdot \mathbf{E}^0)] \\ &= \tfrac{1}{2}[\tfrac{1}{2}i\epsilon_{klm}(\mathbf{I}_2 - \bar{\mathbf{I}}_2) \cdot \boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m - \boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k]. \end{aligned} \quad (129)$$

In order to obtain further simplification, we compute the quantity $\boldsymbol{\Gamma}^5$ defined, in analogy to the usual way, by

$$\boldsymbol{\Gamma}^5 = \boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^1 \cdot \boldsymbol{\Gamma}^2 \cdot \boldsymbol{\Gamma}^3. \quad (130)$$

To this end we make use of Eqs. (52) and

$$\boldsymbol{\Gamma}^\mu \cdot \boldsymbol{\Gamma}^\nu = -2(\mathbf{E}^\mu \cdot \tilde{\mathbf{E}}^\nu + \tilde{\mathbf{E}}^\mu \cdot \mathbf{E}^\nu), \quad (131)$$

which when substituted into Eq. (130) result in

$$\begin{aligned}\boldsymbol{\Gamma}^5 &= 4(\mathbf{E}^0 \cdot \tilde{\mathbf{E}}^1 \cdot \mathbf{E}^2 \cdot \tilde{\mathbf{E}}^3 + \tilde{\mathbf{E}}^0 \cdot \mathbf{E}^1 \cdot \tilde{\mathbf{E}}^2 \cdot \mathbf{E}^3) \\ &= i(\mathbf{I}_2 - \bar{\mathbf{I}}_2). \end{aligned} \quad (132)$$

Thus we have, with the help of Eq. (102),

$$\begin{aligned}-(\mathbf{I}_2 - \bar{\mathbf{I}}_2) \cdot \boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k &= i\boldsymbol{\Gamma}^5 \cdot \boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k \\ &= \tfrac{1}{2}i\epsilon_{klm}\boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m, \end{aligned} \quad (133)$$

and multiplying Eq. (133) by $(\mathbf{I}_2 - \bar{\mathbf{I}}_2)$ immediately yields

$$\boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k = -\tfrac{1}{2}i\epsilon_{klm}(\mathbf{I}_2 - \bar{\mathbf{I}}_2) \cdot \boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m. \quad (134)$$

Finally, inserting Eqs. (133) and (134) into Eqs. (128) and (129), respectively, we obtain

$$\boldsymbol{\sigma}_k - \bar{\boldsymbol{\sigma}}_k = \tfrac{1}{2}i\epsilon_{klm}\boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m, \quad (135)$$

$$\boldsymbol{\sigma}_k + \bar{\boldsymbol{\sigma}}_k = -\boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k. \quad (136)$$

With the use of the above results, Eq. (124) becomes

$$\boldsymbol{\Lambda} = \tau \exp(\tfrac{1}{4}a_k\epsilon_{klm}\boldsymbol{\Gamma}^l \cdot \boldsymbol{\Gamma}^m + \tfrac{1}{2}b_k\boldsymbol{\Gamma}^0 \cdot \boldsymbol{\Gamma}^k). \quad (137)$$

This last expression establishes the desired two-valued homomorphism $\mathsf{L} \leftrightarrow \boldsymbol{\Lambda}$.

8. FOUR-DIMENSIONAL SPIN REPRESENTATION OF THE IMPROPER HOMOGENEOUS LORENTZ TRANSFORMATIONS

The double-valued homomorphism derived in Sec. 5 gave us the law of transformation of spinors for proper homogeneous Lorentz transformations. Furthermore, we saw that this correspondence exhausted all possible spinor transformations satisfying the requirements:

(1) $\mathbf{S}$ is linear;

(2) $\mathbf{S}$ maps $\mathcal{S}_2$ into itself;

(3) $\tilde{\mathbf{S}} \cdot \mathbf{S} = -\mathbf{I}_2$, or equivalently, $\mathbf{S}$ preserves inner products.

Thus, in order to represent improper transformations, we must somehow relax the above conditions. The following generalizations are plausible:

(1′) $\mathbf{S}$ is allowed to be either linear or antilinear;

(2′) $\mathbf{S}$ maps $\mathcal{S}_2$ either into $\mathcal{S}_2$ or $\bar{\mathcal{S}}_2$;

(3′) $\tilde{\mathbf{S}} \cdot \mathbf{S} = -\mathbf{I}_2$ or, equivalently, the linear factor $\mathbf{A}$ of $\mathbf{S}$ (i.e., $\mathbf{A} = \mathbf{S}$ if $\mathbf{S}$ is linear, or $\mathbf{A} = \mathbf{SC}$ if $\mathbf{S}$ is antilinear, where $\mathbf{C}$ is the antiunitary[13] complex conjugation operator defined by $\mathbf{C}\boldsymbol{u} = \bar{\boldsymbol{u}}$, $\mathbf{C}\bar{\boldsymbol{u}} = \boldsymbol{u}$) preserves inner products.

In view of the above generalizations, it is clear that $\mathsf{L}^\ddagger$ cannot always be written in the form of Eq. (87), i.e., $\mathsf{L}^\ddagger = \bar{\mathbf{S}} \otimes \mathbf{S}$. The following possibilities must also be taken into consideration:

$$\mathsf{L}^\ddagger = -\bar{\mathbf{S}} \otimes \mathbf{S}, \quad (138a)$$

$$\mathsf{L}^\ddagger = \mathbf{P}_{13}\bar{\mathbf{S}} \otimes \mathbf{S}, \quad (138b)$$

$$\mathsf{L}^\ddagger = -\mathbf{P}_{13}\bar{\mathbf{S}} \otimes \mathbf{S}. \quad (138c)$$

The permutation operator $\mathbf{P}_{13}$ on a tetradic, which transposes the first and third elements according to

$$\mathbf{P}_{13}(\boldsymbol{uvwz}) = \boldsymbol{wvuz},$$

arises as a consequence of condition (2′). The minus signs occur in the case of nonorthochronous Lorentz transformations. This can be simply seen by recalling the argument preceding Eq. (66), where it was shown that a dyadic of the form $\bar{\boldsymbol{u}} \otimes \boldsymbol{u}$ must be on the future light cone. Thus, $\bar{\varphi}\varphi$ and

$$\begin{aligned}\bar{\mathbf{S}} \otimes \mathbf{S} \diamond \bar{\varphi}\varphi &= \overline{(\mathbf{S} \cdot \varphi)} \otimes (\mathbf{S} \cdot \varphi) \quad (\text{if } \mathcal{S}_2 \to \mathcal{S}_2), \\ \mathbf{P}_{13}\bar{\mathbf{S}} \otimes \mathbf{S} \diamond \bar{\varphi}\varphi &= (\mathbf{S} \cdot \varphi) \otimes \overline{(\mathbf{S} \cdot \varphi)} \\ &= \overline{(\bar{\mathbf{S}} \cdot \bar{\varphi})} \otimes (\bar{\mathbf{S}} \cdot \bar{\varphi}) \quad (\text{if } \mathcal{S}_2 \to \bar{\mathcal{S}}_2) \end{aligned}$$

are all on the future light cone. Therefore, the minus

[13] A. Messiah, *Quantum Mechanics* (North-Holland Publishing Company, Amsterdam, 1962), Vol. II.

sign is necessary to reverse the sense of time. Consequently, nonorthochronous transformations of $\mathcal{S}_2$ and $\bar{\mathcal{S}}_2$ must have opposite signs, i.e.,

$$\varphi \to \mathbf{S} \cdot \varphi,$$
$$\bar{\chi} \to -\bar{\mathbf{S}} \cdot \bar{\chi}. \tag{139}$$

Since the full homogeneous Lorentz group is obtained by adjoining the inversions to the restricted group, it only remains to consider the spinor representations of the space inversion $\mathfrak{I}$, the time inversion $\mathfrak{T}$, and the total inversion $\mathfrak{J}$.

A. Space Inversion

The dyadic in $\mathcal{M}_4$ representing the space inversion (or parity operator) is

$$\mathsf{L}(\mathfrak{I}) = -\mathbf{E}_k\mathbf{E}^k + \mathbf{E}_0\mathbf{E}^0. \tag{140}$$

Substituting into this equation the expression for $\mathbf{E}^\mu$ given in Eqs. (52), and performing some straightforward algebra, we obtain

$$\mathsf{L}(\mathfrak{I}) = -(\bar{h}_1h_2\bar{h}_2h_1 + \bar{h}_2h_1\bar{h}_1h_2 + \bar{h}_1h_1\bar{h}_1h_1 + \bar{h}_2h_2\bar{h}_2h_2)$$

and

$$\begin{aligned}\mathsf{L}(\mathfrak{I})^\ddagger &= (\bar{h}_1\bar{h}_2h_2h_1 + \bar{h}_2\bar{h}_1h_1h_2 \\ &\quad + \bar{h}_1\bar{h}_1h_1h_1 + \bar{h}_2\bar{h}_2h_2h_2) \\ &= \mathbf{P}_{13}(h_1\bar{h}_1 + h_2\bar{h}_2)(\bar{h}_1h_1 + \bar{h}_2h_2) \\ &= 2\mathbf{P}_{13}\tilde{\mathbf{E}}^0\mathbf{E}^0.\end{aligned} \tag{141}$$

Thus, $\mathsf{L}(\mathfrak{I})^\ddagger$ is of the form of Eq. (138b). Moreover, since the parity operator preserves the position–momentum commutation relation $[\mathbf{r}, \mathbf{p}] = i\hbar\mathsf{I}_3$, it must be a linear operator. Consequently, in view of condition (3′), $\mathbf{S}(\mathfrak{I})$ is defined within a sign by

$$\mathbf{S}(\mathfrak{I}) = -\sqrt{2}\,\tau\mathbf{E}^0, \tag{142}$$

with $\tau = \pm 1$.

Note that under this transformation a spinor in $\mathcal{S}_2$ is mapped into a spinor in $\bar{\mathcal{S}}_2$, i.e., $\mathcal{S}_2$ is not invariant under $\mathbf{S}(\mathfrak{I})$. Hence, there is no two-dimensional representation for parity. Going to the direct sum space $\mathcal{S}_2 + \bar{\mathcal{S}}_2$, we obtain the four-dimensional representation of parity by noting that in this case

$$\psi' = \mathbf{S}(\mathfrak{I}) \cdot \varphi + \bar{\mathbf{S}}(\mathfrak{I}) \cdot \bar{\chi} = \mathbf{\Lambda}(\mathfrak{I}) \cdot \psi, \tag{143}$$

where

$$\mathbf{\Lambda}(\mathfrak{I}) = \mathbf{S}(\mathfrak{I}) + \bar{\mathbf{S}}(\mathfrak{I}) = -\sqrt{2}\,\tau(\mathbf{E}^0 + \bar{\mathbf{E}}^0) = i\tau\mathbf{\Gamma}^0. \tag{144}$$

B. Total Inversion

A vector in $\mathcal{M}_4$ transforms under space–time inversion according to

$$\mathsf{L}(\mathfrak{J}) = -\mathsf{I}_4 = -\mathbf{E}_\mu\mathbf{E}^\mu. \tag{145}$$

Proceeding as in the previous case, we substitute the expressions in Eqs. (52) to obtain

$$\mathsf{L}(\mathfrak{J}) = -\bar{h}_1h_2\bar{h}_2h_1 - \bar{h}_2h_1\bar{h}_1h_2 + \bar{h}_1h_1\bar{h}_2h_2 + \bar{h}_2h_2\bar{h}_1h_1$$

and

$$\mathsf{L}(\mathfrak{J})^\ddagger = -(\bar{h}_2\bar{h}_1 - \bar{h}_1\bar{h}_2)(h_2h_1 - h_1h_2) = -\bar{\mathsf{I}}_2\mathsf{I}_2. \tag{146}$$

Thus,

$$\mathsf{L}(\mathfrak{J})^\ddagger \diamond \bar{\boldsymbol{u}} \otimes \boldsymbol{u} = -(\tau'\bar{\mathsf{I}}_2 \cdot \bar{\boldsymbol{u}}) \otimes (\tau'\mathsf{I}_2 \cdot \boldsymbol{u}), \tag{147}$$

where $\tau' = \pm 1$.

Noting, however, that the total inversion $\mathfrak{J}$ must be an antilinear operator (to preserve the commutation relation $[\mathbf{r}, \mathbf{p}] = i\hbar\mathsf{I}_3$), we write the above equation as

$$\begin{aligned}\mathsf{L}(\mathfrak{J})^\ddagger \diamond \bar{\boldsymbol{u}} \otimes \boldsymbol{u} &= -(\tau'\bar{\mathsf{I}}_2\mathsf{C} \cdot \boldsymbol{u}) \otimes (\tau'\mathsf{I}_2\mathsf{C} \cdot \bar{\boldsymbol{u}}) \\ &= -\mathbf{P}_{13}[(\tau'\mathsf{I}_2) \otimes (\tau'\bar{\mathsf{I}}_2)] \\ &\quad \diamond [(\mathsf{C}\bar{\boldsymbol{u}}) \otimes (\mathsf{C}\boldsymbol{u})] \\ &\equiv -\mathbf{P}_{13}[(\tau'\mathsf{I}_2\mathsf{C}) \otimes (\tau'\bar{\mathsf{I}}_2\mathsf{C})] \\ &\quad \diamond (\bar{\boldsymbol{u}} \otimes \boldsymbol{u}),\end{aligned}$$

i.e.,

$$\mathsf{L}(\mathfrak{J})^\ddagger = -\mathbf{P}_{13}[(\tau'\mathsf{I}_2\mathsf{C}) \otimes (\tau'\bar{\mathsf{I}}_2\mathsf{C})]. \tag{148}$$

Equation (148) is of the form of (138c), with

$$\mathbf{S}(\mathfrak{J}) = \tau'\bar{\mathsf{I}}_2\mathsf{C},$$
$$\bar{\mathbf{S}}(\mathfrak{J}) = \tau'\mathsf{I}_2\mathsf{C}, \tag{149}$$

and the two-dimensional spinors transform according to Eq. (139). As in the previous case, $\mathcal{S}_2$ is not invariant under $\mathbf{S}(\mathfrak{J})$, and there is no two-dimensional representation for total inversion. The four-dimensional representation in $\mathcal{S}_2 + \bar{\mathcal{S}}_2$ is obtained by

$$\psi' = \mathbf{S}(\mathfrak{J}) \cdot \varphi - \bar{\mathbf{S}}(\mathfrak{J}) \cdot \bar{\chi} = \mathbf{\Lambda}(\mathfrak{J}) \cdot \psi, \tag{150}$$

where

$$\mathbf{\Lambda}(\mathfrak{J}) = \mathbf{S}(\mathfrak{J}) - \bar{\mathbf{S}}(\mathfrak{J}) = -\tau'(\mathsf{I}_2 - \bar{\mathsf{I}}_2)\mathsf{C} = i\tau'\mathbf{\Gamma}^5\mathsf{C}. \tag{151}$$

C. Time Reversal

Since time reversal is simply given by

$$\mathsf{L}(\mathfrak{T}) = \mathbf{E}_k\mathbf{E}^k - \mathbf{E}_0\mathbf{E}^0 = \mathsf{L}(\mathfrak{I}) \circ \mathsf{L}(\mathfrak{J}), \tag{152}$$

we immediately obtain the four-dimensional spinor representation

$$\mathbf{\Lambda}(\mathfrak{T}) = \mathbf{\Lambda}(\mathfrak{I}) \cdot \mathbf{\Lambda}(\mathfrak{J}) = \tau''\mathbf{\Gamma}^5 \cdot \mathbf{\Gamma}^0\mathsf{C}, \tag{153}$$

where $\tau'' = \pm 1$.

It is apparent from Eq. (153) that $\mathbf{\Lambda}(\mathfrak{T})$ is antilinear. Furthermore, in contradistinction to the previous cases, there is also a two-dimensional spinor representation for $\mathfrak{T}$. It is easily derived by noting that

$$\begin{aligned}\mathbf{\Lambda}(\mathfrak{T}) = \mathbf{\Lambda}(\mathfrak{I}) \cdot \mathbf{\Lambda}(\mathfrak{J}) &= [\mathbf{S}(\mathfrak{I}) + \bar{\mathbf{S}}(\mathfrak{I})] \cdot [\mathbf{S}(\mathfrak{J}) - \bar{\mathbf{S}}(\mathfrak{J})] \\ &= \bar{\mathbf{S}}(\mathfrak{I}) \cdot \mathbf{S}(\mathfrak{J}) - \mathbf{S}(\mathfrak{I}) \cdot \bar{\mathbf{S}}(\mathfrak{J}) \\ &= \mathbf{S}(\mathfrak{T}) - \bar{\mathbf{S}}(\mathfrak{T}),\end{aligned} \tag{154}$$

where

$$\begin{aligned}\mathbf{S}(\mathfrak{C}) &= \bar{\mathbf{S}}(\mathfrak{F}) \cdot \mathbf{S}(\mathfrak{J}) = -\sqrt{2}\,\tau\bar{\mathbf{E}}^0 \cdot \tau'\bar{\mathbf{I}}_2\mathbf{C} \\ &= -\tau''\sqrt{2}\,\bar{\mathbf{E}}^0\mathbf{C}, \\ \bar{\mathbf{S}}(\mathfrak{C}) &= \mathbf{S}(\mathfrak{F}) \cdot \bar{\mathbf{S}}(\mathfrak{J}) = -\sqrt{2}\,\tau\mathbf{E}^0 \cdot \tau'\mathbf{I}_2\mathbf{C} \\ &= -\tau''\sqrt{2}\,\mathbf{E}^0\mathbf{C}. \end{aligned} \tag{155}$$

Thus $\mathcal{S}_2$ and $\bar{\mathcal{S}}_2$ are invariant under $\mathbf{S}(\mathfrak{C})$ and $\bar{\mathbf{S}}(\mathfrak{C})$ and transform according to

$$\begin{aligned}\varphi &\to -\tau''\sqrt{2}\,\bar{\mathbf{E}}^0\mathbf{C} \cdot \varphi, \\ \bar{\chi} &\to +\tau''\sqrt{2}\,\mathbf{E}^0\mathbf{C} \cdot \bar{\chi}, \end{aligned} \tag{156}$$

in agreement with Eq. (139).

ACKNOWLEDGMENTS

The authors wish to express their appreciation to General Electric Company, and in particular to Dr. Roy W. Hendrick, Jr., for their interest and support of this work.

APPENDIX

In this appendix we prove two theorems which apply to dyadics attached to $\mathcal{S}_2$.

Theorem 1: For any dyadic $\mathbf{A}$, if its scalar is zero ($\mathbf{A}_s = 0$), then $\mathbf{A}$ is symmetric ($\tilde{\mathbf{A}} = \mathbf{A}$).

Proof: First write $\mathbf{A}$ as the sum of an antisymmetric and a symmetric term:

$$\mathbf{A} = \tfrac{1}{2}(\mathbf{A} - \tilde{\mathbf{A}}) + \tfrac{1}{2}(\mathbf{A} + \tilde{\mathbf{A}}). \tag{A1}$$

Since $\mathbf{I}_2$ is the only independent antisymmetric dyadic attached to $\mathcal{S}_2$, we have

$$\tfrac{1}{2}(\mathbf{A} - \tilde{\mathbf{A}}) = \alpha\mathbf{I}_2 \tag{A2}$$

for some constant α. Then

$$\mathbf{A} = \alpha\mathbf{I}_2 + \mathbf{M}, \tag{A3}$$

where $\mathbf{M}$ is the symmetric part

$$\mathbf{M} = \tfrac{1}{2}(\mathbf{A} + \tilde{\mathbf{A}}) = \tilde{\mathbf{M}}. \tag{A4}$$

If we now make use of the identity

$$(\mathbf{M})_s = -(\tilde{\mathbf{M}})_s, \tag{A5}$$

which is valid for all dyadics, then it follows from (A4) that

$$(\mathbf{M})_s = -(\mathbf{M})_s,$$

i.e.,

$$(\mathbf{M})_s = 0. \tag{A6}$$

Consequently, after recalling that $(\mathbf{I}_2)_s = -2$, the scalar of Eq. (A3) yields

$$\mathbf{A}_s = -2\alpha. \tag{A7}$$

Thus if $\mathbf{A}_s = 0$, then $\alpha = 0$, and

$$\mathbf{A} = \mathbf{M}. \qquad \text{Q.E.D.} \tag{A8}$$

Theorem 2: Any antisymmetric dyadic ($\tilde{\mathbf{A}} = -\mathbf{A}$) can always be written as

$$\mathbf{A} = -\tfrac{1}{2}\mathbf{A}_s\mathbf{I}_2. \tag{A9}$$

Proof: Since $\mathbf{A}$ is antisymmetric, then $\mathbf{M}$ in Eq. (A3) vanishes and we have

$$\mathbf{A} = \alpha\mathbf{I}_2. \tag{A10}$$

Substituting for α from Eq. (A7) immediately gives Eq. (A9). Q.E.D.

We can summarize the above results in the useful identity;

$$\mathbf{A} = -\tfrac{1}{2}\mathbf{A}_s\mathbf{I}_2 + \tfrac{1}{2}(\mathbf{A} + \tilde{\mathbf{A}}). \tag{A11}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Theory of the Small-Angle X-Ray Scattering from Randomly Oriented Plane Laminas*

HSIN-I WU† AND PAUL W. SCHMIDT
Physics Department, University of Missouri, Columbia, Missouri

(Received 14 March 1968)

In a calculation of the intensity of small-angle x-ray scattering from dilute suspensions of noninteracting randomly oriented particles with uniform electron density, the effect of the particle shape can often be conveniently described by a function $G(M)$ called the intersect distribution. An intersect is defined to be a line which has both ends on the boundary of the particle and which passes through a given point in the particle. For an intersect with length M, the intersect distribution $G(M)$ is the probability-density function, averaged over all allowed orientations of the intersect and over all points of the particle through which an intersect with length M can be drawn. Then $G(M)\,dM$ represents the probability that an intersect has a length between the values M and $M + dM$. Since the calculation of $G(M)$ for a three-dimensional particle appeared too complicated for the first part of an investigation of the properties of $G(M)$, the intersect distribution has been studied for the simpler case of a randomly oriented plane lamina with a smooth convex boundary. Emphasis has been given to a determination of the properties of $G(M)$ which affect the intensity of small-angle x-ray scattering in the outer part of the scattering curve. In this angular region the intensity is determined by the behavior of $G(M)$ at small M and in the neighborhood of M values at which $G(M)$ or its derivatives are discontinuous. An approximate expression for $G(M)$ for small M has been obtained. Discontinuities of $G(M)$ are found to be associated with some special values of the function $M(t_1, t_2)$, which gives the length M of the intersect as a function of the two boundary points at which the ends of the intersect are located, with the end points being specified by the arc length t_i along the boundary from a fixed reference point to the end point i. When $M(t_1, t_2)$ has a maximum, a saddle point, or a double point, $G(M)$ has been found to be discontinuous. [The function $M(t_1, t_2)$ can have no minima.] For a maximum of $M(t_1, t_2)$, $G(M)$ has a finite discontinuity, while the discontinuity is logarithmic for a saddle point. For two types of double points which have been studied, $G(M)$ has discontinuities proportional to $|D - M|^{-\frac{1}{2}}$ and $|D - M|^{-\frac{1}{4}}$, where D is the value of M at which $M(t_1, t_2)$ has the double point. For a plane lamina, the approximate expression for $G(M)$ shows that the form of $G(M)$ for small M has no effect on the outer parts of the scattered intensity curve. [For three-dimensional particles, the small-M behavior of $G(M)$ does affect this part of the scattering curve.] The effect of the discontinuities in $G(M)$ on the outer portion of the scattering curve has been calculated. An expression is developed for $G(M)$ and the scattered intensity for an elliptical lamina, and the results of this calculation verify the properties of $G(M)$ for a plane lamina with an arbitrary smooth convex boundary.

1. INTRODUCTION

The determination of the dimensions and shape of colloidal particles in dilute suspensions is one of the most important and frequent uses of small-angle x-ray scattering. This information is obtained from the scattering data by use of equations from small-angle x-ray scattering theory The relation between the measured intensity and the dimensions and form of a colloidal particle is so complex that a complete, exact treatment of this question is not yet possible. However, a number of approximate expressions have been obtained, and certain special cases, such as the scattering from particles with simple shapes like ellipsoids and right circular cylinders, can be treated in detail. These results have been found to be sufficient for analysis of almost all data obtained from dilute colloidal suspensions.

Nevertheless, further investigation of the theory of small-angle x-ray scattering is important, since a better understanding of the theory can broaden the range of applicability of small-angle x-ray scattering techniques and increase the amount of information obtainable by analysis of the data from a given experimental investigation.

In this discussion, the colloidal suspension will be assumed to satisfy several conditions. First, the suspension will be considered so dilute that interparticle interactions will not affect the observed scattering. The intensity scattered by N identical particles is then N times the intensity scattered by a single particle, and so only the scattering from a single particle need be considered in a study of the theory of the small-angle x-ray scattering from a dilute suspension of identical colloidal particles. Also, the particles can be assumed to be randomly oriented, and the measured scattering is the scattering averaged over all particle orientations. Finally, at small angles, the atomic and molecular structure does not affect the small-angle x-ray scattering,[1] and so the

* Work supported by the National Science Foundation. Further details of this investigation are contained in a thesis presented by H. Wu in partial fulfillment of the requirements for the Ph.D. degree, University of Missouri (1967). (Copies available from University Microfilms, Ann Arbor, Michigan.)

† Present address: Physics Department, Southeast Missouri State College, Cape Girardeau, Missouri.

[1] A. Guinier, G. Fournet, C. B. Walker, and K. L. Yudowitch, *Small Angle Scattering of X-Rays* (J. Wiley & Sons, Inc., New York, 1955), p. 4.

particles can be considered to have a uniform electron density and to be suspended in a solvent with a constant, though different, electron density.

Thus, the sample will be assumed to be a dilute suspension of independent randomly oriented particles with uniform electron density, suspended in a solvent with uniform electron density. Under these assumptions, a discussion of the theory of small-angle x-ray scattering need consider only the scattering from a single randomly oriented particle with uniform electron density.

The scattered intensity $I(h)$ then can be expressed as[2]

$$I(h) = NI_e(h)\overline{F^2(h)}, \tag{1}$$

where $h = 4\pi\lambda^{-1}\sin(\phi/2)$, λ is the x-ray wavelength, ϕ is the scattering angle, N is the number of particles in the sample, $I_e(h)$ is the intensity that would be scattered by a single electron under the same experimental conditions,

$$\overline{F^2(h)} = 4\pi\rho^2 V\int_0^{D_{\max}} r^2\gamma_0(r)\frac{\sin hr}{hr}\,dr, \tag{2}$$

ρ is the difference between the electron densities of the particle and the solvent in which the particle is suspended, V is the particle volume, $\gamma_0(r)$ is a function called the characteristic function (which is determined by the dimensions and form of the particle), and $D_{\max}$ is the length of the longest straight line that can be contained in the particle. The quantity $D_{\max}$ will be referred to as the maximum diameter. The characteristic function $\gamma_0(r)$, which, at least in principle, can be obtained from the experimental data by Fourier transformation, contains all information about the particle obtainable by x-ray studies. An investigation of the relation between the x-ray scattering and the dimensions and shape of the particle thus need consider only the connection between $\gamma_0(r)$ and the particle shape and dimensions, and the problem therefore reduces to a study of the effect of the particle size and shape on the characteristic function $\gamma_0(r)$.

The function $\gamma_0(r)$ represents the probability, averaged over all particle orientations and over all points of the particle, that if one point is in a particle, a second point at a distance r from the first point is also in the particle.[3]

While the characteristic function can be calculated explicitly for relatively simple shapes, such as spheres or ellipsoids of revolution, a general discussion of the characteristic function should, when possible, be concerned with properties of $\gamma_0(r)$ which can be obtained without assumption of a specific particle shape.

Porod[4] has shown that information equivalent to that obtainable from $\gamma_0(r)$ can also be found from a function $G(M)$, called the intersect-distribution function. An intersect is defined to be a line with length M passing through a given point in the particle and with both ends terminating on the particle boundary. The intersect distribution $G(M)$ is the probability density, averaged over all orientations of the intersect and over all points of the particle through which an intersect with length M can be drawn. Thus $G(M)\,dM$ represents the average probability that an intersect will have a length in the interval between M and $(M + dM)$.

Porod[4] has shown that for a three-dimensional particle

$$\gamma_0(r) = (\bar{M})^{-1}\int_r^{D_{\max}} dM(M - r)G(M), \tag{3}$$

where

$$\bar{M} = \frac{\int_0^{D_{\max}} MG(M)\,dM}{\int_0^{D_{\max}} G(M)\,dM}. \tag{4}$$

The intersect distribution is assumed to satisfy the normalization condition

$$1 = \int_0^{D_{\max}} G(M)\,dM.$$

By differentiation of (3) and use of the normalization condition,

$$\bar{M} = -1/\gamma_0'(0). \tag{5}$$

From (3),

$$\gamma_0''(r) = (\bar{M})^{-1}G(r). \tag{6}$$

Thus, if either the characteristic function or the intersect distribution is known, the other function can be calculated.

Since both ends of an intersect lie on the particle surface, the intersect-distribution function is more directly connected with the properties of the boundary than is $\gamma_0(r)$. As this property seemed to simplify the calculations, our recent studies have dealt with the intersect distribution, rather than with the characteristic, function.

Many of our investigations of small angle x-ray scattering theory[5–9] have been primarily concerned

[2] Reference 1, p. 7, and p. 12, Eq. (21).
[3] Reference 1, p. 12.

[4] G. Porod, in *Proceedings of the Conference on Small Angle Scattering of X-Rays*, H. Brumberger, Ed. (Gordon and Breach Science Publishers, New York, 1968), pp. 1–15.
[5] P. W. Schmidt, in *Proceedings of the Conference on Small Angle Scattering of X-Rays*, H. Brumberger, Ed. (Gordon and Breach Science Publishers, New York, 1968), pp. 17–31.
[6] P. W. Schmidt and R. Hight, Jr., J. Appl. Phys. **30**, 866 (1959).
[7] A. Miller and P. W. Schmidt, J. Math. Phys. **3**, 92 (1962).
[8] P. W. Schmidt, J. Math. Phys. **6**, 424 (1965).
[9] P. W. Schmidt, J. Math. Phys. **7**, 1295 (1966).

with the form of the scattered intensity for relatively large values of h. (Even though h is relatively large, the corresponding scattering angles are no greater than a few degrees.) The scattered intensity is a Fourier transform and, according to the theory of the asymptotic expansion of Fourier transforms,[10] the intensity at large h is determined by the form of the characteristic function and thus by the behavior of $G(M)$ near $M = 0$ and in the neighborhood of the points at which $G(M)$ or its derivatives are discontinuous. A calculation of the scattered intensity at large h thus requires knowledge of $G(M)$ for small M and in the neighborhood of M values at which the intersect distribution function or its derivatives are discontinuous.

Because of the complexity of the calculation, we have investigated the intersect distribution for a convex-plane lamina instead of considering the more complex three-dimensional case. (A convex particle is defined to be a particle for which the entire length of every intersect lies within the particle.) In a previous publication[11] we have described some general properties of $G(M)$, and $G(M)$ has been calculated for a circle and also for small M for an arbitrary convex plane lamina.

Below, we make a higher-order approximate calculation of $G(M)$ for small M for a convex-plane lamina. The form of $G(M)$ in the neighborhood of its discontinuities is also considered. These results are then verified for an elliptical lamina, and the form of the intensity scattered by a plane lamina is calculated for large h.

2. THE METHOD FOR CALCULATING $G(M)$

In analogy to the expression (2) for three dimensions, for a plane lamina with area A the particle structure factor $\overline{F^2(h)}$ can be written

$$\overline{F^2(h)} = 2\pi\rho^2 A \int_0^{D_{\max}} r\beta_0(r) \frac{\sin hr}{hr}\, dr, \qquad (7)$$

where $\beta_0(r)$ is the characteristic function for a plane lamina. For a plane lamina, (4) is valid, and the analog of (3) is[11]

$$\beta_0(r) = (\bar{M})^{-1} \int_r^{D_{\max}} dM(M - r)G(M). \qquad (8)$$

Thus

$$\beta_0''(r) = (\bar{M})^{-1} G(r). \qquad (9)$$

In Ref. 11, a method is developed for calculating $G(M)$ for a plane lamina.

Let $\mathbf{p}$ be a vector from a fixed origin to a point in the lamina, and let M be the length of an intersect passing through this point, which will be called "point $\mathbf{p}$." This intersect will make an angle θ with a fixed axis. The first step in the calculation of $G(M)$ is to express θ as a function of M for a given point $\mathbf{p}$. For a plane lamina, there will, in general, be more than one solution of the equation giving θ in terms of M. (Usually, there are two solutions.)

The longest possible intersect will have a length equal to $D_{\max}$. For a given value of M in the interval $0 \leq M \leq D_{\max}$, an intersect with a given length M will not pass through all points of the lamina. For example, when M is small, the intersect will pass only through points near the boundary.

After θ has been expressed as a function of M, $d\theta/dM$ can be evaluated at point $\mathbf{p}$. Also, the distance $R(\mathbf{p}, M)$ from point $\mathbf{p}$ to one end of the intersect can be calculated. The distance from point $\mathbf{p}$ to the other end of the intersect thus is $M - R(\mathbf{p}, M)$. Then[11]

$$G(M) = \frac{3\bar{M}}{2\pi M A} \sum_{i=1}^{j} \int_A dA \left| \frac{d\theta}{dM} \right|_i P_i(\mathbf{p}, M), \qquad (10)$$

where

$$P_i(\mathbf{p}, M) = 1 - \frac{2[M - R_i(\mathbf{p}, M)]R_i(\mathbf{p}, M)}{M^2}. \qquad (11)$$

In (10) and (11), $|d\theta/dM|_i$, $P_i(\mathbf{p}, M)$, and $R_i(\mathbf{p}, M)$ are the values of these quantities for solution i of the j possible solutions of the equation giving θ as a function of M. The surface integration in (10) extends over all points $\mathbf{p}$ through which it is possible to pass an intersect with length M.

3. THE INTERSECT DISTRIBUTION FUNCTION FOR SMALL M

In the evaluation of (10), points on the boundary will be specified by giving the arclength from a reference point on the boundary. The boundary point for which this arclength equals t will be referred to as "point t." For the calculations it is convenient to construct a Cartesian coordinate system with its origin at a point T on the boundary, with the positive y axis lying along the inward normal at point T, and with the positive x axis corresponding to the direction of increasing t.

The vector $\mathbf{p}$ defines a point which will be referred to as "point $\mathbf{p}$." Point T is chosen so that point $\mathbf{p}$ lies on the positive y axis. Then point $\mathbf{p}$ can be described by giving the value of T and the distance b between point $\mathbf{p}$ and point T, and point $\mathbf{p}$ has the coordinates $(0, b)$.

The intersect, which has a length M, passes through point $\mathbf{p}$, and the ends of the intersect lie on the boundary

[10] A. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), p. 49.
[11] P. W. Schmidt, J. Math. Phys. **8**, 475 (1967).

at points t_2 and t_1, which are labeled so that $t_2 \geq T \geq t_1$. The Cartesian coordinates (x_2, y_2) and (x_1, y_1) of points t_2 and t_1 can be expressed by the series[12]

$$x_i = (t_i - T) + R(T)X_i,$$
$$y_i = \frac{(t_i - T)^2}{2R(T)} + R(T)Y_i, \tag{12}$$

where $i = 1, 2$, $R(T)$ is the radius of curvature at point T, and

$$X_i = \frac{1}{R(T)} \sum_{n=0}^{\infty} C_n(T)(t_i - T)^{n+3},$$
$$Y_i = \frac{1}{R(T)} \sum_{n=0}^{\infty} D_n(T)(t_i - T)^{n+3},$$
$$C_0(T) = -\frac{1}{6[R(T)]^2},$$
$$D_0(T) = -\frac{R'(T)}{6[R(T)]^2},$$
$$D_1(T) = -\frac{1 + R(T)R''(T) - 2[R'(T)]^2}{24[R(T)]^3}.$$

The higher-order coefficients in these series, which were developed in a study of the small-angle x-ray scattering from filaments,[5] can be calculated by the Frenet–Serret equations.[13]

For all three points (x_2, y_2), $(0, b)$, and (x_1, y_1) to lie on the intersect, we must have

$$y_i - b = x_i \tan\theta, \tag{13}$$

where $i = 1, 2$, and θ is the angle between the intersect and the x axis.

For each value of i, (12) and (13) give

$$0 = (t_i - T)^2 - 2\tan\theta R(T)(t_i - T) - 2bR(T) + 2Q_i[R(T)]^2,$$

where

$$Q_i = Y_i - X_i \tan\theta.$$

The above equation for $(t_i - T)$ can be put in the form

$$t_i - T = R(T)\tan\theta + (-1)^i E_i, \tag{14}$$

where

$$E_i = \{2bR(T) + [R(T)]^2\tan^2\theta - 2[R(T)]^2 Q_i\}^{\frac{1}{2}}.$$

Points t_2 and t_1 satisfy the condition

$$x_2 - x_1 = M\cos\theta.$$

Thus, from (12) and (14),

$$M\cos\theta - R(T)(X_2 - X_1) = (t_2 - T) - (t_1 - T) = E_2 + E_1.$$

By rearrangement of this equation, one obtains the expression

$$\left\{1 + \frac{M^2}{4[R(T)]^2}\right\}\tan^2\theta = \frac{M^2 - 8bR(T)}{4[R(T)]^2} + Q_1 + Q_2 + W, \tag{15}$$

where

$$W = \frac{M^2\tan^4\theta}{4[R(T)]^2[1 + \tan^2\theta]} - \frac{M(X_2 - X_1)}{2R(T)(1 + \tan^2\theta)^{\frac{1}{2}}} + \frac{(X_2 - X_1)^2}{4} + \frac{(Q_2 - Q_1)^2}{\{[M/R(T)](1 + \tan^2\theta)^{-\frac{1}{2}} - (X_2 - X_1)\}^2}.$$

When (15) is substituted in the E_i in (14), the latter equation becomes

$$t_i - T = R(T)\tan\theta + (-1)^i\{\tfrac{1}{4}M^2 + [R(T)]^2 V_i\}^{\frac{1}{2}}, \tag{16}$$

where

$$V_i = -(-1)^i(Q_2 - Q_1) + W - \frac{M^2}{4[R(T)]^2}\tan^2\theta.$$

Equations (15) and (16) can be used to calculate a sequence of successive approximations for θ and $(t_i - T)$.

According to (15) and (16), M, θ, $(t_2 - T)$, and $(t_1 - T)$ are all of the same order of magnitude. In the discussion below, the "nth-order approximation" is defined to the approximation obtained by neglecting all terms in (15) with magnitude M^j for which $j > n$.

In the second-order approximation, W, the V_i, and the Q_i can be neglected in (15) and (16). Then

$$t_i - T = R(T)\tan\theta + (-1)^i(M/2), \tag{17}$$

$$\theta = \pm\frac{[M^2 - 8bR(T)]^{\frac{1}{2}}}{2R(T)}. \tag{18}$$

Equation (18) is equivalent to the approximate expression for θ obtained in Sec. IV of Ref. 11.

In (16), V_i is of the same order or magnitude as $(t_i - T)^3$ and M^3. Consequently, V_i/M^2 is of the same magnitude as M. By use of this result, (16) can be approximated by the expression

$$t_i - T = R(T)\tan\theta + (-1)^i \times \frac{M}{2}\left\{1 + 2\frac{V_i[R(T)]^2}{M^2} - 2\frac{[R(T)]^4[V_i]^2}{M^4} + \cdots\right\}. \tag{19}$$

[12] Reference 5, pp. 21–22.

[13] D. V. Widder, *Advanced Calculus* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1947), p. 84.

Equation (19) can be used to obtain a "third-order" expression for evaluating the $(t_i - T)$ in the Q_i in (15); the $(t_i - T)$ in W in (15) and in the V_i in (19) can be obtained from (17), since for these terms a lower-order approximation is sufficient.

When all terms with order higher than four are dropped, (15) becomes

$$\theta^2 = c + 2g_0\theta + 2g_2\theta^3 + 2g_3\theta^4, \tag{20}$$

where

$$g_0 = -R'(T)M^2/8[R(T)]^2,$$
$$g_2 = -R'(T)/6,$$
$$g_3 = -\{5 + R(T)R''(T) + [R'(T)]^2\}/24,$$
$$c = \frac{2}{R(T)} \times \frac{\frac{M^2}{8R(T)} - b + M^4\frac{9 - 3R(T)R''(T) + [R'(T)]^2}{1152[R(T)]^3}}{1 - M^2\frac{1 - R(T)R''(T) + [R'(T)]^2}{8[R(T)]^2}}.$$

The solution of this equation is given by Eq. (A16) of Appendix A and can be written in the form

$$\theta = \theta_a + \theta_b, \tag{21}$$

where

$$\theta_a = -R'(T)[M^2 - 2bR(T)]/6[R(T)]^2,$$
$$\theta_b = \pm\left[\frac{2(b_m - b)}{R(T)}\right]^{\frac{1}{2}} \times \left[1 + b\frac{15 + 3R(T)R''(T) - 2[R'(T)]^2}{36R(T)} + M^2\frac{3 - 21R(T)R''(T) + 38[R'(T)]^2}{288[R(T)]^2}\right],$$
$$b_m = \frac{M^2}{8R(T)} + M^4\frac{9 - 3R(T)R''(T) + 10[R'(T)]^2}{1152[R(T)]^3}. \tag{22}$$

In (21), real solutions for θ exist only for $0 \leq b \leq b_m$. Thus, for small M, the region of integration in (10) does not extend over the entire lamina but instead is limited to a narrow band extending a distance b_m inward from the boundary.

The function $P(\mathbf{p}, M)$ in (10) can be expressed

$$P(\mathbf{p}, M) = 1 + \frac{2x_1x_2\sec^2\theta}{M^2},$$

since the coordinate system has been chosen so that

$$R(\mathbf{p}, \theta) = -x_1\sec\theta,$$
$$M - R(\mathbf{p}, \theta) = x_2\sec\theta.$$

From (12), (17), and (18),

$$x_1x_2\sec^2\theta = (t_1 - T)(t_2 - T)\sec^2\theta \times \left\{1 - \frac{(t_1 - T)^2 + (t_2 - T)^2}{6[R(T)]^2} + \cdots\right\}$$
$$\approx (t_1 - T)(t_2 - T) \times \left\{1 + \frac{M^2}{12[R(T)]^2} - \frac{4}{3}\frac{b}{R(T)}\right\}.$$

In a fourth-order approximation, from (19) and the definition of the V_i given below (16), we have

$$(t_2 - T)(t_1 - T) \approx -2bR(T) \times \left\{1 + \frac{2[R(T)]^2}{M}\frac{(t_1 - T)Q_2 - (t_2 - T)Q_1}{(t_2 - T)(t_1 - T)}\right\}^{-1}.$$

When the $t_i - T$ and θ are expressed in terms of M,

$$P(\mathbf{p}, M) = P_2 + P_3 + P_4, \tag{23}$$

where

$$P_2 = 1 - \frac{4bR(T)}{M^2},$$
$$P_3 = \pm 4bR'(T)[M^2 - 8bR(T)]^{\frac{1}{2}}/3M^2,$$
$$P_4 = 2(b/M)^2 - \frac{bR''(T)}{3M^2}[M^2 - 6bR(T)] + \frac{2b[R'(T)]^2}{9R(T)M^2}[M^2 + 6bR(T)].$$

By differentiation of (21),

$$\left|\frac{d\theta}{dM}\right| = \pm\frac{MR'(T)}{3[R(T)]^2} + \frac{M(1 + D_2)}{4R(T)[2R(T)(b_m - b)]^{\frac{1}{2}}}, \tag{24}$$

where

$$D_2 = M^2\frac{45 - 75R(T)R''(T) + 154[R'(T)]^2}{288[R(T)]^2} + b\frac{9 + 45R(T)R''(T) - 78[R'(T)]^2}{36R(T)}.$$

In (10), a convenient choice of the area element dA is

$$dA = [1 - b/R(T)]\,db\,dT.$$

The two signs shown for the first term on the right side of (24) and for P_3 in (23) correspond to the two possible solutions for θ in terms of M. These two solutions give the two terms in the sum in (10). In the expressions for $|d\theta/dM|$ and $P(\mathbf{p}, M)$, one sign corresponds to each value of j in the sum.

When (23) and (24) are substituted in (10), the expressions for $G(M)$ become

$$G(M) = \frac{3\bar{M}}{2\pi MA}\sum_{j=1}^{2}\int_0^L dT \times \int_0^{b_m} db\left[1 - \frac{b}{R(T)}\right]\left|\frac{d\theta}{dM}\right|_j P_j(\mathbf{p}, M), \tag{25}$$

where L is the total arclength of the boundary.

After the integration in (25) is carried out, one obtains

$$G(M) = \frac{\bar{M}ML}{4\pi A}\overline{\left(\frac{1}{R}\right)^2} + \frac{\bar{M}M^3L}{32\pi A}\overline{\left(\frac{1}{R}\right)^4} - \frac{11\bar{M}M^4}{240\pi A} \times \int_0^L \frac{R''(T)\,dT}{[R(T)]^3} + \frac{79\bar{M}M^3}{720\pi A}\int_0^L \frac{[R'(T)]^2}{[R(T)]^4}\,dT,$$

where

$$\overline{\left(\frac{1}{R}\right)^n} = \frac{1}{L}\int_0^L \frac{dT}{[R(T)]^n}.$$

By partial integration,

$$\int_0^L \frac{R''(T)\,dT}{[R(T)]^3} = 3\int_0^L dT\,\frac{[R'(T)]^2}{[R(T)]^4}.$$

The expression for $G(M)$ can therefore be written

$$G(M) = \frac{\bar{M}LM}{4\pi A}\overline{\left(\frac{1}{R}\right)^2} + \frac{\bar{M}LM^3}{32\pi A}\overline{\left(\frac{1}{R}\right)^4} - \frac{\bar{M}LM^3}{36\pi A}\overline{\left[\frac{d}{dT}\left(\frac{1}{R(T)}\right)\right]^2}, \quad (26)$$

where

$$\overline{\left[\frac{d}{dT}\left(\frac{1}{R(T)}\right)\right]^2} = \frac{1}{L}\int_0^L dT\,\frac{[R'(T)]^2}{[R(T)]^4}.$$

4. DISCONTINUITIES OF $G(M)$

As mentioned in the Introduction, the scattered intensity at large h is determined in part by the form of $G(M)$ in the neighborhood of M values at which $G(M)$ or its derivatives are discontinuous.

Jones and Kline[14] have developed a procedure for asymptotic expansion of double Fourier integrals. Even though the calculation of the scattered intensity from a plane lamina is equivalent to evaluation of a fourfold Fourier integral (rather than a double integral), many of the results obtained by Jones and Kline can still be expected to be applicable to the scattering from a plane lamina.

In particular, the work of Jones and Kline suggests consideration of the function $M(t_1, t_2)$, which gives the length M of the intersect which has its ends at points t_1 and t_2 on the lamina boundary. Discontinuities of $G(M)$ or its derivatives can be expected for M values at which $M(t_1, t_2)$ satisfies the conditions

$$\partial M/\partial t_1 = 0,$$
$$\partial M/\partial t_2 = 0.$$

Our calculations have verified that $G(M)$ will have discontinuities for M values equal to the values of $M(t_1, t_2)$ satisfying the above conditions—that is, when $M(t_1, t_2)$ has a maximum, a saddle point, or a double point. [The function $M(t_1, t_2)$ can be shown to have no minimum.] Analogous to the conclusions of Jones and Kline, we will assume that discontinuities occur when and only when $M(t_1, t_2)$ has a maximum, minimum, or double point. The investigation of the discontinuities of $G(M)$ thus involves a study of $M(t_1, t_2)$ for points t_1 and t_2 in the neighborhood of points where $M(t_1, t_2)$ has maxima, saddle points, and double points.

Let T_1 and T_2 be two points such that $M(T_1, T_2) = D$ represents a maximum, a saddle point, or a double point. Let $\mathbf{r}_i$ be the vector from point T_i to a nearby point t_i on the boundary, and let $\mathbf{D} = \mathbf{d}_0 D$ be the vector from point T_1 to point T_2. Then

$$M = |\mathbf{D} + \mathbf{r}_2 - \mathbf{r}_1| = D\left[1 + 2\frac{\mathbf{d}_0 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{D} + \frac{|\mathbf{r}_2 - \mathbf{r}_1|^2}{D^2}\right]^{\frac{1}{2}}. \quad (27)$$

The origin of the coordinate system is chosen to be at point T_1, with the x axis in the direction of increasing t and with the y axis directed inward, along the vector $\mathbf{D}$. Without loss of generality, the points designated as T_1 and T_2 can be selected so that $R(T_1) \geq R(T_2)$. When $M(t_1, t_2)$ has a maximum, a saddle point, or a double point, the unit vector $\mathbf{d}_0$ is normal to the boundary at T_1 and T_2 (see Ref. 5, p. 24). From (12) we obtain

$$\mathbf{r}_i = (-1)^{i+1} \times \left\{\mathbf{e}[(t_i - T_i) + R_i X_i] + \mathbf{d}_0\left[\frac{(t_i - T_i)^2}{2R_i} + R_i Y_i\right]\right\}, \quad (28)$$

where $R_i = R(T_i)$ and $\mathbf{e}$ is a unit vector in the direction of positive x axis. The X_i and Y_i are evaluated at point T_i. The coordinate system has been chosen so that $\mathbf{d}_0$ is directed along the positive y axis. Since $|\mathbf{r}_2 - \mathbf{r}_1| \ll D$, $M(t_1, t_2)$ can be approximated by the expression

$$M(t_1, t_2) = D + \frac{[(t_2 - T_2) + (t_1 - T_1)]^2}{2D} - \frac{(t_1 - T_1)^2}{2R_1} - \frac{(t_2 - T_2)^2}{2R_2} + \cdots. \quad (29)$$

According to (29), $M(T_1, T_2)$ is a saddle point when $D - (R_1 + R_2) < 0$. When $D - (R_1 + R_2) > 0$, $M(T_1, T_2)$ will either be a minimum or a maximum. Further study shows that there can be only a maximum, since the existence of a minimum requires that

[14] D. S. Jones and M. Kline, J. Math. & Phys. 37, 1 (1958).

both the inequalities $R_1 > D$ and $R_2 > D$ be satisfied. Since R_1, R_2, and D are always positive, these two inequalities can never be fulfilled when $D - (R_1 + R_2) > 0$. Therefore when $D - (R_1 + R_2) > 0$, $M(T_1, T_2)$ cannot be a minimum and thus will be a maximum.

When $D = R_1 + R_2$, $M(T_1, T_2)$ is a double point. In the approximation represented by (29), for a double point the boundary is made up of arcs of two *concentric* circles with radii R_1 and R_2. In this case, (29) is not sufficient for determining the region of integration in (10), and a higher-order approximation for $M(t_1, t_2)$ is required.

For a plane lamina with a finite area and a continuous boundary, there must be a maximum value of M. Let A and B represent the respective values of T_1 and T_2 for which this maximum value of M occurs. Then, by symmetry, there will also be a maximum for $t_1 = B$ and $t_2 = A$. If L is the total length of the boundary, maxima of M will then be obtained for $t_1 = A + mL$ and $t_2 = B + nL$ and also, by symmetry, for $t_1 = B + mL$ and $t_2 = A + nL$, where m and n are any integers. Along the lines $t_2 = t_1 \pm mL$, $M(t_1, t_2) = 0$. From the values of t_1 and t_2 at which $M(t_1, t_2)$ has maxima and at which it is zero, one can expect that, ordinarily, the number of saddle points will equal the number of maxima.

When $D = R_1 + R_2$, there is a double point and higher-order terms must be considered in (28) and in (27). For a double point, $M(t_1, t_2)$ has the form

$$M = D - \frac{D\alpha^2}{2R_1R_2} + \frac{1}{6}\left[\frac{R'}{(R_1)^2}(t_1 - T_1)^3 + \frac{R_2'}{(R_2)^2}(t_2 - T_2)^3\right], \quad (30)$$

where

$$\alpha = [R_2(t_1 - T_1) - R_1(t_2 - T_2)]/D.$$

If $R'(T_1)$ and $R'(t_2)$ are both zero, an even higher-order expansion must be used instead of (30). Then

$$M = D - \frac{D\alpha^2}{2R_1R_2} + D\frac{(R_1)^2 - R_1R_2 + (R_2)^2}{24R_1^3R_2^3}\alpha^4 + \frac{1}{24}\left[\frac{R_1''(t_1 - T_1)^4}{(R_1)^2} + \frac{R_2''(t_2 - T_2)^4}{(R_2)^2}\right], \quad (31)$$

where $R_i'' = R''(T_i)$, with $i = 1, 2$.

5. MAXIMA AND SADDLE POINTS

The angle θ will be near 90° for a maximum, saddle point, or double point, unlike the calculation for small M, when θ is small. Let (x_i, y_i) represent the Cartesian coordinates of point t_i. Then, for the coordinate system which has been selected for the discussion of the form of $G(M)$ for this case, with (28) one obtains

$$\begin{aligned} x_1 &= (t_1 - T_1) + R_1X_1, \\ x_2 &= -(t_2 - T_2) - R_2X_2, \\ y_1 &= [(t_1 - T_1)^2/2R_1] + R_1Y_1, \\ y_2 &= D - [(t_2 - T_2)^2/2R_2] - R_2Y_2. \end{aligned} \quad (32)$$

For all the points (x, y), (x_1, y_1), and (x_2, y_2) to lie on a straight line, we must have

$$(y_i - y)\cot\theta = x_i - x.$$

From (32),

$$\begin{aligned} t_1 - T_1 &= x - y\cot\theta + R_1(Y_1\cot\theta - X_1) + [(t_1 - T_1)^2/2R_1]\cot\theta, \\ t_2 - T_2 &= -x - (D - y)\cot\theta + R_2(Y_2\cot\theta - X_2) + [(t_2 - T_2)^2/2R_2]\cot\theta. \end{aligned} \quad (33)$$

When the lowest-order approximations from (33) are substituted in (29), one obtains an approximate equation expressing θ in terms of M:

$$0 = \frac{D - R_1 - R_2}{R_1 + R_2}a(y)\cot^2\theta - 2x\left(y - \frac{DR_1}{R_1 + R_2}\right)\cot\theta + x^2 - \frac{2R_1R_2(D - M)}{R_1 + R_2}, \quad (34)$$

where

$$a(y) = \frac{R_1 + R_2}{D - R_1 - R_2}\left[y - \frac{R_1D}{R_1 + R_2}\right]^2 + \frac{R_1R_2D}{R_1 + R_2}. \quad (35)$$

When M is in the neighborhood of a maximum, saddle point, or double point of $M(t_1, t_2)$, the region of integration is restricted to points in the neighborhood of the y axis, and θ is always near 90°. By differentiation of the expression for $\cot\theta$ obtained by solving (34), for $D \neq R_1 + R_2$ one obtains

$$\left|\frac{d\theta}{dM}\right| \approx \left|\frac{d\cot\theta}{dM}\right| = \frac{(R_1R_2)^{\frac{1}{2}}}{\left\{D(D - R_1 - R_2)\left[\frac{2(D - M)}{D}a(y) - x^2\right]\right\}^{\frac{1}{2}}}. \quad (36)$$

In the weighting function $P(\mathbf{p}, M)$ in (10) for points near the y axis, the quantities $R(\mathbf{p}, M)$ and M can be

approximated by y and D, respectively, giving

$$P(\mathbf{p}, M) \simeq 1 - [2y(D - y)/D^2]. \tag{37}$$

In the sum in (10) there will be two terms corresponding to the two solutions of (34). According to (36) and (37), both terms in the sum have the same value. Then (10) can be written

$$G(M) \approx \frac{3\bar{M}}{\pi DA}\left(\frac{R_1R_2}{D}\right)^{\frac{1}{2}} \times \int_A dA \frac{\left[1 - \frac{2y(D-y)}{D^2}\right]}{\left\{(D - R_1 - R_2)\left[\frac{2(D-M)}{D}a(y) - x^2\right]\right\}^{\frac{1}{2}}}. \tag{38}$$

For the integration in (38), it is convenient to let $dA = dx\,dy$. In this approximation the integration over y extends over the interval $0 \leq y \leq D$, and the limits for x are given by the condition

$$(D - R_1 - R_2)\left[\frac{2(D-M)}{D}a(y) - x^2\right] \geq 0. \tag{39}$$

When M is in the neighborhood of a maximum, $D - R_1 - R_2$ is positive, and, according to (35), $a(y)$ is positive when $D - R_1 - R_2 > 0$. Thus there are no allowed values of x for $M > D$. (This condition merely states that there are no M values greater than the maximum value D.) For $D > M$, the limits for x are given by

$$-x_m \leq x \leq x_m,$$

where

$$x_m = \left[\frac{2(D-M)}{D}a(y)\right]^{\frac{1}{2}}.$$

Therefore, from (38), for $D > M$,

$$G(M) = \frac{2\bar{M}}{A}\left[\frac{R_1R_2}{D(D - R_1 - R_2)}\right]^{\frac{1}{2}}, \tag{40}$$

while for $D < M$, $G(M) = 0$. Thus $G(M)$ has a finite discontinuity when D is a maximum of $M(t_1, t_2)$.

When D is a saddle point, $R_1 + R_2 - D > 0$, and, according to (39), the allowed values of x are given by the condition

$$x^2 - \frac{2(D-M)}{D}a(y) \geq 0.$$

Thus, unlike the region of integration when D is a maximum, for which the allowed x values are restricted to a small region near the y axis, for a saddle point the integration extends over the entire lamina, except for a small forbidden region in the neighborhood of the y axis. This forbidden region is determined by the condition $x^2 \geq x_m^2$, where

$$x_m = \left[\frac{2(D-M)}{D}a(y)\right]^{\frac{1}{2}}, \quad (D - M)a(y) \geq 0,$$
$$x_m = 0, \quad (D - M)a(y) \leq 0.$$

It is convenient to write

$$G(M) = G_c(M) + G_d(M),$$

where $G_c(M)$ is a function that is continuous at $M = D$, while $G_d(M)$ is discontinuous. The function $G_c(M)$ need not be evaluated, since it does not contribute to the asymptotic expansion of the intensity at large h. If X is greater than the largest value of $|x_m|$ throughout the entire interval $0 \leq y \leq D$, there will be no contribution to $G_d(M)$ from x values for which $|x| > X$. Let X be chosen so that, although $X^2 > x_m^2$, X is still small enough that the approximations used when D was a maximum can also be made for the saddle point. Then, from (38), the contribution to $G(M)$ from x values for which $|x| \leq X$ is

$$G_d(M) + G_{c1}(M) \approx \frac{6\bar{M}}{\pi DA}\left[\frac{R_1R_2}{D(R_1 + R_2 - D)}\right]^{\frac{1}{2}} \times \int_0^D dy\left[1 - \frac{2y(D-y)}{D^2}\right]I_1(y),$$

where $G_{c1}(M)$ is a continuous function of M at $D = M$, and

$$I_1(y) = \int_{x_m}^{X} \frac{dx}{\{x^2 - [2(D-M)/D]a(y)\}^{\frac{1}{2}}}.$$

The integral $I_1(y)$ can be written

$$I_1(y) = I_{1X} - I_{1m},$$

where

$$I_{1X} = \log_e\left\{X + \left[X^2 - 2\frac{D-M}{D}a(y)\right]^{\frac{1}{2}}\right\}$$

and

$$I_{1m} = \log_e\left\{x_m + \left[(x_m)^2 - 2\frac{D-M}{D}a(y)\right]^{\frac{1}{2}}\right\}.$$

The quantity x_m is defined so that

$$I_{1m} = \tfrac{1}{2}\log_e\left|\frac{D-M}{D}\right| + \tfrac{1}{2}\log_e[2\,|a(y)|],$$

regardless of whether or not x_m is different from zero. Since I_{1X} is continuous at $D = M$ and $\log_e|2a(y)|$ is independent of M, the discontinuous part $G_d(M)$ of $G(M)$ can be obtained by considering only the term

$-\frac{1}{2}\log_e|(D-M)/D|$ in $I_1(y)$. Thus

$$G_d(M) = -\frac{2\bar{M}}{\pi A}\left[\frac{R_1R_2}{D(R_1+R_2-D)}\right]^{\frac{1}{2}}\log_e\left|\frac{D-M}{D}\right|. \tag{41}$$

Equation (41) shows that when D is a saddle point, $G(M)$ has a logarithmic discontinuity for $M = D$. The integrals in $G_{c1}(M)$ need not be evaluated, since $G_{c1}(M)$ does not affect the intensity at large h.

6. DOUBLE POINTS

When $D = R_1 + R_2$, (34) can be written in the form

$$M = D - (D/2R_1R_2)\alpha_0^2, \tag{42}$$

where

$$\alpha_0 = x + (R_1 - y)\cot\theta.$$

For double points, the value of $|d\theta/dM|$ calculated from (42) is independent of x, and so the existence of real values of this approximation for $|d\theta/dM|$ cannot serve as a condition defining the region of integration in (10). A higher-order approximation is therefore needed for finding $|d\theta/dM|$ for a double point.

When R_1' and R_2' are not both equal to 0, (30) can be used to compute $|d\theta/dM|$ for a double point. The approximations for $(t_1 - T_1)$ and $(t_2 - T_2)$ obtained from (33) for the maxima and saddle points can also be used in (30). The resulting expression is

$$\alpha_0^2 = \frac{2R_1R_2(D-M)}{D} - \frac{R_2^2R_1'(\alpha_0 y - R_1x)^3 + R_1^2R_2'[(D-y)\alpha_0 - R_2x]^3}{3DR_1R_2(R_1-y)^3}. \tag{43}$$

This equation is too complicated for an exact solution. It can be solved approximately by noting that the term proportional to x^3 must have an appreciable effect on the solution of (43) if the existence of solutions of (43) is to determine the region of integration in (10). If this term is to be one of the dominant terms in (43), it must be of the same order of magnitude as $(D-M)$ and α_0^2. Then α_0 and x are proportional to $|D-M|^{\frac{1}{2}}$ and $(D-M)^{\frac{1}{3}}$, respectively. The magnitude of the other terms in (43) can then be determined. If only terms with the same magnitude as $(D-M)$ are retained in (43), one obtains

$$\alpha_0^2 \approx \frac{2R_1R_2(D-M)}{D} + \frac{R_1R_2[R_1R_1' + R_2R_2']}{3(R_1-y)^3D}x^3.$$

From the solution of this quadratic equation for α_0,

$$\left|\frac{d\theta}{dM}\right| \approx \frac{1}{|R_1-y|}\left|\frac{d\alpha_0}{dM}\right| = \frac{(3R_1R_2)^{\frac{1}{2}}}{|R_1-y|(D^{\frac{1}{2}})}\left[6(D-M) + \frac{R_1R_1' + R_2R_2'}{(R_1-y)^3}x^3\right]^{-\frac{1}{2}}. \tag{44}$$

The allowed values of x in the surface integration in (10) are determined by the condition that $|d\theta/dM|$ must be real.

Just as in the discussion of $G(M)$ for M in the neighborhood of a saddle point, for a double point, $G(M)$ can be written as the sum of a continuous function $G_c(M)$ and a function $G_d(M)$ which is discontinuous for $M = D$. Since $G_c(M)$ does not contribute to the asymptotic expansion of the scattered intensity, only $G_d(M)$ need be evaluated explicitly. Let X be a positive quantity small enough that, for $|x| < X$, all approximations needed to obtain (44) are valid, but with X large enough that

$$\left|\frac{R_1R_1' + R_2R_2'}{(R_1-y)^3}\right| X^3 > 6|D-M|.$$

Then in (10), only x values for which $|x| \leq X$ will contribute to $G_d(M)$, and only these values of x need be considered in the calculation of $G_d(M)$. Just as in the study of the form of $G(M)$ for M in the neighborhood of a saddle point, the integration limits $y = 0$ and $y = D$ can be used for a double point.

In (10) the integral has the same value for both values of the summation index, and therefore by analogy with (40), one can write

$$G_d(M) + G_{c1}(M) = \frac{3\bar{M}}{\pi DA}\left(\frac{3R_1R_2}{D}\right)^{\frac{1}{2}}\int_0^D \frac{dy}{|R_1-y|} \times \int_{x_1(y)}^{x_2(y)} \frac{dx\left[1 - \frac{2y(D-y)}{D^2}\right]}{\left[6(D-M) + \frac{R_1R_1' + R_2R_2'}{(R_1-y)^3}x^3\right]^{\frac{1}{2}}}, \tag{45}$$

where $G_{c1}(M)$ is continuous at $M = D$, and where the $x_i(y)$ are determined from the conditions that $x_2(y) \geq x_1(y)$, $|x_i(y)| \leq X$, and

$$6(D-M) + \frac{R_1R_1' + R_2R_2'}{(R_1-y)^3}x^3 \geq 0.$$

In approximating $G_d(M)$, it is convenient to make the change of variables

$$x = \frac{(R_1-y)[6(D-M)]^{\frac{1}{3}}}{[R_1R_1' + R_2R_2']^{\frac{1}{3}}}t. \tag{46}$$

The discontinuous part $G_d(M)$ of $G(M)$ is then found to be expressible in the form

$$G_d(M) = \frac{2\bar{M}}{\pi A(D^{\frac{1}{2}})} \frac{(3R_1R_2)^{\frac{1}{2}}}{(6\,|D-M|)^{\frac{1}{3}}(|R_1R_1' + R_2R_2'|)^{\frac{1}{3}}} \times \int_{-1}^{\infty} \frac{dt}{(1+t^3)^{\frac{1}{2}}} \quad (47)$$

for $D - M > 0$. Similarly, for $D - M < 0$,

$$G_d(M) = \frac{2\bar{M}}{\pi A(D^{\frac{1}{2}})} \frac{(3R_1R_2)^{\frac{1}{2}}}{(6\,|D-M|)^{\frac{1}{3}}(|R_1R_1' + R_2R_2'|)^{\frac{1}{3}}} \times \int_{-\infty}^{-1} \frac{dt}{(-1-t^3)^{\frac{1}{2}}}. \quad (48)$$

The integrals in (47) and (48) can be expressed in terms of gamma functions. By introducing the function

$$K_3(x) = \frac{\sqrt{3}}{2}\left[1 + \frac{x}{|x|}\right] + \frac{1}{2}\left[1 - \frac{x}{|x|}\right],$$

the results for $D - M > 0$ and $D - M < 0$ can be combined to give

$$G_d(M) = \frac{\bar{M}(R_1R_2)^{\frac{1}{2}}\Gamma(\frac{1}{3})\Gamma(\frac{1}{6})K_3(D-M)}{\pi^{\frac{3}{2}}A(D^{\frac{1}{2}})(6\,|D-M|)^{\frac{1}{3}}(|R_1R_1' + R_2R_2'|)^{\frac{1}{3}}}. \quad (49)$$

When there is a double point for which $R_1' = R_2' = 0$ but $R_1^2R_1'' + R_2^2R_2'' \neq 0$, (31) must be used to find $|d\theta/dM|$ and the region of integration in (10). The approximations used for the $(t_i - T_i)$ in (30) are not sufficient for (31), which requires higher-order approximations from (33). For (31), the $(t_i - T_i)$ can be obtained from the relations

$$t_1 - T_1 = \beta + \frac{\beta^2 \cot\theta}{2R_1} + \frac{\beta^3}{6R_1^2},$$

$$t_2 - T_2 = -\gamma + \frac{\gamma^2 \cot\theta}{2R_2} - \frac{\gamma^3}{6R_2^2}, \quad (50)$$

where

$$\beta = x - y\cot\theta,$$

$$\gamma = x + (D-y)\cot\theta.$$

When Eqs. (50) are substituted in (31), a quartic equation for $\cot\theta$ is obtained, which can be simplified by a method analogous to the technique used to obtain the approximation of (43). The resulting equation is

$$\alpha_0^2 = \left[\frac{2R_1R_2}{D}\right]\left[(D-M) + \frac{R_1^2R_1'' + R_2^2R_2''}{24(R_1-y)^4}x^4\right]. \quad (51)$$

Then

$$\left|\frac{d\theta}{dM}\right| \approx \frac{1}{|R_1 - y|}\left|\frac{d\alpha_0}{dM}\right| = \frac{(R_1R_2)^{\frac{1}{2}}}{|R_1-y|\left\{2D\left[D - M + x^4\frac{R_1^2R_1'' + R_2^2R_2''}{24(R_1-y)^4}\right]\right\}^{\frac{1}{2}}}. \quad (52)$$

As in previous calculations, only the discontinuous part of $G(M)$ need be evaluated. Let X be a positive number such that X is small enough to permit the use of all approximations employed previously and satisfying the condition

$$|D-M| + \frac{|R_1^2R_1'' + R_2^2R_2''|}{24(R_1-y)^4}X^4 > 0.$$

Then the part $G_d(M)$ of $G(M)$ that is discontinuous at $M = D$ will come from the region of the surface integration in (10) for which $|x| \leq X$. Since (52) is an even function of x, the integral for negative x values in (10) is equal to the integral for $x > 0$, and so only positive x values need be considered in the approximate evaluation of (10). Let $G_{c1}(M)$ be the continuous function contributed by x values for which $|x| \leq X$. Then

$$G_d(M) + G_{c1}(M) = \frac{6\bar{M}}{\pi DA}\left(\frac{R_1R_2}{2D}\right)^{\frac{1}{2}}\int_0^D \frac{dy}{|R_1-y|} \times \int_{x_1(y)}^{x_2(y)} \frac{dx\left[1 - \frac{2y(D-y)}{D^2}\right]}{\left[(D-M) + \frac{R_1^2R_1'' + R_2^2R_2''}{24(R_1-y)^4}x^4\right]^{\frac{1}{2}}}. \quad (53)$$

The integration limits $x_2(y)$ and $x_1(y)$ are determined by the conditions that: (a) $X \geq x_2(y) \geq x_1(y) \geq 0$, and (b) the integration over x in (53) must extend over all x values for which the integrand is real. A calculation then shows that

$$G_d(M) = \frac{\bar{M}}{\pi A}\left(\frac{6R_1R_2}{\pi D}\right)^{\frac{1}{2}} \times \frac{[\Gamma(\frac{1}{4})]^2K_4(D-M, R_1^2R_1'' + R_2^2R_2'')}{(24\,|R_1^2R_1'' + R_2^2R_2''|\,|D-M|)^{\frac{1}{4}}}, \quad (54)$$

where

$$K_4(x,y) = \frac{1}{4}\left[\sqrt{2}\left(1 + \frac{x}{|x|}\right)\left(1 + \frac{y}{|y|}\right) + \left(1 - \frac{x}{|x|}\right)\left(1 + \frac{y}{|y|}\right) + \left(1 + \frac{x}{|x|}\right)\left(1 - \frac{y}{|y|}\right)\right].$$

Equation (49) indicates that for a double point for which R_1' and R_2' are not both equal to zero, $G(M)$ has an infinite discontinuity proportional to $|D - M|^{-\frac{1}{2}}$. According to (54), when $R_1' = R_2' = 0$ but $R_1^2 R_1'' + R_2^2 R_2'' \neq 0$, the discontinuity is proportional to $|D - M|^{-\frac{1}{2}}$.

7. THE INTERSECT DISTRIBUTION FOR AN ELLIPSE

The characteristic function $\beta_0(r)$ for an ellipse with semimajor axis va and semiminor axis a (thus $v > 1$) is given by Eq. (B7) of Appendix B. The intersect distribution function $G(M)$ for an ellipse can be obtained by substituting (B7) in (9). The resulting expression is

$$G(M) = \frac{32a^2v\bar{M}M}{\pi^2} \times \int_{X(M)}^{2av} \frac{dx}{x^3[x^2 - (2a)^2]^{\frac{1}{2}}(x^2 - M^2)^{\frac{1}{2}}[(2av)^2 - x^2]^{\frac{1}{2}}}, \tag{55}$$

where

$$X(M) = 2a, \quad 0 \leq M \leq 2a,$$
$$X(M) = M, \quad 2a \leq M \leq 2va.$$

For small M, (55) can be approximated by

$$G(M) = \frac{\bar{M}M}{\pi^2 a^3 v^4} I_{\frac{5}{2}} + \frac{\bar{M}M^3}{8\pi^2 a^5 v^6} I_{\frac{7}{2}}, \tag{56}$$

where

$$I_n = \int_0^1 \frac{du}{(1 - u^2)^{\frac{1}{2}}[1 - (1 - v^{-2})u^2]^n}.$$

In the neighborhood of the maximum values $M = 2va$, (55) has the approximate form

$$G(M) \simeq \frac{\bar{M}}{\pi a^2 v^2 (v^2 - 1)^{\frac{1}{2}}}. \tag{57}$$

There is a saddle point at $M = 2a$, and the discontinuous part $G_d(M)$ is given approximately by

$$G_d(M) \simeq -\frac{v\bar{M}}{\pi^2 a^2 (v^2 - 1)^{\frac{1}{2}}} \log_e \left| \frac{M - 2a}{2a} \right|. \tag{58}$$

When the boundary of the ellipse is expressed in terms of the Cartesian coordinates x and y, the radius of curvature R is given by

$$R = \frac{[a^2v^4 - (v^2 - 1)x^2]^{\frac{3}{2}}}{a^2v^4}. \tag{59}$$

Also

$$dT = \left[\frac{a^2v^4 - (v^2 - 1)x^2}{a^2v^4 - v^2x^2}\right]^{\frac{1}{2}} |dx|,$$

where T is the arclength measured from the point with Cartesian coordinates $(av, 0)$. Thus

$$|dR/dT| = 3(v^2 - 1)\,|x|\,(a^2v^2 - x^2)^{\frac{1}{2}}/a^2v^3. \tag{60}$$

The maximum value $M = 2va$ corresponds to the pair of points with Cartesian coordinates $(va, 0)$ and $(-va, 0)$, while the saddle point at $M = 2a$ occurs for the points $(0, a)$ and $(0, -a)$. For the maximum for an ellipse, $R_1 = R_2$. At the saddle point, the two radii of curvature also are equal. Equations (57) and (58) verify the results obtained by substitution of (59) into (40) and (41).

From (59), (60), and (26), for small M, $G(M)$ can be written

$$G(M) = \frac{\bar{M}M}{\pi^2 a^3 v^4} I_{\frac{5}{2}} + \frac{\bar{M}M^3}{8\pi^2 a^5 v^8} I_{1\frac{1}{2}} - \frac{\bar{M}M^3}{\pi^2 a^5 v^{10}} I_G, \tag{61}$$

where

$$I_G = (v^2 - 1)^2 \int_0^1 \frac{u^2(1 - u^2)^{\frac{1}{2}}\,du}{[1 - (1 - v^{-2})u^2]^{1\frac{1}{2}}}.$$

Since

$$I_{n+\frac{1}{2}} = v^{-2} I_{n+\frac{5}{2}} - \frac{2(n+1)(v^2 - 1)^2}{v^4} \times \int_0^1 \frac{u^2(1 - u^2)^{\frac{1}{2}}\,du}{[1 - (1 - v^{-2})u^2]^{n+\frac{5}{2}}},$$

Eq. (61) is equivalent to (56). The approximations for $G(M)$ calculated from the exact expression for the scattered intensity from an ellipse are thus in agreement with the approximate expressions obtained for the plane lamina with a convex boundary.

8. THE SCATTERED INTENSITY FOR LARGE h

For a randomly oriented plane lamina, the scattered intensity can be expressed[15]

$$I(h) = \frac{2\pi A\rho^2}{h^2}[1 - T(h)], \tag{62}$$

where

$$T(h) = \frac{1}{\bar{M}h} \int_0^{D\max} G(M) \sin hM\,dM. \tag{63}$$

[Equations (62) and (63) use a different notation from that employed in Ref. 15. The quantity $G(r)$ in Ref. 15 is $\beta_0(r)$ in Eq. (7), and the variable of integration in (63) has been changed to M. Also, the quantity σ in Ref. 15 corresponds to ρ in (62).]

According to (62) and (63), the form of the intensity for large h is determined by the asymptotic form of the Fourier integral (63). For large h, $T(h)$ can be approximated by use of Erdélyi's theorem for asymptotic

[15] Reference 6, Eq. (9).

expansion of Fourier integrals[10] and the theorem of Jones and Kline[16] for asymptotic expansion of Fourier integrals with logarithmic discontinuities.

Erdélyi's theorem states that if a function $\phi(t)$ is N times continuously differentiable for $\alpha \leq t \leq \beta$, and $0 < \lambda \leq 1$, $0 < \mu \leq 1$, then

$$\int_\alpha^\beta e^{ixt}(t-\alpha)^{\lambda-1}(\beta-t)^{\mu-1}\phi(t)\,dt = B_N(x) - A_N(x) + O(x^{-N})$$

as $x \to \infty$, where

$$A_N(x) = -\sum_{n=0}^{N-1} \frac{\Gamma(n+\lambda)}{n!} e^{\frac{1}{2}i\pi(n+\lambda)} x^{-(n+\lambda)} \times e^{ix\alpha} \frac{d^n}{d\alpha^n}[(\beta-\alpha)^{\mu-1}\phi(\alpha)],$$

$$B_N(x) = \sum_{n=0}^{N-1} \frac{\Gamma(n+\mu)}{n!} e^{\frac{1}{2}i\pi(n-\mu)} x^{-(n+\mu)} \times e^{ix\beta} \frac{d^n}{d\beta^n}[(\beta-\alpha)^{\lambda-1}\phi(\beta)],$$

and where $O(x^{-N})$ may be replaced by $o(x^{-N})$ if $\lambda = \mu = 1$. [The order symbols $O(y)$ and $o(y)$ are explained in Ref. 10, p. 5.]

According to Erdélyi's theorem, (26) makes no contribution to the asymptotic expansion of $T(h)$. The first terms in the approximate expression for $G(M)$ for small M thus do not affect the scattered intensity for a plane lamina with a convex boundary.

The asymptotic expansion of $T(h)$ will consist of a sum of terms corresponding to the values of M at which $G(M)$ is discontinuous. If there is a maximum at $M = D_i$, according to Erdélyi's theorem and (40), there will be a term

$$-\frac{2}{A}\left[\frac{R_{1_i}R_{2_i}}{D_i(D_i - R_{1_i} - R_{2_i})}\right]^{\frac{1}{2}} \frac{\cos hD_i}{h^2} \tag{64}$$

in $T(h)$, where R_{1_i} and R_{2_i} are the values of R_1 and R_2 corresponding to the maximum value D_i. The terms in $T(h)$ resulting from saddle points at $M = D_i$, which can be evaluated from (41) and the theorem of Jones and Kline,[16] are found to be

$$+\frac{2}{A}\left[\frac{R_{1_i}R_{2_i}}{D_i(R_{1_i} + R_{2_i} - D_i)}\right]^{\frac{1}{2}} \frac{\sin hD_i}{h^2}. \tag{65}$$

Erdélyi's theorem can be employed to find the terms in $T(h)$ resulting from double points. When the discontinuity is given by (49), the corresponding term in $T(h)$ has the form

$$\frac{2(R_1R_2)^{\frac{1}{2}}\Gamma(\frac{1}{3})}{A(\pi D)^{\frac{1}{2}}6^{\frac{1}{3}}(|R_1R_1' + R_2R_2'|)^{\frac{1}{3}}} \frac{\sin(hD - \pi/4)}{h^{\frac{1}{6}}}. \tag{66}$$

[16] Reference 14, p. 27.

For a discontinuity given by (54), there will be a term

$$\frac{(R_1R_2)^{\frac{1}{2}}6^{\frac{1}{4}}\Gamma(\frac{1}{4})}{A(\pi D)^{\frac{1}{2}}(|R_1^2R_1'' + R_2^2R_2''|)^{\frac{1}{4}}} \frac{\sin\left(hD - \frac{\pi}{8} - \nu\right)}{h^{\frac{7}{4}}} \tag{67}$$

in $T(h)$, where

$$\nu = \pi/4, \quad R_1^2R_1'' + R_2^2R_2'' < 0,$$
$$\nu = 0, \quad R_1^2R_1'' + R_2^2R_2'' > 0.$$

From (62), (63), (57), and (58), the asymptotic expression for the scattered intensity from an elliptical lamina is

$$I(h) \approx \frac{2\pi^2a^2v\rho^2}{h^2}\left[1 - \frac{1}{\pi(v^2-1)^{\frac{1}{2}}} \times \frac{(v^3\sin 2ha - \cos 2hva)}{(hva)^2} + \cdots\right].$$

9. DISCUSSION

Equations (62) and (63) show that for a plane lamina with a smooth convex boundary, the form of the scattered intensity for large h is determined by the asymptotic expansion of the Fourier integral $T(h)$. The terms in this expansion can be calculated when the form of $G(M)$ is known for small M and in the neighborhood of points where $G(M)$ or its derivatives are discontinuous. The first step in finding the asymptotic expansion of $T(h)$ thus is an examination of the function $M(t_1, t_2)$ to determine its maxima, saddle points, and double points. After these points have been located, the radii of curvature can be evaluated and substituted in (64)–(67), which give the least rapidly vanishing terms in the asymptotic expansion of $T(h)$.

More terms in the expansion could be obtained by developing a higher-order approximation for $G(M)$ in the neighborhood of its discontinuities. This calculation, however, would almost certainly be quite complicated.

For every boundary curve, $M(t_1, t_2)$ can be expected to have maxima and saddle points, and so terms with the form of (64) and (65) should always appear in $T(h)$. In the less frequent cases when there are double points, their contribution to $T(h)$ can be obtained from (66) and (67).

As has been mentioned previously, the terms given in (26) in the approximation for $G(M)$ for small M do not contribute to the asymptotic expansion of $T(h)$. In a higher-order approximation than (26), only coefficients of even powers of M will appear in the asymptotic expansion of $T(h)$. As (26) contains only odd powers of M, very possibly there will be no even

powers of M in a higher-order approximation of $G(M)$ for small M, and consequently the form of $G(M)$ for small M may have no effect on the asymptotic expansion of $T(h)$.

It is interesting that this result is quite similar to the behavior of the asymptotic expansion of the scattered intensity from filaments.[17] Calculations strongly suggest that, for a filament in the form of a loop without bends or points where the filament crosses itself, the analog of the asymptotic expansion of $T(h)$ will contain no nonoscillatory terms.

While the possibility of unforeseen complications cannot be excluded, the methods used for obtaining (26) appear to be applicable to finding higher-order approximations of $G(M)$ for small M.

The coefficients of the terms in (26) are relatively simple, while considerably more complicated expressions appear in intermediate steps of the development of (26). This simplification of the final result, after a fairly long and complex calculation, quite strongly suggests that there must be a simpler way to approximate $G(M)$ for small M. However, our efforts to simplify the calculation have been unsuccessful.

Kirste and Porod[18] have found $\beta_0(r)$ for small r for polygons. Their expression for $\beta_0(r)$ indicates that for a polygon the first term in the approximation of $G(M)$ for small M will be a constant instead of being proportional to M, as in (26). Consequently, for polygons Erdélyi's theorem shows that $T(h)$ will contain a term proportional to h^{-2}, which arises from the behavior of $G(M)$ for small M. Since the boundary of a polygon has sharp corners and thus is not smooth, the calculation of Kirste and Porod suggests that, in order for $T(h)$ to be unaffected by the behavior of $G(M)$ for small M, the boundary must be smooth and cannot have sharp corners. In principle, the presence of corners on the boundary thus could be detected by measurements of the intensity of scattering for large h.

When the boundary is a circle, the procedure used for finding θ as a function of M when M is in the neighborhood of a maximum, a saddle point, or a double point cannot be used, and at $M = D$ the discontinuity in $G(M)$ for a circle is different from the discontinuities for maxima, saddle points, and double points. An exact calculation[11] shows that when M is in the neighborhood of the diameter D of the circle, $G(M)$ is proportional to $(D - M)^{-\frac{1}{2}}$ for $M < D$, and $G(M) = 0$ for $M > D$. Thus $G(M)$ has a sharper discontinuity for a circle than for the other cases which have been considered.

For a double point the boundary can be considered to approach a circular boundary more closely than for a saddle point or a maximum. Moreover, the boundary for a double point for which $R_1' = R_2' = 0$ is more nearly like a circle than the boundary for the case of a double point for which both of these quantities are not zero. The degree of similarity to a circular boundary for these two types of double points is reflected in the exponents of $|D - M|$ in $G_d(M)$. Thus, in (54) the exponent of $|D - M|$ is closer to the value for a circle than in (49), and the boundary approaches a circle more closely in the former case than in the latter.

The results of this investigation of $G(M)$ probably have little direct application to the analysis of experimental data, since it is not possible to prepare a suspension of colloidal particles which are identical thin platelets with a convex boundary. Nevertheless, the information about $G(M)$ which has been obtained has some important consequences for the theory of small-angle x-ray scattering.

First, the expressions for the form of the intensity for large h indicate, in a qualitative way, the type of information that can be obtained from measurements of the outer part of the small-angle x-ray scattering curve for any type of suspension of identical particles. In the expression for the intensity at large h, there will be damped oscillatory terms resulting from the M values at which $G(M)$ or its derivatives are discontinuous. In addition, the intensity will contain nonoscillatory terms proportional to negative powers of h, and from the coefficients of these terms, information can be obtained about some properties of the particles. Although for a plane lamina with a smooth convex boundary the asymptotic expression for the intensity at large h will probably contain only one nonoscillatory term, for three-dimensional particles[19] more nonoscillatory terms will appear in the expression for the intensity at large h and therefore more information will be obtainable in a study of the outer part of the scattering curve from three-dimensional particles than from plane laminas.

Second, the approximate expressions for $G(M)$ can be used in previously developed relations[7-9] giving the scattered intensity for cylinders with arbitrary cross section. In these expressions, the intensity from the three-dimensional cylinder was given in terms of $\beta_0(r)$, the two-dimensional characteristic function of the cross section. When these calculations were made, sufficient information about $\beta_0(r)$ was not available to permit complete evaluation of the expressions for the scattering from the cylinders. The properties of $G(M)$ which have now been obtained

[17] Reference 5, pp. 30–31.

[18] R. Kirste and G. Porod, Kolloid-Z. **184**, 4, (1962).

[19] Reference 6, Eq. (2).

give enough information to evaluate the first terms in the asymptotic expansions for scattering from three-dimensional cylinders. These expressions can now be used for analysis of experimental data.

Finally, the methods developed for studying $G(M)$ for a plane lamina will be useful in finding procedures for attacking the more important, but also more complex, calculation of $G(M)$ for a three-dimensional particle with a smooth convex boundary. For example, preliminary investigations indicate that some of the results for a plane lamina can be directly used as a starting point for approximations of $G(M)$ for a smooth convex three-dimensional particle.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Louise Heine for typing the manuscript and to T. S. Storvick and the University of Missouri Fluid and Transport Properties Research Center, supported by the Department of Defense under Project Themis, for providing funds for some of the costs of preparation of the manuscript.

APPENDIX A: APPROXIMATE SOLUTION OF AN EQUATION

The calculation of $G(M)$ for small M requires an expression giving θ as a function of M. This expression is obtained by solving Eq. (15) for θ. This equation can be put in the form

$$x^2 = 2g(x) + c, \tag{A1}$$

where c is a constant and $g(x)$ is a monotonically increasing function of x such that $g(0) = 0, g''(0) = 0$, and $g'(0) > 0$. The assumption is made that $g(x)$ can be expanded in a power series

$$g(x) = \sum_{n=0}^{\infty} g_n x^{n+1}, \tag{A2}$$

which converges for all x values needed for the solution of the equation. If (A1) is to have a single root at $x = x_m$, the curve $y = x^2$ must be tangent to the curve $y = 2g(x) + c$ at $x = x_m$. The condition defining x_m then is

$$x_m = g'(x_m), \tag{A3}$$

and c must satisfy the condition

$$c = c_m = x_m^2 - 2g(x_m).$$

If $c > c_m$, the equation can have two real roots.

Equation (A1) can be written

$$(x - x_m)^2 = \frac{1 - B(x_m)}{1 - B(x)} A^2, \tag{A4}$$

where

$$B(x) = \frac{2g(x) - g(x_m) - g'(x_m)(x - x_m)}{(x - x_m)^2}, \tag{A5}$$

$$A^2 = \frac{c + 2g(x_m) - x_m^2}{1 - B(x_m)}. \tag{A6}$$

The expression for A^2 in (A6) is obtained by making use of (A3). The assumption is made that $B(x) \ll 1$. From (A5),

$$B(x_m) = g''(x_m). \tag{A7}$$

Equation (A4) can be written

$$x - x_m = \pm A\left[\frac{1 - B(x_m)}{1 - B(x)}\right]^{\frac{1}{2}}. \tag{A8}$$

Since $B(x)$ has been assumed to be small compared to 1, the quantity $[1 - B(x)]^{-\frac{1}{2}}$ in (A8) can be expanded in a power series. Then

$$x - x_m = \pm A \sum_{n=0}^{\infty} b_n (x - x_m)^n, \tag{A9}$$

where

$$b_n = \frac{[1 - B(x_m)]^{\frac{1}{2}}}{n!} \frac{d^n}{dx_m^n} [1 - B(x_m)]^{-\frac{1}{2}}.$$

Since $B(x) \ll 1$, and since, according to (A5), $B(x)$ can be expanded in a Taylor series about $x_{m'}$, the Taylor series expansion (A9) must converge.

Equation (A9) can be used to express $(x - x_m)$ as a sum of powers of $\pm A$. In this series, both A and the b_n depend on $x_{m'}$, and from (A3) x_m can be expressed in terms of the g_n in (A2). By this procedure, a solution of (A1) is obtained.

Let

$$b_n^k = \sum_{j=0}^{n+1} b_j b_{n+1-j}^{k-1}, \tag{A10}$$

for $k \geq 1$, and

$$b_0^0 = 1,$$
$$b_n^0 = 0, \quad n \geq 1.$$

Since $b_0 = 1$, (A9) can be written

$$x = x_m \pm A \pm \left[A(x - x_m) \sum_{n=0}^{\infty} b_n^1 (x - x_m)^n\right]$$
$$= x_m + \sum_{k=0}^{1} b_0^k (\pm A)^{k+1} + (\pm A)^2 (x - x_m) \sum_{n=0}^{\infty} b_n^2 (x - x_m)^n.$$

Similarly, by induction, for $j \geq 1$,

$$x = x_m + \left[\sum_{k=0}^{j-1} b_0^k (\pm A)^{k+1}\right] + A_j, \tag{A11}$$

where

$$A_j = (\pm A)^j \sum_{n=0}^{\infty} b_n^j (x - x_m)^{n+1}.$$

The quantity A_j in (A11) is proportional to $(\pm A)^{j+1}$ and thus is smaller than any of the terms in the sum in (A11). When A is small enough that (A11) is a convenient approximation for x, the value of j is chosen to be large enough that A_j can be neglected. The resulting approximate expression for x then has the form of a series in powers of $\pm A$.

From (A2) and (A3)

$$x_m = g_0 + \sum_{n=0}^{\infty}(n+3)g_{n+2}(x_m)^{n+2} \qquad \text{(A12)}$$

because the condition $g''(0) = 0$ implies that $g_1 = 0$. Since x_m will be assumed to be small, a rough approximation to the solution of (A3) is $x_m = g_0$. This approximate result suggests that the exact solution can be conveniently written

$$x_m = \sum_{n=0}^{\infty} a_n (g_0)^{n+1},$$

with $a_0 = 1$. Let

$$(x_m)^k = \sum_{n=0}^{\infty} a_n^k (g_0)^{n+k}. \qquad \text{(A13)}$$

Then $a_n^1 = a_n$ and

$$a_n^k = \sum_{j=0}^{n} a_{n-j}^{k-1} a_j. \qquad \text{(A14)}$$

When (A13) is substituted in (A12),

$$\sum_{n=0}^{\infty} a_n (g_0)^n = 1 + \sum_{n=0}^{\infty} (g_0)^{n+1} \sum_{j=0}^{n} (n+3-j) g_{n+2-j} a_j^{n+2-j}.$$

Thus

$$a_{n+1} = \sum_{j=0}^{n} (n+3-j) g_{n+2-j} a_j^{n+2-j}. \qquad \text{(A15)}$$

Equations (A14) and (A15) can be used for successive calculation of the a_n. After the a_j have been evaluated for $j \leq n$, a_{n+1} can be obtained from (A15). According to (A14), the a_n^k can be found when the a_j are known for $0 \leq j \leq n$.

Thus, for example,

$$a_1 = 3g_2,$$
$$a_2 = 4g_3 + 18(g_2)^2.$$

When g_0 and c are of the same order of magnitude, and when terms with magnitude less than $(g_0)^{\frac{3}{2}}$ can be neglected,

$$x = x_m \pm b_0^0 A + b_0^1 A^2 \pm b_0^2 A^3 + \cdots,$$
$$x_m \simeq g_0,$$
$$b_0^0 = 1, \quad b_0^1 = b_1 = g_2,$$
$$b_0^2 = b_2 + (b_1)^2$$
$$= \frac{1}{4}\frac{B''(x_m)}{1 - B(x_m)} + \frac{5}{8}\frac{[B'(x_m)]^2}{[1 - B(x_m)]^2}$$
$$\simeq g_3 + (\tfrac{5}{2})(g_2)^2,$$
$$A \simeq (c + g_0^2)^{\frac{1}{2}}[1 + 3g_2 g_0].$$

Thus

$$x \simeq g_0 \pm (1 + 3g_2 g_0)(c + g_0^2)^{\frac{1}{2}} + cg_2 \pm [g_3 + (\tfrac{5}{2})g_2^2]c^{\frac{3}{2}}. \qquad \text{(A16)}$$

APPENDIX B: THE SCATTERED INTENSITY FOR AN ELLIPSE

The average structure factor $\overline{F^2(h)}$ for a randomly oriented elliptical lamina with uniform electron density ρ, semimajor axis va, and semiminor axis a is given by the expression[20]

$$\overline{F^2(h)} = \frac{1}{4\pi}\int_0^{2\pi} d\phi \int_0^{\pi} d\alpha \sin\alpha [A(\mathbf{h})]^2, \qquad \text{(B1)}$$

where

$$\mathbf{h} = [\mathbf{i}\sin\alpha\cos\phi + \mathbf{j}\sin\alpha\sin\phi + \mathbf{k}\cos\alpha]h,$$
$$A(\mathbf{h}) = \rho\int_{-a}^{a} dx \int_{-v(a^2-x^2)^{\frac{1}{2}}}^{v(a^2-x^2)^{\frac{1}{2}}} dy\, e^{i(\mathbf{h}\cdot\mathbf{r})}, \qquad \text{(B2)}$$
$$\mathbf{r} = x\mathbf{i} + y\mathbf{j},$$

and where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in a Cartesian coordinate system with the x and y axes coinciding with the semiminor and semimajor axes of the ellipse, respectively, and with the z axis perpendicular to the plane of the ellipse. The quantity $A(\mathbf{h})$ is proportional to the scattering amplitude for an arbitrary orientation of the ellipse, and (B1) states that the intensity is the average of $[A(\mathbf{h})]^2$ over all orientations.

When the integration over y in (B2) is carried out and use is made of an integral relation between Bessel functions,[21] $A(\mathbf{h})$ can be expressed as

$$A(\mathbf{h}) = 2\pi\rho a^2 v \frac{J_1[ha\sin\alpha(\cos^2\phi + v^2\sin^2 a)^{\frac{1}{2}}]}{ha\sin\alpha(\cos^2\phi + v^2\sin^2\phi)^{\frac{1}{2}}},$$

where $J_1(x)$ is the first-order Bessel function of the first kind. Let

$$K(2x) = \tfrac{1}{2}\int_0^{\pi} d\alpha \sin\alpha \left[\frac{2J_1(x\sin\alpha)}{x\sin\alpha}\right]^2.$$

Then,[22]

$$K(2x) = \frac{8}{\pi x^2}\int_0^{\pi/2}\frac{d\alpha}{\sin\alpha}\int_0^{\pi/2} d\theta J_2(2x\sin\alpha\cos\theta).$$

Let

$$\theta = \cos^{-1}\left[\frac{u}{\sin\alpha}\right].$$

Then

$$K(2x) = \frac{8}{\pi x^2}\int_0^{\pi/2}\frac{d\alpha}{\sin\alpha}\int_0^{\sin\alpha}\frac{du J_2(2xu)}{(\sin^2\alpha - u^2)^{\frac{1}{2}}}$$
$$= \frac{8}{\pi x^2}\int_0^1 du J_2(2xu)\int_{\sin^{-1}u}^{\pi/2}\frac{d\alpha}{\sin\alpha(\sin^2\alpha - u^2)^{\frac{1}{2}}}.$$

[20] Reference 1, p. 10.
[21] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematics Physics* (Chelsea Publ. Co., New York, 1954), p. 30.
[22] Reference 21, p. 28.

Thus

$$K(2x) = \frac{8}{(2x)^2}\left[1 - \frac{2J_1(2x)}{2x}\right]. \quad \text{(B3)}$$

From (B1)

$$\overline{F^2(h)} = 2\pi\rho^2 a^4 v^2 \int_0^{\pi/2} d\phi K(w), \quad \text{(B4)}$$

where

$$w = 2ha(\cos^2\phi + v^2\sin^2\phi)^{\frac{1}{2}}.$$

By a change of the variable of integration,

$$\overline{F^2(h)} = 2\pi\rho^2 a^4 v^2 \int_{2a}^{2va} \frac{x\,dx\,K(hx)}{(x^2 - 4a^2)^{\frac{1}{2}}(4v^2a^2 - x^2)^{\frac{1}{2}}}. \quad \text{(B5)}$$

The two-dimensional characteristic function $\beta_0(r)$ will now be obtained. By use of an integral representation[23] for $J_1(x)$, $K(hx)$ can be written in the form

$$K(hx) = \frac{32}{\pi h^3 x^4} \int_0^x \frac{u\,du}{(x^2 - u^2)^{\frac{1}{2}}} [hu - \sin hu].$$

[23] Reference 21, p. 26.

By integration by parts,

$$K(hx) = \frac{32}{\pi h^2 x^4} \int_0^x du(x^2 - u^2)^{\frac{1}{2}}[1 - \cos hu]$$

$$= \frac{32}{\pi h x^4} \int_0^x du(x^2 - u^2)^{\frac{1}{2}} \int_0^u dr \sin hr.$$

Thus

$$K(hx) = \frac{16}{\pi^2 x^4} \int_0^x 2\pi r\,dr \frac{\sin hr}{hr} \int_r^x du(x^2 - u^2)^{\frac{1}{2}}. \quad \text{(B6)}$$

When (B6) is substituted in (B5) and the order of the x and r integrations are interchanged, $\overline{F^2(h)}$ can be written

$$\overline{F^2(h)} = \pi\rho^2 a^2 v \int_0^{2va} 2\pi r\,dr \frac{\sin hr}{hr} \beta_0(r),$$

where

$$\beta_0(r) = \frac{32a^2v}{\pi^2} \int_{x(r)}^{2av} \frac{dx}{x^3(x^2 - 4a^2)^{\frac{1}{2}}(4v^2a^2 - x^2)^{\frac{1}{2}}} \times \int_r^x du(x^2 - u^2)^{\frac{1}{2}} \quad \text{(B7)}$$

and

$$x(r) = 2a, \quad 0 \leq r \leq 2a,$$
$$x(r) = r, \quad 2a \leq r \leq 2va.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Transmission Properties of an Isotopically Disordered One-Dimensional Harmonic Crystal

ROBERT J. RUBIN
National Bureau of Standards, Washington, D.C.

(Received 1 March 1968)

The amplitude $\mathcal{C}_N(\omega)$ of a wave of frequency ω which is transmitted by a disordered array of N isotopic defects in a one-dimensional crystal has been investigated in the limit in which $N \to \infty$ while the over-all concentration of the defects in the array remains fixed. The transmitted amplitude $\mathcal{C}_N(\omega)$ is proportional to the reciprocal of the magnitude of an Nth-order determinant whose elements depend explicitly upon the spacings between defects, the incident frequency ω, and the relative mass difference $Q = (M - m)/m$ between the defect particles and the particles of the host crystal. $\mathcal{C}_N(\omega)$ is represented as $\exp[-N\alpha_N(\omega, Q, C)]$, where C is the over-all fractional concentration of defects; two types of estimates of $\alpha_N(\omega, Q, C)$ are obtained. First, assuming that the spacings between nearest-neighbor pairs of defects are independent random variables, upper and lower bounds are obtained on $\alpha_N(\omega, Q, C)$ which are independent of N. Provided that C is sufficiently small, the lower bound is positive. Second, Monte Carlo estimates of $\alpha_N(\omega, Q, C)$ are obtained in the cases $Q = 1$, $C = 0.1$ and $Q = 1$, $C = 0.5$, for arrays of 3×10^4 defects. These Monte Carlo estimates are compared with the previously obtained bounds. It is also shown that at the special frequencies of Matsuda and for $Q \geq Q_{\text{crit}}$, the limiting value of $\alpha_N(\omega, Q, C)$ is positive in the entire concentration range $0 < C < 1$. Explicit upper and lower bounds are obtained on $\alpha(\sin(\pi/4), 1, C)$.

1. INTRODUCTION

This paper is devoted to the study of the transmission characteristics of a disordered array of isotopic defects substituted in an otherwise perfect one-dimensional harmonic crystal. The transmitted amplitude of a wave of frequency ω incident on a segment of crystal containing N defects is shown to be the reciprocal of the magnitude of an Nth-order determinant. The determinantal expression for the transmitted amplitude is investigated in the limit in which the number of defects in the disordered region approaches infinity while the over-all concentration of defects

Thus

$$K(2x) = \frac{8}{(2x)^2}\left[1 - \frac{2J_1(2x)}{2x}\right]. \quad \text{(B3)}$$

From (B1)

$$\overline{F^2(h)} = 2\pi\rho^2 a^4 v^2 \int_0^{\pi/2} d\phi K(w), \quad \text{(B4)}$$

where

$$w = 2ha(\cos^2\phi + v^2\sin^2\phi)^{\frac{1}{2}}.$$

By a change of the variable of integration,

$$\overline{F^2(h)} = 2\pi\rho^2 a^4 v^2 \int_{2a}^{2va} \frac{x\,dx\,K(hx)}{(x^2 - 4a^2)^{\frac{1}{2}}(4v^2a^2 - x^2)^{\frac{1}{2}}}. \quad \text{(B5)}$$

The two-dimensional characteristic function $\beta_0(r)$ will now be obtained. By use of an integral representation[23] for $J_1(x)$, $K(hx)$ can be written in the form

$$K(hx) = \frac{32}{\pi h^3 x^4} \int_0^x \frac{u\,du}{(x^2 - u^2)^{\frac{1}{2}}} [hu - \sin hu].$$

[23] Reference 21, p. 26.

By integration by parts,

$$K(hx) = \frac{32}{\pi h^2 x^4} \int_0^x du(x^2 - u^2)^{\frac{1}{2}}[1 - \cos hu]$$

$$= \frac{32}{\pi h x^4} \int_0^x du(x^2 - u^2)^{\frac{1}{2}} \int_0^u dr \sin hr.$$

Thus

$$K(hx) = \frac{16}{\pi^2 x^4} \int_0^x 2\pi r\,dr \frac{\sin hr}{hr} \int_r^x du(x^2 - u^2)^{\frac{1}{2}}. \quad \text{(B6)}$$

When (B6) is substituted in (B5) and the order of the x and r integrations are interchanged, $\overline{F^2(h)}$ can be written

$$\overline{F^2(h)} = \pi\rho^2 a^2 v \int_0^{2va} 2\pi r\,dr \frac{\sin hr}{hr} \beta_0(r),$$

where

$$\beta_0(r) = \frac{32a^2v}{\pi^2} \int_{x(r)}^{2av} \frac{dx}{x^3(x^2 - 4a^2)^{\frac{1}{2}}(4v^2a^2 - x^2)^{\frac{1}{2}}} \times \int_r^x du(x^2 - u^2)^{\frac{1}{2}} \quad \text{(B7)}$$

and

$$x(r) = 2a, \quad 0 \leq r \leq 2a,$$
$$x(r) = r, \quad 2a \leq r \leq 2va.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Transmission Properties of an Isotopically Disordered One-Dimensional Harmonic Crystal

ROBERT J. RUBIN
National Bureau of Standards, Washington, D.C.

(Received 1 March 1968)

The amplitude $\mathcal{C}_N(\omega)$ of a wave of frequency ω which is transmitted by a disordered array of N isotopic defects in a one-dimensional crystal has been investigated in the limit in which $N \to \infty$ while the over-all concentration of the defects in the array remains fixed. The transmitted amplitude $\mathcal{C}_N(\omega)$ is proportional to the reciprocal of the magnitude of an Nth-order determinant whose elements depend explicitly upon the spacings between defects, the incident frequency ω, and the relative mass difference $Q = (M - m)/m$ between the defect particles and the particles of the host crystal. $\mathcal{C}_N(\omega)$ is represented as $\exp[-N\alpha_N(\omega, Q, C)]$, where C is the over-all fractional concentration of defects; two types of estimates of $\alpha_N(\omega, Q, C)$ are obtained. First, assuming that the spacings between nearest-neighbor pairs of defects are independent random variables, upper and lower bounds are obtained on $\alpha_N(\omega, Q, C)$ which are independent of N. Provided that C is sufficiently small, the lower bound is positive. Second, Monte Carlo estimates of $\alpha_N(\omega, Q, C)$ are obtained in the cases $Q = 1$, $C = 0.1$ and $Q = 1$, $C = 0.5$, for arrays of 3×10^4 defects. These Monte Carlo estimates are compared with the previously obtained bounds. It is also shown that at the special frequencies of Matsuda and for $Q \geq Q_{\text{crit}}$, the limiting value of $\alpha_N(\omega, Q, C)$ is positive in the entire concentration range $0 < C < 1$. Explicit upper and lower bounds are obtained on $\alpha(\sin(\pi/4), 1, C)$.

1. INTRODUCTION

This paper is devoted to the study of the transmission characteristics of a disordered array of isotopic defects substituted in an otherwise perfect one-dimensional harmonic crystal. The transmitted amplitude of a wave of frequency ω incident on a segment of crystal containing N defects is shown to be the reciprocal of the magnitude of an Nth-order determinant. The determinantal expression for the transmitted amplitude is investigated in the limit in which the number of defects in the disordered region approaches infinity while the over-all concentration of defects

remains fixed. The nearest-neighbor spacings between defects are assumed to be independent, identically distributed random variables. Bounds are then obtained on the logarithm of the Nth root of the transmitted amplitude, a quantity which is the average attenuation of the incident wave per defect. In Sec. 7, Monte Carlo estimates of this attenuation constant are compared with the attenuation constant of an array of independently scattering defects.

From an abstract point of view, the problem investigated in this paper is equivalent to the one-dimensional problem of wave propagation in stochastic media[1] and to the problem of the transmission characteristics of statistically inhomogeneous waveguides.[2-4] In all of these problems a functional equation for the transmission or reflection coefficient plays a central role. The functional equation is an "addition" law relating the coefficient for a combination of two inhomogeneities to the coefficients for the separate inhomogeneities. The functional equation for the reflection coefficient is well known in transmission-line theory and network theory.[5-8] Bellman and Kalaba[1] propose two different methods for investigating the properties of the reflection coefficient which are based on its "addition" law: The first is a Monte Carlo method; the second is a method based on the derivation of a functional equation for the distribution function of values of the reflection coefficient. Using essentially the second method, Gertsenshtein and Vasil'ev[2,3] have given an elegant solution of the inhomogeneous waveguide problem in a Brownian-motion limit by assuming that the reflection coefficient of individual inhomogeneities is infinitesimal and that their number is infinite. In this paper, we obtain Monte Carlo estimates of the transmission coefficient of an isotopically disordered crystal and obtain conditions sufficient to ensure that the transmission coefficient is an exponentially decreasing function of the number of defects. In a future publication we will treat the problem of deriving and numerically solving the functional equation for the distribution function of values of the transmitted amplitude. This functional equation is an analog of the functional equations obtained by Dyson,[9] Schmidt,[10] and Dean[11] in determining the frequency distribution function of isotopically disordered crystals.

The problem treated in this paper is an example of a large class of problems concerned with multiple scattering by an array of elementary scatterers in one, two, and three dimensions. There is an enormous literature on this subject.[12] Recently Kay and Silverman[13] and Bazer[14] have considered one-dimensional multiple-scattering problems which are similar to the one considered in this paper. In both investigations[13,14] the Neumann series solution of the integral equation for the scattering problem plays a central role. Kay and Silverman[13] noted that the radius of absolute convergence of their series for the transmission (or reflection) coefficient was proportional to N^{-1}, the reciprocal of the number of elementary scatterers. In this paper, by dealing directly with the exact determinantal expression for the transmission coefficient, we are able to treat the limit $N \to \infty$.

2. MODEL AND FORMAL SOLUTION

We consider an infinite, one-dimensional harmonic crystal with nearest-neighbor interactions. The particles are labeled consecutively by the index r, $-\infty < r < \infty$; and all particles have the mass m, except for N isotopic defect particles at random lattice positions $r = A_j, j = 1, \cdots, N$. The mass of each of the defect particles is M. It is assumed that $A_1 = 0$, that all other defect particles lie to the right of $r = 0$, and that the subscript j on A_j specifies the order of the defects, i.e., $0 = A_1 < \cdots < A_j < \cdots < A_N$. The spacings between adjacent pairs of defects are assumed to be independent, identically distributed random variables. The nearest-neighbor force constant is assumed to be equal to f everywhere in the crystal. The transmission coefficient of the array of N defects is determined by solving the following initial-value problem: A semi-infinite wave of frequency ω and unit amplitude is

[1] R. Bellman and R. Kalaba in *Electromagnetic Wave Propagation*, M. Desirant and J. L. Michiels, Eds. (Academic Press Inc., New York, 1960), p. 243.

[2] M. E. Gertsenshtein and V. B. Vasil'ev, Radiotekhn. i Elektron. **4**, 611 (1959) [English transl.: Radio Eng. Electron. USSR **4**, 74 (1959)].

[3] M. E. Gertsenshtein and V. B. Vasil'ev, Teoriya Veroyatnostei Primeneniya **4**, 424 (1959) [English transl.: Theory Probability Appl. USSR **4**, 391 (1959)].

[4] F. I. Karpelevich, V. N. Tutubalin, and M. G. Shur, Teoriya Veroyatnostei Primeneniya **4**, 432 (1959) [English transl.: Theory Probability Appl. USSR **4**, 399 (1959)].

[5] R. M. Redheffer, *Technique of Microwave Measurements*, C. G. Montgomery, Ed. (Massachusetts Institute of Technology Radiation Laboratory Series, McGraw-Hill Book Co., New York, 1947), Vol. 11, Chap. 10.

[6] R. M. Redheffer, J. Math. & Phys. **28**, 237 (1950).

[7] R. Redheffer, J. Math. & Phys. **41**, 1 (1962).

[8] N. G. Parke, J. Math. & Phys. **28**, 131 (1949).

[9] F. J. Dyson, Phys. Rev. **92**, 1331 (1953).

[10] H. Schmidt, Phys. Rev. **105**, 425 (1957).

[11] P. Dean, Proc. Phys. Soc. **73**, 413 (1959).

[12] We list several papers, which contain extensive bibliographies or compare classes of approximations: M. Lax, Rev. Mod. Phys. **23**, 287 (1951) and Phys. Rev. **88**, 621 (1952); V. Twersky, J. Res. Nat. Bur. Std. **D64**, 715 (1960) and **D68**, 500 (1964); J. B. Keller, Proc. Symp. Appl. Math. Am. Math. Soc. **13**, 227 (1960) and **16**, 145 (1964); U. Frisch, Ann. Astrophys. **29**, 645 (1966) and **30**, 565 (1967); P. Lloyd, Proc. Phys. Soc. (London) **91**, 678 (1967).

[13] I. Kay and R. A. Silverman, Nuovo Cimento Suppl. **9**, 626 (1958).

[14] J. Bazer, J. Soc. Ind. Appl. Math. **12**, 539 (1964).

incident from the left on the array of defects; the asymptotic value (in time) of particle amplitudes to the right of the defect at A_N is, by definition, the transmission coefficient of the array.

In this section we obtain the exact formal solution to the foregoing time-dependent problem. The solution could also be obtained by treating the time-independent problem and using the method of transfer matrices.[5-8,10,15] However, our method of solution of the time-dependent problem is a simple generalization of that used in a one-defect model[16] and is also applicable to the analogous two- and three-dimensional transmission (scattering) problems.[17] The equations of motion of the one-dimensional crystal are

$$(m_r/m)x_{\tau\tau}(r,\tau) = \tfrac{1}{4}[x(r-1,\tau) - 2x(r,\tau) + x(r+1,\tau)], \quad -\infty < r < \infty, \quad (1)$$

where $x(r,\tau)$ is the displacement of particle r from its equilibrium position and m_r is the mass of the particle at lattice site r. In Eq. (1), τ is a dimensionless time, $\tau = 2(f/m)^{\frac{1}{2}}t$, and each subscript τ denotes differentiation with respect to τ. The initial values of the particle positions and velocities corresponding to the initial condition in which a semi-infinite wave of frequency ω moves from left to right with its front initially at lattice site $-R < 0$ are the real parts of

$$x(r,0) = \begin{cases} 0, & r > -R, \\ e^{-ikr}, & r \le -R, \end{cases} \quad \text{where} \quad R > 0,$$
$$x_\tau(r,0) = \begin{cases} 0, & r > -R, \\ i\omega e^{-ikr}, & r \le -R. \end{cases} \quad (2)$$

The relation between frequency ω and wavenumber k is the perfect lattice expression

$$\omega = \sin(k/2). \quad (3)$$

As a first step in obtaining the solution of (1) for the initial conditions (2), we determine the solution of (1) for the initial condition in which only particle n is in motion and all particles are at their equilibrium positions:

$$x(r,0) = 0, \quad \text{all } r,$$
$$x_\tau(r,0) = \begin{cases} 1, & r = n, \\ 0, & r \ne n, \end{cases} \quad \text{for} \quad n \ne A_j,\ j = 1, \cdots, N. \quad (4)$$

[15] H. A. Kramers, Physica **2**, 483 (1935); H. M. James, Phys. Rev. **76**, 1602 (1949); J. M. Luttinger, Philips Res. Rep. **6**, 303 (1951); H. M. James and A. S. Ginzbarg, J. Phys. Chem. **57**, 840 (1953); E. H. Kerner, Proc. Phys. Soc. (London) **69**, 234 (1956); J. Hori and T. Asahi, Progr. Theoret. Phys. (Kyoto) **17**, 523 (1957); H. Matsuda, Progr. Theoret. Phys. (Kyoto) Suppl. **23**, 22 (1962).

[16] R. J. Rubin, J. Math. Phys. **1**, 309 (1960).

[17] Although the method of solution is applicable in the two- and three-dimensional problems, the formal result cannot be simplified significantly.

Introduce the generating function

$$G(\phi,\tau) = \sum_{r=-\infty}^{\infty} x(r,\tau)e^{i\phi r} \quad \text{with} \quad -\pi \le \phi \le \pi, \quad (5)$$

by multiplying the equation of motion for the rth particle, Eq. (1), by $e^{i\phi r}$ and summing the entire set of equations with respect to r, one obtains the following second-order inhomogeneous differential equation:

$$G_{\tau\tau}(\phi,\tau) + Q\sum_{j=1}^{N} x_{\tau\tau}(A_j,\tau)e^{i\phi A_j} = -\tfrac{1}{2}(1-\cos\phi)G(\phi,\tau), \quad (6)$$

where $Q = (M-m)/m$. Now take the Laplace transform of Eq. (6) to obtain

$$P^2\Gamma(\phi,P) - G_\tau(\phi,0) - PG(\phi,0) + P^2Q\sum_{j=1}^{N}\xi(A_j,P)e^{i\phi A_j} = -\tfrac{1}{2}(1-\cos\phi)\Gamma(\phi,P), \quad (7)$$

where

$$\Gamma(\phi,P) = \int_0^\infty e^{-P\tau}G(\phi,\tau)\,d\tau,$$
$$\xi(A,P) = \int_0^\infty e^{-P\tau}x(A,\tau)\,d\tau,$$

and where, for the initial condition (4),

$$G(\phi,0) = 0$$

and

$$G_\tau(\phi,0) = e^{i\phi n}.$$

Solving Eq. (7) for $\Gamma(\phi,P)$, one obtains the following implicit equation for the N unknown quantities $\xi(A_j,P)$:

$$\Gamma(\phi,P) = \frac{e^{i\phi n} - QP^2\sum_{j=1}^{N}\xi(A_j,P)e^{i\phi A_j}}{P^2 + \tfrac{1}{2}(1-\cos\phi)}. \quad (8)$$

Multiplying Eq. (8) by the factor $(2\pi)^{-1}e^{-i\phi A_k}$ and integrating with respect to ϕ between the limits $-\pi$ and π, one can obtain N independent linear equations involving the unknown $\xi(A_k,P)$'s:

$$\xi(A_k,P) + QP^2\sum_{j=1}^{N}\zeta(A_j - A_k,P)\xi(A_j,P) = \zeta(n-A_k,P), \quad k = 1,\cdots,N, \quad (9)$$

where

$$\zeta(A,P) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\phi A}}{P^2 + \tfrac{1}{2}(1-\cos\phi)}\,d\phi = P^{-1}(P^2+1)^{-\frac{1}{2}}[P + (P^2+1)^{\frac{1}{2}}]^{-2|A|}. \quad (10)$$

The solution of Eqs. (9) is

$$\xi(A_j,P) = P^{-1}(P^2+1)^{-\frac{1}{2}}D_N^{(j)}(P)/D_N(P), \quad j = 1,\cdots,N, \quad (11)$$

where $D_N^{(j)}(P)$ and $D_N(P)$ are $N \times N$ determinants. The denominator $D_N(P)$ is the determinant of the matrix of the coefficients of the $\xi(A_k, P)$'s in Eq. (9), which we denote by

$$D_N(P) = d(\delta_{r,s} + QP(P^2+1)^{-\frac{1}{2}}E^{2|A_r - A_s|}), \quad (12)$$

where the argument of $d(\ )$ is the (r, s) element of the coefficient matrix, where $E = [P + (P^2+1)^{\frac{1}{2}}]^{-1}$, and where $\delta_{r,s}$ is the Kronecker delta. The determinant $D_N^{(j)}(P)$ is formed from $D_N(P)$ by replacing the jth column of $D_N(P)$ by the column vector

$$(E^{2|n-A_1|}, E^{2|n-A_2|}, \cdots, E^{2|n-A_N|}).$$

Having obtained an explicit expression for $\xi(A_j, P)$, Eq. (11), we substitute it in Eq. (8) for $\Gamma(\phi, P)$ and then extract from $\Gamma(\phi, P)$ the following expression for $\xi(S, P)$:

$$\xi(S, P) = P^{-1}(P^2+1)^{-\frac{1}{2}}E^{2|n-S|} - QP(P^2+1)^{-\frac{1}{2}}\sum_{j=1}^{N} E^{2|A_j - S|}D_N^{(j)}(P)/D_N(P). \quad (13)$$

In Appendix A it is shown that Eq. (13) can be greatly simplified. For example, when $n < 0$ and $S > A_N$, one obtains

$$\xi(S, P) = P^{-1}(P^2+1)^{-\frac{1}{2}}E^{2(S+|n|)}/D_N(P), \quad n < 0 \text{ and } S > A_N. \quad (14)$$

There is also some simplification of (13) in case particle S lies inside the array of defects, i.e., $A_1 < S < A_N$ and $S \neq A_j, j = 2, \cdots, N-1$ [see Eq. (A10)]. The expression for $\xi(S, P)$ in Eq. (14) is the Laplace transform of the position of particle S corresponding to the initial condition (4), in which each particle is at its equilibrium position and all particles are at rest except particle n. It can be shown in an identical way that the corresponding expression for $\xi(S, P)$ in the case of the initial condition in which only particle n is displaced and all particles are at rest,

$$x(r, 0) = \begin{cases} 1, & r = n \text{ and } n \neq A_j, \ j = 1, \cdots, N, \\ 0, & r \neq n, \end{cases}$$
$$x_\tau(r, 0) = 0, \quad \text{all } r, \quad (15)$$

merely differs by a factor P from (14), namely,

$$\xi(S, P) = (P^2+1)^{-\frac{1}{2}}E^{2(S+|n|)}/D_N(P), \quad n < 0 \text{ and } S > A_N. \quad (16)$$

From the results in (14) and (16), the Laplace transform of the displacement of particle S in the case of the semi-infinite incident wave (2) can be written as

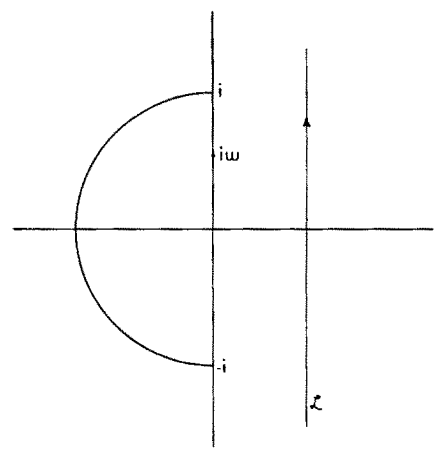


FIG. 1. Path of integration in the complex P plane and the semicircular cut connecting $+i$, $-i$.

$$\xi(S, P) = \frac{i\omega + P}{P(P^2+1)^{\frac{1}{2}}}\frac{E^{2S}}{D_N(P)}\sum_{r=-\infty}^{-R} e^{-ikr}E^{2|r|} = \frac{i\omega + P}{P(P^2+1)^{\frac{1}{2}}}\left(\frac{E^{2R+2S}}{D_N(P)}\right)\left(\frac{e^{ikR}}{1 - E^2e^{ik}}\right). \quad (17)$$

The amplitude of particle S at time τ is determined from (17),

$$x(S, \tau) = \text{Re}\left\{\frac{1}{2\pi i}\int_{\mathfrak{L}} e^{P\tau}\frac{i\omega + P}{P(P^2+1)^{\frac{1}{2}}} \times \left(\frac{E^{2R+2S}}{D_N(P)}\right)\frac{e^{ikR}}{1 - E^2e^{ik}}\,dP\right\}, \quad (18)$$

where Re $\{\ \}$ denotes the real part of the quantity in braces, and $\mathfrak{L}$, the path of integration, is parallel to and to the right of the imaginary P axis and is shown in Fig. 1. Except for poles on the imaginary P axis or in the left half-plane, the integrand in Eq. (18) is analytic off the semicircular cut connecting i and $-i$ in Fig. 1. The poles on the imaginary P axis, which determine the asymptotic time behavior of the integral in (18), are of one or possibly two kinds. The important pole for the present problem arises from a zero of $1 - E^2e^{ik}$ at $P = i\omega = i \sin(k/2)$ and represents the persistent action of the incident wave. Other poles may be present on the imaginary P axis outside the interval $(-i, i)$ in case the defects are light, i.e., $M < m$. These other poles, which are zeros of $D_N(P)$, correspond to the frequencies of localized modes. The number and position of these poles depends upon the number of defects, the details of the configuration of defects, and the value of the relative mass[18] $M < m$. The contribution of this second type of pole to the

[18] M. D. Bacon, P. Dean, and J. L. Martin, Proc. Phys. Soc. (London) **80**, 174 (1962).

asymptotic time behavior of $x(S, \tau)$ can be made arbitrarily small by taking R sufficiently large (R is the number of lattice sites between the initial position of the wavefront and particle zero). The foregoing assertion is based on the fact that the localized-mode particle amplitudes must decrease exponentially with an increase in the distance from the region containing the defects because the localized-mode frequencies are greater than the maximum frequency which can be propagated in the host crystal.

It can be verified that for the semicircular cut shown in Fig. 1 there is only one zero of $1 - E^2e^{ik} = (1 - Ee^{ik/2})(1 + Ee^{ik/2})$ on the imaginary P axis, and that zero is a zero of

$$1 - E^{ik/2} = 1 - \frac{e^{ik/2}}{P + (P^2 + 1)^{\frac{1}{2}}} = 0, \qquad (19)$$

namely,

$$P = i \sin (k/2) = i\omega. \qquad (20)$$

It follows from the above discussion that the asymptotic time dependence of the contour integral in Eq. (18) is given by the residue of the integrand at $P = i\omega$:

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{\mathcal{L}} e^{P\tau}\xi(S, P)\, dP \\
&\sim \left[e^{P\tau}\frac{i\omega + P}{P(P^2 + 1)^{\frac{1}{2}}}\left(\frac{E^{2R+2S}}{D_N(P)}\right)\right. \\
&\quad \left. \times \frac{e^{ikR}}{[(d/dP)(1 - Ee^{ik/2})](1 + E^{ik/2})}\right]_{P=i\omega} \\
&\sim e^{i\omega\tau - ikS}/d(\delta_{r,s} + iQ\omega(1 - \omega^2)^{-\frac{1}{2}}e^{-ik|A_r - A_s|}). \qquad (21)
\end{aligned}$$

In obtaining Eq. (21), we have used Eqs. (12), (19), and (20). The reciprocal of the determinant in Eq. (21) can be written as $\mathcal{C}_N(\omega)e^{i\psi_N}$, where

$$\mathcal{C}_N(\omega) = |d(\delta_{r,s} + iQ\omega(1 - \omega^2)^{-\frac{1}{2}}e^{-ik|A_r - A_s|})|^{-1}. \qquad (22)$$

Thus we have

$$\frac{1}{2\pi i}\int_{\mathcal{L}} e^{P\tau}\xi(S, P)\, dp \sim \mathcal{C}_N(\omega)e^{i(\omega\tau - kS + \psi_N)}.$$

Since the amplitude of the incident wave is one, we define $\mathcal{C}_N(\omega)$ as the transmission coefficient of the array of N defects at the frequency ω. The phase ψ_N is the phase shift of the array of defects. As $M \to m$, $Q \to 0$, and $D_N(i\omega) \to 1$ [see Eq. (12)]. Thus $\mathcal{C}_N(\omega) \to 1$ and $\psi_N \to 0$.

Equation (22) is an explicit formal expression for the transmission coefficient $\mathcal{C}_N(\omega)$. In Secs. 3 and 4 we obtain bounds on the limiting form of $\mathcal{C}_N(\omega)$ as N approaches infinity in such a way that the fractional concentration of defects $C = N/A_N$ remains constant and there is no correlation in the spacings between defects.

3. A GENERAL BOUND AND AN ESTIMATE FOR THE TRANSMISSION COEFFICIENT IN THE LIMIT $C \to 0$

The value of the transmission coefficient, Eq. (22), depends explicitly on the set of numbers $\{A_j\}$, $j = 1, \cdots, N$, which specifies the configuration of the array of defects. For periodic configurations of defects, $\mathcal{C}_N(\omega)$ exhibits band structure which is characteristic of the infinite periodic array.[19] Our principal problem is to estimate the value of $\mathcal{C}_N(\omega)$ for a random configuration of defects. For this investigation we consider the following representation of the transmission coefficient

$$\mathcal{C}_N(\omega) = \exp\,[-N\alpha_N(\omega, Q, C)] \qquad (23)$$

and obtain estimates of the value of

$$\alpha_N(\omega, Q, C) = -N^{-1} \ln\,[\mathcal{C}_N(\omega)] \qquad (24)$$

in the limit in which $N \to \infty$ and the over-all concentration of defects remains fixed, i.e., $N/A_N = C$. The quantity $\alpha_N(\omega, Q, C)$ is the attenuation of the incident wave per defect. On physical grounds, it is expected that when there is no correlation in the spacings between defects and when the average spacing is sufficiently large compared to the wavelength of the incident wave, then the attenuation per defect is equal to $-\ln\,[\mathcal{C}_1(\omega)]$, the attenuation of a single isolated defect. In the following, we verify this expectation and show that there is a range of concentration $0 < C < \bar{C}(\omega, Q)$ for which

$$\lim_{\substack{N\to\infty \\ N/A_N = C}} \alpha_N(\omega, Q, C) = \alpha(\omega, Q, C) > 0, \qquad (25)$$

provided only that the spacings between successive pairs of defects are statistically independent. Conservation of energy insures that $\alpha_N(\omega, Q, C) \geq 0$. Therefore, the most significant aspect of the result stated in (25) is that the right-hand side is independent of N and strictly positive. It should be noted further that the limitation on the range of concentration is a sufficient condition which insures exponential attenuation. This sufficient condition is not the best possible.

In order to obtain bounds on $\alpha_N(\omega, Q, C)$, we start with the following tridiagonal form of

$$D_N = d(\delta_{r,s} + iQ\omega(1 - \omega^2)^{-\frac{1}{2}}e^{ik|A_r - A_s|}),$$

[19] In Appendix B it is shown that the asymptotic dependence of $\mathcal{C}_N(\omega)$ on N for periodic configurations of defects is different depending upon whether the frequency of the incident wave lies in a band, in a band gap, or at a band edge of the periodic lattice.

which is obtained in Appendix A, Eq. (A3):

$$D_N = \begin{vmatrix} 1+i\Delta & -\exp(-ka_2 i) & 0 & 0 & \cdots & 0 \\ -\exp(-ka_2 i) & 1+i\Delta+(1-i\Delta)\exp(-2ka_2 i) & -\exp(-ka_3 i) & 0 & \cdots & 0 \\ 0 & -\exp(-ka_3 i) & 1+i\Delta+(1-i\Delta)\exp(-2ka_3 i) & \cdots & & 0 \\ 0 & 0 & & & & \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ & & & & & \\ & & & & & \cdot \\ & & & & & \cdot \\ & & & & & \cdot \\ & & & & \cdots & 0 \\ \cdot & & \cdot & \cdot & \cdots & -\exp(-ka_N i) \\ \cdot & & \cdot & \cdot & & \\ \cdot & & \cdot & \cdot & & \\ 0\cdots & & \cdots 0 & -\exp(-ka_N i) & & 1+i\Delta+(1-i\Delta)\exp(-2ka_N i) \end{vmatrix}, \tag{26}$$

where $\Delta = Q\omega(1-\omega^2)^{-\frac{1}{2}}$ and $a_n = A_n - A_{n-1}$. Thus the determinant D_N satisfies the two-term recurrence relation

$$D_N = [1 + i\Delta + (1 - i\Delta)\exp(-2a_N ki)]D_{N-1} - \exp(-2a_N ki)D_{N-2}, \tag{27}$$

with $D_0 = 1$ and $D_1 = 1 + i\Delta$. The sequence of determinants generated by the relation (27) leads to a sequence of values of the transmission coefficient. This connection has been used as a basis for obtaining Monte Carlo estimates of $\alpha_N(\omega, Q, C)$; the results of some of these calculations are presented in Sec. 7.

For our present purpose we obtain a more useful form of the recurrence equation (27) if we introduce the ratio

$$g_n = e^{\phi i} D_{n-1}/D_n, \tag{28}$$

where $1 + i\Delta = \delta e^{i\phi}$. Then we can rewrite Eq. (27) as

$$g_N = [\delta + (\delta - g_{N-1})\exp(-2a_N ki - 2\phi i)]^{-1}, \tag{29}$$

with $g_1 = \delta^{-1}$. Equation (29) is an example of a linear fractional (or Möbius) transformation[20] by which the complex number g_{N-1} is transformed into g_N. Since the difference $\delta - g_{N-1}$ appears explicitly in (29), it is convenient in discussing the properties of this transformation to introduce still another variable through the definition

$$g_n = \delta + |\Delta| h_n. \tag{30}$$

In terms of h_n, Eq. (29) becomes[21]

$$g_N = [\delta + |\Delta| h_{N-1}\exp(\Omega_N i)]^{-1} \tag{31}$$

or

$$h_N = -\frac{|\Delta| + \delta h_{N-1}\exp(\Omega_N i)}{\delta + |\Delta| h_{N-1}\exp(\Omega_N i)}, \tag{32}$$

where $\Omega_n = \pi - 2(a_N k + \phi)$. It is seen in Eq. (31) that the vector $|\Delta| h_{N-1}$ is rotated through an angle Ω_N, added to δ, and the resultant is reciprocated or inverted in the unit circle to produce the new vector $g_N = \delta + |\Delta| h_N$. Various stages in the transformation are shown in Fig. 2. The set of values of Ω_N, which corresponds to different spacings between the Nth and $(N-1)$th defects, results in a set of vectors $\delta + |\Delta| h_{N-1}\exp(i\Omega_N)$ which all lie on the dashed circle K in Fig. 2. As the result of inversion, the circle K is transformed into the dotted circle K' on which the vectors $\delta + |\Delta| h_N$ lie. Since the starting vector $g_1 = \delta^{-1} = \delta - \Delta^2\delta^{-1}$ lies inside the circle K_0 whose center lies at δ and whose radius is $|\Delta|$, and since K_0 is its own inverse with respect to the unit circle[20] C, all g_n, $n = 2, \cdots$, lie inside K_0. Consequently, the magnitude of every h_n is less than unity.

In terms of the g_n's, the expression for the attenuation per defect is

$$\alpha_N(\omega, Q, C) = -N^{-1}\ln\left(\left|\frac{D_{N-1}}{D_N}\cdot\frac{D_{N-2}}{D_{N-1}}\cdots\frac{D_1}{D_2}\cdot\frac{D_0}{D_1}\right|\right) = -N^{-1}\sum_{n=1}^{N}\ln(|g_n|). \tag{33}$$

Since the g_n's are confined to the interior of the circle

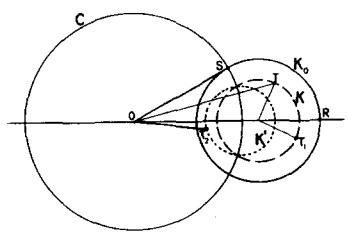


FIG. 2. The complex g plane showing the unit circle C with its center at the origin and the circles K and K_0 with their centers at $\delta = (1+\Delta^2)^{\frac{1}{2}}$. The radius of K_0 is $|\Delta|$, and the angle SOR is $\phi = \tan^{-1}(|\Delta|)$. The three vectors OT, OT_1, and OT_2 are, respectively, $\delta + h_{n-1}|\Delta|$, $\delta + h_{n-1}|\Delta|\exp(i\Omega_n)$, and $\delta + |\Delta| h_n = [\delta + h_{n-1}|\Delta|\exp \times (i\Omega_n)]^{-1}$.

[20] H. Schwerdtfeger, *Geometry of Complex Numbers* (The University of Toronto Press, Toronto, Canada, 1962).

[21] In Sec. 4 it is shown that $|h_n|$ is the amplitude of the wave reflected by an array of n defects. It should also be noted that the linear fractional transformation, Eq. (32), is an example of an "addition" law for the reflection coefficient of two inhomogeneities. See Refs. 1, 5–8.

K_0, we conclude that

$$0 \leq \alpha_N(\omega, Q, C) \leq \ln(\delta + |\Delta|). \tag{34}$$

The foregoing inequality is valid for *any* set of spacings of the defects, periodic or not.

Expression for $\alpha_N(\omega, Q, C)$ in Case of Random Spacing between Defects

In order to improve the bounds in (34) in the case where the spacings a_n are assumed to be independent, identically distributed, random integer variables, we substitute Eq. (31) in (33):

$$\alpha_N(\omega, Q, C) = N^{-1}\ln\delta + N^{-1}\sum_{n=2}^{N}\ln(|\delta + |\Delta| h_{n-1}\exp(\Omega_n i)|). \tag{35}$$

Subdivide the interior of K_0 into small elements $d\sigma_{jl}$ and group the terms of the sum in Eq. (35) according to the element of area $d\sigma_{jl}$ in which $\delta + |\Delta| h_{n-1}$ lies. It is convenient to use a polar-coordinate representation

$$\mathbf{h} = re^{i\chi}$$

for this purpose, where $0 \leq r < 1$ and $0 \leq \chi < 2\pi$. The indices on $d\sigma_{jl}$ then refer to r_j and χ_l and the element of area $d\sigma_{jl}$ is $r_j\,dr_j\,d\chi_l$. We denote by $f_C^{(N)}(j, l)$ the fraction of terms in (35) for which $\delta + |\Delta| h_{n-1}$ lies in $d\sigma_{jl}$. The group of terms which constitutes each $f_C^{(N)}(j, l)$ can be subdivided further according to the fraction of this group of terms for which the spacing has the value a. We assume that the spacings between defects are independent identically distributed, random integer variables with the probability distribution $\mathfrak{W}(a)$. It follows from this assumption that, for sufficiently large N, the fraction of terms in (35) for which $\delta + |\Delta| h_{n-1}$ lies in $d\sigma_{jl}$ and for which the spacing parameter is a is $f_C^{(N)}(j, l)\,d\sigma_{jl}\mathfrak{W}(a)$. Using $f_C^{(N)}(j,l)\,d\sigma_{jl}\mathfrak{W}(a)$, we replace Eq. (35) by the following approximate expression:

$$\alpha_N(\omega, Q, C) \cong \sum_{j,l}\sum_{a=1}^{\infty} f_C^{(N)}(j, l)\,d\sigma_{jl}\mathfrak{W}(a) \times \ln(|\delta + |\Delta| h(j, l)\exp[\pi i - 2(ak + \phi)i]|), \tag{36}$$

where

$$\sum_{j,l} f_C^{(N)}(j, l)\,d\sigma_{jl} = 1$$

and

$$\sum_{a=1}^{\infty}\mathfrak{W}(a) = 1.$$

In the following discussion, we assume that in the limit $N \to \infty$ the weight function $f_C^{(N)}(j, l)$ approaches a well-behaved limit $f_C(j, l)$. In particular, we assume that in the limit $N = \infty$ and as the subdivision of K_0 is refined, the sum over j, l can be replaced by an integral over $\mathbf{h}$ (the interior of K_0). On this assumption Eq. (36) is an approximate version of the following exact expression for $\alpha(\omega, Q, C) = \lim_{N\to\infty}\alpha_N(\omega, Q, C)$:

$$\alpha(\omega, Q, C) = \iint_{K_0} d\mathbf{h}\sum_{a=1}^{\infty} f_C(\mathbf{h})\mathfrak{W}(a) \times \ln(|\delta + |\Delta|\,\mathbf{h}\exp[\pi i - 2(ak + \phi)i]|), \tag{37}$$

where

$$\iint_{K_0} d\mathbf{h}\,f_C(\mathbf{h}) = 1.$$

Estimate of $\lim_{C\to 0}\alpha(\omega, Q, C)$

We first consider Eq. (37) in the limiting case $C = 0$. Two different forms of (37) are then appropriate, depending upon whether k is an irrational or a rational fraction of π. In the limit $C = 0$, all values of the spacing a are equally probable; and if k is an irrational fraction of π, the sum over a in Eq. (37) can be replaced by an integral[22] over Ω, the argument of the exponential function:

$$\lim_{C\to 0}\sum_{a=1}^{\infty}\mathfrak{W}(a)\ln(|\delta + |\Delta|\,\mathbf{h}\exp[\pi i - 2(ak + \phi)i]|) = (2\pi)^{-1}\int_0^{2\pi} d\Omega\ln(|\delta + |\Delta|\,\mathbf{h}\exp(i\Omega)|] = \ln\delta. \tag{38}$$

Substituting (38) in (37), we obtain

$$\lim_{C\to 0}\alpha(\omega, Q, C) = \ln\delta. \tag{39}$$

This limiting value of $\alpha(\omega, Q, C)$ is identical with the attenuation coefficient of a single, isolated defect. The second modified form of Eq. (37) in the limiting case $C = 0$ arises if $k = \pi r/s$, where r and s are relatively prime with $r < s$. Then the exponential in Eq. (37) assumes s values, each with the frequency s^{-1}. Consequently, in the limit $C = 0$, Eq. (37) becomes

$$\lim_{C\to 0}\alpha[\sin(\pi r/2s), Q, C] = \iint_{K_0} d\mathbf{h}\,f_0(\mathbf{h}) \times s^{-1}\sum_{a=1}^{s}\ln(|\delta + |\Delta|\,\mathbf{h}\exp[\pi i - 2(\phi + ar\pi/s)i]|)$$

or

$$\alpha[\sin(\pi r/2s), Q, 0] = \iint_{K_0} d\mathbf{h}\,f_0(\mathbf{h})s^{-1}\ln(|\delta^s - (|\Delta|\,\mathbf{h}e^{-2\phi i})^s|). \tag{40}$$

The following upper and lower bounds on

$$\alpha[\sin(\pi r/2s), Q, 0]$$

[22] G. Polya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis* (Springer-Verlag, Berlin, 1954), 2nd ed., Vol. 1, pp. 70, 71.

can be obtained from Eq. (40) by replacing the logarithmic term on the right-hand side by its largest and smallest values:

$$s^{-1}\ln(\delta^s - |\Delta|^s) \leq \alpha[\sin(\pi r/2s), Q, 0] \leq s^{-1}\ln(\delta^s + |\Delta|^s), \quad s \geq 2. \quad (41)$$

For $s > 2$, the lower bound in (41) is positive and is therefore an improvement on the general lower bound in Eq. (34). For $s = 2$, it should be noted that the minimum value of the logarithmic factor in the integrand of (40) is zero and is attained for only one value of $\mathbf{h}$, namely, $\mathbf{h} = e^{2\phi i}$. Therefore, even in the case $s = 2$, the minimum value of (40) must be positive.

4. THE DISTRIBUTION FUNCTION $f_C(\mathbf{h})$

The foregoing estimate of the attenuation per defect in the limit $C \to 0$ as well as the general expression for $\alpha(\omega, Q, C)$ in Eq. (37) are obtained on the implicit assumption that $f_C(\mathbf{h})$, the limiting distribution of the g_n's inside the circle K_0 in Fig. 2, exists and that the sum on $f_C^{(N)}(j, l)\, d\sigma_{jl}$ in Eq. (36) approaches the integral over $f_C(\mathbf{h})$ in Eq. (37). In this section we show that under conditions where $\alpha(\omega, Q, C) > 0$ the limiting distribution $f_C(\mathbf{h})$ is one-dimensional in the sense that $f_C(\mathbf{h})$ is zero everywhere except on the circumference of K_0. Consider Eq. (32) and form the quantity $1 - |h_n|^2$, where $0 \leq |h_n| < 1$ is the fraction of the distance of g_n from the center of K_0. The expression for $1 - |h_n|^2$ can be written as

$$\frac{1 - |h_n|^2}{1 - |h_{n-1}|^2} = |\delta + |\Delta|\, h_{n-1} \exp(i\Omega_n)|^{-2} = |g_n|^2. \quad (42)$$

Form the product

$$\prod_{n=2}^{N} \frac{1 - |h_n|^2}{1 - |h_{n-1}|^2} = \prod_{n=2}^{N} |g_n|^2. \quad (43)$$

Using the relation $1 - |h_1|^2 = \delta^{-2} = |g_1|^2$, Eq. (43) simplifies to

$$1 - |h_N|^2 = \mathcal{T}_N^2(\omega). \quad (44)$$

Thus h_N, which was introduced as an auxiliary variable in Eqs. (30)–(32), is directly related to the amplitude of the reflected wave. Equation (44) is simply a statement of the conservation of energy, i.e., the sum of the squares of the magnitudes of the transmitted and reflected waves is unity. It is clear in Eq. (44) that, if $\mathcal{T}_N^2(\omega) \to 0$ exponentially as $N \to \infty$, h_N approaches the boundary of K_0 exponentially in the same limit. Consequently, the limiting distribution of the g_n's for the isotopically disordered lattice is nonzero only along the boundary of K_0, provided that $\alpha(\omega, Q, C) > 0$. In Sec. 6 it will be shown that if the wavenumber k is a rational fraction of π and if Q is sufficiently large, then the limiting distribution of the g_n's is nonzero only along that portion of the circumference of K_0 lying inside the unit circle. Much stronger bounds can be obtained on the attenuation per defect in such a case.

5. ESTIMATE OF $\alpha(\omega, Q, C)$ FOR $C > 0$

We now determine bounds[23] on $\alpha(\omega, Q, C)$ for $C > 0$. In order to be explicit, we adopt the following form for the spacing distribution function:

$$\mathfrak{W}(a) = C(1 - C)^{a-1}, \quad a = 1, 2, \cdots, \quad (45)$$

for which the average spacing is

$$\langle a \rangle = C\sum_{a=1}^{\infty} a(1 - C)^{a-1} = C^{-1}.$$

This mean spacing is consistent with the limit in which $N \to \infty$ while $N/A_N = C$. The method which we use to obtain bounds on $\alpha(\omega, Q, C)$ in Eq. (37) is based on the observation that in the case $C = 0$ and $k = (r/s)\pi$, the sum of s consecutive terms in (37) is positive for $s > 2$. Therefore, when $C > 0$ and k is expressed as $k = [(r/s) + \epsilon]\pi$ with $|\epsilon| < s^{-1}$, the sum of s consecutive terms will be positive for sufficiently small C and $|\epsilon|$ for $s > 2$. The regrouped form for $\alpha(\omega, Q, C)$ is

$$\alpha(\omega, Q, C) = \iint_{K_0} d\mathbf{h}\, f_C(\mathbf{h}) \Big\{ C\sum_{n=0}^{\infty}(1 - C)^{ns} \sum_{a=1}^{s}(1 - C)^{a-1} \times \ln(|\delta + |\Delta|\, \mathbf{h} \exp\{\pi i - 2a\pi[(r/s) + \epsilon]i - 2\pi n\epsilon i - 2\phi i\}|) \Big\}. \quad (46)$$

We have already shown that the a sum in (46) is positive if $\epsilon = C = 0$. Thus it is a straightforward matter to show that the a sum in (46) is positive for sufficiently small $|\epsilon|$ and C. We denote the smallest value of the a sum by

$$l_s(\omega, Q, C) = \min \Big\{ \sum_{a=1}^{s}(1 - C)^{a-1} \times \ln(|\delta + |\Delta|\, \mathbf{h} \exp\{\pi i - 2a\pi[(r/s) + \epsilon] - 2\pi n\epsilon i - 2\phi i\}|) \Big\}, \quad (47)$$

where the minimum is taken with respect to all values of $\mathbf{h} \exp(-2\pi n\epsilon i)$ for fixed s and C. Similarly, we denote the maximum value of the a sum for fixed s

[23] Some of the results in this section were reported in R. J. Rubin, Bull. Am. Phys. Soc. (Series II) **12**, 117 (1967).

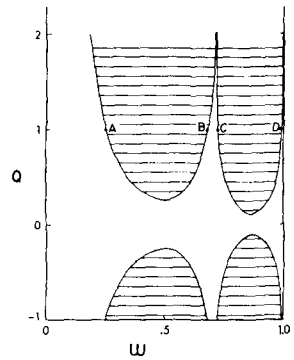


FIG. 3. The Q-ω plane. The shaded area indicates those combinations of Q and ω for which the lower bound in Eq. (49) is positive when $C = 0.1$, i.e., for which $\lambda(\omega, Q, 0.1) > 0$.

and C by $L_s(\omega, Q, C)$. In terms of $l_s(\omega, Q, C)$ and $L_s(\omega, Q, C)$, we obtain the following bounds on $\alpha(\omega, Q, C)$:

$$\frac{Cl_s(\omega, Q, C)}{1-(1-C)^s} \leq \alpha(\omega, Q, C) \leq \frac{CL_s(\omega, Q, C)}{1-(1-C)^s}. \quad (48)$$

The optimum values of $l_s(\omega, Q, C)$ and $L_s(\omega, Q, C)$ for use in the inequality (48) are obtained by direct evaluation of the minimum in Eq. (47) for $l_s(\omega, Q, C)$ and of the corresponding maximum for $L_s(\omega, Q, C)$. However, it should be noted that our representation of k by $[(r/s) + \epsilon]\pi$, with $|\epsilon| < s^{-1}$ and $r < s$ where r and s are relatively prime integers, is not unique. Therefore the best possible bounds in the inequality (48) are obtained by determining that value of s for which $Cl_s(\omega, Q, C)[1-(1-C)^s]^{-1}$ is largest at a particular value of ω (or k) and that value of s for which $CL_s(\omega, Q, C)[1-(1-C)^s]^{-1}$ is smallest. We denote these optimum bounds in (48) by $\lambda(\omega, Q, C)$ and $\Lambda(\omega, Q, C)$, respectively. Then the inequalities (48) and (34) can be combined in the single expression

$$\max\{0, \lambda(\omega, Q, C)\} \leq \alpha(\omega, Q, C) \leq \min\{\ln(\delta + |\Delta|), \Lambda(\omega, Q, C)\}. \quad (49)$$

In the limit $C \to 0$ and $\epsilon = 0$, the inequality (49) reduces to the inequality (41) in which the lower bound is positive except at $\omega = 2^{-\frac{1}{2}}$. For concentrations $C > 0$, the question arises as to when the lower bound in (49) is positive. A qualitative answer to this question is presented in a plot of the (ω, Q) plane in Fig. 3. The shaded areas represent those points in the (ω, Q) plane for which $\lambda(\omega, Q, 0.1) > 0$. It follows from the dependence of the expression for $l_s(\omega, Q, C)$ on Δ in Eq. (47) that the shaded areas in the negative Q region are the mirror images of those in the positive Q region below the line $Q = 1$. In addition, the boundary curves of the shaded areas approach asymptotically the vertical lines through $\omega = 0, 2^{-\frac{1}{2}}$, and 1 as $Q \to \infty$. As the concentration decreases, the shaded areas expand subject to the above constraints so as to fill the entire figure. In the $C = 0$ limit, the ω and Q axes are approached asymptotically.

According to the above discussion, when $Q = 1$ and $C = 0.1$ the lower bound in the inequality (49) is positive for all frequencies between A and B and between C and D. Consequently, in these frequency ranges the limiting value of $-N^{-1}\ln[\mathcal{C}_N(\omega)] = \alpha_N(\omega, Q, C)$ is positive and independent of N. The bounds in (49) have been evaluated explicitly for the case $Q = 1$ and $C = 0.1$. The results of these calculations are presented in Fig. 5 along with Monte Carlo estimates of $\alpha(\omega, 1, 0.1)$ and will be discussed in Sec. 7. A slightly different way of summarizing the qualitative behavior of the lower bound in (49) is that for fixed Q and $\omega \neq 2^{-\frac{1}{2}}$ there is a range of concentration $0 < C < \bar{C}(\omega, Q)$ for which $\alpha(\omega, Q, C) > 0$. The bounds implied in (49) are not the best possible and further improvement is desirable. We conjecture that for a random array of defects

$$\lim_{N\to\infty}\{-N^{-1}\ln[\mathcal{C}_N(\omega)]\} > 0$$

for any $Q \neq 0$ and $C < 1$.

6. THE TRANSMISSION COEFFICIENT IN THE CASE OF THE SPECIAL FREQUENCIES OF MATSUDA

The existence of special frequencies of isotopically disordered harmonic crystals has been pointed out by Matsuda.[24] He has shown that, for infinite disordered crystals containing finite concentrations of two isotopes, the frequency $\omega = \sin(\pi r/2s)$, where r and s are relatively prime, is not a normal-mode frequency of the crystal if

$$Q \geq Q_{\text{crit}}(r, s), \quad (50)$$

where

$$Q_{\text{crit}}(r, s) = \cot(\pi/2s)\cot(\pi r/2s). \quad (51)$$

In the context of our problem of determining the limiting value of the attenuation per defect of an isotopically disordered section of harmonic crystal, we have already seen in Sec. 5 that at frequencies of the form $\omega = \sin(\pi r/2s)$, with $s > 2$, there is a finite range of concentration in which there is exponential attenuation of an incident wave. In this section we show that if Q satisfies the restriction $Q \geq Q_{\text{crit}}(r, s)$, then $\alpha[\sin(\pi r/2s), Q, C] > 0$ in the entire concentration range $0 < C < 1$. The result that there is exponential attenuation for any value of the concentration, provided that (50) is satisfied, is consistent with Matsuda's result that $\omega = \sin(\pi r/2s)$ is not a

[24] H. Matsuda, Progr. Theoret. Phys. (Kyoto) 31, 161 (1964).

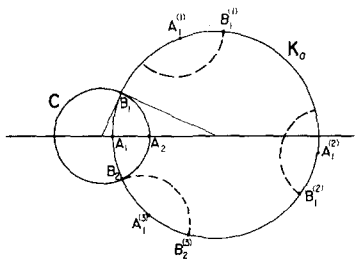


FIG. 4. The complex g plane showing the unit circle C and the circle K_0 in the case $2\pi s^{-1} > 2(\pi - 2\phi)$ when $\omega = \sin(\pi r/2s)$, and $r/s = \frac{2}{3}$.

normal-mode frequency of the infinite disordered crystal.

Consider the transformation in Eq. (31), by which successive g_n's are generated; and consider in Fig. 4 the region of intersection $A_1B_1A_2B_2$ between the unit circle and the circle K_0. Only s images of a point in region $A_1B_1A_2B_2$ are generated by the rotation of Δh_{n-1} when the spacing a_n ranges over all possible values, namely,

$$\delta + \Delta h_{n-1} \exp\,[i\pi - 2(\phi + a\pi/s)i], \qquad (52)$$

for $a = 1, \cdots, s$. The point A_1 is at $\delta + \Delta e^{i\pi}$; and we denote the s images of A_1 by $A_1^{(a)}$ where a appears in the angle of rotation in (52). The point $A_1^{(s)}$ which is located at $\delta + \Delta e^{-2\phi i}$ is shown in Fig. 4. Since $\pi - 2\phi$ is the angle subtended by region $A_1B_1A_2B_2$ at the center of K_0, the image region $A_1^{(s)}B_1^{(s)}A_2^{(s)}B_2^{(s)}$ is tangent to $A_1B_1A_2B_2$ at B_2. Moving clockwise around K_0 from $A_1^{(s)}B_1^{(s)}A_2^{(s)}B_2^{(s)}$, the next image region is $A_1^{(1)}B_1^{(1)}A_2^{(1)}B_2^{(1)}$. If the angle of rotation $2\pi/s$ between successive image regions exceeds $2(\pi - 2\phi)$, region $A_1^{(1)}B_1^{(1)}A_2^{(1)}B_2^{(1)}$ will not overlap $A_1B_1A_2B_2$. Thus, if

$$2\pi/s \geq 2(\pi - 2\phi), \qquad (53)$$

all image regions lie outside the unit circle; and when the reciprocal in Eq. (31) is formed, every g_n which is generated from a g_{n-1} which lies in $A_1B_1A_2B_2$ also lies inside $A_1B_1A_2B_2$. Since $g_1 = \delta^{-1}$ is inside $A_1B_1A_2B_2$, all g_n's lie inside $A_1B_1A_2B_2$, provided that condition (53) is satisfied. If the definition of ϕ which was introduced in (28),

$$\phi = \tan^{-1}\Delta = \tan^{-1}\,[Q \tan\,(\pi r/2s)],$$

is substituted in (53), the condition which is obtained is identical with Eqs. (50) and (51), the condition given by Matsuda for the existence of special frequencies. We now show that when the above condition is satisfied, the attenuation per defect $\alpha[\sin(\pi r/2s), Q, C]$ is positive for all C, $0 < C < 1$. First recall that when g_n is represented as $\delta + \Delta h_n$ where $0 \leq |h_n| < 1$, the magnitudes $|h_n|$ and $|h_{n-1}|$ satisfy Eq. (42), from which follows the inequality

$$|h_n|^2 > |h_{n-1}|^2. \qquad (54)$$

Hence g_n lies closer to the circumference of K_0 in Fig. 4 than does g_{n-1}. Consequently as $N \to \infty$, $f_C(h)$, the distribution function of the g_n's inside K_0 is zero, except in the neighborhood of the circumference of K_0 between B_1 and B_2. The expression for $\alpha(\omega, Q, C)$ in Eq. (46) in the present case reduces to

$$\alpha[\sin(\pi r/2s), Q, C] = \int_0^{2\pi} d\chi\, f_C(\chi)\Big\{C[1-(1-C)^s]^{-1}\sum_{a=1}^{s}(1-C)^{a-1} \times \ln\,(|\delta + \Delta \exp\,[i\chi + \pi i - 2(\phi + a\pi r/s)i]|)\Big\}, \qquad (55)$$

where $f_C(\chi)\,d\chi$ is the limiting fraction of g_n's in the interval $(\chi, \chi + d\chi)$ and $\mathbf{h} = e^{i\chi}$. The a sum in Eq. (55) is strictly positive because the argument of the logarithm corresponds to a partially transformed g which lies outside the unit circle. Therefore

$$\alpha[\sin(\pi r/2s), Q, C] > 0 \quad \text{for all } C, \quad 0 < C < 1.$$

We now obtain explicit bounds on $\alpha(\sin(\pi/4), 1, C)$. This is a simple case to treat and the procedure which we use can be adapted to other values of $Q > 1$ as well as other special frequencies. The expression for the attenuation per defect is

$$\alpha(2^{-\frac{1}{2}}, 1, C) = \int_{3\pi/4}^{5\pi/4} d\chi\, f_C(\chi)\Big\{\left(\frac{1}{2-C}\right)\ln\,(|\sqrt{2} + e^{i(\chi-\pi/2)}|) + \left(\frac{1-C}{2-C}\right)\ln\,(|\sqrt{2} + e^{i(\chi+\pi/2)}|)\Big\}, \qquad (56)$$

where $\Delta = 1$, $\delta = \sqrt{2}$, and $\phi = \pi/4$. The following bounds for $\alpha(2^{-\frac{1}{2}}, 1, C)$ can be obtained from Eq. (56) by replacing the term in braces by its minimum and maximum values:

$$\frac{1}{2}\left(\frac{1-C}{2-C}\right)\ln 5 \leq \alpha(2^{-\frac{1}{2}}, 1, C) \leq \tfrac{1}{2}\ln\left(\frac{6}{2-C}\right) + \frac{1}{2}\left(\frac{1-C}{2-C}\right)\ln(1-C). \qquad (57)$$

Comparison of the zero-concentration limit of the bounds obtained in Eq. (57) with those obtained in Eq. (41) for this case ($\omega = 2^{-\frac{1}{2}}$, $s = 2$, $Q = 1$) shows that the upper bounds are identical, but that the lower bound in Eq. (57), $\frac{1}{4}\ln 5$, is significantly larger than the lower bound in Eq. (41). Furthermore, the value of $\alpha(\omega, 1, 0^+)$ for frequencies arbitrarily close to $\omega = 2^{-\frac{1}{2}}$ is arbitrarily close to $\frac{1}{2}\ln 2$ according to Eqs. (38) and (41); and this value lies outside the range of possible values obtained in (57). Thus we conclude that in

the zero-concentration limit $\alpha(\omega, 1, 0^+)$ is a discontinuous function of frequency at the special frequency $\omega = 2^{-\frac{1}{2}}$.

7. MONTE CARLO CALCULATION OF $\alpha_N(\omega, Q, C)$

In this section we present some numerical results of the direct calculation of

$$\alpha_N(\omega, Q, C) = -N^{-1} \ln [\mathcal{C}_N(\omega)] = -N^{-1} \sum_{n=1}^{N} \ln (|g_n|) \qquad (33)$$

for randomly generated configurations of defects. The g_n's which appear on the right-hand side of Eq. (33) are formed recursively using Eq. (29):

$$g_n = [\delta + (\delta - g_{n-1}) \exp (-2a_n ki - 2\phi i)]^{-1}, \quad n = 2, \cdots, \qquad (29)$$

with $g_1 = \delta^{-1}$. The calculations were performed on a CDC 6600 at the Los Alamos Scientific Laboratory. The computing procedure requires the choosing of three parameters: an incident frequency $\omega = \sin(k/2)$, a reduced-mass ratio $Q = (M/m) - 1$, and a value of the concentration C. Then a sequence of random, integer values of the spacings a_n of the defects is generated in which each a_n has the frequency distribution $C(1 - C)^{a_n}$. As each value of a_n, $n = 2, \cdots$, is generated, a value of g_n is determined from Eq. (29) and a value of $\alpha_n(\omega, Q, C)$ from Eq. (33). The results of some of the calculations are presented in Figs. 5 and 6 in the case $Q = 1$. The results in Fig. 5 correspond to the concentration $C = 0.1$ and those in Fig. 6

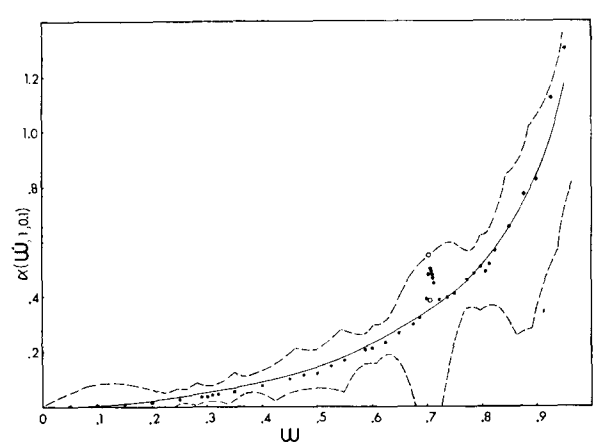


FIG. 5. Plot of the Monte Carlo estimates of the attenuation constant $\alpha(\omega, 1, 0.1)$ based on arrays of defects consisting of 3×10^4 defects. The dots represent the values of $\alpha_{30000}(\omega, 1, 0.1)$. The solid curve is a plot of $\ln \delta$, the attenuation constant for a single isolated defect. The dashed curves are a plot of the upper and nonzero lower bounds of the attenuation constant determined from Eq. (49) but with the restriction that $s \leq 7$. The bounds on $\alpha(2^{-\frac{1}{2}}, 1, 0.1)$ obtained from Eq. (57) are indicated by the pair of open circles.

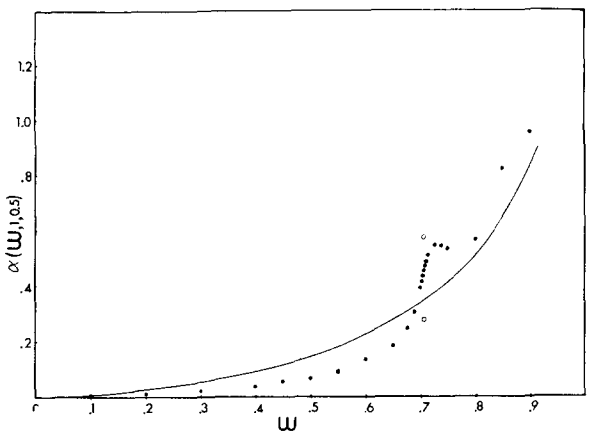


FIG. 6. Plot of the Monte Carlo estimates of $\alpha(\omega, 1, 0.5)$ based on arrays of defects consisting of 3×10^4 defects. The dots represent the values of $\alpha_{30000}(\omega, 1, 0.5)$. The solid curve is a plot of $\ln \delta$. The bounds on $\alpha(2^{-\frac{1}{2}}, 1, 0.5)$ are indicated by a pair of open circles.

to the concentration $C = 0.5$. Figure 5 presents calculated values of the attenuation constant $\alpha_N(\omega, 1, 0.1)$ as a function of frequency for arrays of $N = 3 \times 10^4$ defects. Each computed value of $\alpha_{30000}(\omega, 1, 0.1)$ corresponds to a different array of 3×10^4 defects. When values of $\alpha_{30000}(\omega, 1, 0.1)$, for different arrays of defects but the same frequency ω, are compared, there is no visible difference on the scale of Fig. 5. The solid curve plotted in Fig. 5 is $\ln \delta$, the attenuation constant of a single, isolated defect. It is seen that this curve lies close to the computed values of $\alpha_{30000}(\omega, 1, 0.1)$ over most of the frequency range. However, a striking anomaly in the values of $\alpha_{30000}(\omega, 1, 0.1)$ can be seen in the vicinity of the special frequency $\omega = \sin(\pi/4)$. This anomaly is the remnant of the discontinuity in

$$\lim_{C \to 0} \lim_{N \to \infty} \alpha_N(\omega, 1, C)$$

in the vicinity of $\omega = \sin(\pi/4)$ which was deduced in Sec. 6. The upper and lower bounds on $\alpha(\omega, 1, 0.1)$ obtained from Eq. (49) are indicated by the dashed curves in Fig. 5. The bounds on $\alpha(2^{-\frac{1}{2}}, 1, 0.1)$ obtained from Eq. (57) are indicated by open circles. There is additional structure evident in the Monte Carlo estimates of $\alpha(\omega, Q, C)$ in the vicinity of $\omega \cong 0.8$.

Figure 6 presents calculated values of $\alpha_N(\omega, 1, 0.5)$ for arrays of 3×10^4 defects. The solid curve is a plot of $\alpha_1 = \ln \delta$. At this higher concentration, the computed values of $\alpha_{30000}(\omega, 1, 0.5)$ are noticeably less than $\ln \delta$ in the frequency range $0.2 < \omega < 0.66$. There is a peak in the value of the attenuation constant in the vicinity of $\omega = 2^{-\frac{1}{2}}$ similar to the one found at the concentration $C = 0.1$. Compared to the sharp resonancelike peak in the plot of $\alpha_{30000}(\omega, 1, 0.1)$ in the vicinity of $\omega = 2^{-\frac{1}{2}}$, the present peak is much broader. This broadening is presumably due to an

interaction between groups of defects. The bounds on $\alpha(2^{-\frac{1}{2}}, 1, 0.5)$ from Eq. (57) are indicated by the open circles in Fig. 6. The analogs of the nonzero lower bounds at the other frequencies which are plotted in Fig. 5 cannot be obtained from Eq. (49). However, it appears from an examination of the Monte Carlo estimates of $\alpha_N(\omega, 1, 0.5)$ as a function of N that $\alpha_N(\omega, 1, 0.5) > 0$ and independent of N at all these frequencies.

8. SUMMARY AND REMARKS

The amplitude $\mathcal{C}_N(\omega)$ of a wave of frequency ω which is transmitted by a disordered array of N isotopic defects in a one-dimensional crystal has been investigated. In particular, the limiting value

$$\alpha(\omega, Q, C) = \lim_{N\to\infty} \{-N^{-1} \ln [\mathcal{C}_N(\omega)]\}$$

has been studied, where $Q = (M/m) - 1$ is the reduced mass difference between the defect and host particles, and where the spacings between successive defects are independent identically distributed integer random variables with the mean value C^{-1}.

There are two aspects of this investigation. First, it is established that the limiting value $\alpha(\omega, Q, C)$ satisfies the inequality

$$\max \{0, \lambda(\omega, Q, C)\} \leq \alpha(\omega, Q, C) \leq \min \{\Lambda(\omega, Q, C), \ln (\delta + |\Delta|)\}, \quad (49)$$

where $\delta = (1 + \Delta^2)^{\frac{1}{2}}$, $\Delta = Q\omega(1 - \omega^2)^{-\frac{1}{2}}$ and where $\lambda(\omega, Q, C)$ and $\Lambda(\omega, Q, C)$ are defined preceding Eq. (49) in Sec. 5. In addition, it is shown that the bounds in (49) can be considerably improved at the special frequencies of Matsuda provided that $Q \geq Q_{\text{crit}}$, where Q_{crit} is defined in Eq. (51). Second, Monte Carlo estimates of $-N^{-1} \ln [\mathcal{C}_N(\omega)]$ are obtained as a function of ω for $N = 3 \times 10^4$ in the cases $Q = 1$, $C = 0.1$ and $Q = 1$, $C = 0.5$. In the former case, these estimates are presented in Fig. 5 along with the values of the bounds obtained from Eq. (49).

Finally, we list some remarks and some questions which have not been answered in this paper and which are of considerable interest in themselves.

(1) In establishing bounds on $\alpha_N(\omega, Q, C)$, we have not considered the question of how the values of $\alpha_N(\omega, Q, C)$ are distributed, as a function of N, between the bounds. The analysis in this paper is based on the representation of $\alpha_N(\omega, Q, C)$ as the average of a sum of N terms

$$\alpha_N(\omega, Q, C) = -N^{-1} \sum_{n=1}^{N} \ln (|g_n|),$$

where the g_n's are recursively related complex numbers confined to the interior of the circle K_0 in Fig. 2. The linear fractional transformation of g_{n-1} to g_n is a random transformation depending upon the independent random variable a_n. In general, a sequence of two successive transformations do not commute.

(2) It follows from Sec. 4 that $|\Delta|\,(1 - |h_n|)$, the distance of g_n from the circle K_0, satisfies the inequality

$$0 < |\Delta|\,(1 - |h_n|) \leq |\Delta|\,\mathcal{C}_N^2(\omega).$$

When the limiting value

$$\alpha(\omega, Q, C) = -\lim_{N\to\infty} \{N^{-1} \ln [\mathcal{C}_N(\omega)]\}$$

is positive, so that $\mathcal{C}_N(\omega) \to 0$ and g_n approaches the boundary circle K_0, an interesting question still remains as to the form of the limiting distribution of the g_n's along the circumference of K_0. A functional equation for the approximately one-dimensional g distribution can be derived by a method analogous to that used by Dyson,[9] Schmidt,[10] and Dean.[11] Approximate solutions of the functional equation will be compared with the distribution of the g_n's obtained in our Monte Carlo calculations in a separate paper. In this connection it should be noted that $\hat{\mathcal{C}}_N(\omega)$, the ratio of the amplitude of the Nth defect to the amplitude of the first defect, can be obtained by using Eq. (A9) and the appropriate modification of the integrand in Eq. (18). The result, which is almost identical to the expression for $\mathcal{C}_N(\omega)$, is

$$\hat{\mathcal{C}}_N(\omega) = |\hat{D}_N|^{-1},$$

where $\hat{D}_N$ is the following $(N-1)$th-order tridiagonal determinant:

$$\hat{D}_N = \begin{vmatrix} 1+i\Delta-i\Delta\exp(-2ka_2i) & -\exp(-ka_3i) & \cdots & & \cdots 0 \\ -\exp(-ka_3i) & 1+i\Delta+(1-i\Delta)\exp(-2ka_3i) & -\exp(-ka_4i) & \cdots & \cdot \\ 0 & -\exp(-ka_4i) & \cdots & & \cdot \\ 0 & 0 & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ & & & & \cdots 0 \\ \cdot & & \cdot & & \cdots -\exp(-ka_Ni) \\ \cdot & & \cdot & & \\ \cdot & & \cdot & & \\ 0 \cdots & & \cdots\ 0 & -\exp(-ka_Ni) & 1+i\Delta+(1-i\Delta)\exp(-2ka_Ni) \end{vmatrix}.$$

The only difference between $\hat{D}_N$ and the determinant in the definition of the transmitted amplitude $\mathcal{C}_N(\omega)$ is in their starting values:

$$\hat{D}_0 = 1 \quad \text{and} \quad \hat{D}_1 = 1 + i\Delta - i\Delta \exp(-2ka_2 i),$$

rather than

$$D_0 = 1 \quad \text{and} \quad D_1 = 1 + i\Delta.$$

Otherwise, $\hat{D}_N$ and D_N and $\hat{g}_N = e^{\phi i}\hat{D}_{N-1}/\hat{D}_N$ and g_n satisfy the same recurrence equations. Since the starting value $\hat{g}_1$ lies on the circle K_0, all $\hat{g}_n$'s lie on K_0, so that the $\hat{g}$ distribution function is exactly one-dimensional. In investigating the limiting transmission properties of a disordered array of defects, one can expect that there is no significant difference between $\alpha(\omega, Q, C)$ and

$$\hat{\alpha}(\omega, Q, C) = -\lim_{N\to\infty}\{N^{-1}\ln[\hat{\mathcal{C}}_N(\omega)]\},$$

and consequently that the limiting g_n and $\hat{g}_n$ distributions should be identical.

(3) It should be emphasized that the bounds in Eq. (49) are not the best possible ones. For example, these bounds are independent of the algebraic sign of Q for $|Q| < 1$. However, Monto Carlo estimates of $\alpha(\omega, |Q|, C)$ and $\alpha(\omega, -|Q|, C)$, while consistent with these bounds, are significantly different from each other.

(4) From a physical point of view, there is a similarity between the two limiting cases $C \to 0$ and $1 - C \to 0$. In the first case, the region containing the defect particles consists of widely spaced defect particles in a background of host particles. In the second case, the region containing the defects consists of widely spaced host particles in a background of defect particles. This similarity or symmetry in the two limits is not evident in the determinantal expression for $\mathcal{C}_N(\omega)$. Nevertheless, it is worthwhile to demonstrate its existence explicitly in the dependence of $\alpha_N(\omega, Q, C)$ on C.

(5) Although we have only treated the transmission problem for a single type of isotopic defect, the analysis can be carried through when there are several types of isotopes. The result for the transmitted amplitude, which is a simple generalization of Eq. (22), is

$$\mathcal{C}_N(\omega) = |d[\delta_{r,s} + iQ_r\omega(1-\omega^2)^{-\frac{1}{2}} \times \exp(-ik|A_r - A_s|)]|^{-1},$$

where $Q_r = (M_r/m) - 1$ and M_r is the mass of the rth defect located at lattice position A_r. This determinant can also be written in a tridiagonal form.

(6) We have not developed explicitly the connection between the present work and recent investigations of the localized nature of eigenmodes in disordered systems.[25]

ACKNOWLEDGMENTS

Part of the work reported here was started in 1963–64 during the tenure of a National Science Foundation Senior Postdoctoral Fellowship at the University of California, San Diego; and the author wishes to thank Professor W. Kohn and Professor H. Suhl for their hospitality. The Monto Carlo calculations were carried out with the assistance of Mrs. Mary Menzel during the summer of 1967 at the Los Alamos Scientific Laboratory; and the author wishes to thank Dr. W. M. Visscher and Dr. J. Gammel for their hospitality and interest.

APPENDIX A: REDUCTION OF FORMULA $\xi(S, P)$ IN EQ. (13)

The expression for the response of particle S to the initial disturbance (4) at particle n when $n \neq A_j$, $j = 1, \cdots, N$ is given in Eq. (13):

$$\xi(S,P) = P^{-1}(P^2+1)^{-\frac{1}{2}}E^{2|n-S|} \times \left(D_N - E^{-2|n-S|}\sum_{j=1}^{N} E^{2|A_j-S|}D_N^{(j)}\right)\Big/ D_N, \tag{A1}$$

where $E = [P + (P^2+1)^{\frac{1}{2}}]^{-1}$, D_N denotes the determinant $d(\delta_{r,s} + QP(P^2+1)^{-\frac{1}{2}}E^{2|A_r-A_s|}]$, and $D_N^{(j)}$ is the determinant formed from D_N by replacing the jth column of D_N by the column vector

$$\{\gamma E^{2|n-A_1|}, \gamma E^{2|n-A_2|}, \cdots, \gamma E^{2|n-A_N|}\}$$

with $\gamma = QP(P^2+1)^{-\frac{1}{2}}$. There is a major simplification in the expression for $\xi(S, P)$ in Eq. (A1) when $n < 0$ and $S > A_N$: the expression in braces is equal to unity. To show this, transform the determinant

$$d(\delta_{r,s} + \gamma E^{2|A_r-A_s|}) = \begin{vmatrix} 1+\gamma & \gamma E^{2(A_2-A_1)} & \gamma E^{2(A_3-A_1)} & \cdots & \gamma E^{2(A_N-A_1)} \\ \gamma E^{2(A_2-A_1)} & 1+\gamma & \gamma E^{2(A_3-A_2)} & \cdots & \gamma E^{2(A_N-A_2)} \\ \gamma E^{2(A_3-A_1)} & \gamma E^{2(A_3-A_2)} & 1+\gamma & \cdots & \gamma E^{2(A_N-A_3)} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \gamma E^{2(A_N-A_1)} & \gamma E^{2(A_N-A_2)} & \gamma E^{2(A_N-A_3)} & \cdots & 1+\gamma \end{vmatrix} \tag{A2}$$

[25] N. F. Mott and W. D. Twose, Advan. Phys. **10**, 137 (1961); A. P. Roberts and R. E. B. Makinson, Proc. Phys. Soc. (London) **79**, 630 (1962); R. E. Borland, Proc. Roy. Soc. (London) **274**, 529 (1963); N. F. Mott, Advan. Phys. **16**, 49 (1967); J. Hori and S. Minami, *Proceedings of International Conference on Localized Excitations in Solids, University of California, Irvine, 1967* (Plenum Press, Inc., New York, 1968), p. 611.

by means of the following operations: (1) Multiply the rth row by $E^{2(A_{r+1}-A_r)}$ and subtract from the $r+1$th row; (2) then repeat this procedure for the rth and $r+1$th columns. If these operations are performed in the order $r = N-1, N-2, \cdots, 1$, the determinant $d(\delta_{r,s} + \gamma E^{2|A_r - A_s|})$ assumes the continuant or tridiagonal form

$$d(\delta_{r,s}+\gamma E^{2(A_r-A_s)}) = \begin{vmatrix} 1+\gamma & -E^{2(A_2-A_1)} & 0 & 0 & \cdots & 0 \\ -E^{2(A_2-A_1)} & 1+\gamma+(1-\gamma)E^{4(A_2-A_1)} & -E^{2(A_3-A_2)} & 0 & \cdots & \vdots \\ 0 & -E^{2(A_3-A_2)} & 1+\gamma+(1-\gamma)E^{4(A_3-A_2)} & -E^{2(A_4-A_3)} & \cdots & 0 \\ 0 & 0 & -E^{2(A_4-A_3)} \cdots & & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ & & & & \cdots & -E^{2(A_N-A_{N-1})} \\ 0 \cdots & & & \cdots -E^{2(A_N-A_{N-1})} & & 1+\gamma+(1-\gamma)E^{4(A_N-A_{N-1})} \end{vmatrix}. \quad \text{(A3)}$$

Next consider the determinant $D_N^{(j)}$ and carry out the same pair of operations (1) and (2) in the order $r = N-1, \cdots, j+1$. Then perform operation (1) for $r = j$. At this stage, the jth column has the form $\{\gamma E^{2(A_1-A_j)}, \cdots, \gamma E^{2(A_{j-1}-A_j)}, \gamma, 0, \cdots, 0\}$, and the $(j+1)$th through the Nth elements of the first column through the $(j-1)$th column are all zero. Finally, perform the following set of operations: multiply column j by $E^{2(A_j-A_r)}$ and subtract from the rth column for $r = 1, \cdots, j-1$. The diagonal elements of the first $j-1$ columns are all equal to one, the diagonal element of column j is γ, and all elements of these columns below the diagonal are zero. Consequently the determinant $D_N^{(j)}$ is reduced to tridiagonal form

$$D_N^{(j)} = \gamma t_{N-j} = \gamma \begin{vmatrix} 1+\gamma-\gamma E^{4(A_{j+1}-A_j)} & -E^{2(A_{j+2}-A_{j+1})} & 0 & \cdots & 0 \\ -E^{2(A_{j+2}-A_{j+1})} & 1+\gamma+(1-\gamma)E^{4(A_{j+2}-A_{j+1})} & \cdots & & \vdots \\ 0 & & & & \\ \vdots & & & & \\ \vdots & & & & \\ & & & \cdots & 0 \\ \vdots & & 1+\gamma+(1-\gamma)E^{4(A_{N-1}-A_{N-2})} & & -E^{2(A_N-A_{N-1})} \\ 0 \cdots & \cdots 0 & -E^{2(A_N-A_{N-1})} & & 1+\gamma+(1-\gamma)E^{4(A_N-A_{N-1})} \end{vmatrix}, \quad \text{(A4)}$$

where $N-j$ is the order of the reduced determinant t_{N-j} and $N-j > 1$. In the cases $N-j = 1$ and $N-j = 0$, we have $t_1 = 1+\gamma-\gamma E^{4(A_N-A_{N-1})}$ and $t_0 = 1$.

The term in braces in Eq. (A1) has thus been transformed to

$$\{\ \} = D_N - \gamma \sum_{j=1}^{N} t_{N-j}. \quad \text{(A5)}$$

Now consider the difference $D_N - \gamma t_{N-1}$. It is a simple matter to show that

$$D_N - \gamma t_{N-1} = \tilde{D}_{N-1}, \quad \text{(A6)}$$

where $\tilde{D}_{N-1}$ is an $(N-1)$th order determinant identical in form with D_N but with all reference to the defect at A_1 missing. As a result Eq. (A5) becomes

$$\{\ \} = \tilde{D}_{N-1} - \gamma \sum_{j=2}^{N} t_{N-j}. \quad \text{(A7)}$$

Equation (A7) now refers to the defects 2 through N. The foregoing reduction can be repeated $N-2$ more times and

$$\{\ \} = \tilde{D}_1 - \gamma t_0 = 1.$$

Therefore the expression for the response $\xi(S, P)$ at particle $S > A_N$ to the initial disturbance (4) at particle $n < 0$ is

$$\xi(S, P) = P^{-1}(P^2+1)^{-\frac{1}{2}} E^{2(S+|n|)}/D_N. \quad \text{(A8)}$$

The response at defect A_j to the initial disturbance (4) for $n < 0$ is

$$\xi(A_j, P) = P^{-1}(P^2+1)^{-\frac{1}{2}} E^{2(A_j+|n|)} t_{N-j}/D_N. \quad \text{(A9)}$$

In a similar fashion, it can be shown that the response at S to the initial disturbance (4) is

$$\xi(S, P) = P^{-1}(P^2+1)^{-\frac{1}{2}} E^{2(S+|n|)} \times \left\{\tilde{D}_{N-j} - \gamma \sum_{k=j+1}^{N} E^{4(A_k-S)} t_{N-k}\right\} \Big/ D_N \quad \text{(A10)}$$

if $n < 0$ and $A_j < S < A_{j+1}$; and

$$\xi(S, P) = P^{-1}(P^2 + 1)^{-\frac{1}{2}}E^{2(|n|+|S|)} \times \left\{D_N - \gamma \sum_{k=1}^{N} E^{4(A_k+|S|)}t_{N-k}\right\} \Big/ D_N \quad \text{(A11)}$$

if $n < 0$ and $S < 0$.

APPENDIX B: TRANSMISSION COEFFICIENT OF A PERIODIC ARRAY OF DEFECTS

The transmission coefficient of an array of N defects is expressed in terms of the magnitude of the determinant D_N in Eq. (26). This determinant is a continuant and satisfies the two-term recurrence relation (27):

$$D_N = [1 + i\Delta + (1 - i\Delta)\exp(-2ka_Ni)]D_{N-1} - \exp(-2ka_Ni)D_{N-2}, \quad \text{(B1)}$$

with $D_0 = 1$ and $D_1 = 1 + i\Delta$. In this appendix, we derive an explicit expression for D_N for the case of a periodic array of defects in which all nearest-neighbor spacings of defects are the same, $a_k \equiv A_k - A_{k-1} = a$, $k = 2, \cdots, N$. In this case the difference equation (B1) can be solved by introducing the generating function $\mathfrak{F}(z) = \sum_{n=0}^{\infty} D_n z^n$. Multiplying the equation for D_n by z^n and summing with respect to n, one obtains an algebraic equation for $\mathfrak{F}(z)$ which, when solved, yields

$$\mathfrak{F}(z) = \frac{1 - z(1 - i\Delta)\exp(-2kai)}{1 - [1 + i\Delta + (1 - i\Delta)\exp(-2kai)]z + z^2\exp(-2kai)}. \quad \text{(B2)}$$

The coefficient of z^n in the expansion of (B2) is

$$D_n = [U_n(x) - (1 - i\Delta)\exp(-kai)U_{n-1}(x)] \times \exp(nkai), \quad \text{(B3)}$$

where $U_n(x) = \sin[(n + 1)\eta]/\sin\eta$ is a Tchebycheff polynomial, $\eta = \cos^{-1}x$, and

$$x = \tfrac{1}{2}(1 + i\Delta)\exp(kai) + \tfrac{1}{2}(1 - i\Delta)\exp(-kai)$$

is real. If $|x| < 1$, the two Tchebycheff polynomials in braces in (B3) are oscillatory functions of the order. However, when $|x| > 1$, these polynomials grow exponentially with N. We illustrate this behavior by considering the periodic array in which $a = 2$. When $\Delta = Q\omega(1 - \omega^2)^{-\frac{1}{2}}$ and $\exp(ik/2) = i\omega + (1 - \omega^2)^{\frac{1}{2}}$ are substituted in the expression for x, it can be reduced to

$$x = 8(Q + 1)\omega^4 - 4(Q + 2)\omega^2 + 1. \quad \text{(B4)}$$

At the transition values $x = \pm 1$, the relations obtained between Q and ω are

$$\omega_2 = (2 + Q)/2(1 + Q) = \tfrac{1}{2}(1 + mM^{-1})$$

and

$$\omega^2 = \begin{cases} \frac{1}{2}, \\ 2^{-1}(1 + Q)^{-1} = 2^{-1}mM^{-1}. \end{cases} \quad \text{(B5)}$$

The frequencies in (B5) define the boundaries between frequency intervals where D_N is either an oscillatory or exponentially growing function of N. These three frequencies correspond to the upper edge of the acoustical branch and the two edges of the optical branch of a diatomic lattice. When the defects are heavy, $mM^{-1} < 1$ and all critical frequencies are less than unity. It can be readily verified that the frequencies outside the band correspond to values of $|x| > 1$, and hence to exponentially decreasing values of the transmission coefficient. When the defects are light, either the frequency of the upper or the frequencies of the upper and lower edges of the optical band are "inaccessible." They are "inaccessible" in the sense that these frequencies exceed unity and hence cannot be propagated in the host crystal.

At the critical frequencies, the transmission coefficient exhibits a transitional behavior. For example, at $\omega = 2^{-\frac{1}{2}}$ or $k = \pi/2$, Eq. (B3) for D_N reduces to

$$D_N = 1 + NQi,$$

where the relation $U_N(-1) = (-1)^N(N + 1)$ has been used. In this case the transmission coefficient is

$$\mathfrak{C}_N(2^{-\frac{1}{2}}) = (1 + N^2Q^2)^{-\frac{1}{2}}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Electro-Optical Effects. III

M. M. CARROLL
Division of Applied Mechanics, University of California, Berkeley, California
AND
R. S. RIVLIN
Center for the Application of Mathematics, Lehigh University, Bethlehem, Pennsylvania

(Received 26 January 1968)

The theory of electro-optical effects developed in previous papers is applied to the study of the reflection–refraction of an electromagnetic wave at the interface between two isotropic centrosymmetric materials to which a static electric field is applied. The theory is developed initially for an arbitrary angle of incidence and arbitrary direction of the static field. Detailed implications are obtained in the case when the static field is normal to the interface and when it is parallel to the interface and in the plane formed by the normal to the interface and the direction of propagation.

1. INTRODUCTION

The theory of the propagation of electromagnetic waves in the presence of static electric and magnetic induction fields was formulated by Toupin and Rivlin.[1] They considered materials which are isotropic and centrosymmetric (i.e., holohedral isotropic) in the absence of any fields and assumed that the constitutive equations are linear with respect to the electromagnetic fields, but not necessarily so with respect to the static fields. This theory provides a basis for the study of electro-optical effects, if the impressed static magnetic induction field is taken as zero. In Paper I of this series,[2] we have discussed the refraction–reflection problem for a plane electromagnetic wave incident normally from free space on the plane interface between free space and a half-space occupied by a centrosymmetric isotropic material to which a static electric field in an arbitrary direction is applied. It was found that, in general, there are two refracted rays and their directions are not normal to the interface.

In this paper, we again discuss the refraction–reflection problem. We consider the somewhat more general case of a plane interface between two material half-spaces to which uniform static electric fields of arbitrary direction are applied. Both materials are considered to be centrosymmetric isotropic in the absence of applied fields. The case of refraction–reflection at a plane interface between free space and a material half-space results as a special case. Also, the analysis of this paper is broader than that of the previous paper,[2] in that arbitrary directions of incidence of the wave are considered.

In Sec. 2, the basic equations of the theory are set down and developed with the degree of generality required for the present analysis.

In Sec. 3, the refraction–reflection problem is considered for arbitrary direction of incidence and arbitrary direction of the static field.

For a specified direction of the forward-drawn normal to the wavefront of the incident wave, there are generally two possible waves which have different velocities. Each of these is, in general, elliptically polarized both as regards its electric and magnetic vectors. In the case when the direction of the static field lies in the plane formed by the normal to the interface and the direction of propagation, one of these waves becomes a wave with its electric vector polarized linearly in a direction normal to this plane and its magnetic vector elliptically polarized. We call such a wave *transverse* and, in the more general case when its electric vector is not linearly polarized, we call it a *wave of the first kind.* The other wave becomes a wave with its magnetic vector linearly polarized in a direction normal to the interface and the direction of wave propagation and, its electric vector elliptically polarized. We call such a wave *planar*, and in the more general case when the magnetic vector is not linearly polarized, we call it a *wave of the second kind.*

In Sec. 3 we assume that corresponding to an incident wave which is either of the first or second kind there are, in general, two reflected waves and two transmitted waves. (This assumption is proven under restricted conditions later in the paper.) One of the reflected waves is of the first kind and the other is of the second kind, and one of the transmitted waves is of the first kind and the other is of the second kind. Equations are obtained for the calculation of the electromagnetic fields associated with these waves.

In Secs. 4 and 5 these equations are solved for the cases when the static electric field is normal to the interface and when it is parallel to the interface and in the plane formed by the direction of propagation of the incident wave and the normal to the interface.

[1] R. A. Toupin and R. S. Rivlin, Arch. Ratl. Mech. Anal. 7, 434 (1961).
[2] M. M. Carroll and R. S. Rivlin, J. Math. Phys. **8**, 2088 (1967).

In these cases waves of the first kind become transverse waves and those of the second kind become planar waves.

In each case results are obtained for transverse and planar incident waves. It is found that in these cases there is only one reflected wave and one refracted wave and these are transverse or planar accordingly as the incident wave is transverse or planar.

In Secs. 6 and 7 we discuss the ray directions of the waves (determined by their Poynting vectors) for each of the cases discussed in Secs. 4 and 5. The procedure is to calculate the components of the Poynting vector for an electromagnetic wave (transverse or planar) in the incident material, the electromagnetic field for which is specified as regards its dependence on position in any plane parallel to the interface. In each case it is found that there are two ray directions. One of these corresponds to the incident wave and the other to the reflected wave, provided that their components normal to the interface have opposite signs. In the case when the static electric field is parallel to the interface, as described in Sec. 5, this is found to be the case quite generally, and the angle of reflection is found to be equal to the angle of incidence. On the other hand, when the static electric field is normal to the interface, the angle of reflection is not, in general, equal to the angle of incidence. It is, however, equal, when the material is nondissipative. The inequality between the angles of incidence and reflection thus appears to be associated with the existence in the material of a dissipation mechanism and one would presumably look for this effect at a wavelength lying in an absorption band for the material.

In Sec. 7 we calculate the relations between the angles of incidence, refraction, and reflection explicitly in the cases when the material carrying the refracted wave is replaced by free space and when the material carrying the incident wave is replaced by free space.

2. BASIC EQUATIONS

We consider the propagation of a plane electromagnetic wave of angular frequency ω, in which the complex electric, magnetic induction, magnetic intensity, and electric displacement fields **E**, **B**, **H**, **D** may vary over the wavefront in the manner expressed by

$$(\mathbf{E}, \mathbf{B}, \mathbf{H}, \mathbf{D}) = (\mathbf{e}, \mathbf{b}, \mathbf{h}, \mathbf{d})e^{\imath\omega(\boldsymbol{\eta}\cdot\mathbf{x}-t)}, \tag{2.1}$$

where **e**, **b**, **h**, **d**, and $\boldsymbol{\eta}$ are complex vectors independent of **x** and t. We note that the surfaces of constant phase (the wavefronts) are planes normal to the direction[3] $\boldsymbol{\eta}^+$ and the surfaces of constant amplitude are planes normal to the direction $\boldsymbol{\eta}^-$. In the case when **E**, **B**, **H**, and **D** are constant on a wavefront, $\boldsymbol{\eta}$ takes the form

$$\boldsymbol{\eta} = \eta\mathbf{n}, \tag{2.2}$$

where **n** is the unit normal to the wavefront, η is the complex slowness, and $\boldsymbol{\eta}$ the complex vector slowness.

Introducing (2.1) into Maxwell's equations, we obtain

$$\boldsymbol{\eta}\times\mathbf{e} - \mathbf{b} = 0 \quad \text{and} \quad \boldsymbol{\eta}\times\mathbf{h} + \mathbf{d} = 0. \tag{2.3}$$

We consider the propagation of the wave (2.1) in a centrosymmetric isotropic material to which a static electric field $\boldsymbol{\mathcal{E}}$ is applied. It has been shown[1,2] that the constitutive equations relating **d** and **h** with **e** and **b** are

$$\mathbf{d} = \boldsymbol{\Phi}\cdot\mathbf{e} + \boldsymbol{\Psi}\cdot\mathbf{b}, \quad \mathbf{h} = \boldsymbol{\Omega}\cdot\mathbf{e} + \boldsymbol{\Lambda}\cdot\mathbf{b}, \tag{2.4}$$

where $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Lambda}$ are 3×3 matrices defined in a rectangular Cartesian system x by

$$\Phi_{ij} = \alpha_1\delta_{ij} + \alpha_7\mathcal{E}_i\mathcal{E}_j, \tag{2.5a}$$

$$\Lambda_{ij} = \beta_1\delta_{ij} + \beta_7\mathcal{E}_i\mathcal{E}_j, \tag{2.5b}$$

$$\Psi_{ij} = -\alpha_3\epsilon_{ijk}\mathcal{E}_k, \tag{2.5c}$$

$$\Omega_{ij} = -\beta_2\epsilon_{ijk}\mathcal{E}_k, \tag{2.5d}$$

where[4] $\mathcal{E}_i$ are the components of $\boldsymbol{\mathcal{E}}$ in the system x. In (2.5), the α's and β's are functions of $\imath\omega$ and $\mathcal{E}^2$, where we employ the notation $\mathcal{E}^2 = \boldsymbol{\mathcal{E}}\cdot\boldsymbol{\mathcal{E}}$.

It is easily seen from (2.3) and (2.4) that **e** satisfies the equation

$$\chi_{ij}e_j = 0, \tag{2.6}$$

where

$$\chi_{ij} = \Phi_{ij} + (\epsilon_{jrs}\Psi_{ir}\eta_s + \epsilon_{ipq}\Omega_{qj}\eta_p) + \epsilon_{ipq}\epsilon_{jrs}\Lambda_{qr}\eta_p\eta_s. \tag{2.7}$$

From (2.6) we have

$$|\chi_{ij}| = 0. \tag{2.8}$$

From $(2.3)_1$ and $(2.4)_2$ we readily obtain

$$h_i = (\Omega_{ik} + \Lambda_{ij}\epsilon_{jpk}\eta_p)e_k. \tag{2.9}$$

Introducing (2.5) into (2.7), we have

$$\chi_{ij} = P\mathcal{E}_i\mathcal{E}_j + Q\eta_i\eta_j + R\eta_i\mathcal{E}_j + S\mathcal{E}_i\eta_j + T\delta_{ij}, \tag{2.10}$$

where

$$\begin{aligned} &P = \alpha_7 + \beta_7\boldsymbol{\eta}\cdot\boldsymbol{\eta}, \quad Q = \beta_1 + \beta_7\mathcal{E}^2, \\ &R = \alpha_3 - \beta_7\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta}, \quad S = \beta_2 - \beta_7\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta}, \\ &T = \alpha_1 - (\alpha_3 + \beta_2)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta} + \beta_7(\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2 - (\beta_1 + \beta_7\mathcal{E}^2)\boldsymbol{\eta}\cdot\boldsymbol{\eta}. \end{aligned} \tag{2.11}$$

[3] We denote the real and imaginary parts of a complex quantity by using the superscripts + and −, respectively.

[4] Throughout this paper we will use an analogous notation for the components in the system x of other vectors occurring in the theory.

Introducing (2.10) into (2.8) we obtain

$$\epsilon_{ijk}(P\mathcal{E}_i\mathcal{E}_1 + Q\eta_i\eta_1 + R\eta_i\mathcal{E}_1 + S\mathcal{E}_i\eta_1 + T\delta_{i1}) \times (P\mathcal{E}_j\mathcal{E}_2 + Q\eta_j\eta_2 + R\eta_j\mathcal{E}_2 + S\mathcal{E}_j\eta_2 + T\delta_{j2}) \times (P\mathcal{E}_k\mathcal{E}_3 + Q\eta_k\eta_3 + R\eta_k\mathcal{E}_3 + S\mathcal{E}_k\eta_3 + T\delta_{k3}) = 0. \tag{2.12}$$

Carrying out the multiplication on the left-hand side it is seen that (2.12) may be rewritten as

$$T[T^2 + T\{P\mathcal{E}^2 + Q\boldsymbol{\eta}\cdot\boldsymbol{\eta} + (R+S)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta}\} + (PQ - RS)\{\mathcal{E}^2(\boldsymbol{\eta}\cdot\boldsymbol{\eta}) - (\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2\}] = 0. \tag{2.13}$$

Equation (2.13) yields

$$T = 0 \tag{2.14a}$$

or

$$T^2 + T\{P\mathcal{E}^2 + Q\boldsymbol{\eta}\cdot\boldsymbol{\eta} + (R+S)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta}\} + (PQ - RS)\{\mathcal{E}^2(\boldsymbol{\eta}\cdot\boldsymbol{\eta}) - (\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2\} = 0. \tag{2.14b}$$

From (2.11) it is seen that (2.14a) is an equation of second degree in $\boldsymbol{\eta}$. Also, with (2.11), it can be seen that in (2.14b) the coefficients of the terms of fourth and third degrees in $\boldsymbol{\eta}$ are zero, so that (2.14b) is also of second degree in $\boldsymbol{\eta}$.

Introducing (2.11) into (2.14), we obtain

$$(\beta_1 + \beta_7\mathcal{E}^2)\boldsymbol{\eta}\cdot\boldsymbol{\eta} - \beta_7(\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2 + (\alpha_3 + \beta_2)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta} - \alpha_1 = 0 \tag{2.15a}$$

or

$$(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2)\boldsymbol{\eta}\cdot\boldsymbol{\eta} + (\alpha_7\beta_1 - \alpha_3\beta_2)(\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2 + (\alpha_1 + \alpha_7\mathcal{E}^2)[(\alpha_3 + \beta_2)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta} - \alpha_1] = 0. \tag{2.15b}$$

Introducing (2.5) into (2.9), we obtain

$$h_i = \epsilon_{jpk}[(\beta_2\mathcal{E}_p + \beta_1\eta_p)\delta_{ij} + \beta_7\mathcal{E}_i\mathcal{E}_j\eta_p]e_k. \tag{2.16}$$

Results having essentially the same physical significance as those obtained above can be derived through a slightly different path. We show in the Appendix that the constitutive equations (2.4) can be inverted to give

$$\mathbf{e} = \tilde{\boldsymbol{\Phi}}\cdot\mathbf{d} + \tilde{\boldsymbol{\Psi}}\cdot\mathbf{h}, \quad \mathbf{b} = \tilde{\boldsymbol{\Omega}}\cdot\mathbf{d} + \tilde{\boldsymbol{\Lambda}}\cdot\mathbf{h}, \tag{2.17}$$

where $\tilde{\boldsymbol{\Phi}}$, $\tilde{\boldsymbol{\Psi}}$, $\tilde{\boldsymbol{\Omega}}$, and $\tilde{\boldsymbol{\Lambda}}$ are 3×3 matrices defined in the system x by [cf. Eqs. (8.7) and (8.9)]:

$$\tilde{\Phi}_{ij} = \tilde{\alpha}_1\delta_{ij} + \tilde{\alpha}_7\mathcal{E}_i\mathcal{E}_j, \quad \tilde{\Psi}_{ij} = -\tilde{\alpha}_3\epsilon_{ijk}\mathcal{E}_k,$$
$$\tilde{\Lambda}_{ij} = \tilde{\beta}_1\delta_{ij} + \tilde{\beta}_7\mathcal{E}_i\mathcal{E}_j, \quad \tilde{\Omega}_{ij} = -\tilde{\beta}_2\epsilon_{ijk}\mathcal{E}_k, \tag{2.18}$$

and the $\tilde{\alpha}$'s and $\tilde{\beta}$'s are functions of $\mathcal{E}^2$ and $\iota\omega$ related to the α's and β's by (8.8) and (8.10).

From (2.3) and (2.17), we readily show that $\mathbf{h}$ satisfies the equation

$$\tilde{\chi}_{ij}h_j = 0, \tag{2.19}$$

where

$$\tilde{\chi}_{ij} = \tilde{\Lambda}_{ij} - (\epsilon_{jrs}\tilde{\Omega}_{ir}\eta_s + \epsilon_{ipq}\tilde{\Psi}_{qj}\eta_p) + \epsilon_{ipq}\epsilon_{jrs}\tilde{\Phi}_{qr}\eta_p\eta_s. \tag{2.20}$$

We now have, from (2.19),

$$|\tilde{\chi}_{ij}| = 0. \tag{2.21}$$

From (2.17a) and (2.3b), we obtain

$$e_i = (\tilde{\Psi}_{ik} - \epsilon_{pjk}\tilde{\Phi}_{ip}\eta_j)h_k. \tag{2.22}$$

Introducing the expressions (2.18) into (2.20), we obtain

$$\tilde{\chi}_{ij} = \tilde{P}\mathcal{E}_i\mathcal{E}_j + \tilde{Q}\eta_i\eta_j + \tilde{R}\eta_i\mathcal{E}_j + \tilde{S}\mathcal{E}_i\eta_j + \tilde{T}\delta_{ij}, \tag{2.23}$$

where

$$\tilde{P} = \tilde{\beta}_7 + \tilde{\alpha}_7\boldsymbol{\eta}\cdot\boldsymbol{\eta}, \quad \tilde{Q} = \tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2,$$
$$\tilde{R} = -\tilde{\beta}_2 - \tilde{\alpha}_7\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta}, \quad \tilde{S} = -\tilde{\alpha}_3 - \tilde{\alpha}_7\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta},$$
$$\tilde{T} = \tilde{\beta}_1 + (\tilde{\alpha}_3 + \tilde{\beta}_2)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta} + \tilde{\alpha}_7(\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2 - (\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\boldsymbol{\eta}\cdot\boldsymbol{\eta}. \tag{2.24}$$

Using this expression for $\tilde{\chi}_{ij}$, it can be shown in a manner similar to that used to derive (2.15) that Eq. (2.21) consists of two sheets given by

$$(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\boldsymbol{\eta}\cdot\boldsymbol{\eta} - \tilde{\alpha}_7(\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2 - (\tilde{\alpha}_3 + \tilde{\beta}_2)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta} - \tilde{\beta}_1 = 0 \tag{2.25a}$$

and

$$(\tilde{\alpha}_1\tilde{\beta}_1 + \tilde{\alpha}_3\tilde{\beta}_2\mathcal{E}^2)\boldsymbol{\eta}\cdot\boldsymbol{\eta} + (\tilde{\alpha}_1\tilde{\beta}_7 - \tilde{\alpha}_3\tilde{\beta}_2)(\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2 - (\tilde{\beta}_1 + \tilde{\beta}_7\mathcal{E}^2)[(\tilde{\alpha}_3 + \tilde{\beta}_2)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta} + \tilde{\beta}_1] = 0. \tag{2.25b}$$

Using relations (A.8) and (A.10) to substitute for the $\tilde{\alpha}$'s and $\tilde{\beta}$'s in terms of the α's and β's, it is quite easy to show that Eq. (2.25a) is the same as (2.15b) and (2.25b) is the same as (2.15a).

From (2.18) and (2.22), we obtain

$$e_i = \epsilon_{jpk}[(\tilde{\alpha}_3\mathcal{E}_p - \tilde{\alpha}_1\eta_p)\delta_{ij} - \tilde{\alpha}_7\mathcal{E}_i\mathcal{E}_j\eta_p]h_k. \tag{2.26}$$

In this paper we shall be largely concerned with the case when the direction of the static field $\boldsymbol{\mathcal{E}}$ is perpendicular to the line of intersection of the planes of constant phase and constant amplitude. Then, we may choose the coordinate system x so that the x_2 direction is in the direction of this line of intersection and the vector $\boldsymbol{\mathcal{E}}$ is in the x_1x_3 plane. We may then write

$$\mathcal{E}_i = \mathcal{E}_1\delta_{i1} + \mathcal{E}_3\delta_{i3} \quad \text{and} \quad \eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3}. \tag{2.27}$$

Introducing (2.27) into (2.10), we see that

$$\chi_{12} = \chi_{21} = \chi_{23} = \chi_{32} = 0, \quad \chi_{22} = T,$$

and

$$\chi_{11}\chi_{33} - \chi_{13}\chi_{31} = T^2 + T\{P\mathcal{E}^2 + Q\boldsymbol{\eta}\cdot\boldsymbol{\eta} + (R+S)\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta}\} + (PQ - RS) \times \{\mathcal{E}^2(\boldsymbol{\eta}\cdot\boldsymbol{\eta}) - (\boldsymbol{\mathcal{E}}\cdot\boldsymbol{\eta})^2\}. \quad (2.28)$$

Introducing (2.28) into (2.6), we obtain

$$\begin{aligned} \chi_{11}e_1 + \chi_{13}e_3 &= 0, \\ \chi_{22}e_2 &= 0, \\ \chi_{31}e_1 + \chi_{33}e_3 &= 0, \end{aligned} \quad (2.29)$$

while from (2.28) and (2.14), we have either

$$\chi_{22} = 0 \quad (2.30a)$$

or

$$\chi_{11}\chi_{33} - \chi_{13}\chi_{31} = 0. \quad (2.30b)$$

We see from (2.29) and (2.30) that for the waves corresponding to values of $\boldsymbol{\eta}$ satisfying (2.30a), $e_1 = e_3 = 0$ and consequently they are polarized with their electric vectors parallel to the x_2 direction. We call such waves *transverse waves*. On the other hand, for the waves corresponding to values of $\boldsymbol{\eta}$ given by (2.30b), $e_2 = 0$. Their electric vectors are therefore, in general, elliptically polarized in the x_1x_3 plane. We shall call these waves *planar waves*. It can be shown that the magnetic vectors $\mathbf{h}$ for these waves are polarized in the x_2 direction, while for the transverse waves they are elliptically polarized in the x_1x_3 plane.

When the static field $\boldsymbol{\mathcal{E}}$ is not perpendicular to the line of intersection of the planes of constant phase and of constant amplitude, the vectors $\mathbf{e}$ and $\mathbf{h}$ for the waves corresponding to values of $\boldsymbol{\eta}$ satisfying (2.15) are, in general, elliptically polarized. We shall refer to the waves for which $\boldsymbol{\eta}$ satisfies (2.15a) as *waves of the first kind* and those for which $\boldsymbol{\eta}$ satisfies (2.15b) as *waves of the second kind.*

3. REFLECTION–REFRACTION OF A PLANE WAVE

We consider the reflection and refraction of a plane electromagnetic wave at the plane boundary $x_3 = 0$, in a rectangular Cartesian coordinate system x, between two centrosymmetric isotropic media occupying the half-spaces $x_3 < 0$ and $x_3 > 0$, respectively, to which a strong static electric field is applied. Let us suppose that the static electric field and associated electric displacement field are $\boldsymbol{\mathcal{E}}$ and $\mathfrak{D}$, respectively, for $x_3 < 0$, and $\bar{\boldsymbol{\mathcal{E}}}$ and $\bar{\mathfrak{D}}$, respectively, for $x_3 > 0$. Since the tangential component of the electric field and the normal component of the electric displacement field are continuous at the interface $x_3 = 0$, we have

$$\mathcal{E}_\alpha = \bar{\mathcal{E}}_\alpha\ (\alpha = 1, 2) \quad \text{and} \quad \mathfrak{D}_3 = \bar{\mathfrak{D}}_3, \quad \text{when} \quad x_3 = 0. \quad (3.1)$$

We suppose that the electromagnetic wave is incident on $x_3 = 0$ from the region $x_3 < 0$ and the complex electromagnetic fields $\mathbf{E}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{D}$ associated with the incident wave are given, at the point $\mathbf{x}$ and time t, by

$$(\mathbf{E}, \mathbf{B}, \mathbf{H}, \mathbf{D}) = (\mathbf{e}, \mathbf{b}, \mathbf{h}, \mathbf{d})e^{i\omega(\boldsymbol{\eta}\cdot\mathbf{x}-t)}, \quad (3.2)$$

where $\mathbf{e}$, $\mathbf{b}$, $\mathbf{h}$, and $\mathbf{d}$ are constant complex vectors.

We assume that there are P reflected waves in the region $x_3 < 0$ and Q refracted waves in the region $x_3 > 0$. Let

$$\mathbf{E}^{(A)}, \mathbf{B}^{(A)}, \mathbf{H}^{(A)}, \mathbf{D}^{(A)} \quad \text{and} \quad \bar{\mathbf{E}}^{(B)}, \bar{\mathbf{B}}^{(B)}, \bar{\mathbf{H}}^{(B)}, \bar{\mathbf{D}}^{(B)}$$

be the complex electromagnetic fields associated with these waves in $x_3 < 0$ and $x_3 > 0$, respectively. Then, we may write these in the forms

$$\begin{aligned} &(\mathbf{E}^{(A)}, \mathbf{B}^{(A)}, \mathbf{H}^{(A)}, \mathbf{D}^{(A)}) \\ &\qquad = (\mathbf{e}^{(A)}, \mathbf{b}^{(A)}, \mathbf{h}^{(A)}, \mathbf{d}^{(A)})e^{i\omega(\boldsymbol{\eta}^{(A)}\cdot\mathbf{x}-t)} \\ &\text{and} \\ &(\bar{\mathbf{E}}^{(B)}, \bar{\mathbf{B}}^{(B)}, \bar{\mathbf{H}}^{(B)}, \bar{\mathbf{D}}^{(B)}) \\ &\qquad = (\bar{\mathbf{e}}^{(B)}, \bar{\mathbf{b}}^{(B)}, \bar{\mathbf{h}}^{(B)}, \bar{\mathbf{d}}^{(B)})e^{i\omega(\bar{\boldsymbol{\eta}}^{(B)}\cdot\mathbf{x}-t)}, \end{aligned} \quad (3.3)$$

where

$$\mathbf{e}^{(A)}, \mathbf{b}^{(A)}, \mathbf{h}^{(A)}, \mathbf{d}^{(A)} \quad \text{and} \quad \bar{\mathbf{e}}^{(B)}, \bar{\mathbf{d}}^{(B)}, \bar{\mathbf{h}}^{(B)}, \bar{\mathbf{d}}^{(B)}$$

are constant complex vectors.

Since the tangential components of the electric and magnetic intensity fields and the normal components of the electric displacement and magnetic induction fields, at the interface $x_3 = 0$, are continuous, we have

$$\begin{aligned} E_\alpha + \sum_{A=1}^{P} E_\alpha^{(A)} &= \sum_{B=1}^{Q} \bar{E}_\alpha^{(B)}, \\ H_\alpha + \sum_{A=1}^{P} H_\alpha^{(A)} &= \sum_{B=1}^{Q} \bar{H}_\alpha^{(B)}, \\ D_3 + \sum_{A=1}^{P} D_3^{(A)} &= \sum_{B=1}^{Q} \bar{D}_3^{(B)}, \\ B_3 + \sum_{A=1}^{P} B_3^{(A)} &= \sum_{B=1}^{Q} \bar{B}_3^{(B)}, \end{aligned} \quad (3.4)$$

where[5] the E's, B's, H's, and D's in (3.4) are given by (3.2) and (3.3) with $x_3 = 0$. Since these relations are valid for all x_1 and x_2, it follows that

$$\eta_\alpha^{(A)} = \eta_\alpha \quad \text{and} \quad \bar{\eta}_\alpha^{(B)} = \eta_\alpha. \quad (3.5)$$

We then obtain, from (3.2)–(3.5),

$$e_\alpha + \sum_{A=1}^{P} e_\alpha^{(A)} = \sum_{B=1}^{Q} \bar{e}_\alpha^{(B)}, \quad (3.6a)$$

$$h_\alpha + \sum_{A=1}^{P} h_\alpha^{(A)} = \sum_{B=1}^{Q} \bar{h}_\alpha^{(B)} \quad (3.6b)$$

[5] We employ the convention that Greek subscripts take the values 1, 2.

and

$$d_3 + \sum_{A=1}^{P} d_3^{(A)} = \sum_{B=1}^{Q} \bar{d}_3^{(B)}, \tag{3.7a}$$

$$b_3 + \sum_{A=1}^{P} b_3^{(A)} = \sum_{B=1}^{Q} \bar{b}_3^{(B)}. \tag{3.7b}$$

The incident wave and each reflected and transmitted wave satisfies Maxwell's equations. Thus, from (2.3), we have

$$\begin{aligned} b_3 &= \eta_1 e_2 - \eta_2 e_1, \\ d_3 &= -(\eta_1 h_2 - \eta_2 h_1), \\ b_3^{(A)} &= \eta_1^{(A)} e_2^{(A)} - \eta_2^{(A)} e_1^{(A)}, \\ d_3^{(A)} &= -(\eta_1^{(A)} h_2^{(A)} - \eta_2^{(A)} h_1^{(A)}), \\ \bar{b}_3^{(B)} &= \bar{\eta}_1^{(B)} \bar{e}_2^{(B)} - \bar{\eta}_2^{(B)} \bar{e}_1^{(B)}, \\ \bar{d}_3^{(B)} &= -(\bar{\eta}_1^{(B)} \bar{h}_2^{(B)} - \bar{\eta}_1^{(B)} \bar{h}_1^{(B)}). \end{aligned} \tag{3.8}$$

It is evident, from (3.8) and (3.5), that the conditions (3.6) imply the conditions (3.7).

For the region $x_3 < 0$, the constitutive equations are given by (2.4) and for the region $x_3 > 0$, we adopt the constitutive equations

$$\begin{aligned} \bar{\mathbf{d}}^{(B)} &= \bar{\boldsymbol{\Phi}} \cdot \bar{\mathbf{e}}^{(B)} + \bar{\boldsymbol{\Psi}} \cdot \bar{\mathfrak{b}}^{(B)}, \\ \bar{\mathfrak{h}}^{(B)} &= \bar{\boldsymbol{\Omega}} \cdot \bar{\mathbf{e}}^{(B)} + \bar{\boldsymbol{\Lambda}} \cdot \bar{\mathfrak{b}}^{(B)}, \end{aligned} \tag{3.9}$$

where $\bar{\boldsymbol{\Phi}}$, $\bar{\boldsymbol{\Psi}}$, $\bar{\boldsymbol{\Omega}}$, and $\bar{\boldsymbol{\Lambda}}$ are defined by relations similar to (2.5) with the α's, β's, and $\mathfrak{E}$ replaced by $\bar{\alpha}$'s, $\bar{\beta}$'s, and $\bar{\mathfrak{E}}$, respectively.

For each of the reflected waves, an equation analogous to (2.6) is applicable. Thus, we have

$$\chi_{ij}^{(A)} e_j^{(A)} = 0 \quad (A = 1, \cdots, P), \tag{3.10}$$

where

$$\chi_{ij}^{(A)} = \Phi_{ij} + (\epsilon_{jrs}\Psi_{ir}\eta_s^{(A)} \quad \epsilon_{ipq}\Omega_{qj}\eta_p^{(A)}) + \epsilon_{ipq}\epsilon_{jrs}\Lambda_{qr}\eta_p^{(A)}\eta_s^{(A)}. \tag{3.11}$$

Again, for the transmitted waves, we have

$$\bar{\chi}_{ij}^{(B)} \bar{e}_j^{(B)} = 0 \quad (B = 1, \cdots, Q), \tag{3.12}$$

where $\bar{\chi}_{ij}^{(B)}$ is defined by

$$\bar{\chi}_{ij}^{(B)} = \bar{\Phi}_{ij} + (\epsilon_{jrs}\bar{\Psi}_{ir}\bar{\eta}_s^{(B)} + \epsilon_{ipq}\bar{\Omega}_{qj}\bar{\eta}_p^{(B)}) + \epsilon_{ipq}\epsilon_{jrs}\bar{\Lambda}_{qr}\bar{\eta}_p^{(B)}\bar{\eta}_s^{(B)}. \tag{3.13}$$

Equation (3.10) enables us to determine the ratios between the components of $\mathbf{e}^{(A)}$ $(A = 1, \cdots, P)$, and Eq. (3.12) enables us to determine those for

$$\bar{\mathbf{e}}^{(B)} \quad (B = 1, \cdots, Q).$$

We may therefore write

$$\bar{\mathbf{e}}^{(A)} = \lambda^{(A)}\boldsymbol{\epsilon}^{(A)}, \quad \bar{\mathbf{e}}^{(B)} = \bar{\lambda}^{(B)}\bar{\boldsymbol{\epsilon}}^{(B)}, \tag{3.14}$$

where $\boldsymbol{\epsilon}^{(A)}$ and $\bar{\boldsymbol{\epsilon}}^{(B)}$ are considered determined from (3.10) and (3.12).

Also, a relation of the type (2.16) must apply to the incident wave and to each of the reflected and transmitted waves. We have, therefore, with (3.14) and the notation $e_n = \lambda\epsilon_n$,

$$\begin{aligned} h_i &= \epsilon_{jkn}[(\beta_2\mathfrak{E}_k + \beta_1\eta_k)\delta_{ij} + \beta_7\mathfrak{E}_i\mathfrak{E}_j\eta_k]\lambda\epsilon_n, \\ h_i^{(A)} &= \epsilon_{jkn}[(\beta_2\mathfrak{E}_k + \beta_1\eta_k^{(A)})\delta_{ij} + \beta_7\mathfrak{E}_i\mathfrak{E}_j\eta_k^{(A)}]\lambda^{(A)}\epsilon_n^{(A)}, \\ \bar{h}_i^{(B)} &= \epsilon_{jkn}[(\bar{\beta}_2\bar{\mathfrak{E}}_k + \bar{\beta}_1\bar{\eta}_k^{(B)})\delta_{ij} + \bar{\beta}_7\bar{\mathfrak{E}}_i\bar{\mathfrak{E}}_j\bar{\eta}_k^{(B)}]\bar{\lambda}^{(B)}\bar{\epsilon}_n^{(B)}. \end{aligned} \tag{3.15}$$

We make the assumption that there are two reflected waves and two transmitted waves, i.e., $P = Q = 2$. Then, introducing (3.15) into (3.6b), we obtain

$$\begin{aligned} \epsilon_{jkn}\{&[(\beta_2\mathfrak{E}_k + \beta_1\eta_k)\delta_{\alpha j} + \beta_7\mathfrak{E}_\alpha\mathfrak{E}_j\eta_k]\lambda\epsilon_n \\ &+ \sum_{A=1}^{2}[(\beta_2\mathfrak{E}_k + \beta_1\eta_k^{(A)})\delta_{\alpha j} + \beta_7\mathfrak{E}_\alpha\mathfrak{E}_j\eta_k^{(A)}]\lambda^{(A)}\epsilon_n^{(A)} \\ &- \sum_{B=1}^{2}[(\bar{\beta}_2\bar{\mathfrak{E}}_k + \bar{\beta}_1\bar{\eta}_k^{(B)})\delta_{\alpha j} + \bar{\beta}_7\bar{\mathfrak{E}}_\alpha\bar{\mathfrak{E}}_j\bar{\eta}_k^{(B)}]\bar{\lambda}^{(B)}\bar{\epsilon}_n^{(B)}\} = 0. \end{aligned} \tag{3.16}$$

Also introducing (3.14) into (3.6a), we obtain

$$\lambda\epsilon_\alpha + \sum_{A=1}^{2}\lambda^{(A)}\epsilon_\alpha^{(A)} = \sum_{B=1}^{2}\bar{\lambda}^{(B)}\bar{\epsilon}_\alpha^{(B)}. \tag{3.17}$$

Equations (3.16) and (3.17), with $\alpha = 1, 2$, provide four equations which can be used to determine the four quantities $\lambda^{(A)}$ $(A = 1, 2)$ and $\lambda^{(B)}$ $(B = 1, 2)$ in terms of λ, provided that the values of η_k, $\eta_k^{(A)}$, and $\bar{\eta}_k^{(B)}$ are known. These can be determined in the following manner.

In the interests of explicitness, we shall consider an incident wave in which the surfaces of constant field at fixed time are planes normal to the x_3 axis. This requires, from (3.2), that η_1 and η_2 be real. It follows from (3.5) that $\eta_1^{(A)}$, $\eta_2^{(A)}$ and $\bar{\eta}_1^{(B)}$, $\bar{\eta}_2^{(B)}$ are all real.

For the material occupying the region $x_3 < 0$, the relation (2.15a) between the components of $\boldsymbol{\eta}$ for waves of the first kind may be rewritten as an equation for η_3 if η_1 and η_2 are known; viz.,

$$\begin{aligned} &[(\beta_1 + \beta_7\mathfrak{E}^2) - \beta_7\mathfrak{E}_3^2]\eta_3^2 + [(\alpha_3 + \beta_2) - 2\beta_7\mathfrak{E}_\alpha\eta_\alpha]\mathfrak{E}_3\eta_3 \\ &\quad + \{[(\beta_1 + \beta_7\mathfrak{E}^2)\delta_{\alpha\beta} - \beta_7\mathfrak{E}_\alpha\mathfrak{E}_\beta]\eta_\alpha\eta_\beta \\ &\quad + (\alpha_3 + \beta_2)\mathfrak{E}_\alpha\eta_\alpha - \alpha_1\} = 0. \end{aligned} \tag{3.18}$$

The corresponding equation for waves of the second kind in the region $x_3 < 0$, equation (2.25a), may similarly be rewritten as

$$\begin{aligned} &[(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathfrak{E}^2) - \tilde{\alpha}_7\mathfrak{E}_3^2]\eta_3^2 - [(\tilde{\alpha}_3 + \tilde{\beta}_2) + 2\tilde{\alpha}_7\mathfrak{E}_\alpha\eta_\alpha]\mathfrak{E}_3\eta_3 \\ &\quad + \{[(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathfrak{E}^2)\delta_{\alpha\beta} - \tilde{\alpha}_7\mathfrak{E}_\alpha\mathfrak{E}_\beta]\eta_\alpha\eta_\beta \\ &\quad - (\tilde{\alpha}_3 + \tilde{\beta}_2)\mathfrak{E}_\alpha\eta_\alpha - \tilde{\beta}_1\} = 0. \end{aligned} \tag{3.19}$$

The corresponding equations for waves of the first and second kinds in the region $x_3 > 0$ may be written by analogy with (3.18) and (3.19), respectively, by replacing the α's, β's, and $\mathcal{E}$'s by $\bar{\alpha}$'s, $\bar{\beta}$'s, and $\bar{\mathcal{E}}$'s and the $\tilde{\alpha}$'s and $\tilde{\beta}$'s by quantities defined in terms of the $\bar{\alpha}$'s and $\bar{\beta}$'s in precisely the same way as the $\tilde{\alpha}$'s and $\tilde{\beta}$'s are defined in terms of the α's and β's.

Suppose η_1 and η_2 are given values appropriate to the incident wave. Then (3.18) has two solutions for η_3. We shall assume that the Poynting vector corresponding to one of these solutions has its 3-component in the positive direction of the x_3 axis, while that corresponding to the other solution has its 3-component in the negative direction. Again, for specified values of η_1 and η_2, Eq. (3.19) yields two solutions for η_3. We shall assume that the 3-component of the Poynting vector corresponding to one of these is in the positive direction of the x_3 axis, while that corresponding to the other solution is in the negative direction. We make precisely similar assumptions regarding the solutions of the corresponding equations for waves of the first and second kinds in the material occupying the region $x_3 > 0$. These assumptions, which lead to $P = Q = 2$, are proven in Secs. 6 and 7 for the case when the material is nondissipative and stable and the static field $\mathcal{E}$ is normal to the interface, or in the direction of the x_1 axis.

If the incident wave is of the first kind, then the value of η_3 appropriate to the incident wave is the solution of (3.18) for which the 3-component of the Poynting vector is positive. The values of $\eta_3^{(1)}$ and $\eta_3^{(2)}$ appropriate to the reflected waves are the solutions of (3.18) and (3.19) for which the 3-components of the Poynting vectors are negative. The values of $\bar{\eta}_3^{(1)}$ and $\bar{\eta}_3^{(2)}$ appropriate to the transmitted waves are the solutions of the equations corresponding to (3.18) and (3.19) for the material occupying $x_3 > 0$, for which the 3-components of the Poynting vectors are positive.

If the incident wave is of the second kind, the value of η_3 appropriate to it is the solution of (3.19) for which the 3-component of the Poynting vector is positive, while the values appropriate to the two reflected and two transmitted waves remain the same as in the case when the incident wave is of the first kind.

In Secs. 4 and 5 we carry out in detail the analyses indicated in this section, for the cases when $\mathcal{E}$ is normal to the interface (i.e., parallel to x_3) and when it is parallel to x_1. In both cases, the direction of propagation is in the x_1x_3 plane. For these cases the waves of the first kind become transverse waves and those of the second kind become planar waves.

4. STATIC FIELD NORMAL TO INTERFACE

We choose the rectangular Cartesian system x so that the x_3 axis is normal to the interface between the two media, as in Sec. 3, the direction of propagation lies in the x_1x_3 plane, and the planes of constant amplitude are parallel to the interface. The static field $\mathcal{E}$ is taken normal to the interface. Then we have

$$\eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3}, \quad \mathcal{E}_i = \mathcal{E}\delta_{i3}. \tag{4.1}$$

Introducing (4.1) into (2.10) and (2.11) and using the resulting expression for χ_{ij} in (2.6), we obtain

$$\begin{aligned}\{&(\alpha_1\delta_{ij} + \alpha_7\mathcal{E}^2\delta_{i3}\delta_{j3}) + \mathcal{E}[\eta_1(\alpha_3\delta_{i1}\delta_{j3} + \beta_2\delta_{i3}\delta_{j1}) \\ &- \eta_3(\alpha_3 + \beta_2)(\delta_{ij} - \delta_{i3}\delta_{j3})] + [\beta_1\{\eta_1^2(\delta_{i1}\delta_{j1} - \delta_{ij}) \\ &+ \eta_3^2(\delta_{i3}\delta_{j3} - \delta_{ij}) + \eta_1\eta_3(\delta_{i1}\delta_{j3} + \delta_{i3}\delta_{ji})\} \\ &- \beta_7\mathcal{E}^2\eta_1^2\delta_{i2}\delta_{j2}]\}e_j = 0.\end{aligned} \tag{4.2}$$

Equation (4.2) may be written as the three equations

$$[\alpha_1 - (\alpha_3 + \beta_2)\mathcal{E}\eta_3 - \beta_1\eta_3^2]e_1 + \eta_1(\alpha_3\mathcal{E} + \beta_1\eta_3)e_3 = 0, \tag{4.3a}$$

$$[\alpha_1 - (\alpha_3 + \beta_2)\mathcal{E}\eta_3 - \{\beta_1(\eta_1^2 + \eta_3^2) + \beta_7\mathcal{E}^2\eta_1^2\}]e_2 = 0, \tag{4.3b}$$

$$\eta_1(\beta_2\mathcal{E} + \beta_1\eta_3)e_1 + (\alpha_1 + \alpha_7\mathcal{E}^2 - \beta_1\eta_1^2)e_3 = 0. \tag{4.3c}$$

Introducing (4.1) into (3.18) and (3.19), we see that for the transverse waves (waves of the first kind) η_3 is given in terms of η_1 by[6]

$$\beta_1\eta_3^2 + (\alpha_3 + \beta_2)\mathcal{E}\eta_3 + (\beta_1 + \beta_7\mathcal{E}^2)\eta_1^2 - \alpha_1 = 0, \tag{4.4}$$

and for the planar waves (waves of the second kind) by

$$\tilde{\alpha}_1\eta_3^2 - (\tilde{\alpha}_3 + \tilde{\beta}_2)\mathcal{E}\eta_3 + (\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\eta_1^2 - \tilde{\beta}_1 = 0. \tag{4.5}$$

It is evident from the discussion of transverse waves at the end of Sec. 2 that, for them, we may take

$$\epsilon_j = (0, 1, 0). \tag{4.6}$$

Similarly, from the discussion of planar waves, we may take, for them,

$$\epsilon_j = (\epsilon_1, 0, \epsilon_3), \tag{4.7}$$

where, from (4.3c),

$$\frac{\epsilon_3}{\epsilon_1} = \frac{e_3}{e_1} = -\frac{\eta_1(\beta_2\mathcal{E} + \beta_1\eta_3)}{\alpha_1 + \alpha_7\mathcal{E}^2 - \beta_1\eta_1^2}. \tag{4.8}$$

Analogous results apply to the waves in the material occupying the region $x_3 > 0$.

[6] These results may also be obtained directly from (4.3) by equating its discriminant to zero and, in the case of Eq. (4.5), using (8.8) and (8.10).

We can now write Eq. (3.16) as

$$(\beta_2\mathcal{E}+\beta_1\eta_3)\epsilon_2\lambda+\sum_{A=1}^{2}(\beta_2\mathcal{E}+\beta_1\eta_3^{(A)})\epsilon_2^{(A)}\lambda^{(A)}=\sum_{B=1}^{2}(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(B)})\bar\epsilon_2^{(B)}\bar\lambda^{(B)},$$

$$[(\beta_2\mathcal{E}+\beta_1\eta_3)\epsilon_1-\beta_1\eta_1\epsilon_3]\lambda+\sum_{A=1}^{2}[(\beta_2\mathcal{E}+\beta_1\eta_3^{(A)})\epsilon_1^{(A)}-\beta_1\eta_1^{(A)}\epsilon_3^{(A)}]\lambda^{(A)}=\sum_{B=1}^{2}[(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(B)})\bar\epsilon_1^{(B)}-\bar\beta_1\bar\eta_1^{(B)}\bar\epsilon_3^{(B)}]\bar\lambda^{(B)}. \quad (4.9)$$

We have also Eqs. (3.17), viz.,

$$\epsilon_1\lambda+\sum_{A=1}^{2}\epsilon_1^{(A)}\lambda^{(A)}=\sum_{B=1}^{2}\bar\epsilon_1^{(B)}\bar\lambda^{(B)},$$

$$\epsilon_2\lambda+\sum_{A=1}^{2}\epsilon_2^{(A)}\lambda^{(A)}=\sum_{B=1}^{2}\bar\epsilon_2^{(B)}\bar\lambda^{(B)}. \quad (4.10)$$

We have seen in Sec. 3 that one of the reflected waves is transverse and one is planar. We will take these to be given by $A=1$ and 2, respectively. Then,

$$\epsilon_1^{(1)}=\epsilon_3^{(1)}=0,\quad \epsilon_2^{(1)}=1,\quad \text{and}\quad \epsilon_2^{(2)}=0. \quad (4.11)$$

Similarly, one of the transmitted waves is transverse and one is planar. We take these to be given by $B=1$ and 2, respectively. Then,

$$\bar\epsilon_1^{(1)}=\bar\epsilon_3^{(1)}=0,\quad \bar\epsilon_2^{(1)}=1,\quad \text{and}\quad \bar\epsilon_2^{(2)}=0. \quad (4.12)$$

We now consider separately the cases when the incident wave is transverse and when it is planar.

A. Incident Wave Transverse

In this case

$$\epsilon_1=\epsilon_3=0,\quad \epsilon_2=1. \quad (4.13)$$

Introducing (4.11), (4.12), and (4.13) into (4.9) and (4.10), we obtain

$$(\beta_2\mathcal{E}+\beta_1\eta_3)\lambda+(\beta_2\mathcal{E}+\beta_1 n_3^{(1)})\lambda^{(1)}=(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(1)})\bar\lambda^{(1)}, \quad (4.14a)$$

$$[(\beta_2\mathcal{E}+\beta_1 n_3^{(2)})\epsilon_1^{(2)}-\beta_1\eta_1^{(2)}\epsilon_3^{(2)}]\lambda^{(2)}=[(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(2)})\bar\epsilon_1^{(2)}-\bar\beta_1\bar\eta_1^{(2)}\bar\epsilon_3^{(2)}]\bar\lambda^{(2)} \quad (4.14b)$$

and

$$\epsilon_1^{(2)}\lambda^{(2)}=\bar\epsilon_1^{(2)}\bar\lambda^{(2)}, \quad (4.15a)$$

$$\lambda+\lambda^{(1)}=\bar\lambda^{(1)}. \quad (4.15b)$$

From (4.14b) and (4.15a), we obtain

$$\lambda^{(2)}=\bar\lambda^{(2)}=0, \quad (4.16)$$

unless

$$\bar\epsilon_1^{(2)}[(\beta_2\mathcal{E}+\beta_1\eta_3^{(2)})\epsilon_1^{(2)}-\beta_1\eta_1^{(2)}\epsilon_3^{(2)}]=\epsilon_1^{(2)}[(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(2)})\bar\epsilon_1^{(2)}-\bar\beta_1\bar\eta_1^{(2)}\bar\epsilon_3^{(2)}]. \quad (4.17)$$

From (4.14a) and (4.15b), we obtain

$$\lambda^{(1)}=-\frac{(\beta_2\mathcal{E}+\beta_1 n_3)-(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(1)})}{(\beta_2\mathcal{E}+\beta_1\eta_3^{(1)})-(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(1)})}\lambda$$

and

$$\bar\lambda^{(1)}=-\frac{\beta_1(\eta_3-\eta_3^{(1)})}{(\beta_2\mathcal{E}+\beta_1\eta_3^{(1)})-(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(1)})}\lambda. \quad (4.18)$$

B. Incident Wave Planar

In this case

$$\epsilon_2=0. \quad (4.19)$$

Introducing (4.11), (4.12), and (4.19) into (4.9) and (4.10), we obtain

$$(\beta_2\mathcal{E}+\beta_1 n_3^{(1)})\lambda^{(1)}=(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(1)})\bar\lambda^{(1)}, \quad (4.20a)$$

$$[(\beta_2\mathcal{E}+\beta_1\eta_3)\epsilon_1-\beta_1\eta_1\epsilon_3]\lambda+[(\beta_2\mathcal{E}+\beta_1\eta_3^{(2)})\epsilon_1^{(2)}-\beta_1\eta_1^{(2)}\epsilon_3^{(2)}]\lambda^{(2)}=[(\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(2)})\bar\epsilon_1^{(2)}-\bar\beta_1\bar\eta_1^{(2)}\bar\epsilon_3^{(2)}]\bar\lambda^{(2)} \quad (4.20b)$$

and

$$\epsilon_1\lambda+\epsilon_1^{(2)}\lambda^{(2)}=\bar\epsilon_1^{(2)}\bar\lambda^{(2)}, \quad (4.21a)$$

$$\lambda^{(1)}=\bar\lambda^{(1)}. \quad (4.21b)$$

From (4.8) and analogous relations for the other planar waves, we can rewrite (4.20b) as

$$(\alpha_1+\alpha_7\mathcal{E}^2)\left(\frac{1}{\eta_1}\epsilon_3\lambda+\frac{1}{\eta_1^{(2)}}\epsilon_3^{(2)}\lambda^{(2)}\right)=(\bar\alpha_1+\bar\alpha_7\bar{\mathcal{E}}^2)\frac{1}{\bar\eta_1^{(2)}}\bar\epsilon_3^{(2)}\bar\lambda^{(2)}. \quad (4.22)$$

From (4.20a) and (4.21b), we obtain

$$\lambda^{(1)}=\bar\lambda^{(1)}=0, \quad (4.23)$$

unless

$$\beta_2\mathcal{E}+\beta_1\eta_3^{(1)}=\bar\beta_2\bar{\mathcal{E}}+\bar\beta_1\bar\eta_3^{(1)}. \quad (4.24)$$

From (4.22) and (4.21a), we obtain, with $\eta_1=\eta_1^{(2)}=\bar\eta_1^{(2)}$ [cf. Eqs. (3.5)],

$$\lambda^{(2)}\epsilon_1^{(2)}=-\lambda\epsilon_1\frac{(\alpha_1+\alpha_7\mathcal{E}^2)\frac{\epsilon_3}{\epsilon_1}-(\bar\alpha_1+\bar\alpha_7\bar{\mathcal{E}}^2)\frac{\bar\epsilon_3^{(2)}}{\bar\epsilon_1^{(2)}}}{(\alpha_1+\alpha_7\mathcal{E}^2)\frac{\epsilon_3^{(2)}}{\epsilon_1^{(2)}}-(\bar\alpha_1+\bar\alpha_7\bar{\mathcal{E}}^2)\frac{\bar\epsilon_3^{(2)}}{\bar\epsilon_1^{(2)}}},$$

$$\bar\lambda^{(2)}\bar\epsilon_1^{(2)}=-\lambda\epsilon_1\frac{(\alpha_1+\alpha_7\mathcal{E}^2)\left(\frac{\epsilon_3}{\epsilon_1}-\frac{\epsilon_3^{(2)}}{\epsilon_1^{(2)}}\right)}{(\alpha_1+\alpha_7\mathcal{E}^2)\frac{\epsilon_3^{(2)}}{\epsilon_1^{(2)}}-(\bar\alpha_1+\bar\alpha_7\bar{\mathcal{E}}^2)\frac{\bar\epsilon_3^{(2)}}{\bar\epsilon_1^{(2)}}}, \quad (4.25)$$

where ϵ_3/ϵ_1, $\epsilon_3^{(2)}/\epsilon_1^{(2)}$, and $\bar\epsilon_3^{(2)}/\bar\epsilon_1^{(2)}$ are given by expressions of the type (4.8).

5. STATIC FIELD IN x_1 DIRECTION

We again choose the rectangular Cartesian reference system x so that the x_3 axis is normal to the interface

between the two media and we take the static field $\mathcal{E}$ parallel to the x_1 direction. We take the direction of propagation to lie in the x_1x_3 plane and the planes $x_3 = \text{const}$ to be planes of constant amplitude. Then,

$$\eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3} \quad \text{and} \quad \mathcal{E}_i = \mathcal{E}\delta_{i1}. \tag{5.1}$$

Introducing (5.1) into (2.10) and (2.11) and using the resulting expression for χ_{ij} in (2.6), we obtain

$$\{(\alpha_1\delta_{1j} + \alpha_7\mathcal{E}^2\delta_{i1}\delta_{j1}) + \mathcal{E}[\eta_3(\alpha_3\delta_{i3}\delta_{j1} + \beta_2\delta_{i1}\delta_{j3}) - \eta_1(\alpha_3 + \beta_2)(\delta_{ij} - \delta_{i1}\delta_{j1})] + [\beta_1\{\eta_1^2(\delta_{i1}\delta_{j1} - \delta_{ij}) + \eta_3^2(\delta_{i3}\delta_{j3} - \delta_{ij}) + \eta_1\eta_3(\delta_{i3}\delta_{j1} + \delta_{i1}\delta_{j3})\} - \beta_7\mathcal{E}^2\eta_3^2\delta_{i2}\delta_{j2}]\}e_j = 0. \tag{5.2}$$

Equation (5.2) may be written as the three equations

$$(\alpha_1 + \alpha_7\mathcal{E}^2 - \beta_1\eta_3^2)e_1 + \eta_3(\beta_2\mathcal{E} + \beta_1\eta_1)e_3 = 0, \tag{5.3a}$$

$$[\alpha_1 - (\alpha_3 + \beta_2)\mathcal{E}\eta_1 - \{\beta_1(\eta_1^2 + \eta_3^2) + \beta_7\mathcal{E}^2\eta_3^2\}]e_2 = 0, \tag{5.3b}$$

$$\eta_3(\alpha_3\mathcal{E} + \beta_1\eta_1)e_1 + [\alpha_1 - (\alpha_3 + \beta_2)\mathcal{E}\eta_1 - \beta_1\eta_1^2]e_3 = 0. \tag{5.3c}$$

Introducing (5.1) into (3.18) and (3.19), we see that for the transverse waves η_3 is given in terms of η_1 by

$$(\beta_1 + \beta_7\mathcal{E}^2)\eta_3^2 + \beta_1\eta_1^2 + (\alpha_3 + \beta_2)\mathcal{E}\eta_1 - \alpha_1 = 0 \tag{5.4}$$

and for the planar waves by

$$(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\eta_3^2 + \tilde{\alpha}_1\eta_1^2 - (\tilde{\alpha}_3 + \tilde{\beta}_2)\mathcal{E}\eta_1 - \tilde{\beta}_1 = 0. \tag{5.5}$$

It is again evident, as in Sec. 4, that for the transverse waves we may take

$$\epsilon_j = (0, 1, 0) \tag{5.6}$$

and for the planar waves

$$\epsilon_j = (\epsilon_1, 0, \epsilon_3), \tag{5.7}$$

where, from (5.3a),

$$\frac{\epsilon_3}{\epsilon_1} = -\frac{\alpha_1 + \alpha_7\mathcal{E}^2 - \beta_1\eta_3^2}{\eta_3(\beta_2\mathcal{E} + \beta_1\eta_1)}. \tag{5.8}$$

From (5.1) and (3.16), we obtain

$$(\beta_1 + \beta_7\mathcal{E}^2)\eta_3\epsilon_2\lambda + \sum_{A=1}^{2}(\beta_1 + \beta_7\mathcal{E}^2)\eta_3^{(A)}\epsilon_2^{(A)}\lambda^{(A)} = \sum_{B=1}^{2}(\bar{\beta}_1 + \bar{\beta}_7\bar{\mathcal{E}}^2)\bar{\eta}_3^{(B)}\bar{\epsilon}_2^{(B)}\bar{\lambda}^{(B)}, \tag{5.9a}$$

$$[\beta_1\eta_3\epsilon_1 - (\beta_2\mathcal{E} + \beta_1\eta_1)\epsilon_3]\lambda + \sum_{A=1}^{2}[\beta_1\eta_3^{(A)}\epsilon_1^{(A)} - (\beta_2\mathcal{E} + \beta_1\eta_1^{(A)})\epsilon_3^{(A)}]\lambda^{(A)} = \sum_{B=1}^{2}[\bar{\beta}_1\bar{\eta}_3^{(B)}\bar{\epsilon}_1^{(B)} - (\bar{\beta}_2\bar{\mathcal{E}} + \bar{\beta}_1\bar{\eta}_1^{(B)})\bar{\epsilon}_3^{(B)}]\bar{\lambda}^{(B)}. \tag{5.9b}$$

We have also Eqs. (3.17), viz.,

$$\epsilon_1\lambda + \sum_{A=1}^{2}\epsilon_1^{(A)}\lambda^{(A)} = \sum_{B=1}^{2}\bar{\epsilon}_1^{(B)}\bar{\lambda}^{(B)}, \tag{5.10a}$$

$$\epsilon_2\lambda + \sum_{A=1}^{2}\epsilon_2^{(A)}\lambda^{(A)} = \sum_{B=1}^{2}\bar{\epsilon}_2^{(B)}\bar{\lambda}^{(B)}. \tag{5.10b}$$

Again, we adopt the notation of Sec. 4 to distinguish the transverse and planar reflected and transmitted waves. This implies (4.11) and (4.12).

A. Incident Wave Transverse

In this case

$$\epsilon_1 = \epsilon_3 = 0, \quad \epsilon_2 = 1 \tag{5.11}$$

and we obtain, from (5.9b), (5.10a), (4.11), and (4.12), the result that

$$\lambda^{(2)} = \bar{\lambda}^{(2)}, \tag{5.12}$$

unless

$$\bar{\epsilon}_1^{(2)}[\beta_1\eta_3^{(2)}\epsilon_1^{(2)} - (\beta_2\mathcal{E} + \beta_1\eta_1^{(2)})\epsilon_3^{(2)}] = \epsilon_1^{(2)}[\bar{\beta}_1\bar{\eta}_3^{(2)}\bar{\epsilon}_1^{(2)} - (\bar{\beta}_2\bar{\mathcal{E}} + \bar{\beta}_1\bar{\eta}_1^{(2)})\bar{\epsilon}_3^{(2)}]. \tag{5.13}$$

Equations (5.9a) and (5.10b) then yield

$$(\beta_1 + \beta_7\mathcal{E}^2)(\eta_3\lambda + \eta_3^{(1)}\lambda^{(1)}) = (\bar{\beta}_1 + \bar{\beta}_7\bar{\mathcal{E}}^2)\bar{\eta}_3^{(1)}\bar{\lambda}^{(1)}$$

and

$$\lambda + \lambda^{(1)} = \bar{\lambda}^{(1)}. \tag{5.14}$$

Whence

$$\lambda^{(1)} = -\lambda\frac{(\beta_1 + \beta_7\mathcal{E}^2)\eta_3 - (\bar{\beta}_1 + \bar{\beta}_7\bar{\mathcal{E}}^2)\bar{\eta}_3^{(1)}}{(\beta_1 + \beta_7\mathcal{E}^2)\eta_3^{(1)} - (\bar{\beta}_1 + \bar{\beta}_7\bar{\mathcal{E}}^2)\bar{\eta}_3^{(1)}},$$

$$\bar{\lambda}^{(1)} = -\lambda\frac{(\beta_1 + \beta_7\mathcal{E}^2)(\eta_3 - \eta_3^{(1)})}{(\beta_1 + \beta_7\mathcal{E}^2)\eta_3^{(1)} - (\bar{\beta}_1 + \bar{\beta}_7\bar{\mathcal{E}}^2)\bar{\eta}_3^{(1)}}. \tag{5.15}$$

B. Incident Wave Planar

In this case

$$\epsilon_2 = 0, \tag{5.16}$$

and it follows from (5.9a), (5.10b), (4.11), and (4.12), that

$$\lambda^{(1)} = \bar{\lambda}^{(1)}, \tag{5.17}$$

unless

$$(\beta_1 + \beta_7\mathcal{E}^2)\eta_3^{(1)} = (\bar{\beta}_1 + \bar{\beta}_7\bar{\mathcal{E}}^2)\bar{\eta}_3^{(1)}. \tag{5.18}$$

Equations (5.9b) and (5.10a) then yield

$$[\beta_1\eta_3\epsilon_1 - (\beta_2\mathcal{E} + \beta_1\eta_1)\epsilon_3]\lambda + [\beta_1\eta_3^{(2)}\epsilon_1^{(2)} - (\beta_2\mathcal{E} + \beta_1\eta_1^{(2)})\epsilon_3^{(2)}]\lambda^{(2)} = [\bar{\beta}_1\bar{\eta}_3^{(2)}\bar{\epsilon}_1^{(2)} - (\bar{\beta}_2\bar{\mathcal{E}} + \bar{\beta}_1\bar{\eta}_1^{(2)})\bar{\epsilon}_3^{(2)}]\bar{\lambda}^{(2)} \tag{5.19a}$$

and

$$\epsilon_1\lambda + \epsilon_1^{(2)}\lambda^{(2)} = \bar{\epsilon}_1^{(2)}\bar{\lambda}^{(2)}. \tag{5.19b}$$

Employing (5.8) and analogous equations for the reflected and transmitted waves, we can rewrite (5.19a) as

$$(\alpha_1 + \alpha_7\mathcal{E}^2)\left(\frac{\epsilon_1\lambda}{\eta_3} + \frac{\epsilon_1^{(2)}\lambda^{(2)}}{\eta_3^{(2)}}\right) = (\bar{\alpha}_1 + \bar{\alpha}_7\bar{\mathcal{E}}^2)\frac{\bar{\epsilon}_1^{(2)}\bar{\lambda}^{(2)}}{\bar{\eta}_3^{(2)}}. \tag{5.20}$$

Then, from (5.20) and (5.19b), we obtain

$$\lambda^{(2)}\epsilon_1^{(2)} = -\lambda\epsilon_1 \frac{\bar{\eta}_3^{(2)}(\alpha_1 + \alpha_7\mathcal{E}^2) - \eta_3(\bar{\alpha}_1 + \bar{\alpha}_7\bar{\mathcal{E}}^2)}{\bar{\eta}_3^{(2)}(\alpha_1 + \alpha_7\mathcal{E}^2) - \eta_3^{(2)}(\bar{\alpha}_1 + \bar{\alpha}_7\bar{\mathcal{E}}^2)} \cdot \frac{\eta_3^{(2)}}{\eta_3},$$

$$\bar{\lambda}^{(2)}\bar{\epsilon}_1^{(2)} = -\lambda\epsilon_1 \frac{(\alpha_1 + \alpha_7\mathcal{E}^2)(\eta_3^{(2)} - \eta_3)}{\bar{\eta}_3^{(2)}(\alpha_1 + \alpha_7\mathcal{E}^2) - \eta_3^{(2)}(\bar{\alpha}_1 + \bar{\alpha}_7\bar{\mathcal{E}}^2)} \cdot \frac{\bar{\eta}_3^{(2)}}{\eta_3}. \tag{5.21}$$

6. RAY DIRECTIONS

The ray direction for the propagation of a wave with electromagnetic field given by (2.1) is in the direction of the vector obtained from the Poynting vector by averaging it over a cycle. The averaged Poynting vector **S** is given by

$$\mathbf{S} = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} [(\mathcal{E} + \mathbf{E}^+) \times \mathbf{H}^+]\, dt. \tag{6.1}$$

As explained in Sec. 3, we shall consider waves for which $\eta_1^- = \eta_2^- = 0$. Since we are concerned only with the direction of the Poynting vectors associated with the various waves and not with their magnitudes, we shall calculate them at the interface $x_3 = 0$. Then, from (2.1) and (6.1), we have

$$\mathbf{S} = \tfrac{1}{2}(\mathbf{e}^+ \times \mathbf{h}^+ + \mathbf{e}^- \times \mathbf{h}^-). \tag{6.2}$$

For waves which are polarized with either their **e** or **h** vectors in a fixed direction, we may, without loss of generality and with considerable simplification of the analysis, take **e** or **h** to be real. In either case, Eq. (6.2) becomes

$$\mathbf{S} = \tfrac{1}{2}(\mathbf{e}^+ \times \mathbf{h}^+). \tag{6.3}$$

We consider separately the case when **e** is taken real and that when **h** is taken real.

If **e** is real, we have from (2.16)

$$h_i^+ = \epsilon_{jkn}[(\beta_2\mathcal{E}_k + \beta_1\eta_k)\delta_{ij} + \beta_7\mathcal{E}_i\mathcal{E}_j\eta_k]^+ e_n. \tag{6.4}$$

Introducing this into (6.3), we obtain

$$S_p = \tfrac{1}{2}\epsilon_{pqr}\epsilon_{jkn}[(\beta_2\mathcal{E}_k + \beta_1\eta_k)\delta_{rj} + \beta_7\mathcal{E}_r\mathcal{E}_j\eta_k]^+ e_n e_q. \tag{6.5}$$

If **h** is real, we have from (2.26):

$$e_i^+ = \epsilon_{jkn}[(\tilde{\alpha}_3\mathcal{E}_k - \tilde{\alpha}_1\eta_k)\delta_{ij} - \tilde{\alpha}_7\mathcal{E}_i\mathcal{E}_j\eta_k]^+ h_n. \tag{6.6}$$

Introducing this into (6.3), we obtain

$$S_p = -\tfrac{1}{2}\epsilon_{pqr}\epsilon_{jkn}[(\tilde{\alpha}_3\mathcal{E}_k - \tilde{\alpha}_1\eta_k)\delta_{rj} - \tilde{\alpha}_7\mathcal{E}_r\mathcal{E}_j\eta_k]^+ h_n h_q. \tag{6.7}$$

We now consider the cases discussed in Secs. 4 and 5 when the electric field is in the x_3 and x_1 directions, respectively, the direction of wave propagation lying in the x_1x_3 plane. In each case we consider separately the transverse and planar waves. The main object of the analysis is to relate the directions of the reflected and transmitted waves to that of the incident wave. However, the calculations in this section are preliminary and these relations are derived from them in Sec. 7.

A. Static Field Parallel to the x_3 Direction

For the transverse waves, we introduce the relations into (6.5):

$$\mathcal{E}_i = \mathcal{E}\delta_{i3}, \quad \eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3}, \quad e_i = e_2\delta_{i2}. \tag{6.8}$$

We obtain

$$S_p = \tfrac{1}{2}[(\beta_1 + \beta_7\mathcal{E}^2)\eta_1\delta_{p1} + (\beta_2\mathcal{E} + \beta_1\eta_3)\delta_{p3}]^+(e_2)^2. \tag{6.9}$$

The planar waves are polarized with their **h** vectors in the x_2 direction. We have therefore, instead of (6.8), the relations

$$\mathcal{E}_i = \mathcal{E}\delta_{i3}, \quad \eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3}, \quad h_i = h_2\delta_{i2}. \tag{6.10}$$

Introducing (6.10) into (6.7), we obtain, for the planar waves,

$$S_p = \tfrac{1}{2}[(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\eta_1\delta_{p1} - (\tilde{\alpha}_3\mathcal{E} - \tilde{\alpha}_1\eta_3)\delta_{p3}]^+(h_2)^2. \tag{6.11}$$

We have seen in Sec. 4 that if η_1 is specified for the transverse waves, then the two possible values of η_3 are given by Eq. (4.4) and if η_1 is specified for the planar waves, the two possible values of η_3 are given by Eq. (4.5).

Equation (4.4) yields

$$\eta_3 = (1/2\beta_1)[-(\alpha_3 + \beta_2)\mathcal{E} \pm B^{\frac{1}{2}}], \tag{6.12}$$

where

$$B = (\alpha_3 + \beta_2)^2\mathcal{E}^2 - 4\beta_1[(\beta_1 + \beta_7\mathcal{E}^2)\eta_1^2 - \alpha_1]. \tag{6.13}$$

Introducing (6.12) into the expression (6.9) for the averaged Poynting vector appropriate to transverse waves, we have,[7] bearing in mind that η_1 is real,

$$S_p = [\tfrac{1}{2}(\beta_1^+ + \beta_7^+\mathcal{E}^2)\eta_1, 0, \tfrac{1}{4}\{(\beta_2 - \alpha_3)\mathcal{E} \pm B^{\frac{1}{2}}\}^+](e_2)^2. \tag{6.14}$$

Similarly, Eq. (4.5) yields

$$\eta_3 = \frac{1}{2\tilde{\alpha}_1}\{(\tilde{\alpha}_3 + \tilde{\beta}_2)\mathcal{E} \pm \tilde{B}^{\frac{1}{2}}\}, \tag{6.15}$$

where

$$\tilde{B} = (\tilde{\alpha}_3 + \tilde{\beta}_2)^2\mathcal{E}^2 - 4\tilde{\alpha}_1[(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\eta_1^2 - \tilde{\beta}_1]. \tag{6.16}$$

Introducing (6.15) into the expression (6.11) for the averaged Poynting vector appropriate to planar

[7] The values of e_2 appropriate in (6.14) are, of course, those for the waves considered and are generally different accordingly as the + and − signs are taken. Equations (6.17), (6.24), and (6.27) below must be interpreted analogously.

waves, we have

$$S_p = [\tfrac{1}{2}(\tilde{\alpha}_1^+ + \tilde{\alpha}_7^+\mathcal{E}^2)\eta_1,\ 0,\ \tfrac{1}{4}\{(\tilde{\beta}_2 - \tilde{\alpha}_3)\mathcal{E} \pm \tilde{B}^{\frac{1}{2}}\}^+](h_2)^2. \tag{6.17}$$

The result (6.17) could, of course, be read off from (6.14) in view of the formal similarity between Eqs. (6.9) and (6.11) and between Eqs. (4.4) and (4.5).

B. Static Field Parallel to x_1 Direction

For the transverse waves, we now have

$$\mathcal{E}_i = \mathcal{E}\delta_{i1}, \quad \eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3}, \quad e_i = e_2\delta_{i2}. \tag{6.18}$$

Introducing (6.18) into (6.5), we obtain

$$S_p = \tfrac{1}{2}[(\beta_2\mathcal{E} + \beta_1\eta_1)\delta_{p1} + (\beta_1 + \beta_7\mathcal{E}^2)\eta_3\delta_{p3}]^+(e_2)^2. \tag{6.19}$$

Again for the planar waves we take, in (6.7),

$$\mathcal{E}_i = \mathcal{E}\delta_{i1}, \quad \eta_i = \eta_1\delta_{i1} + \eta_3\delta_{i3}, \quad h_i = h_2\delta_{i2}. \tag{6.20}$$

We obtain

$$S_p = \tfrac{1}{2}[-(\tilde{\alpha}_3\mathcal{E} - \tilde{\alpha}_1\eta_1)\delta_{p1} + (\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\eta_3\delta_{p3}]^+(h_2)^2. \tag{6.21}$$

We have seen in Sec. 5 that if η_1 is specified for the transverse waves, then the two possible values of η_3 are given by Eq. (5.4) and if η_1 is specified for the planar waves, the two possible values of η_3 are given by Eq. (5.5).

Equation (5.4) yields

$$\eta_3 = \pm C^{\frac{1}{2}}, \tag{6.22}$$

where

$$C = \frac{\alpha_1 - (\alpha_3 + \beta_2)\mathcal{E}\eta_1 - \beta_1\eta_1^2}{\beta_1 + \beta_7\mathcal{E}^2}. \tag{6.23}$$

Introducing (6.22) into the expression (6.19) for the averaged Poynting vector appropriate to transverse waves, we obtain

$$S_p = [\tfrac{1}{2}(\beta_2^+\mathcal{E} + \beta^+\eta_1),\ 0,\ \pm\{(\beta_1 + \beta_7\mathcal{E}^2)C^{\frac{1}{2}}\}^+](e_2)^2. \tag{6.24}$$

Again, Eq. (5.5) yields

$$\eta_3 = \pm\tilde{C}^{\frac{1}{2}}, \tag{6.25}$$

where

$$\tilde{C} = \frac{\tilde{\beta}_1 + (\tilde{\alpha}_3 + \tilde{\beta}_2)\mathcal{E}\eta_1 - \tilde{\alpha}_1\eta_1^2}{\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2}. \tag{6.26}$$

Introducing (6.25) into the expression (6.21) for the averaged Poynting vector appropriate to planar waves, we obtain

$$S_p = [-\tfrac{1}{2}(\tilde{\alpha}_3^+\mathcal{E} - \tilde{\alpha}_1^+\eta_1),\ 0,\ \pm\tfrac{1}{2}\{(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2)\tilde{C}^{\frac{1}{2}}\}^+](h_2)^2. \tag{6.27}$$

Let α be the angle between the ray direction and the positive direction of the x_1 axis. Then, in the case when $\mathcal{E}$ is parallel to the x_1 direction, we see from (6.24) that, for a given value of η_1, the values of α corresponding to the ray directions for the two transverse waves are equal and opposite. Similarly, from (6.27), the values of α for the two planar waves are equal and opposite. Thus, whether the incident wave is transverse or planar, the angle of incidence and the angle of reflection are equal. If we take the components S_1 to be positive, then the positive signs in the expression for S_3 will apply to the incident waves and the negative signs to the reflected waves.

On the other hand, in the case when $\mathcal{E}$ is normal to the interface, it does not appear, from (6.14) and (6.17), that there is any compelling reason, on purely phenomenological grounds, why the values of α for the two transverse waves should be equal and opposite, or why the values of α for the two planar waves should be equal and opposite. A measurement of the difference between the angle of incidence and the angle of reflection for the case of a transverse incident wave would, in fact, provide a measure of $\beta_2^+ - \alpha_3^+$, since equality of these angles implies $\beta_2^+ = \alpha_3^+$. In the case when the incident wave is planar, the difference between the angle of incidence and angle of reflection would provide a measure of $\tilde{\beta}_2^+ - \tilde{\alpha}_3^+$, since equality of these angles implies $\beta_2^+ = \alpha_3^+$. We note, however, from (8.10), that if $\beta_2^+ = \alpha_3^+$ and $\tilde{\beta}_2^+ = \tilde{\alpha}_3^+$, then $\alpha_3^- = \beta_2^-$ unless $(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2)^- = 0$.

We have seen, in a previous paper,[8] that for a nondissipative, stable material (at angular frequency ω and static field $\mathcal{E}$), α_1, β_1, α_7, β_7 are real and α_3 and β_2 are complex conjugates. In this case we have, of course, $\alpha_3^+ = \beta_2^+$, and it follows that the angles of incidence and reflection are equal for transverse waves, both when $\mathcal{E}$ is normal to the interface and parallel to the x_1 axis. Since, in this case, $\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2$ is real, it follows from (8.10) that $\tilde{\alpha}_3^+ = \tilde{\beta}_2^+$ and hence the angles of incidence and reflection are equal for planar waves.

We have made the assumption in this paper that corresponding to a given value of η_1, the two values of S_3 for transverse waves are of opposite signs and the two values of S_3 for planar waves are of opposite signs. The above discussion proves this to be the case for a stable, nondissipative material in both the cases when $\mathcal{E}$ is normal to the interface and parallel to the x_1 axis. In the latter case it is proven to be true generally. In the former case when $\mathcal{E}$ is normal to the interface, it does not appear to be necessarily valid, unless $\beta_2^+ = \alpha_3^+$ and $\tilde{\beta}_2^+ = \tilde{\alpha}_3^+$, from purely phenomenological reasoning, although it seems likely to be valid from heuristic considerations.

[8] M. M. Carroll and R. S. Rivlin, J. Math. Phys. 9, 1701 (1968).

7. REFLECTION-REFRACTION AT BOUNDARY WITH FREE SPACE

In this section we discuss the reflection–refraction problem at the interface between a centrosymmetric isotropic material, to which a static electric field is applied, and free space. As before, the interface is the plane $x_3 = 0$. We will consider the two cases:

(i) the material medium occupies the semi-infinite domain $x_3 < 0$ and the domain $x_3 > 0$ is free space;

(ii) the material medium occupies the domain $x_3 > 0$ and the domain $x_3 < 0$ is free space.

In both cases the wave is incident on the interface from $x_3 < 0$ and we shall consider the cases when the static field applied to the material medium is parallel to x_3 and when it is parallel to x_1, the direction of propagation lying in the x_1x_3 plane.

When the material medium occupies the region $x_3 < 0$ (Case i), we shall consider the cases when the incident wave is transverse and when it is planar. When the material medium occupies the region $x_3 > 0$ (Case ii), we shall consider the corresponding cases when the electric vector of the incident wave is polarized in the direction of x_2 and when it is polarized in a direction in the x_1x_3 plane, perpendicular to the direction of propagation.

Case i

The analysis of the previous sections applies in this case if we take

$$\bar{\alpha}_1 = \bar{\beta}_1 = 1, \quad \bar{\alpha}_7 = \bar{\beta}_7 = \bar{\alpha}_3 = \bar{\beta}_2 = 0. \quad (7.1)$$

We have seen that in both the cases when the incident wave is transverse and when it is planar, there is one refracted wave and one reflected wave. In general, these are transverse if the incident wave is transverse and planar if the incident wave is planar. When the refracted wave is in free space and the incident wave is planar, the electric vector associated with the refracted wave will not, of course, be elliptically polarized, as is the more general case. However, its electric vector will be linearly polarized in a direction perpendicular to the direction of propagation in a plane containing the direction of propagation and the static field $\mathfrak{E}$ in the material medium.

We note that since $x_3 > 0$ is free space, the vectors $\bar{\eta}^{(1)}$ and $\bar{\eta}^{(2)}$ of the previous analysis are the same and we shall denote them $\bar{\eta}$ $(= \bar{\eta}_1, 0, \bar{\eta}_3)$.

We then have

$$\bar{\eta}_1^2 + \bar{\eta}_3^2 = 1 \quad (7.2)$$

and, of course [see Eqs. (3.5)],

$$\bar{\eta}_1 = \eta_1. \quad (7.3)$$

We denote the angle of incidence by θ, the angle of refraction by ϕ, and the angle of reflection by ψ.

We note that in free space the ray direction and direction of propagation are the same and in the case under consideration are both in the direction of $\bar{\eta}$. Then, from (7.2) and (7.3), it follows that

$$\sin\phi = \eta_1. \quad (7.4)$$

We now consider four cases.

$\mathfrak{E}$ *Parallel to* x_3, *Incident Wave Transverse*

From (6.14) and (7.4), we have

$$\tan\theta = \frac{2(\beta_1^+ + \beta_7^+\mathfrak{E}^2)\sin\phi}{[(\beta_2 - \alpha_3)\mathfrak{E} + B^{\frac{1}{2}}]^+}, \quad (7.5)$$

and

$$\tan\psi = \frac{2(\beta_1^+ + \beta_7^+\mathfrak{E}^2)\sin\phi}{[-(\beta_2 - \alpha_3)\mathfrak{E} + B^{\frac{1}{2}}]^+}, \quad (7.6)$$

where B is given by (6.13) and hence, with (7.4), by

$$B = (\alpha_3 + \beta_2)^2\mathfrak{E}^2 - 4\beta_1[(\beta_1 + \beta_7\mathfrak{E}^2)\sin^2\phi - \alpha_1]. \quad (7.7)$$

With (7.7), Eq. (7.5) provides a relation between the angle of refraction and the angle of incidence and Eq. (7.6) provides a relation between the angle of reflection and the angle of refraction.

$\mathfrak{E}$ *Parallel to* x_3, *Incident Wave Planar*

From (6.17) and (7.4), we have

$$\tan\theta = \frac{2(\tilde{\alpha}_1^+ + \tilde{\alpha}_7^+\mathfrak{E}^2)\sin\phi}{[(\tilde{\beta}_2 - \tilde{\alpha}_3)\mathfrak{E} + \tilde{B}^{\frac{1}{2}}]^+} \quad (7.8)$$

and

$$\tan\psi = \frac{2(\tilde{\alpha}_1^+ + \tilde{\alpha}_7^+\mathfrak{E}^2)\sin\phi}{[-(\tilde{\beta}_2 - \tilde{\alpha}_3)\mathfrak{E} + \tilde{B}^{\frac{1}{2}}]^+}, \quad (7.9)$$

where $\tilde{B}$ is given by (6.16) and hence, with (7.4), by

$$\tilde{B} = (\tilde{\alpha}_3 + \tilde{\beta}_2)^2\mathfrak{E}^2 - 4\tilde{\alpha}_1[(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathfrak{E}^2)\sin^2\phi - \tilde{\beta}_1]. \quad (7.10)$$

$\mathfrak{E}$ *Parallel to* x_1, *Incident Wave Transverse*

In this case we have, from (6.24), (6.23), and (7.4),

$$\tan\theta = \tan\psi = \frac{\beta_2^+\mathfrak{E} + \beta_1^+\sin\phi}{[(\beta_1 + \beta_7\mathfrak{E}^2)C^{\frac{1}{2}}]^+}, \quad (7.11)$$

where

$$C = \frac{\alpha_1 - (\alpha_3 + \beta_2)\mathfrak{E}\sin\phi - \beta_1\sin^2\phi}{\beta_1 + \beta_7\mathfrak{E}^2}. \quad (7.12)$$

$\mathfrak{E}$ *Parallel to* x_1, *Incident Wave Planar*

Again, from (6.27), (6.26), and (7.4), we have

$$\tan\theta = \tan\psi = \frac{\tilde{\alpha}_1^+\sin\phi - \tilde{\alpha}_3^+\mathfrak{E}}{[(\tilde{\alpha}_1 + \tilde{\alpha}_7\mathfrak{E}^2)\tilde{C}^{\frac{1}{2}}]^+}, \quad (7.13)$$

where

$$\tilde{C} = \frac{\tilde{\beta}_1 + (\tilde{\alpha}_3 + \tilde{\beta}_2)\mathcal{E}\sin\phi - \tilde{\alpha}_1\sin^2\phi}{\tilde{\alpha}_1 + \tilde{\alpha}_7\mathcal{E}^2}. \quad (7.14)$$

Case ii

In this case we take

$$\alpha_1 = \beta_1 = 1, \quad \alpha_7 = \beta_7 = \alpha_3 = \beta_2 = 0 \quad (7.15)$$

for free space (i.e., in the region $x_3 < 0$) and for simplicity we replace $\tilde{\alpha}$'s and $\tilde{\beta}$'s by α's and β's and $\tilde{\eta}$'s by η's. It is evident that the angle of incidence and angle of reflection are equal, whether the static field applied to the material medium is parallel to x_3 or to x_1 and whether the incident wave is polarized with its electric vector in the direction x_2 or in the plane x_1x_3. It is also evident that the relations between the angle of incidence and angle of refraction are given by (7.5), (7.8), (7.11), and (7.13), if we take ϕ to be the angle of incidence and θ the angle of refraction. Equations (7.5) and (7.8) apply to the cases when the static field applied to the material medium is parallel to x_3 and the incident wave is polarized parallel to x_2, or in the plane x_1x_3. Equations (7.11) and (7.13) apply to the corresponding cases with the static field parallel to x_1.

ACKNOWLEDGMENT

The research described in this paper was supported by a grant from the National Science Foundation to Lehigh University.

APPENDIX: INVERSION OF THE CONSTITUTIVE EQUATIONS

We take as our starting point the constitutive equations (2.4):

$$\mathbf{d} = \boldsymbol{\Phi}\cdot\mathbf{e} + \boldsymbol{\Psi}\cdot\mathbf{b}, \quad \mathbf{h} = \boldsymbol{\Omega}\cdot\mathbf{e} + \boldsymbol{\Lambda}\cdot\mathbf{b}, \quad (A1)$$

where $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Lambda}$ are given by (2.5). If we invert these, they will take the form

$$\mathbf{e} = \tilde{\boldsymbol{\Phi}}\cdot\mathbf{d} + \tilde{\boldsymbol{\Psi}}\cdot\mathbf{h}, \quad \mathbf{b} = \tilde{\boldsymbol{\Omega}}\cdot\mathbf{d} + \tilde{\boldsymbol{\Lambda}}\cdot\mathbf{h}. \quad (A2)$$

By using (A1) to substitute for $\mathbf{d}$ and $\mathbf{h}$ in (A2) and comparing coefficients, we see that

$$\tilde{\boldsymbol{\Phi}}\boldsymbol{\Phi} + \tilde{\boldsymbol{\Psi}}\boldsymbol{\Omega} = \mathbf{I}, \quad \tilde{\boldsymbol{\Omega}}\boldsymbol{\Psi} + \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Lambda} = \mathbf{I},$$
$$\tilde{\boldsymbol{\Phi}}\boldsymbol{\Psi} + \tilde{\boldsymbol{\Psi}}\boldsymbol{\Lambda} = 0, \quad \tilde{\boldsymbol{\Omega}}\boldsymbol{\Phi} + \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Omega} = 0. \quad (A3)$$

From these equations we readily obtain

$$\tilde{\boldsymbol{\Psi}} = -(\boldsymbol{\Phi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega})^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}^{-1}, \quad (A4a)$$
$$\tilde{\boldsymbol{\Phi}} = (\boldsymbol{\Phi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega})^{-1}, \quad (A4b)$$
$$\tilde{\boldsymbol{\Omega}} = -(\boldsymbol{\Lambda} - \boldsymbol{\Omega}\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Omega}\boldsymbol{\Phi}^{-1}, \quad (A4c)$$
$$\tilde{\boldsymbol{\Lambda}} = (\boldsymbol{\Lambda} - \boldsymbol{\Omega}\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi})^{-1}. \quad (A4d)$$

From (2.5a) and (2.5b), we obtain

$$(\boldsymbol{\Phi}^{-1})_{ij} = \frac{1}{\alpha_1}\left(\delta_{ij} - \frac{\alpha_7}{\alpha_1 + \alpha_7\mathcal{E}^2}\mathcal{E}_i\mathcal{E}_j\right)$$
and (A5)
$$(\boldsymbol{\Lambda}^{-1})_{ij} = \frac{1}{\beta_1}\left(\delta_{ij} - \frac{\beta_7}{\beta_1 + \beta_7\mathcal{E}^2}\mathcal{E}_i\mathcal{E}_j\right),$$

where $\mathcal{E}^2 = \mathcal{E}_j\mathcal{E}_j$. Employing (A5) and (2.5c) and (2.5d) we obtain

$$(\boldsymbol{\Phi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega})_{ij} = (1/\beta_1)[(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2)\delta_{ij} + (\alpha_7\beta_1 - \alpha_3\beta_2)\mathcal{E}_i\mathcal{E}_j]$$
and (A6)
$$(\boldsymbol{\Lambda} - \boldsymbol{\Phi}\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi})_{ij} = (1/\alpha_1)[(\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2)\delta_{ij} + (\beta_7\alpha_1 - \alpha_3\beta_2)\mathcal{E}_i\mathcal{E}_j].$$

From (A6), (A4b), and (A4d) we have

$$\tilde{\Phi}_{ij} = \{(\boldsymbol{\Phi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega})^{-1}\}_{ij} = \tilde{\alpha}_1\delta_{ij} + \tilde{\alpha}_7\mathcal{E}_i\mathcal{E}_j,$$
$$\tilde{\Lambda}_{ij} = \{(\boldsymbol{\Lambda} - \boldsymbol{\Omega}\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi})^{-1}\}_{ij} = \tilde{\beta}_1\delta_{ij} + \tilde{\beta}_7\mathcal{E}_i\mathcal{E}_j, \quad (A7)$$

where

$$\tilde{\alpha}_1 = \frac{\beta_1}{\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2}, \quad \tilde{\alpha}_7 = -\tilde{\alpha}_1\frac{\alpha_7\beta_1 - \alpha_3\beta_2}{\beta_1(\alpha_1 + \alpha_7\mathcal{E}^2)},$$
$$\tilde{\beta}_1 = \frac{\alpha_1}{\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2}, \quad \tilde{\beta}_7 = -\tilde{\beta}_1\frac{\beta_7\alpha_1 - \alpha_3\beta_2}{\alpha_1(\beta_1 + \beta_7\mathcal{E}^2)}. \quad (A8)$$

With (A4a) and (A4c), (A5) and (2.5), we obtain

$$\tilde{\Psi}_{ij} = -\tilde{\alpha}_3\epsilon_{ijk}\mathcal{E}_k \quad \text{and} \quad \tilde{\Omega}_{ij} = -\tilde{\beta}_2\epsilon_{ijk}\mathcal{E}_k, \quad (A9)$$

where

$$\tilde{\alpha}_3 = -\frac{\alpha_3}{\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2}, \quad \tilde{\beta}_2 = -\frac{\beta_2}{\alpha_1\beta_1 + \alpha_3\beta_2\mathcal{E}^2}. \quad (A10)$$

We note that the $\tilde{\alpha}$'s and $\tilde{\beta}$'s are functions of $\mathcal{E}^2$ and $\iota\omega$ only.

In a previous paper,[8] it was shown that if the material is nondissipative and stable, α_1, β_1, α_7, and β_7, are real, and α_3 and β_2 are complex conjugates. It follows from (A8) and (A10) that for such a material, $\tilde{\alpha}_1$, $\tilde{\beta}_1$, $\tilde{\alpha}_7$, and $\tilde{\beta}_7$ are real, and $\tilde{\alpha}_3$ and $\tilde{\beta}_2$ are complex conjugates.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 9, NUMBER 12 DECEMBER 1968

Cayley Transform of the S Matrix

P. V. LANDSHOFF
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England

(Received 29 September 1966)

The requirement that the S matrix be unitary is expressed in simple form by use of its Cayley transform (K matrix), which is also of value for practical computations. Here the physical-region structure of the K matrix is analyzed, and it is used to derive the nature of the "simple" physical-region singularities of the S matrix.

1. INTRODUCTION

In a previous paper,[1] here referred to as I, two methods were given of analyzing the physical-region singularity structure of S-matrix elements. The first of the methods was less complete than the second, in that it involved an explicit assumption, related to a causality requirement, about the Riemann-sheet properties of the singularities. The second method also had the advantage over the first that it avoided the introduction of off-mass-shell amplitudes, but was considerably more complex. In this paper a third method is given, which is rather simpler than either of those in I, but which does, to some extent, make use of off-mass-shell amplitudes. The assumptions involved in this method are exactly the same as those of the second method of I, namely unitarity, connectedness structure, and some weak local analyticity properties.

The complexity of the previous methods arises from the complexity of the unitarity equations when account is taken of the connectedness structure. Here the requirement of the unitarity of S is imposed by using its Cayley transform K, defined by

$$S(1 - iK) = 1 + iK \tag{1a}$$

or, formally,

$$iK = (S - 1)(S + 1)^{-1}. \tag{1b}$$

The unitarity of the operator S corresponds to a requirement that the operator K be Hermitian.

It was suggested in I that the operator K is of some practical interest, in that it might provide a method of performing numerical computations that would be rather simpler than the usual dispersion-relation approach. Some calculations on these lines have now been carried out, using two-particle/two-particle matrix elements of K, with very encouraging results.[2] Hence, a further motivation for this paper is that further knowledge of the properties of K might be hoped to lead to further interesting calculations.

It must be stressed that a matrix element of K between a given pair of states is not an analytic function.[3] Rather, it is piecewise analytic, the different pieces being separated by the physical-region Landau curves. This is easy to see for the case of the normal thresholds. On taking matrix elements, we obtain from (1)

$$\langle B|\, S - 1\,|A\rangle = i \sum_C \langle B|\, K\,|C\rangle\langle C|\, S + 1\,|A\rangle, \tag{2}$$

where we have inserted a complete set of intermediate states $|C\rangle$. The equations (2) are a set of integral equations for the matrix elements of K in terms of those of S. As the energy is increased to include a new state $|C\rangle$ in any of Eqs. (2), that integral equation changes form and its solution is not related to its previous solution by analytic continuation. We see below that similar nonanalytic behavior is also associated with physical-region singularities other than normal thresholds.

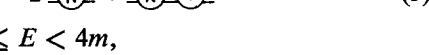
This nonanalytic behavior should not be a bar to calculations, since, as we find below, its nature can be explicitly displayed, at least for a large class of singularities. This is the class of "simple" singularities, namely, those whose Landau–Cutkosky diagram consists of single internal lines joining its vertices. This class of singularities has also been analyzed[4] by the second, highly complicated, method of I. The nonsimple singularities are much more difficult to cope with.[5]

In Sec. 2 a derivation is given of the physical-region one-particle singularities. In Sec. 3 this derivation is simplified by making a series expansion. The latter method is extended, in Sec. 4, to triangle singularities, and applies equally to all the simple singularities at any energy.

[1] P. V. Landshoff and D. I. Olive, J. Math. Phys. **7**, 1464 (1966).
[2] M. O. Taha, Nuovo Cimento **42**, 201 (1966); J. Cordes, Phys. Rev. **156**, 1707 (1967).
[3] Throughout this paper the term *analytic* is taken to mean *almost everywhere* analytic, that is, analytic except for certain discrete singularities, in the same way as the connected parts of S-matrix elements are supposed to be analytic.
[4] M. Boxham, D. I. Olive and J. C. Polkinghorne, in preparation.
[5] P. V. Landshoff, D. I. Olive and J. C. Polkinghorne, J. Math. Phys. **7**, 1600 (1966); J. K. Storrow, Nuovo Cimento **48**, 593 (1967).

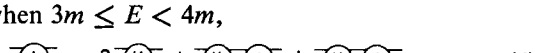
2. ONE-PARTICLE SINGULARITIES

Using the notation introduced by Olive,[6] we write (2) in diagrammatic form. For simplicity, we consider a single type of particle having mass m and no intrinsic quantum numbers. We impose the usual connectedness structure on the S-matrix elements.[7] Then, when the total energy E satisfies $2m \leq E < 3m$, we have

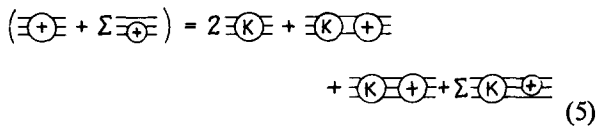


(3)

and when $3m \leq E < 4m$,



(4)



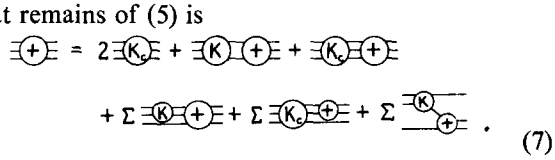
(5)

and so on.

The three terms in the sum on the left-hand side of (5) each contain a δ function corresponding to one of the particles having its four-momentum unchanged by interaction. The right-hand side must have a similar δ function structure in order for the equation to balance, and this leads us to impose on the K-matrix element the connectedness structure:

(6)

Here the amplitude K_c is supposed free of the straight-through δ functions explicitly displayed in (6). We readily see that with (6) and (3) the terms in (5) containing straight-through δ functions do balance. What remains of (5) is

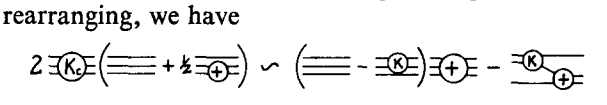


(7)

Our notation implies that each of the terms in the last sum in (7) represents a straight product of the K and (+) amplitudes there appearing, times the factor[8]

$$-2\pi i\delta(q^2 - m^2), \tag{8}$$

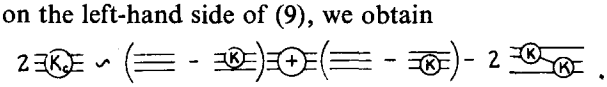
where q is the four-momentum of the internal particle. Since the (+) amplitude on the left of (7) is, by supposition, an analytic function,[3] it must be that the K_c amplitude contains some nonanalyticity at $q^2 = m^2$ to make the right-hand side of the equation analytic. Let us concentrate on the singularity associated with the term displayed explicitly in the last sum in (7). Picking out[9] the terms that are singular at $q^2 = m^2$ and rearranging, we have



(9)

If we "postmultiply" by

which, by (3), is the inverse of the factor in brackets on the left-hand side of (9), we obtain



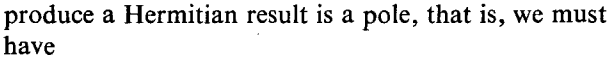
(10)

So far we have made no use of the unitarity of S, that is, the Hermiticity of the operator K. When S and thus K is symmetric, as might result from PT invariance, this means that the matrix elements of K are real functions. Then also, by the definition (6), the amplitude K_c is real. More generally, K_c suffers complex conjugation when initial and final states are interchanged. One property of a Hermitian amplitude we use below is that, if it is analytic,[3] it can have no singularity. In the case where the matrix element is real, this follows very simply from the reflection principle of complex variable theory: If $f(z)$ is an analytic function of $z = x + iy$ and is real for real z, the discontinuity associated with a possible singularity on the real axis would be

$$\begin{aligned}\lim_{y\to 0+} &[f(x + iy) - f(x - iy)] \\ &= \lim_{y\to 0+} [f(x + iy) - f^*(x + iy)] \\ &= 0.\end{aligned}$$

In the more general case, the property can be seen as follows: If $f = g^*$ on the real axis, and $f(z)$ and $g(z)$ are both analytic functions of z, the above argument applied separately to the analytic functions $[f(z) + g(z)]$ and $i[f(z) - g(z)]$ shows that neither can have a singularity on the real axis.

Since the (+) amplitude is supposed to be analytic,[3] the first term on the right of (10) is analytic[9] in the variable q^2. The other term on the right of (10) is not Hermitian, because of the factor i in (8). The only analytic[3] factor that can be combined with (8) to produce a Hermitian result is a pole, that is, we must have



(11)

[6] R. J. Eden, P. V. Landshoff, D. I. Olive and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, New York 1966).

[7] Reference 6, Sec. 4.2.

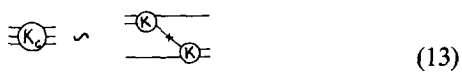
[8] Cf. Ref. 6, Eq. (4.5.2).

[9] The singularities of unitaritylike integrals are discussed in Ref. 6, Sec. 4.10. This work applies equally to the cases here where the integrand is piecewise analytic, provided account is taken of the boundary surfaces separating the pieces.

where

$$\text{—•—} = \pm \frac{1}{q^2 - m^2 \pm i\epsilon}, \tag{12}$$

so that



(13)

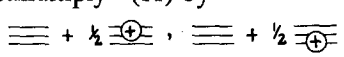
where

$$= \pm\left\{\frac{1}{q^2 - m^2 + i\epsilon} + \frac{1}{q^2 - m^2 - i\epsilon}\right\}$$

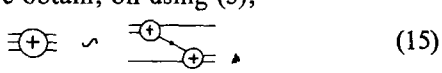
$$= \pm 2P \frac{1}{q^2 - m^2}. \tag{14}$$

A discussion of the apparent necessity of departing from the mass-shell in Eq. (11) may be found in Ref. 1.

It has been known for some time that the one-particle singularities of the K-matrix elements correspond to principal parts of poles.[10] If we "premultiply" and "postmultiply" (11) by



respectively, we obtain, on using (3),



(15)

The ambiguity of sign in (12) is resolved essentially by convention,[11] and the upper sign is usually chosen. This means taking also the upper sign in (14).

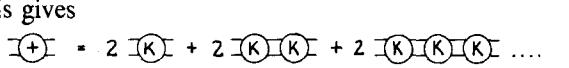
3. SERIES EXPANSION

The foregoing analysis is scarcely simpler than Olive's original derivation[11] of the one-particle singularity structure. To achieve greater simplicity, we derive the result again using the iteration solution to (7). From (1) we have

$$S = 1 + 2iK + 2i^2K^2 + 2i^3K^3 + \cdots \tag{16a}$$

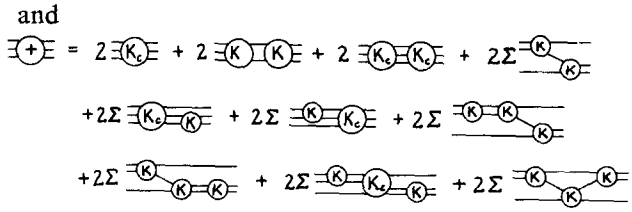
$$= 1 + 2iK + 2i^2 \sum_{\Lambda} K|A\rangle\langle A|K + 2i^3 \sum_{A,B} K|A\rangle\langle A|K|B\rangle\langle B|K + \cdots, \tag{16b}$$

where $A, B, \cdots$ denote physical intermediate states. This gives



(17)

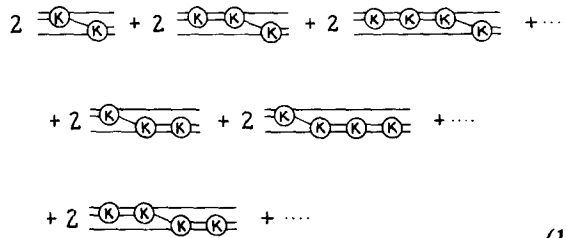
and



(18)

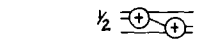
An obvious criticism of this procedure is that the convergence of the series in (16) may be no more than formal. However, we have undertaken the expansion only in order to achieve simplicity, and any results derived from it may, if necessary, be checked by the rather more lengthy methods of the previous section, using the integral equations. Alternatively, if one accepts what is suggested by previous work,[1,5] that the input to the analysis uniquely determines the singularity structure, a sufficient check would be simply to insert the results derived into the integral equations and verify consistency.

An infinite subset of terms on the right-hand side of (18) manifestly contains the one-particle singularity studied above. This subset is



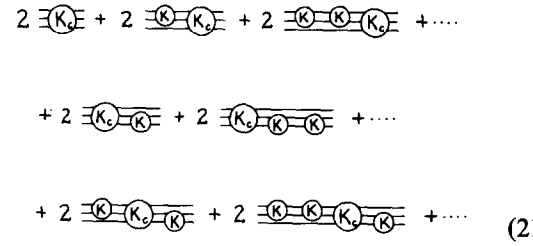
(19)

which, by (17), may be summed to give just



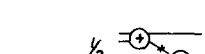
(20)

Now, as before, we see that the K_c amplitude must contain a nonanalytic one-particle structure to compensate this, if the (+) amplitude is to be analytic.[3] In this case a further infinite subset of terms on the right of (18) contain the one-particle singularity, namely,



(21)

By (17), we see that the structure (13) for the K_c amplitude makes this series sum to



(22)

which, together with (20), gives the analytic result (15) for the structure of the (+) amplitude. To see that (17) is the only possibility, note that any additional one-particle singularity we might add to it would have to be nonanalytic, otherwise K_c would not be Hermitian. This would produce an additional nonanalytic term in the sum of the series (21), and so

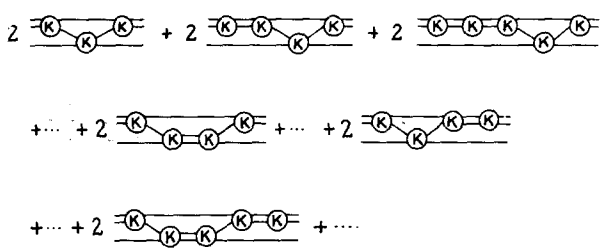
[10] This was proved in perturbation theory by J. C. Polkinghorne, Proc. Cambridge Phil. Soc. **51**, 113 (1955).

[11] Reference 6, Sec. 4.5.

would produce a nonanalytic result for the (+) amplitude.

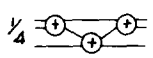
4. THE SIMPLE SINGULARITIES

The foregoing method may be immediately extended to build up all the simple singularities. Consider, as an example, the triangle singularity explicitly displayed in (18). The infinite subset of terms in (18) that manifestly contain this singularity is



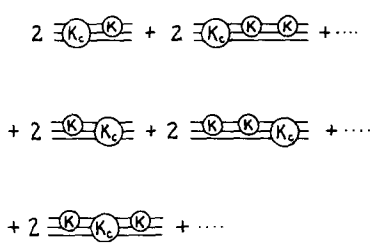
(23)

which, by (17), sums to

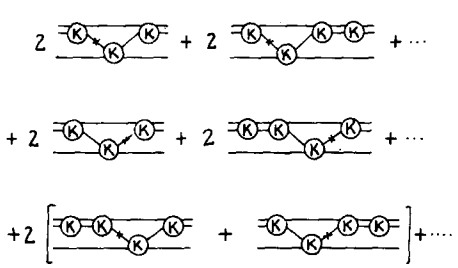


(24)

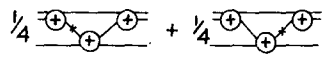
The triangle singularity will further arise in (18) as a consequence of the one-particle singularities of the K_c amplitude. Thus we must also consider the subset of terms



which, because of the result (13), yield the following triangle singularity structure:

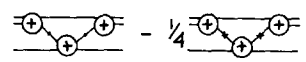


By (17), these terms sum to



(25)

Combining (24) and (25) we have

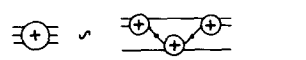


(26)

The first term in (26) is analytic,[3] but the second is not. To compensate for this, if the (+) amplitude is to be analytic, we must give the K_c amplitude a nonanalytic structure

(27)

Again (27) is unique; any further triangle singularity added in would have to be analytic, or K_c could not be Hermitian, and this would yield a nonanalytic contribution to the (+) amplitude. So again we have deduced that



(28)

as in I.

In this way we can evidently build up all the simple physical-region singularities, both for the three-particle/three-particle and the higher amplitudes. As in the example just considered, each singularity will manifestly appear in an infinite subset of terms of the series expansion of the appropriate (+) amplitude. It will also be generated in infinite subsets by the presence of lower singularities in the K_c amplitudes. Then the K_c amplitudes must be given some nonanalyticity corresponding to the singularity itself, in order to make the (+) amplitude analytic.[3] The result is then unique; any extra nonanalyticity added to K_c would make the (+) amplitude nonanalytic, while an analytic singularity would violate the required Hermiticity of K_c.

A particular feature of the method is that it will go through equally well at higher energies. Then hosts of new terms appear in (18), because of the new allowed intermediate states; but the intermediate results (20), (22), (24), and (25) still hold, and so also the final results, because new terms appear similarly in (17).

ACKNOWLEDGMENTS

I am grateful to Dr. I. T. Drummond and Dr. J. C. Polkinghorne for some stimulating discussions.