JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Classical Green's Functions in Response Theory*

JOHN C. HERZEL
Physics Department, University of Melbourne, Parkville, Victoria 3052, Australia

(Received 10 December 1968)

The present paper is concerned with the retarded many-time thermodynamic Green's function method in classical statistical mechanics. In the formulation of the nonlinear response of a dynamical variable, unaveraged Green's functions are specified and used to discuss the physical meaning of many-time Poisson-bracket nests. Equilibrium ensemble averages of these Green's functions are relevant to the classical Kubo response theory, the latter of which is interpreted from the viewpoints of both the classical "Heisenberg" and "Schrödinger" pictures. A method for evaluating the many-time thermodynamic Green's functions based on their definition in terms of the resolvent operator is suggested and discussed in relation to the linear dielectric function and the anharmonic oscillator. The many-time Green's function equations of motion are shown to be equivalent to identities based on the resolvent operator. It is shown how microscopic equations of motion can be used to derive the Bogoliubov–Sadovnikov (B–S) hierarchy from the lowest-order many-time Green's function equation of motion.

I. INTRODUCTION

The employment of thermodynamic Green's functions in classical statistical mechanics has received some attention in recent years.[1–5]

By means of a theorem concerning the linear variation of the mean value of a dynamical quantity, a hierarchy of double-time Green's functions has been derived.[1] The basic Green's function of the hierarchy is $\langle\langle N_X; N_{X'}\rangle\rangle_\omega$, where N_X is the Klimontovich phase-space density operator.[6] (The notation is defined in Appendix A.) A decoupling of this hierarchy corresponds to a particular decoupling of the BBKGY hierarchy.[7] In fact, one can invoke the theorem on variations to transform a kinetic equation for the single-time one-particle distribution function into a corresponding Green's function equation.[1–3] The entire double-time hierarchy is equivalent to a Liouville equation.[4]

The double-time theory enables one to evaluate two-time correlations and these are related to linear transport coefficients.[8,9] Indeed, the theorem on variations is based on the linear response of the expectation value of a dynamical variable to an infinitesimal perturbation. Consideration of nonlinear response shows that multiple-time Green's functions can be invoked to determine nonlinear transport coefficients.[10,11,5] Quantum statistically these Green's functions have been evaluated by means of a hierarchy of equations which couple the n-time Green's function to the $(n+1)$-time Green's functions.[10]

Under certain assumptions the many-time Green's functions can be written as a product of the lowest-order Green's function.[11] One may also determine the many-time Green's functions through their general expressibility in terms of the resolvent operator.[5] This is a generalization of ideas presented for the quantum statistical two-time case.[12] In this connection the many-time equations of motion can be shown to be equivalent to identities based on the resolvent.[5]

Kubo[13] has referred to an important theorem on linear response, which states that "the response of a system in a quantity $B(t)$ to a pulsive force F at $t = 0$ is given by the expectation of the Poisson bracket $\{A(0); B(t)\}$ defined with respect to the natural motion of the system, where A is the conjugate quantity to F." He also pointed out that "a response function may be called a Green's function in its general sense, so that a Poisson bracket essentially represents a Green's function or a response function."

The present paper commences with a generalized qualification of this viewpoint via an examination of the nonlinear response of a dynamical variable in a single system.[14,15] Many-time unaveraged Green's functions are defined. In terms of these functions the physical meaning of multiple-time Poisson-bracket nests is discussed. Unaveraged Green's functions have been mentioned in the literature.[12]

In Sec. III it is shown that the ensemble averages of these classical Green's functions are consistent with those defined in Ref. 10[16] and with the previously defined, classical, double-time ones.[1] The work of Sec. II is seen to lead directly to a classical "Heisenberg" picture of nonlinear response, in contrast with which the procedure analogous to that of Ref. 10 is essentially based on a "Schrödinger" picture. The perturbation of interest in this section is harmonic and turned on adiabatically at $t = -\infty$. The many-time equations of motion of the Green's functions are seen to be derivable in a way analogous to that of the quantum case.[10] Appendix A indicates the method for deriving the B–S hierarchy from the lowest-order many-time equation via microscopic equations of motion.

Section IV derives the general resolvent-operator

form of the Green's functions and demonstrates the equivalence of the equations of motion and the resolvent-operator identities.[17] This viewpoint is linked with earlier work by Ron[8] on the linear dielectric function of a many-body system. Appendix B discusses some relevant mathematical points concerning many-time nests of Poisson brackets.

In Sec. V the same anharmonic-oscillator problem, as discussed in Ref. 10, is treated classically. The equation-of-motion method is contrasted with that based on the resolvent operator.[17]

II. UNAVERAGED GREEN'S FUNCTIONS[18]

Consider a single system governed by a Hamiltonian H which is not explicitly time dependent. Let B be a dynamical variable which also is not explicitly time dependent.

The natural motion of B is then determined by

$$\frac{dB}{dt} = \{B; H\}, \tag{1}$$

where $\{\ \}$ denote Poisson brackets. If one designates the entire set of canonical coordinates of the system $(\mathbf{x}_1, \mathbf{p}_1, \cdots, \cdots, \mathbf{x}_N, \mathbf{p}_N)$ by Γ, then the solution to (1) may be written

$$B(t) = e^{iLt}B, \tag{2}$$

where $B(t) \equiv B[\Gamma(t)]$ and $B \equiv B(0)$.

If a perturbation H_1 is applied to the system, one can use the general perturbation result of Garrido and Gascón,[15] except that here the perturbation is to be turned on at $t = -\infty$. The dynamical variable in the presence of H_1 is then given by

$$B_R(t) = S_1(t)B[\Gamma(t)], \tag{3}$$

where $S_1(t)$ is the solution to the integral equation (the operational procedure of Ref. 15 is not used here)

$$S_1(t) = 1 - \int_{-\infty}^{t} dt' S_1(t')\{H_1[\Gamma(t'); t'];\ \} \tag{4}$$

and

$$H_1 = -AF(t), \tag{5}$$

where A is the dynamical variable conjugate to the force $F(t)$. Iteration of (4) gives

$$S_1(t) = 1 - \int_{-\infty}^{t} dt_1\{H_1[\Gamma(t_1), t_1];\ \} + \cdots + (-1)^n \int_{-\infty}^{t} dt_1 \cdots \int_{-\infty}^{t_{n-1}} \{H_1[\Gamma(t_n), t_n]; \{\cdots\{H_1[\Gamma(t_1); t_1];\ \}\cdots\}\} + \cdots. \tag{6}$$

Thus bearing in mind (3), one can write $\Delta_n B(t)$, the nth-order correction to $B(t)$, as

$$\Delta_n B(t) = (-1)^n \int_{-\infty}^{t} dt_1 \cdots \int_{-\infty}^{t_{n-1}} dt_n \times \{H_1[\Gamma(t_n); t_n]; \{\cdots\{H_1[\Gamma(t_1); t_1]; B(t)\}\cdots\}\}. \tag{7}$$

If one now takes the external potential to be harmonic and adiabatically switched on at $t = -\infty$, one can write[19]

$$H_1 = -Ae^{-i(\omega+i\epsilon)t}. \tag{8}$$

Rearranging the Poisson-bracket nest, and changing the time variables according to $t_{n-1} - t_n = \tau_n$, etc., and $t - t_1 = \tau_1$, one then obtains

$$\Delta_n B(t) = (-1)^n e^{-in(\omega+i\epsilon)t} \int_0^{\infty} d\tau_1 \cdots \int_0^{\infty} d\tau_n \times \{\{\cdots\{B(t); A(t-\tau_1)\}; \cdots\}; A(t-\tau_1\cdots-\tau_n)\} \times \exp[i(\omega + i\epsilon)(n\tau_1 + (n-1)\tau_2 + \cdots + \tau_n)]. \tag{9}$$

Now introducing the Heaviside function $\mathcal{H}$, and then defining the unaveraged Green's function as

$${}^{U}G_{n+1}(B; A(-\tau_1); \cdots A(-\tau_1 \cdots - \tau_n)) \equiv \mathcal{H}(\tau_1)\cdots\mathcal{H}(\tau_n)\{\{\cdots\{B; A(-\tau_1)\}\cdots\}; A(-\tau_1\cdots-\tau_n)\} \tag{10}$$

together with the Fourier transforms

$${}^{U}G_{n+1}(B, A)_{\omega+i\epsilon} \equiv (1/2\pi)^n \int_{-\infty}^{\infty} d\tau_1 \cdots \int_{-\infty}^{\infty} d\tau_n \times {}^{U}G_{n+1}(B; A(-\tau_1); \cdots; A(-\tau_1\cdots\tau_n)) \times \exp[i(\omega + i\epsilon)(n\tau_1 + \cdots + \tau_n)], \tag{11}$$

one has

$$\Delta_n B(t) = (-2\pi)^n e^{-in(\omega+i\epsilon)t} e^{iLt}\, {}^{U}G_{n+1}(B, A)_{\omega+i\epsilon}. \tag{12}$$

The retarded Green's functions defined in Eq. (10) can be interpreted in terms of responses to impulsive forces. In this regard one can generalize the above-mentioned considerations due to Kubo.

Taking $F(t_1) = \delta(t_1)$ in (7), one obtains $\Delta_1 B(t) = \{A; B(t)\}$ as the change in B at time $t > 0$ due to an impulsive force applied at $t = 0$.[19] This result agrees with similar considerations made by Kubo[20] in discussing the physical meaning of the two-time Poisson bracket. Since $\{A; B(t)\}$ is the linear response at time t due to an impulsive force acting at $t = 0$, it should be zero for $t < 0$ in accord with causality requirements. The causality aspect is qualified mathematically through multiplication by the step function

$\mathcal{H}(t)$. Thus the Green's function $\mathcal{H}(t)\{A(-t); B(0)\}$ is the linear response at $t = 0$ due to an impulsive force acting at time $(-t)$. Equation (7) implies that

$$\Delta_n B(0)|_{\text{pulse}} = (-1)^n\{\{\cdots\{\{B(0); A(t_1)\}; A(t_2)\}; \cdots\}; A(t_n)\} \times \mathcal{H}(-t_1)\mathcal{H}(t_1 - t_2)\cdots\mathcal{H}(t_{n-1} - t_n) \quad (13)$$

is the nth-order response, at zero time, due to impulsive forces acting at $t_1, \cdots, t_n$. The causality condition that the nonlinear response $\Delta_n B(0)|_{\text{pulse}}$ is nonzero only if $0 > t_1 > \cdots > t_n$ is imposed by the Heaviside functions. One sees that the result (13) is just the Green's function defined in (10) apart from a phase factor.[21] The term "Green's function" is thus being used here in its general sense.

III. RETARDED THERMODYNAMIC GREEN'S FUNCTIONS

Multiple-time retarded thermodynamic Green's functions can be defined as

$$G_{n+1}(B; A) = \langle {}^{U}G_{n+1}(B, A)\rangle = \mathcal{H}(\tau_1)\cdots\mathcal{H}(\tau_n)\langle\{\{\cdots\{B; A(-\tau_1)\}; \cdots\}; A(-\tau_1 \cdots - \tau_n)\}\rangle, \quad (14)$$

where the average is taken over the equilibrium ensemble $\rho_e = Q^{-1}e^{-H/\theta}$, Q being the partition function. The Fourier transform of (14) then is the ensemble average of (11). The definition (14) is a generalization of the two-time Green's functions defined in (1)[1].

If in Ref. 10 one were not to take $\hbar = 1$, then there the definition of the Green's function would read

$$G_{n+1}(B, A) = (-i/\hbar)^n\mathcal{H}(\tau_1)\cdots\mathcal{H}(\tau_n) \times \langle[[\cdots[B; A(-\tau_1)]; \cdots]; A(-\tau_1 \cdots - \tau_n)]\rangle, \quad (15)$$

and since

$$(-i/\hbar)[A; B] \xrightarrow[\hbar\to 0]{} \{A; B\}, \quad (16)$$

one sees that the functions defined in (14) are the classical limit of (15).

The present viewpoint is essentially that of the "Heisenberg" picture.[22] This is borne out by the fact that the expectation value of B at time t is given by[23]

$$\langle B_R(t)\rangle = \int d\Gamma B_R(t)\rho_e, \quad (17)$$

where ρ_e is the initial (equilibrium) distribution when the perturbation is switched on at $t = -\infty$ and $B_R(t)$ is determined by the classical Heisenberg equation of motion

$$\frac{dB_R}{dt} = \{B_R; H + H_1\}, \quad (18)$$

the solution of which is

$$B_R(t) = B(t) + \sum_{n=1}^{\infty}\Delta_n B(t), \quad (19)$$

where $\Delta_n B(t)$ is given by (7). $B_R(t)$ was defined earlier in (3).

Combining (12), (17), and (19), and using the fact that[23]

$$\int d\Gamma e^{iLt}A = \int d\Gamma A, \quad (20)$$

one sees that the nth-order response to the perturbation, (8), is given by

$$\langle\Delta_n B(t)\rangle = (-2\pi)^n e^{-in(\omega+i\epsilon)t}G_{n+1}(B, A)_{\omega+i\epsilon}, \quad (21)$$

which agrees with the classical limit of (13)[10].

In the equivalent "Schrödinger" picture

$$\langle B_R(t)\rangle = \int d\Gamma B\rho(t), \quad (22)$$

with $\rho(t)$ being determined by the total Hamiltonian $(H + H_I)$ through the Liouville equation $\partial_t\rho = -\{\rho; H + H_I\}$. When the Liouville-operator formalism is used, the derivation of (21) from this direction is similar to that followed in Ref. 10. However, in the derivation of the classical counterpart of (10)[10] from that of (9)[10], one now makes repeated use of

$$\int d\Gamma A\{B; C\} = \int d\Gamma C\{A; B\} \quad (23)$$

in lieu of the trace operation. [An integration by parts readily proves (23).]

Just as in quantum mechanics, one can obtain the classical equations of motion for the Green's functions from the identity

$$n(\omega + i\epsilon)G_{n+1}(B, A)_{\omega+i\epsilon} = (1/2\pi)^n\int_{-\infty}^{\infty}d\tau_1\cdots\int_{-\infty}^{\infty}d\tau_n G_{n+1}(B, A)\frac{1}{i}\frac{\partial}{\partial\tau_1} \times \exp[i(\omega + i\epsilon)(n\tau_1 + \cdots + \tau_n)]. \quad (24)$$

Because of (20) and the fact that ρ_e is stationary with respect to the canonical displacement operator e^{iLt}, the Green's functions are invariant under time translations and (24) leads, via an integration by parts, to

$$-in(\omega + i\epsilon)G_{n+1}(B, A)_{\omega+i\epsilon} = (1/2\pi)G_n(\{B; A\}, A)_{\omega+i\epsilon} + G_{n+1}(\{B; H\}, A)_{\omega+i\epsilon}, \quad (25)$$

where

$$G_1(\{B; A\}, A)_{\omega+i\epsilon} = \langle\{B; A\}\rangle.$$

When one uses appropriate microscopic equations of motion, one sees that the equation of motion given by (25) with $n = 1$ generates the Bogoliubov–Sadovnikov hierarchy. This point is discussed further in Appendix A.

IV. RESOLVENT-OPERATOR FORMS OF THE GREEN'S FUNCTIONS

The explicit form of (14), when Fourier transformed, is

$$\begin{aligned} &G_{n+1}(B, A)_{\omega+i\epsilon} \\ &\quad = (1/2\pi)^n \int_{-\infty}^{\infty} d\tau_1 \cdots \int_{-\infty}^{\infty} d\tau_n \mathcal{H}(\tau_1) \cdots \mathcal{H}(\tau_n) \\ &\qquad \times \langle\{\{\cdots\{B; A(-\tau_1)\}; \cdots\}; A(-\tau_1 \cdots -\tau_n)\}\rangle \\ &\qquad \times \exp\,[i(\omega + i\epsilon)(n\tau_1 + \cdots + \tau_n)]. \end{aligned} \tag{26}$$

Utilizing the invariance of the Green's functions under time translation, one shifts the time by $(\tau_1 + \tau_2 + \cdots + \tau_n)$ so that $A(-\tau_1 \cdots -\tau_n) \to A$, etc. One then has that

$$\begin{aligned} &G_{n+1}(B, A)_{\omega+i\epsilon} \\ &\quad = (1/2\pi)^n \int_{-\infty}^{\infty} d\tau_1 \cdots \int_{-\infty}^{\infty} d\tau_n \mathcal{H}(\tau_1) \cdots \mathcal{H}(\tau_n) \\ &\qquad \times \langle\{e^{iL\tau_n}\{\cdots\{e^{iL\tau_1}B; A\}; \cdots\}; A\}\rangle \\ &\qquad \times \exp\,[i(\omega + i\epsilon)(n\tau_1 + \cdots + \tau_n)]. \end{aligned} \tag{27}$$

Defining the resolvent operator[24]

$$g(\omega) = (L - \omega)^{-1} \tag{28}$$

so that

$$g(\omega) = i\int_{-\infty}^{\infty} d\tau \mathcal{H}(\tau) e^{i\omega\tau} e^{-iL0\tau}, \tag{29}$$

one sees that (27) can be re-expressed as

$$\begin{aligned} &G_{n+1}(B, A)_{\omega} = (-1/2\pi i)^n \\ &\qquad \times \langle\{g(-\omega)\{g(-2\omega) \\ &\qquad \times \{\cdots\{g(-n\omega)B; \bar{A}\}\cdots\}; A\}; A\}\rangle. \end{aligned} \tag{30}$$

In obtaining (30) from (27), one assumes that ϵ is positive definite.

The "Schrödinger"-picture development of the nonlinear response theory leads directly to the equivalent result

$$\begin{aligned} &G_{n+1}(B, A)_{\omega+i\epsilon} = (1/2\pi)^n \int_{-\infty}^{\infty} d\tau_1 \cdots \int_{-\infty}^{\infty} d\tau_n \\ &\quad \times \mathcal{H}(\tau_1) \cdots \mathcal{H}(\tau_n) \int d\Gamma B\{A(-\tau_1); \\ &\quad \{A(-\tau_1 - \tau_2); \cdots \{A(-\tau_1 \cdots -\tau_n); \rho_e\}\cdots\}\} \\ &\quad \times \exp\,[i(\omega + i\epsilon)(n\tau_1 + \cdots + \tau_n)]. \end{aligned} \tag{31}$$

[As implied earlier, the equivalence of (26) and (31) is readily demonstrated via (23).]

Equation (31) is directly expressible as

$$\begin{aligned} &G_{n+1}(B, A)_{\omega} = (1/2\pi i)^n \int d\Gamma B \\ &\qquad \times g(n\omega)\{A; g(\{n-1\}\omega) \\ &\qquad \times \{A; \cdots; g(\omega)\{A; \rho_e\}\cdots\}\}. \end{aligned} \tag{32}$$

On the basis of one or the other of the two forms (30) and (32), one can readily show that the equations of motion (25) are equivalent to identities based on the resolvent operator.

The definition (28) implies that

$$\begin{aligned} &\langle\{g(-\omega)\{g(-2\omega)\{\cdots\{[g(-n\omega)(L + n\omega) - 1] \\ &\qquad \times B; A\}; \cdots\}; A\}; A\}\rangle \equiv 0, \end{aligned} \tag{33}$$

and that

$$\begin{aligned} &\int d\Gamma B[g(n\omega)(L - n\omega) - 1]\{A; g(\{n-1\}\omega) \\ &\qquad \times \{A; \cdots; g(\omega)\{A; \rho_e\}\cdots\}\} \equiv 0. \end{aligned} \tag{34}$$

One obtains (25) from each of these identities by multiplying them out with respect to the square brackets, noting that $L = -i\{\ ; H\}$, and multiplying by the appropriate "$2\pi i$" factors. In the case of (34), one also needs to employ the equation

$$\int d\Gamma ALA' = -\int d\Gamma A'LA,$$

where A and A' are arbitrary functions.

(In Appendix B, a number of mathematical observations concerning many-time Poisson-bracket nests are made.)

The two important forms of the response are abstractly summarized by (17) and (22). Ron[8] has discussed the linear expressions corresponding to each of these, pointing out that the former is more appropriate for physical interpretation and is linked to Kubo's theory of transport coefficients.[13,20] See also Kubo's widely known 1957 paper.[25]

Reference 8 does not mention retarded thermodynamic Green's functions. However, in principle, the form (32), corresponding to (22), was used there to evaluate the linear dielectric function using the Prigogine–Balescu diagram technique.[24]

One can readily show that the linear dielectric function $\epsilon(\mathbf{k}, \omega)$ can be written in terms of

$$\langle\langle n(\mathbf{k}); n(-\mathbf{k})\rangle\rangle,$$

where

$$n(\mathbf{k}, t) = \sum (1 \leq j \leq N) \exp\,[i\mathbf{k} \cdot \mathbf{x}_j(t)]$$

is the spatial Fourier transform[8] of the configuration-space density operator

$$n(\mathbf{x}, t) = \sum (1 \leq j \leq N)\delta(\mathbf{x} - \mathbf{x}_j(t))$$

for an N-body system. (The brackets, $\langle\langle\ \rangle\rangle$, are standard notation for double-time Green's functions.) Thus

$$1/\epsilon(\mathbf{k}, \omega) = 1 + (2\pi i\lambda^2/V)\varphi(\mathbf{k})\langle\langle n(\mathbf{k}); n(-\mathbf{k})\rangle\rangle_\omega,$$

$\lambda^2\varphi$ being the interparticle potential[8] and V the volume of a system of the ensemble.

It is seen that evaluation of many-time Green's functions from the present point of view has the advantage that the already well-developed resolvent-operator techniques can be used.

V. THE ANHARMONIC OSCILLATOR

Here the first-order-in-λ approximation for $G_3(x, x)_\omega$ of the anharmonic oscillator problem[10] will be calculated classically. More specifically, the Green's-function method based on the equations of motion[10] will be contrasted with that based on the resolvent operator approach.

As in Ref. 10, again following Armstrong *et al.*,[26] a simplified Hamiltonian will be employed; namely

$$H = \tfrac{1}{2}\sum_{i=x,y,z}[(p_i^2/m_i) + m_i\omega_{0i}^2x_i^2] - \tfrac{1}{3}\sum_{i=x,y,z}\lambda_i x_i^3 \quad (35)$$

and

$$H_1 = -\sum_{i=x,y,z} e_i x_i E_i e^{-i(\omega+i\epsilon)t}. \quad (36)$$

The notation is essentially that used in Ref. 10.

As mentioned in Sec. III of the present paper,

$$G_n(B, A)_{\omega+i\epsilon}^{\text{Q.M.}} \xrightarrow[\hbar\to 0]{} G_n(B, A)_{\omega+i\epsilon}, \quad (37)$$

so that it is not surprising that, using (25), one obtains equations of exactly the same form as the corresponding quantum-mechanical ones:

$$2\omega G_3(x; x)_\omega = (i/m_x)G_3(p_x; x)_\omega, \quad (38)$$

$$2\omega G_3(p_x; x)_\omega = -im_x\omega_{0x}^2 G_3(x; x)_\omega + i\lambda_x G_3(x^2; x)_\omega. \quad (39)$$

Directly from (38) and (39) one has

$$G_3(x; x)_\omega = (\lambda_x/m_x)[G_3(x^2, x)_\omega/(\omega_{0x}^2 - 4\omega^2)], \quad (40)$$

so that one needs $G_3(x^2; x)_\omega$ only to zeroth order[27] in λ:

$$G_3(x^2, x)_\omega = [(2\pi m_x)^2(\omega^2 - \omega_{0x}^2)^2]^{-1}, \quad (41)$$

and (40) and (41) then give

$$G_3(x; x)_\omega = (2\pi)^{-2}m_x^{-3}\lambda_x[(\omega^2 - \omega_{0x}^2)^2(\omega_{0x}^2 - 4\omega^2)]^{-1}. \quad (42)$$

Thus one sees that the equation-of-motion method leads directly to an expression for $G_3(x; x)_\omega$ in closed form.

Turning to the resolvent-operator approach, one finds that the appropriate form to be employed in the present problem is that of (30). In particular,

$$G_3(x; x)_\omega = -(1/2\pi)^2\langle\{g(-\omega)\{g(-2\omega)x; x\}; x\}\rangle. \quad (43)$$

In the Poisson-bracket nest, one is effectively concerned with the Liouville operator

$$L = L_0 + L',$$

where

$$L_0 = i\left(m\omega_{0x}^2 x\frac{\partial}{\partial p_x} - \frac{p_x}{m_x}\frac{\partial}{\partial_x}\right) \quad (44)$$

and

$$L' = -i\lambda_x x^2\frac{\partial}{\partial p_x}.$$

Then writing

$$g(-\omega) = (L + \omega)^{-1}$$

and

$$g_0(-\omega) = (L_0 + \omega)^{-1}, \quad (45)$$

one obtains the expansion[24]

$$g(-\omega) = \sum_{n=0}^{\infty} g_0(-\omega)\{-L'g_0(-\omega)\}^n. \quad (46)$$

To first order in λ, Eq. (43) then becomes

$$\begin{aligned} G_3(x; x)_\omega = &-(1/2\pi)^2[\langle\{g_0(-\omega)\{g_0(-2\omega)x; x\}; x\}\rangle \\ &- \langle\{g_0(-\omega)\{g_0(-2\omega)L'g_0(-2\omega)x; x\}; x\}\rangle \\ &- \langle\{g_0(-\omega)L'g_0(-\omega)\{g_0(-2\omega)x; x\}; x\}\rangle]. \end{aligned} \quad (47)$$

Since $g_0(-n\omega)x$ and $g_0(-n\omega)p_x$ are first-degree homogeneous polynomials in x and p_x, and since $\{x; x\} = 0$ and $\{p_x; x\} = -1$, it follows that only the second term of (47) gives a nonzero contribution.

In general,

$$\begin{aligned} g_0(-n\omega)x^{n'} = &\ {}^{n'}A_{n'}(n\omega)x^{n'} \\ &+ {}^{n'}A_{n'-1}(n\omega)x^{n'-1}p_x \\ &+ \cdots + {}^{n'}A_0(n\omega){p_x}^{n'} \end{aligned} \quad (48)$$

and, similarly,

$$g_0(-n\omega){p_x}^{n'} = {}^{n'}B_{n'}(n\omega)x^{n'} + \cdots + {}^{n'}B_0(n\omega){p_x}^{n'}, \quad (49)$$

where the A's and B's arise as series in ω, and those relevant to the present approximation are easily expressible in closed forms [see Eq. (42)]. An examination of the relevant elementary Poisson brackets

reveals that the second term of (47) can be written as

$$G_3(x;x)_\omega = [-i\lambda_x/(2\pi)^2]\,{}^1A_0(2\omega) \times [{}^2A_1(2\omega)\,{}^1A_0(\omega) + 2\,{}^2A_0(2\omega)\,{}^1B_0(\omega)]. \quad (50)$$

The A and B coefficients can be obtained by expanding $g_0(-\omega)$ and $g_0(-2\omega)$ as binomial series of differential operators and then summing the terms contributing to a particular coefficient.

In contrast to the equation-of-motion method discussed above, the present approach is thus seen to provide a statement of the problem in the familiar terminology of the resolvent operator. A clear meaning of the first-order-in-λ approximation is afforded by the second term of (47). However, the latter method does not lead immediately to closed-form results.

ACKNOWLEDGMENT

The author is indebted to Dr. K. Hines for his interest in this research.

APPENDIX A

In order to clearly show the relationship between the B–S hierarchy and the first member of the many-time hierarchy (25), the construction of the former through the repeated use of the latter will be indicated here using a method closely analogous to that usually employed in quantum statistical mechanics. See, for example, Zubarev's review article,[28] Secs. 2 and 6.

With the familiar notation $\langle\langle B; A\rangle\rangle_\omega = G_2(B, A)_\omega$, the lowest-order equation of the hierarchy, (25), reads

$$-i\langle\langle B; A\rangle\rangle_\omega = (2\pi)^{-1}\langle\{B; A\}\rangle + \langle\langle\{B; H\}, A\rangle\rangle_\omega. \quad (A1)$$

The dynamical-variable arguments of the Green's functions of the B–S hierarchy are generalizations, $A_{X_1,\cdots,X_s}(t)$, of the Klimontovich phase-space density operator $N(\mathbf{X}, t)$, where $\mathbf{X} = (\mathbf{x}, \mathbf{p})$. Thus,

$$A_{X_1\cdots,X_s}(t) \equiv \sum_{(1\le i_1<\cdots<i_s\le N)} \delta(\mathbf{X}_1 - \mathbf{X}_{i_1}(t))\cdots\delta(\mathbf{X}_s - \mathbf{X}_{i_s}(t)), \quad (A2)$$

with $N(\mathbf{X}, t) \equiv A_{\mathbf{X}}(t)$. As mentioned in Ref. 1, $(\mathbf{x}_1, \mathbf{p}_1, \cdots, \mathbf{x}_s, \mathbf{p}_s)$ are independent variables playing the role of parameters. The lowest-order Green's function of the B–S hierarchy is thus $\langle\langle N(\mathbf{X}); N(\mathbf{X}')\rangle\rangle_\omega$. Its equation of motion can be constructed from that of N as follows.

Taking the Hamiltonian of the system to be

$$H \equiv H_N = \sum_{1\le j\le N} \frac{p_j^2}{2m} + \sum_{1\le j<n\le N} \Phi(|\mathbf{x}_j - \mathbf{x}_n|), \quad (A3)$$

one can write[29]

$$\frac{\partial}{\partial t}N(\mathbf{X}_1) + \frac{\mathbf{p}_1}{m}\cdot\frac{\partial N}{\partial \mathbf{x}_1} = \frac{\partial}{\partial \mathbf{p}_1}N(\mathbf{X}_1)\cdot\frac{\partial}{\partial \mathbf{x}_1}\int d\mathbf{X}_2\Phi(|\mathbf{x}_1 - \mathbf{x}_2|)N(\mathbf{X}_2) = \int d\mathbf{X}_2\left[\frac{\partial}{\partial \mathbf{x}_1}\Phi(|\mathbf{x}_1 - \mathbf{x}_2|)\right]\cdot\frac{\partial}{\partial \mathbf{p}_1} \times \sum_{1\le i<j\le N}\delta(\mathbf{X}_1 - \mathbf{X}_i(t))\delta(\mathbf{X}_2 - \mathbf{X}_j(t)), \quad (A4)$$

or briefly

$$\frac{\partial N}{\partial t} = \{N; H\}. \quad (A5)$$

In obtaining the lowest-order equation of the hierarchy from (A1), one is interested in writing the zero-time Poisson bracket $\{B; H\}$ of (A1) explicitly with $B = N_{\mathbf{X}_1}$. This bracket appears in a phase-space integral wherein the other integrand factors (ρ_e and $N_{\mathbf{X}_1}$) are symmetric under any interchange $\mathbf{X}_i \leftrightarrow \mathbf{X}_j$. Thus, in the present context, one can write

$$\sum_{1\le i<j\le N}\delta(\mathbf{X}_1 - \mathbf{X}_i(t))\delta(\mathbf{X}_2 - \mathbf{X}_j(t)) = 2A_{\mathbf{X}_1\mathbf{X}_2}, \quad (A6)$$

so that (A4) becomes

$$\frac{\partial}{\partial t}N(\mathbf{p}_1, \mathbf{x}_1) = -\frac{\mathbf{p}_1}{m}\cdot\frac{\partial N}{\partial \mathbf{x}_1} + 2\int d\mathbf{X}_1 \times\frac{\partial}{\partial \mathbf{x}_1}\Phi(|\mathbf{x}_1 - \mathbf{x}_2|)\cdot\frac{\partial A_{\mathbf{X}_1\mathbf{X}_2}}{\partial \mathbf{p}_1} = \left\{\frac{\mathbf{p}_1^2}{2m}; N\right\} + 2\int d\mathbf{X}_2\{\Phi(|\mathbf{x}_1 - \mathbf{x}_2|); A_{\mathbf{X}_1\mathbf{X}_2}\}. \quad (A7)$$

When this is substituted into (A1) with $A = A_{\mathbf{X}'}$ and $B = A_{\mathbf{X}_1}$, on the basis of (A5) evaluated at zero time, one obtains

$$-i\langle\langle A_{\mathbf{X}_1}; A_{\mathbf{X}'}\rangle\rangle_\omega = (2\pi)^{-1}\langle\{A_{\mathbf{X}_1}; A_{\mathbf{X}'}\}\rangle + \left\{\frac{\mathbf{p}_1^2}{2m}; \langle\langle A_{\mathbf{X}_1}; A_{\mathbf{X}'}\rangle\rangle_\omega\right\} + 2\int d\mathbf{X}_2\{\Phi(|\mathbf{x}_1 - \mathbf{x}_2|); \langle\langle A_{\mathbf{X}_1\mathbf{X}_2}; A_{\mathbf{X}'}\rangle\rangle_\omega\}. \quad (A8)$$

The first term on the right can be written explicitly in terms of the equilibrium one-particle distribution

function by means of the general formula[4]

$$\langle\{A_{\mathbf{X}_1\cdots\mathbf{X}_s};A_{\mathbf{X}'}\}\rangle=\frac{n^s}{s!}\sum_{1\le j\le s}\{\delta(\mathbf{X}'-\mathbf{X}_j);F_s^0\}. \quad \text{(A9)}$$

From (A8) it is seen that $\langle\langle A_{\mathbf{X}_1};A_{\mathbf{X}'}\rangle\rangle_\omega$ is coupled through the interaction to $\langle\langle A_{\mathbf{X}_1\mathbf{X}_2};A_{\mathbf{X}'}\rangle\rangle_\omega$.

In order to obtain the sth equation of the B–S hierarchy, one substitutes

$$\begin{aligned}&\frac{\partial}{\partial t}A_{\mathbf{X}_1\cdots\mathbf{X}_s}(t)\Big|_{t=0}\\&=\{A_{\mathbf{X}_1\cdots\mathbf{X}_s};H\}\\&=\{H_s(\mathbf{X}_1\cdots\mathbf{X}_s);A_{\mathbf{X}_1\cdots\mathbf{X}_s}\}\\&\quad+(s+1)\sum_{1\le i\le s}\int d\mathbf{X}_{s+1}\{\Phi(|\mathbf{x}_i-\mathbf{x}_{s+1}|);A_{\mathbf{X}_1\cdots\mathbf{X}_{s+1}}\}\end{aligned} \quad \text{(A10)}$$

into (A1) with $B\equiv A_{\mathbf{X}_1\cdots\mathbf{X}_s}$ and $A\equiv A_{\mathbf{X}'}$.

APPENDIX B

In order that the time-displacement properties of e^{iLt} be consistent, one must have, in addition to (2),

$$e^{iLt}\frac{\partial}{\partial\mathbf{p}_j}=\frac{\partial}{\partial\mathbf{p}_j(t)}e^{iLt}, \quad \text{(B1)}$$

and similarly for any $\partial/\partial\mathbf{x}_j$.[14]

The explicit interpretation of (21), for example, is then

$$\begin{aligned}&\langle\Delta_nB(t)\rangle\\&=(-1)^ne^{-in(\omega+i\epsilon)t}\int_{-\infty}^{\infty}d\tau_1\cdots\int_{-\infty}^{\infty}d\tau_n\mathcal{K}(\tau_1)\cdots\mathcal{K}(\tau_n)\\&\quad\times\int d\Gamma B\{A(-\tau_1);\{A(-\tau_1-\tau_2);\cdots\\&\quad\times\{A(-\tau_1\cdots-\tau_n);\rho_e\}_{(-\tau_1\cdots-\tau_n)}\cdots\}_{(-\tau_1-\tau_2)}\}_{(-\tau_1)}\\&\quad\times\exp[i(\omega+i\epsilon)(n\tau_1+\cdots+\tau_n)],\end{aligned} \quad \text{(B2)}$$

where

$$\{A(t);B\}_t\equiv\sum_{1\le j\le N}\left\{\frac{\partial}{\partial\mathbf{x}_j(t)}A[\Gamma(t)]\frac{\partial}{\partial\mathbf{p}_j(t)}B-\frac{\partial}{\partial\mathbf{p}_j(t)}A[\Gamma(t)]\frac{\partial}{\partial\mathbf{x}_j(t)}B\right\}. \quad \text{(B3)}$$

Equations (2) and (B1) imply that the phase-space integral of (B2) may be rewritten as

$$\int d\Gamma B(\Gamma)e^{-iL\tau_1}\{A(\Gamma);\{A(-\tau_2);\cdots\{A(-\tau_2\cdots-\tau_n);\rho_e\}_{(-\tau_2\cdots-\tau_n)}\cdots\}_{(-\tau_2)}\}. \quad \text{(B4)}$$

[The functional form of L is

$$L(\{\partial/\partial_{\mathbf{x}_j}\};\{\partial/\partial_{\mathbf{p}_j}\},\Gamma)$$

where $\mathbf{x}_j\equiv\mathbf{x}_j(\Gamma)$ and $\mathbf{p}_j\equiv\mathbf{p}_j(\Gamma)$.]

Proceeding in this way, one arrives finally at (32) in which the Poisson brackets are to be understood as "referring to" zero-time differential operators $\{\partial/\partial_{\mathbf{x}_j}\}$ and $\{\partial/\partial_{\mathbf{p}_j}\}$.

* This work has been financed with a Melbourne University Research Grant.

[1] N. N. Bogoliubov, Jr., and B. I. Sadovnikov, Zh. Eksp. Teor. Fiz. **43**, 677 (1962) [Sov. Phys.—JETP **16**, 482 (1963)].
[2] B. I. Sadovnikov, Physica **32**, 858 (1966).
[3] V. N. Mel'nikov, Dokl. Akad. Nauk SSSR **171**, 1072 (1966) [Sov. Phys—Dokl. **11**, 1076 (1967)].
[4] J. C. Herzel, J. Math. Phys. **8**, 1650 (1967).
[5] J. C. Herzel, Phys. Letters **27A**, 654 (1968).
[6] Yu. L. Klimontovich, Zh. Eksp. Teor. Fiz. **33**, 982 (1957) [Sov. Phys—JETP **6**, 753 (1958)].
[7] D. C. Montgomery and D. A. Tidman, *Plasma Kinetic Theory* (McGraw-Hill Book Co., Inc., New York, 1964).
[8] A. Ron, J. Math. Phys. **4**, 811 (1963).
[9] A. G. Sitenko, *Electromagnetic Fluctuations in Plasma* (Khar'kov University Press, Khar'kov, USSR, 1965) [English transl.: Academic Press Inc., New York, 1967].
[10] T. Tanaka, K. Moorjani, and T. Morita, Phys. Rev. **155**, 388 (1967).
[11] T. Kawasaki, Progr. Theoret. Phys. (Kyoto) **39**, 331 (1968).
[12] E. R. Pike, Proc. Phys. Soc. (London) **84**, 83 (1964).
[13] R. Kubo, in *Statistical Mechanics of Equilibrium and Non-equilibrium*, J. Meixner, Ed. (North-Holland Publ. Co., Amsterdam, 1965).
[14] L. M. Garrido, Proc. Phys. Soc. (London) **76**, 33 (1960).
[15] L. M. Garrido and F. Gascón, Proc. Phys. Soc. (London) **81**, 1115 (1963).
[16] In the text, (i)^j denotes equation (i) of reference j.
[17] The principal relationships have already been presented in the short communication—Ref. 5.
[18] For later convenience the notation of this section follows that of Ref. 10 rather than that of Refs. 14 and 15.
[19] From now on, following Ref. 10, any constant factor F_0 of F will be included in A for brevity.
[20] R. Kubo, Rept. Progr. Phys. (London) **29**, Pt. I, 255 (1966).
[21] As will be seen from Sec. III, this phase factor has arisen only because of convention.
[22] The classical counterparts of the Heisenberg, Schrödinger, and interaction pictures have been discussed by R. Aronson, J. Math. Phys. **7**, 589 (1966).
[23] Integrations extend over the entire phase space and assume that all distribution functions, Green's functions, and dynamical variables obey periodic boundary conditions in configuration space and homogeneous boundary conditions in momentum space, such that they approach zero exponentially as any canonical momentum variable on which they depend approaches infinity.
[24] P. Résibois, in *Physics of Many-Body Systems*, E. Meeron, Ed. (Gordon & Breach, Science Publishers, Inc., New York, 1966), Vol. I.
[25] R. Kubo, J. Phys. Soc. (Japan) **12**, 570 (1957).
[26] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. **127**, 1918 (1962).
[27] The result (41) is larger by a factor of 2 than that in Ref. 10.
[28] D. N. Zubarev, Usp. Fiz. Nauk **71**, 71 (1960) [Sov. Phys. Usp. **3**, 320 (1960)].
[29] A lucid account of the derivation of (A4) is given by E. P. Gross, J. Nucl. Energy, Pt. C, **2**, 173 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Global and Infinitesimal Nonlinear Chiral Transformations*

A. JOSEPH† AND A. I. SOLOMON‡
Tel Aviv University, Israel

(Received 6 January 1969)

The problem of determining arbitrary nonlinear representations of a given compact Lie group is studied with the object of constructing Lagrangians invariant under the group. To achieve this, expressions for the covariant derivatives are obtained and it is shown how previous treatments based on global and on infinitesimal considerations are related. The noncompact case, the relationship between nonlinearity and zero-mass particles, and the possibility of embedding the representation manifold in a higher-dimensional space are all discussed.

1. INTRODUCTION

The treatment of pion–nucleon scattering in the framework of Lagrangian dynamics[1] has motivated the study of nonlinear representation of Lie groups. This study has involved essentially two different methods of approach. The first of these, initiated by the physical models of Gell-Mann and Lévy,[2] Gürsey and others,[3] has been developed by Coleman, Wess, and Zumino,[4] and Callan *et al.*[5] It may be identified as a global procedure, and the nonlinear representations so derived are closely related to those described by the regular representation. The second method originates from a nonlinear transformation of the pion fields introduced by Schwinger[6] to show that the group structure of a certain Lagrange function was precisely $SU(2) \times SU(2)$. As pointed out by Fairlie and Yoshida,[7] this is just the transformation which is used to describe the extra symmetry of the hydrogen atom in nonrelativistic quantum mechanics.[8] It has been developed by several authors,[9,10] and may be identified as an infinitesimal procedure. Thus the nonlinear representations are described by the solutions of a set of nonlinear partial-differential equations resulting from consistency requirements imposed by the Lie algebraic structure.

In the present work, we interrelate these two approaches and extend the results of both. The discussion is subdivided into several sections. In Sec. 2 we outline briefly the abstract theory underlying nonlinear group representations and discuss the canonical form obtained for these in the case of compact groups by the global procedure. In Sec. 3 we present the representation problem in infinitesimal form and relate this in Sec. 4 to the global problem by use of classical theorems due to Lie.[11] This enables us to derive the infinitesimal form of the canonical representation, which is given explicitly in Sec. 5. In Sec. 6 we outline the infinitesimal analog of the changes of variable used in deriving the canonical form of the global representation. In Sec. 7 the difficulties presented by noncompactness are briefly considered. In Sec. 8 covariant derivatives are derived and invariant Lagrangians constructed. We relate, in Sec. 9, the nonlinearity of the representations to the appearance of massless particles in the theory. In Sec. 10 it is pointed out that linearization of the representation can always be effected by embedding the representation manifold in a sufficiently high-dimensional space. The significance of this result is discussed, and attention is given throughout to the physical applications of the theory.

Notation: Dimension of $G = n_G$: λ, μ, γ go from $1, 2, \cdots, n_G$; dimension of $H = n_H$: $\bar{\alpha}, \bar{\beta}, \bar{\gamma}$ go from $(n_G - n_H + 1), \cdots, n_G$; dimension of $G/H = n_G - n_H$: α, β, γ go from $1, 2, \cdots, (n_G - n_H)$; dimension of $M = n$: i, j, k, go from $1, 2, \cdots, n$; dimension of ξ = same as G/H; dimension of $\psi = n - n_G + n_H$: r, s, t, go from $(n_G - n_H + 1), \cdots, n$.

2. GLOBAL METHOD

Let G be an n_G-parameter Lie group and M a real analytic manifold of dimension n. Then the mapping T from $G \times M$ into M is said to form a group of transformations [Ref. 11(b), Chap. III, p. 66] if, for each $g \in G$, $p \in M$, there is an element $T(g)p \in M$, such that

$$T:(g, p) \to T(g)p \text{ is analytic,} \tag{2.1}$$

$$T(e)p = p, \text{ for all } p \in M, \text{ where } e \text{ is the identity in } G, \tag{2.2}$$

$$T(g)T(g')p = T(gg')p \text{ for all } g, g' \in G, \text{ all } p \in M. \tag{2.3}$$

Implicit in this definition is that $[T(g)]^{-1}$ exists and is equal to $T(g^{-1})$. In physical applications, G is identified with the invariance group (or approximate invariance group) of the Lagrangian and M with the space spanned by the field components.

At each point $p_0 \in M$, local coordinates may be introduced by mapping an open neighborhood of p_0

into an open neighborhood of Euclidean space R^n. Such a map is called a chart [Ref. 11(b), Chap. I, pp. 4–6]. If two charts can be defined at a given point, then the corresponding pair of coordinates must be analytic functions of one another. Let us use $\mathbf{x}$ to denote the coordinates of a general point $p \in M$. Similarly, we write $\mathbf{a}$ for the group parameters of an element $g \in G$ in a neighborhood of e. Then (2.1) expresses the fact that $T(g)p$ can be expressed as an analytic function $\mathbf{h}(\mathbf{x}, \mathbf{a})$ (in general, nonlinear) of both $\mathbf{x}$ and $\mathbf{a}$. It is in this form that group transformations are usually considered in physical problems. Thus we define a nonlinear group representation as the set of analytic functions $\mathbf{h}(\mathbf{x}, \mathbf{a})$ arising from a suitable parametrization of M (and G). Naturally, representations which differ only through a choice of coordinates ought to be regarded as being equivalent. However, in the physical problem with which we are concerned, not all coordinate transformations are allowed.[4] In more detail, it is assumed that there is a special point on M, referred to as the base point, which must be represented by the origin $\mathbf{0}$ in all coordinates. With this restriction we must necessarily confine our attention to neighborhoods of the base point. Similarly, we may consider only those elements g of G which are sufficiently close to the identity so that $T(g)p$ is in some neighborhood of the base point for given p in such a neighborhood. Such elements generate G if and only if G is connected [Ref. 11(b), Chap. II, p. 40], so that it is convenient to assume that G is connected. This restriction is of no great importance since we may always choose the connected component of identity G_0 of G, which, since it is a normal subgroup [Ref. 11(b), Chap. II, p. 39], allows us to recover a general representation of G by combining the representations of the groups G/G_0 and G_0.

With the above points in mind, we define two representations[12] $T(g)$ and $\tilde{T}(g)$ to be locally equivalent if, in terms of admissible coordinates defined at the base point, there exists a possibly nonlinear operator A from $R^n \to R^n$ such that

$$A: \mathbf{x} \to A(\mathbf{x}) \text{ is analytic and has an analytic inverse at } \mathbf{0}, \tag{2.4}$$

$$A[T(g)\mathbf{x}] = \tilde{T}(g)A(\mathbf{x}) \text{ for all } g \in G \text{ in a suitable neighborhood of the identity and all } \mathbf{x} \text{ in a neighborhood of } \mathbf{0}, \tag{2.5}$$

$$A(\mathbf{0}) = \mathbf{0}. \tag{2.6}$$

The mapping A which effects the local equivalence between $T(g)$ and $\tilde{T}(g)$ defines a transformation from one set of admissible coordinates to another. Because it is invertible, the elements of locally equivalent representations must be in one-to-one correspondence. In addition, we shall call a representation linearizable if it is locally equivalent to a linear representation. If both T and $\tilde{T}$ are linear, then, without loss of generality, A may be chosen to be a linear operator. In this case Eq. (2.6) holds trivially and Eq. (2.5) reduces to the usual condition for equivalent representations.

We now derive a canonical form for an arbitrary finite-dimensional nonlinear representation of a compact connected Lie group G. The method used is a modification[13] of that developed by Coleman *et al.* for the case of compact connected semisimple Lie groups. It will be referred to as the global procedure.

Let H denote the set of all $h \in G$ such that $T(h)\mathbf{0} = \mathbf{0}$. This is a closed, and hence compact, subgroup of G (it is known as the isotropy subgroup of the origin). If we then expand $T(h)\mathbf{x}$ in the power series

$$T(h)\mathbf{x} = D(h)\mathbf{x} + O(x^2), \tag{2.7}$$

it follows that $D(h)$ is a linear representation of H and that the change of coordinates defined by

$$\mathbf{y} \equiv A\mathbf{x} = \int_H D(h^{-1})\{T(h)\mathbf{x}\}\, dh, \tag{2.8}$$

where dh is the normalized Haar measure on H, establishes a local equivalence between $D(h)$ and the restriction of $T(g)$ to H. Indeed Eq. (2.8) is just the usual averaging procedure for the construction of intertwining operators [Ref. 11(a), Chap. III, p. 101], and local equivalence is a consequence of the implicit function theorem [Ref. 11(b), Appendix, p. 161] and the fact that the Jacobian of the transformation is nonzero at the origin. A useful illustration of this linearization is given in Appendix A.

Since H is closed, it is generated by a subalgebra L of the Lie algebra K of G [Ref. 11(b), Chap. VI, p. 123]. The adjoint map [Ref. 11(d), Chap. II, p. 116] defines a representation of H on K. Since H is compact and K finite dimensional, this representation is completely reducible [Ref. 11(d), Chap. III, p. 161]. Let the $V_{\bar{\alpha}}$, $\bar{\alpha} = n_G - n_H + 1, \cdots, n_G$, be a basis for L. The subalgebra property

$$[V_{\bar{\alpha}}, V_{\beta}] = c_{\bar{\alpha}\beta\bar{\gamma}} V_{\bar{\lambda}} \tag{2.9}$$

expresses the fact that L is an invariant subspace of K. Let the A_α, $\alpha = 1, 2, \cdots, n_G - n_H$, be a basis for the complementary invariant subspace. Then

$$[V_{\bar{\alpha}}, A_\beta] = c_{\bar{\alpha}\beta\gamma} A_\gamma, \tag{2.10}$$

no terms in $\mathbf{V}$ appearing on the right-hand side. On the other hand, it will not in general be true that

$$[A_\alpha, A_\beta] = c_{\alpha\beta\bar{\gamma}} V_{\bar{\gamma}}, \tag{2.11}$$

though, when this does hold, we shall refer to it as the parity case. It is the situation which has been of most physical interest.

In a neighborhood of the identity we may write a given element $g_0 \in G$ uniquely in the form

$$g_0 = \exp \boldsymbol{\xi} \cdot \mathbf{A} \exp \mathbf{u} \cdot \mathbf{V}, \tag{2.12}$$

where $(\boldsymbol{\xi}, \mathbf{u})$ form a real n_G-component vector. Then, since H is the subgroup which leaves the origin fixed, the orbit N of $\mathbf{0}$ under G separates the G/H cosets defined by $\exp \boldsymbol{\xi} \cdot \mathbf{A}$. Hence the ξ_α variables may be chosen as coordinates for N. However,[14] in doing this it is important to observe that left multiplication of g_0 by an arbitrary element $g \in G$ induces, through the commutation relations of the generators, an action on the function space of the ξ variable and not on the ξ itself. For T to be a representation, we must write

$$(gF)(\mathbf{x}) = F[T(g^{-1})\mathbf{x}], \tag{2.13}$$

so that the above choice of coordinates in N gives the action

$$T(e^{-\boldsymbol{\xi}\cdot\mathbf{A}})(\mathbf{0}) = (\boldsymbol{\xi}). \tag{2.14}$$

We remark that,for the case when $H = e$, this choice of coordinates is known as the canonical chart [Ref. 11(b), Chap. VI, p. 110].

For the coordinates of M we write $\mathbf{x} = (\boldsymbol{\xi}, \boldsymbol{\psi})$, with N all the vectors of the form $(\boldsymbol{\xi}, \mathbf{0})$. The action of H on M is linear, whereas the action by the G/H cosets takes the form

$$T(e^{-\boldsymbol{\xi}\cdot\mathbf{A}})(\mathbf{0}, \boldsymbol{\psi}) = (\boldsymbol{\xi}'(\boldsymbol{\xi}, \boldsymbol{\psi}), \boldsymbol{\psi}'(\boldsymbol{\xi}, \boldsymbol{\psi})), \tag{2.15}$$

where $\boldsymbol{\xi}'(\boldsymbol{\xi}, \boldsymbol{\psi})$ and $\boldsymbol{\psi}'(\boldsymbol{\xi}, \boldsymbol{\psi})$ are analytic functions, and by (2.14)

$$\begin{aligned} &\boldsymbol{\xi}'(\boldsymbol{\xi}, \mathbf{0}) = \boldsymbol{\xi}, \quad \boldsymbol{\psi}'(\boldsymbol{\xi}, \mathbf{0}) = \mathbf{0}, \\ &\boldsymbol{\xi}'(\mathbf{0}, \boldsymbol{\psi}) = \mathbf{0}, \quad \boldsymbol{\psi}'(\mathbf{0}, \boldsymbol{\psi}) = \boldsymbol{\psi}. \end{aligned} \tag{2.16}$$

This admits a coordinate transformation in the sense of (2.4) such that

$$T(e^{-\tilde{\boldsymbol{\xi}}\cdot\mathbf{A}})(\mathbf{0}, \tilde{\boldsymbol{\psi}}) = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\psi}}), \tag{2.17}$$

since the Jacobian of the transformation is the identity at $\mathbf{0}$.

Dropping the tilde, we now define $(\boldsymbol{\xi}', \mathbf{u}')$ as the unique vector determined by the left translation

$$g^{-1}e^{-\boldsymbol{\xi}\cdot\mathbf{A}} = e^{-\boldsymbol{\xi}'\cdot\mathbf{A}}e^{-\mathbf{u}'\cdot\mathbf{A}}, \tag{2.18}$$

where (see above) $g \in G$ is in a suitable neighborhood of the identity.

The action of the whole group on M is now completely defined. We have

$$\begin{aligned} T(g^{-1})(\boldsymbol{\xi}, \boldsymbol{\psi}) &= T(g^{-1})T(e^{-\boldsymbol{\xi}\cdot\mathbf{A}})(\mathbf{0}, \boldsymbol{\psi}) \\ &= T(e^{-\boldsymbol{\xi}'\cdot\mathbf{A}})T(e^{-\mathbf{u}'\cdot\mathbf{V}})(\mathbf{0}, \boldsymbol{\psi}) \\ &= T(e^{-\boldsymbol{\xi}'\cdot\mathbf{A}})(\mathbf{0}, D(e^{-\mathbf{u}'\cdot\mathbf{V}})\boldsymbol{\psi}) \\ &= (\boldsymbol{\xi}', D(e^{-\mathbf{u}'\cdot\mathbf{V}})\boldsymbol{\psi}), \end{aligned} \tag{2.19}$$

where D is an arbitrary real linear finite-dimensional representation of H.

This describes a canonical form with respect to which any finite-dimensional nonlinear representation is locally equivalent. In the subsequent discussion we refer to it as the canonical representation.

It is important to note that, as a consequence of Eq. (2.10), not only the $\boldsymbol{\psi}$-, but also the $\boldsymbol{\xi}$-components of M transforms linearly under H. The latter transformation is defined by

$$e^{\boldsymbol{\xi}'\cdot\mathbf{A}} = h^{-1}e^{\boldsymbol{\xi}\cdot\mathbf{A}}h, \tag{2.20}$$

with $h \in H$. On the other hand, except in special cases, the G/H cosets act nonlinearly on the *whole* manifold, since $\mathbf{u}'$ is a function of $\boldsymbol{\xi}$. [We shall, however, find (cf. Sec. 6) that a further coordinate transformation can be introduced to remove this dependence.] Moreover, the transformation properties of the $\boldsymbol{\xi}$-components are independent of the $\boldsymbol{\psi}$, being determined essentially by the regular representation of G. It is a remarkable fact that the number of components of $\boldsymbol{\xi}$ which describe the nonlinear part of the representation can be chosen to be equal to the difference in the number of generators of G and the subgroup H. In Sec. 6 we attempt to give some insight into the reason for this by studying the group representation in infinitesimal form. It suffices for the present to remark that this separation of variables into $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ components with markedly different transformation properties is not only of mathematical, but also of physical, interest. Thus one may regard these components as representing two different sets of particles, both arising from a single transformation law. Typically the components are identified as the meson field and the baryon field. The physical predictions that this leads to are discussed in subsequent paragraphs.

3. INFINITESIMAL METHOD

Let K be a Lie algebra with elements X_λ, $\lambda = 1, 2, \cdots, n_G$, and M a real analytic manifold of arbitrary dimension n, with coordinates x_i, $i = 1, 2, \cdots, n$, chosen at any given point $p_0 \in M$. Therefore, we have

$$[X_\lambda, X_\mu] = c_{\lambda\mu\nu}X_\nu. \tag{3.1}$$

Let the analytic functions $f_{\lambda i}(\mathbf{x})$ be defined by the following equations:

$$[X_\lambda, x_i] = f_{\lambda i}(\mathbf{x}), \tag{3.2}$$

which hold for all $X_\lambda \in K$ and all x_i. These functions cannot be arbitrary, but must satisfy the following consistency requirement imposed by the Jacobi identity

$$f_{\lambda i,j} f_{\mu j} - f_{\mu i,j} f_{\lambda j} = c_{\mu\lambda\nu} f_{\nu i}, \tag{3.3}$$

where the $c_{\lambda\mu\nu}$ are the structure constants of K above, and the comma denotes differentiation. The integrability requirements of this set of nonlinear differential equations are expressed by the familiar identities [Ref. 11(a), Chap. VIII, p. 301]

$$\begin{aligned} c_{\lambda\mu\nu} &= -c_{\mu\lambda\nu}, \\ c_{\lambda\mu\sigma}c_{\sigma\nu\rho} + c_{\nu\lambda\sigma}c_{\sigma\mu\rho} + c_{\mu\nu\sigma}c_{\sigma\lambda\rho} &= 0. \end{aligned} \tag{3.4}$$

Hence (3.3) may be integrated to yield a solution to (3.2). The differential operators $f_{\lambda i}(\mathbf{x})\partial/\partial x_i$ then close on a Lie algebra homomorphic to K which is said to be a representation of K, and act linearly on M if and only if the $f_{\lambda i}(\mathbf{x})$ are linear functions. We may identify x_i and $\partial/\partial x_i$ as canonically conjugate operators; insofar as the particle creation and annihilation operators derive their transformation laws from these (cf. Sec. 9), then particle number is conserved if and only if the representation is linear. This fact underlies the motivation for the use of nonlinear representations to describe the symmetries of interacting systems.

In solving Eq. (3.3), a change of variable may be admitted. Hence it is natural to identify representations of K related by an invertible coordinate transformation. In more detail, suppose that it is possible to choose a new set of coordinates $\tilde{\mathbf{x}}$ at $p_0 \in M$ (with the same number of components as $\mathbf{x}$) such that

$$[X_\lambda, \tilde{x}_i] = \tilde{f}_{\lambda i}(\tilde{\mathbf{x}}),$$

for all $X_\lambda \in K$ and all $\tilde{x}_i$. Then the representations f and $\tilde{f}$ are said to be equivalent, if the x_i can be expressed as analytic functions of the $\tilde{x}_i$ and vice versa. Following Macfarlane and Weisz,[10] we note that this holds if and only if the pair of equations

$$\begin{aligned} \tilde{x}_{i,j} f_{\lambda j}(\mathbf{x}) &= \tilde{f}_{\lambda i}(\tilde{\mathbf{x}}), \\ x_{i,j}\tilde{f}_{\lambda j}(\tilde{\mathbf{x}}) &= f_{\lambda i}(\mathbf{x}) \end{aligned} \tag{3.5}$$

are integrable [Ref. 11(a), Chap. V, p. 104]. Now, use of Eq. (3.3) and a similar relation which holds for the $\tilde{f}$ shows that

$$\begin{aligned} f_{\lambda s} f_{\mu t}(\tilde{x}_{j,st} - \tilde{x}_{j,ts}) &= 0, \\ \tilde{f}_{\lambda s}\tilde{f}_{\mu t}(x_{j,st} - x_{j,ts}) &= 0. \end{aligned} \tag{3.6}$$

Hence a sufficient condition for equivalence is the invertibility of both the functions f and $\tilde{f}$. Unfortunately, this is not also a necessary condition, though it does prove useful in discussing the canonical representation.

The above argument can be extended if K contains a subalgebra L of order n_H on which the representation is linear. Then, if the change of coordinates leaves the linear subrepresentation unaltered, Eq. (3.6) still holds, but with $(n_G - n_H)$ being the range of Greek indices. Although invertibility depends on the range of indices in the function f being the same, so that Eq. (3.6) may seem to be seldom applicable, this condition is satisfied by the ξ component of the canonical representation (see Secs. 2 and 6).

4. THEOREMS OF LIE

Thus far two nonlinear representation theories have been presented, one for the group and the other for the algebra. These may be interrelated by application of a fundamental result due to Lie described briefly below. It will enable us to derive the solution of Eq. (3.3) corresponding to the canonical representation and to reconstruct from solutions to Eq. (3.3) representations of the Lie group.

Let G be a n_G-parameter Lie group. Assign to each $g \in G$ the n_G-component vector $\mathbf{a}$. Then a representation T of G may be given the form

$$T(\mathbf{a})\mathbf{x} \equiv \mathbf{x}' = \mathbf{h}(\mathbf{x}, \mathbf{a}). \tag{4.1}$$

The functions $\mathbf{h}$ cannot be arbitrary. Besides being analytic in both $\mathbf{x}$ and $\mathbf{a}$, they must satisfy a stringent requirement imposed by the group structure, namely,

$$\mathbf{h}[\mathbf{h}(\mathbf{x}, \mathbf{a}), \mathbf{b}] = \mathbf{h}[\mathbf{x}, \boldsymbol{\phi}(\mathbf{a}, \mathbf{b})], \tag{4.2}$$

where $\boldsymbol{\phi}$ is an analytic function of $\mathbf{a}$ and $\mathbf{b}$ [Ref. 11(a), Chap. 8, p. 286]. From this it may be shown [Ref. 11(a), Chap. 8, pp. 295–301] that

$$\frac{\partial x_i'}{\partial a_\mu} = v_{i\lambda}(\mathbf{x}')\Xi_{\lambda\mu}(\mathbf{a}), \tag{4.3}$$

with

$$\Xi_{\lambda\mu}(\mathbf{0}) = \delta_{\lambda\mu}. \tag{4.4}$$

The functions $v_{i\lambda}(\mathbf{x}')$ and $\Xi_{\lambda\mu}(\mathbf{a})$, which are analytic, must also satisfy

$$v_{i\lambda,j}v_{j\mu} - v_{i\mu,j}v_{j\lambda} = c_{\mu\lambda\nu}v_{i\nu} \tag{4.5}$$

and

$$(\Xi^{-1}_{\nu\lambda})_{,\rho}\Xi^{-1}_{\rho\mu} - (\Xi^{-1}_{\nu\mu})_{,\rho}\Xi^{-1}_{\rho\lambda} = c_{\mu\lambda\nu}\Xi^{-1}_{\nu\rho}, \tag{4.6}$$

where the $c_{\lambda\mu\nu}$ are constants satisfying Eq. (3.4).

This important result shows that both the sets of functions v and Ξ^{-1} provide nonlinear (or linear)

representations of the Lie algebra of G. Conversely, given analytic functions satisfying (4.5) and (4.6), we may reconstruct the nonlinear (or linear) representation of G. This is obtained by integration of Eq. (4.3) for which Eqs. (4.5) and (4.6) are the integrability conditions. For a suitable choice of the $\Xi_{\lambda\mu}$, it may be shown (see Appendix B) that we may choose

$$T(\mathbf{a}) = \exp \mathbf{a} \cdot \mathbf{X}, \tag{4.7}$$

where the X_λ are the infinitesimal generators of the Lie algebra and are given by

$$X_\lambda = \left(\frac{\partial x_i'}{\partial a_\lambda}\right)_{\mathbf{a}=0} \frac{\partial}{\partial x_i} = v_{i\lambda}(\mathbf{x}) \frac{\partial}{\partial x_i}. \tag{4.8}$$

We note that $T(\mathbf{a})$ leaves the origin fixed if and only if $v_{i\lambda}(\mathbf{0}) = 0$, for all i, λ. Also, $T(\mathbf{a})$ is a linear representation if and only if $v_{i\lambda}(\mathbf{x})$ is a linear function.

Finally, we remark that the relationship between Eq. (3.5) and the previously given equivalence condition (2.4) and (2.5) may be established by setting

$$X_\lambda = f_{\lambda j}(\mathbf{x}) \frac{\partial}{\partial x_j}, \quad \tilde{X}_\lambda = \tilde{f}_{\lambda j}(\tilde{\mathbf{x}}) \frac{\partial}{\partial \tilde{x}_j}, \tag{4.9}$$

and defining A as the operator effecting the coordinate transformation

$$A\mathbf{x} = \tilde{\mathbf{x}}(\mathbf{x}), \quad A^{-1}\tilde{\mathbf{x}} = \mathbf{x}(\tilde{\mathbf{x}}) \tag{4.10}$$

involved in the Eqs. (3.5).

From (3.5) and (4.9) we have that

$$\tilde{X}_\lambda = \tilde{f}_{\lambda j}(\tilde{\mathbf{x}}) \frac{\partial}{\partial \tilde{x}_j} = f_{\lambda k}(\mathbf{x}) \frac{\partial \tilde{x}_j}{\partial x_k} \frac{\partial}{\partial \tilde{x}_j}$$
$$= f_{\lambda k}(\mathbf{x}) \frac{\partial}{\partial x_k} = X_\lambda.$$

Hence,

$$(\exp \mathbf{a} \cdot \mathbf{X})\tilde{\mathbf{x}}(\mathbf{x}) = (\exp \mathbf{a} \cdot \tilde{\mathbf{X}})\tilde{\mathbf{x}}(\mathbf{x})$$

holds trivially. Application of Lemma 2 (Appendix B) then gives

$$\tilde{\mathbf{x}}\{(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{x}\} = (\exp \mathbf{a} \cdot \tilde{\mathbf{X}})\tilde{\mathbf{x}}(\mathbf{x}),$$

which may be written, using Eqs. (4.7) and (4.10), in the form

$$A[T(\mathbf{a})\mathbf{x}] = \tilde{T}(\mathbf{a})(A\mathbf{x}), \tag{4.11}$$

as required by Eq. (2.5). Finally, we note that Eq. (2.6) may be chosen as an initial condition in the integration of Eqs. (3.5) [Ref. 11(b), Chap. V, p. 104].

5. CANONICAL REPRESENTATION IN INFINITESIMAL FORM

In the following we derive the infinitesimal form of the canonical representation using the general theory outlined above.

Let us write the general group element effecting the transformation (2.18) in the form

$$g = e^{\mathbf{a}\cdot\mathbf{A}} e^{\mathbf{b}\cdot\mathbf{V}}. \tag{5.1}$$

Transformations of the ξ and ψ components are induced by the action of the generators $\mathbf{V}$ and $\mathbf{A}$ appearing in Eq. (5.1), although they do not themselves act on ξ and ψ directly. The content of the theory outlined in Sec. 4 is that we may define new operators $\mathbf{V}^*$ and $\mathbf{A}^*$ with the same commutation properties as $\mathbf{V}$ and $\mathbf{A}$, but which act directly on ξ and ψ. These new operators are given explicitly by Eq. (4.8), which in the present context takes the form

$$V_{\bar{\alpha}}^* = \left(\frac{\partial \xi_\beta'}{\partial b_{\bar{\alpha}}}\right)_0 \frac{\partial}{\partial \xi_\beta} + \left(\frac{\partial \psi_r'}{\partial b_{\bar{\alpha}}}\right)_0 \frac{\partial}{\partial \psi_r},$$
$$A_\alpha^* = \left(\frac{\partial \xi_\beta'}{\partial a_\alpha}\right)_0 \frac{\partial}{\partial \xi_\beta} + \left(\frac{\partial \psi_r'}{\partial a_\alpha}\right)_0 \frac{\partial}{\partial \psi_r}. \tag{5.2}$$

It is convenient to drop the star, though the distinction between these two sets of generators must be kept in mind, particularly with regard to Eq. (2.13), which interrelates their action. We shall also drop the prime for convenience, as the meaning is otherwise clear.

From Eq. (5.2) we obtain the commutation relations

$$[V_{\bar{\alpha}}, \xi_\beta] = \left(\frac{\partial \xi_\beta}{\partial b_{\bar{\alpha}}}\right)_0; \quad [V_{\bar{\alpha}}, \psi_r] = \left(\frac{\partial \psi_r}{\partial b_{\bar{\alpha}}}\right)_0,$$
$$[A_\alpha, \xi_\beta] = \left(\frac{\partial \xi_\beta}{\partial a_\alpha}\right)_0; \quad [A_\alpha, \psi_r] = \left(\frac{\partial \psi_r}{\partial a_\alpha}\right)_0, \tag{5.3}$$

which describe the representation in infinitesimal form. It is of interest to note that these relations are more general than Eq. (5.2) in that they hold irrespective of the explicit form of the generators $\mathbf{V}$ and $\mathbf{A}$; the generators are merely required to satisfy the correct commutation relations.

In order to compute the differential coefficients appearing in Eq. (5.3), we first note the formulas[15]

$$e^{-X} A e^{X} = e^{-\Delta_X} A \tag{5.4}$$

and

$$e^{-X}\left(\frac{\partial}{\partial y}\right)e^{X} = (1 - e^{-\Delta_X})/\Delta_X \frac{\partial X}{\partial y}, \tag{5.5}$$

where Δ_X denotes the operator $[X, \ ]$. Then, from Eqs. (2.18) and (5.1), by differentiation with respect to $b_{\bar{\alpha}}$ and setting $\mathbf{a} = \mathbf{b} = \mathbf{0}$, we obtain

$$A_\beta \left(\frac{\partial \xi_\beta}{\partial b_{\bar{\alpha}}}\right)_0 = \Delta_{\xi\cdot\mathbf{A}} V_{\bar{\alpha}} \tag{5.6}$$

or

$$\left(\frac{\partial \xi_\beta}{\partial b_{\bar{\alpha}}}\right)_0 = c_{\bar{\alpha}\beta\gamma}\xi_\gamma$$

and

$$\left(\frac{\partial u_\beta}{\partial b_{\bar\alpha}}\right)_0 = \delta_{\bar\alpha\beta}, \tag{5.7}$$

where $\delta_{\bar\alpha\beta}$ is the Kronecker delta. The first of these relations, which shows that the $\boldsymbol{\xi}$-components transform linearly under the subgroup H, is the infinitesimal analog of Eq. (2.20). The second is an elementary consequence of Eqs. (2.10) and (2.18), which together imply that $\mathbf{u}' = \mathbf{b}$ under the subgroups.

Similarly, from Eqs. (2.18) and (5.1), by differentiating with respect to a_α and setting $\mathbf{a} = \mathbf{b} = \mathbf{0}$, we obtain (dropping the subscript $\boldsymbol{\xi}\cdot\mathbf{A}$ on the operator $\Delta_{\boldsymbol{\xi}\cdot\mathbf{A}}$)

$$e^\Delta A_\alpha = \left(\frac{e^\Delta - 1}{\Delta}\right)\left(\frac{\partial \xi_\beta}{\partial a_\alpha}\right)_0 A_\beta + \left(\frac{\partial u_{\bar\alpha}}{\partial a_\alpha}\right)_0 V_{\bar\alpha}. \tag{5.8}$$

To simplify this expression, we define the $\boldsymbol{\xi}$-dependent functions $F_{\alpha\beta}$, $G_{\alpha\bar\alpha}$, $H_{\alpha\beta}$, and $K_{\alpha\bar\alpha}$ by the following relations:

$$\begin{aligned}\Delta^{-1}(e^\Delta - 1)A_\alpha &= F_{\alpha\beta}A_\beta + G_{\alpha\bar\alpha}V_{\bar\alpha},\\ e^\Delta A_\alpha &= H_{\alpha\beta}A_\beta + K_{\alpha\bar\alpha}V_{\bar\alpha}.\end{aligned} \tag{5.9}$$

The leading term of both $F_{\alpha\beta}$ and $H_{\alpha\beta}$ is the Kronecker delta, so these functions are invertible [Ref. 11(b), Appendix, p. 158]. Thus, by substituting Eq. (5.9) in Eq. (5.8) and equating coefficients of $\mathbf{A}$ and $\mathbf{V}$, we obtain

$$\begin{aligned}\left(\frac{\partial \xi_\beta}{\partial a_\alpha}\right)_0 &= H_{\alpha\gamma}F^{-1}_{\gamma\beta},\\ \left(\frac{\partial u_{\bar\alpha}}{\partial a_\alpha}\right)_0 &= K_{\alpha\bar\alpha} - H_{\alpha\beta}F^{-1}_{\beta\gamma}G_{\gamma\bar\alpha}.\end{aligned} \tag{5.10}$$

These expressions take a particularly simple form in the parity case [Eq. (2.11)]. This is because, recalling Eq. (2.10), terms in $\mathbf{A}$ and $\mathbf{V}$ appear alternately in the action of the operator $\Delta \equiv \Delta_{\boldsymbol{\xi}\cdot\mathbf{A}}$ on $\mathbf{A}$. Thus, from Eq. (5.8), we obtain after a little reduction

$$\left(\frac{\partial \xi_\beta}{\partial a_\alpha}\right)_0 A_\beta = \Delta \coth \Delta A_\alpha, \tag{5.11}$$

$$\left(\frac{\partial u_{\bar\alpha}}{\partial a_\alpha}\right)_0 V_{\bar\alpha} = (\sinh \Delta)^{-1}(\cosh \Delta - 1)A_\alpha, \tag{5.12}$$

from which the differential coefficients are readily obtained in series form.

Equations (5.6) and (5.11) define the action on the $\boldsymbol{\xi}$-components. To derive the corresponding result for the $\boldsymbol{\psi}$-components, we note from Eq. (2.19) that

$$\begin{aligned}\left(\frac{\partial \psi}{\partial b_{\bar\alpha}}\right)_0 &= \left\{\frac{\partial}{\partial b_{\bar\alpha}} D(e^{-\mathbf{u}\cdot\mathbf{V}})\right\}_0 \psi\\ &= \left\{\frac{\partial}{\partial b_{\bar\alpha}} e^{-\mathbf{u}\cdot\mathbf{R}}\right\}_0 \psi\\ &= -R_{\bar\alpha}\psi,\end{aligned} \tag{5.13}$$

where R is the linear representation of the subalgebra corresponding to the linear representation D of the subgroup H. Similarly,

$$\left(\frac{\partial \psi}{\partial a_\alpha}\right)_0 = -\left(\frac{\partial u_{\bar\alpha}}{\partial a_\alpha}\right)R_{\bar\alpha}\psi. \tag{5.14}$$

To summarize, we may write down a "standard representation" for the commutation relations (cf. Secs. 6 and 8) as follows:

$$\begin{aligned}&[V_{\bar\alpha}, \xi_\beta] = c_{\bar\alpha\beta\gamma}\xi_\gamma, \quad [A_\alpha, \xi_\beta] = f_{\alpha\beta}(\xi),\\ &[V_{\bar\alpha}, \psi] = -R_{\bar\alpha}\psi, \quad [A_\alpha, \psi] = -g_{\alpha\bar\alpha}(\xi)R_{\bar\alpha}\psi,\end{aligned} \tag{5.15}$$

where the $f_{\alpha\beta}$, $g_{\alpha\bar\alpha}$, and $R_{\bar\alpha}$ are restricted by the various Jacobi identities. We have shown here that in the special case of the canonical representation, when parity is assumed, $f_{\alpha\beta}(\xi) \equiv (\partial\xi_\beta/\partial a_\alpha)_0$ and $g_{\alpha\bar\alpha}(\xi) \equiv (\partial u_{\bar\alpha}/\partial a_\alpha)_0$ are given in the form of infinite series by Eqs. (5.11) and (5.12), respectively.

6. REPRESENTATION PROBLEM FOR THE ALGEBRA

Nonlinear representations in infinitesimal form have been derived by an indirect method involving knowledge of the global form. In the following we attempt a direct solution of the representation problem in infinitesimal form, as described by the set of nonlinear partial-differential equations (3.3). Much of the emphasis is placed on deriving the infinitesimal analogs of the transformations which lead in the global theory to the canonical representation.

The first step is to derive conditions under which linearization of a given representation can be effected. In principle this is straightforward, for it is sufficient to establish the integrability conditions (3.5) corresponding to a change of variable effecting linearization. Unfortunately, owing to the general noninvertibility of the tensor $f_{\alpha i}$ which describes a linear representation, the method outlined in Sec. 3 is not immediately applicable, and we must seek an alternative approach.

By analogy with the global theory, we are led to consider the linearization of the representation on the maximal subset $X_{\bar\alpha}$ for which $f_{\bar\alpha i}(\mathbf{0}) = 0$, for all i. This is a subalgebra L of K, so that we have

$$[X_{\bar\alpha}, X_{\bar\beta}] = c_{\bar\alpha\bar\beta\bar\gamma}X_{\bar\gamma},$$

which imposes a condition analogous to Eq. (3.3) on the $f_{\bar\alpha i}$. From this, on differentiating with respect to x_k and setting $\mathbf{x} = \mathbf{0}$, we obtain

$$f^0_{\bar\alpha i,j}f^0_{\bar\beta j,k} - f^0_{\bar\beta i,j}f^0_{\bar\alpha j,k} = c_{\bar\beta\bar\alpha\bar\gamma}f^0_{\bar\gamma i,k}, \tag{6.1}$$

where the superscript denotes the value of the function at $\mathbf{0}$. Equation (6.1) is just the condition for $f^0_{\bar\alpha i,j}$,

which is the first coefficient in the expansion of $f_{\bar{\alpha}i}$, to be a linear representation of the subalgebra L. This is exactly as obtained in the global analysis, though we have yet to show that the representations described by $f_{\bar{\alpha}i}$ and $f^0_{\bar{\alpha}i,j}$ are equivalent. Indeed, this fails to be true in general, as the example in Sec. 7 shows.

However, in the case when L is compact, we may appeal to the global argument to establish equivalence. This is because the subgroup H generated by L is compact and hence the linearization procedure given in Sec. 2 may be applied. Combining this with the analysis given in the latter part of Sec. IV, the equivalence of the representations described by $f_{\bar{\alpha}i}$ and $f^0_{\bar{\alpha}i,j}$ is easily verified. It would be of interest to give a more direct argument involving only the Lie algebra, though this is apparently less easy. Indeed, it is generally true that changes of variable are more readily carried out globally.

We may write an arbitrary representation, linear on some subalgebra L (not necessarily that defined above), in the form

$$[X_{\bar{\alpha}}, x_i] = -R_{\bar{\alpha}ij}x_j; \quad [X_\beta, x_i] = f_{\beta i}(x).$$

Let us further assume that K is semisimple, with the Killing form diagonal, and that $R_{\bar{\alpha}ij}$ is antisymmetric in the Latin indices. In this case we have the further equation (cf. 2.10):

$$[X_{\bar{\alpha}}, X_\beta] = c_{\bar{\alpha}\beta\gamma}X_\gamma.$$

It then follows from the Jacobi identities that

$$R_{\bar{\alpha}ij}R_{\beta jk} - R_{\beta ij}R_{\bar{\alpha}jk} = c_{\bar{\alpha}\beta\bar{\gamma}}R_{\bar{\gamma}ik}, \tag{6.2}$$

$$f_{\alpha i,j}R_{\beta jk}x_k = R_{\beta ij}f_{\alpha j} - f_{\gamma i}c_{\beta\alpha\gamma}, \tag{6.3}$$

$$f_{\alpha i,j}f_{\beta j} - f_{\beta i,j}f_{\alpha j} = c_{\alpha\beta\bar{\gamma}}R_{\bar{\gamma}ij}x_j + c_{\alpha\beta\gamma}f_{\gamma i}. \tag{6.4}$$

The first of these expressions implies that $R_{\bar{\alpha}ij}$ is a linear representation of L. The second, that $f_{\beta i}$ is a mixed tensor transforming according to the linear representations $R_{\bar{\alpha}ij}$ and $c_{\bar{\alpha}\gamma\beta}$ of L.

Evaluated at the origin, Eq. (6.3) gives

$$R_{\beta ij}f_{\alpha j}(\mathbf{0}) = f_{\gamma i}(\mathbf{0})c_{\beta\alpha\gamma}. \tag{6.5}$$

The simple interpretation of Eq. (6.5) is that the matrix M with components $M_{j\alpha} \equiv f_{\alpha j}(\mathbf{0})$ is an intertwining operator for the representations of L defined by $(R_{\bar{\alpha}})_{ij} \equiv R_{\bar{\alpha}ij}$ and $(C_{\bar{\alpha}})_{\beta\gamma} \equiv c_{\bar{\alpha}\gamma\beta}$. Hence, by Schur's Lemma and the assumed antisymmetry of $R_{\bar{\alpha}}$, if $f_{\alpha j}$ starts with a constant term, $R_{\bar{\alpha}}$ must contain in its reduction at least one of the irreducible subrepresentations of $C_{\bar{\alpha}}$. We use this fact in Sec. 7. This decomposition underlies the separation of the representation manifold M in ξ and ψ components as described in Sec. 2.

For compact Lie algebras, this argument may be considerably strengthened. In this case, L can be chosen to be the maximal subalgebra for which the representation vanishes at the origin. This implies that the $(n_G - n_H)$ vectors $\mathbf{a}_\alpha$ with components $f_{\alpha j}(\mathbf{0})$ are linearly independent. As the Lie bracket is itself linear, we may choose

$$f_{\alpha i}(\mathbf{0}) = \delta_{\alpha i}, \tag{6.6}$$

which corresponds to the orthonormalization of the $\mathbf{a}_\alpha$. Substituting back in (6.3) and setting $\mathbf{x} = (\xi, \psi)$ (identifying the ξ with the Greek indices), we find that the ξ-coordinates form an invariant subspace of the linear representation described by $R_{\bar{\alpha}ij}$. Since L is compact and M is assumed finite dimensional, it follows that the ψ-coordinates also form an invariant subspace [Ref. 11(d), Chap. III, p. 161] and we obtain

$$[X_{\bar{\alpha}}, \xi_\beta] = c_{\bar{\alpha}\beta\gamma}\xi_\gamma, \tag{6.7}$$

$$[X_{\bar{\alpha}}, \psi_r] = -R_{\bar{\alpha}rs}\psi_s, \tag{6.8}$$

$$[X_\alpha, \xi_\beta] = f_{\alpha\beta}(\xi, \psi), \tag{6.9}$$

$$[X_\alpha, \psi_r] = f_{\alpha r}(\xi, \psi), \tag{6.10}$$

where the subscripts have the ranges given under *Notation* in Sec. 1.

Equations (6.7) and (6.8) illustrate the separation of M into ξ- and ψ-components, though the representation is not yet in canonical form. For this, two further coordinate transformations are necessary. First we note that Eq. (6.6) is equivalent, though to first order only in the group parameters, to the choice made in Sec. 2 of the ξ as the coordinates of the orbit N of $\mathbf{0}$ under G/H. For this to hold to all orders, we must also have

$$f_{\alpha r}(\xi, \mathbf{0}) = 0.$$

If this does not hold, we must find new coordinates in which it does. Given such a transformation, it then follows from Eq. (6.4) that $f_{\alpha\beta}(\xi, \mathbf{0})$ is a representation of $K - L$. Moreover, by (6.6) this tensor is invertible at the origin, and hence, by an argument parallel to that given in Sec. 3, is equivalent to any such invertible representation. In particular, we may choose $f_{\alpha\beta}(\xi, \mathbf{0})$ to have the canonical form.

Secondly, the transformation carried out in the global procedure, as described by Eqs. (2.15) and (2.17), indicates that it is possible to choose coordinates such that

$$f_{\alpha\beta}(\xi, \psi) = f_{\alpha\beta}(\xi, \mathbf{0}) \equiv f_{\alpha\beta}(\xi),$$
$$f_{\alpha r}(\xi, \psi) = g_{\alpha rs}(\xi)\psi_s. \tag{6.11}$$

Indeed the canonical representation can be recovered by setting

$$g_{\alpha rs}(\xi) = -g_{\alpha\bar{\alpha}}(\xi)R_{\bar{\alpha}rs}, \tag{6.12}$$

where $g_{\alpha\bar{\alpha}} = (\partial u_{\bar{\alpha}}/\partial a_\alpha)_0$ is given by Eq. (5.10). [When $g_{\alpha rs}(\xi)$ separates in the manner described by Eq. (6.12), we shall say that it is of standard form, the canonical representation being a special case. The importance of this form arises in the construction of invariant Lagrangians (cf. Sec. 8).] Moreover, we may show that any analytic function $\tilde{g}_{\alpha rs}(\xi)$ gives an equivalent representation to Eq. (6.11) above by considering the following coordinate transformation:

$$\tilde{\xi}_\alpha = \xi_\alpha, \quad \tilde{\psi}_r = \Lambda_{rs}(\xi)\psi_s.$$

Substituting in

$$[X_{\bar{\alpha}}, \tilde{\psi}_r] = -R_{\bar{\alpha}rs}\psi_s; \quad [X_\alpha, \psi_r] = \tilde{g}_{\alpha rs}(\xi)\tilde{\psi}_s,$$

we obtain the differential equations for the function $\Lambda_{rs}(\xi)$:

$$\Lambda_{rs,\beta}c_{\bar{\alpha}\beta\gamma}\xi_\gamma = R_{\bar{\alpha}rt}\Lambda_{ts} + R_{\bar{\alpha}ts}\Lambda_{rt}, \tag{6.13}$$

$$\Lambda_{rs,\beta}f_{\alpha\beta} + \Lambda_{rt}g_{\alpha ts} = \tilde{g}_{\bar{\alpha}rt}\Lambda_{ts}. \tag{6.14}$$

The first of these [Eq. (6.13)] merely expresses the fact that Λ_{rs} must be a tensor with respect to the representation $R_{\bar{\alpha}rs}$ of L. As for the second, one may show from the defining equations (6.7)–(6.10) for a representation that

$$(\Lambda_{rs,\alpha\beta} - \Lambda_{rs,\beta\alpha})f_{\gamma\alpha}f_{\delta\beta} = 0.$$

On account of the invertibility of $f_{\alpha\beta}$, this shows that Eq. (6.14) is integrable and thus establishes the equivalence of the representations $g_{\alpha rs}$ and $\tilde{g}_{\alpha rs}$. In particular, we may choose $\tilde{g}_{\alpha rs}$ to be independent of the ξ coordinate. This is a generalization of a result obtained by MacFarlane and Weisz[10] for K_n. However, in general, such a choice for $\tilde{g}_{\alpha rs}$ may not be of standard form, whereas the canonical representation always has this property. Constant $\tilde{g}_{\alpha rs}$ means that the representation manifold separates out into a subspace which transforms linearly under the whole group. Moreover, any two such linear representations are equivalent if and only if they are unitarily equivalent on the subgroup H corresponding to L. This holds even if they are not unitarily equivalent on G. A result of a rather similar nature has been obtained from the global analysis.[4]

Finally, we remark that if we identify the ξ- and ψ-components of the canonical representation with the meson and baryon fields respectively, then, insofar as the creation and annihilation operators for the particles obey similar transformation laws (cf. Secs. 3 and 9), the particle number of the baryons, but not that of the mesons, is conserved by the action of the infinitesimal generators on the state space.

7. NONCOMPACT CASE

In the global procedure, the compactness of G is required only to ensure compactness of the isotropy subgroup H. Hence the results of Sec. 2 also apply to representations of noncompact groups where the isotropy subgroup is compact. However, this is not the most general situation, and consequently noncompact groups may admit representations not equivalent to the canonical one. This is conveniently illustrated by a study of $SO(2, 1)$, the Lie algebra of which takes the form

$$[L_1, L_2] = L_3, \quad [L_2, L_3] = -L_1, \quad [L_3, L_1] = L_2. \tag{7.1}$$

We consider a representation of this algebra linear when restricted to L_3. We suppose further that it is nonzero at the origin. Then, according to the argument given in Sec. 6, its restriction to L_3 must contain an irreducible component of the representation of L_3 induced by the commutation relations (7.1). The latter is 2-dimensional and has irreducible components given by

$$[L_3, L_\pm] = \pm L_\pm,$$

where

$$L_\pm = L_1 \pm L_2.$$

We choose a representation transforming under L_3 like the L_+ component. That is,

$$[L_3, z] = z.$$

For the nonlinear part, we introduce functions $f_\pm(z)$ defined by

$$[L_\pm, z] = f_\pm(z).$$

These must satisfy, on account of the Jacobi identities,

$$f^- \frac{df^+}{dz} - f^+ \frac{df^-}{dz} = 2z, \quad z\frac{df^+}{dz} = 2f^+, \quad z\frac{df^-}{dz} = 0.$$

Hence the representation takes the form

$$[L, z] = \tfrac{1}{2}(\lambda z^2 + 1/\lambda), \quad [L_2, z] = \tfrac{1}{2}(\lambda z^2 - 1/\lambda), \quad [L_3, z] = z, \tag{7.2}$$

where λ is an arbitrary real constant. This representation is not equivalent to the canonical representation. Moreover, it cannot be linearized even on the (noncompact) subalgebra L_3, L_+. This is in spite of the fact that on this subalgebra the representation vanishes at the origin.

By complexification we may obtain from Eq. (7.2) a representation of $SO(3)$. This takes the form

$$[L_1, z] = \tfrac{1}{2}(\lambda z^2 + 1/\lambda), \quad [L_2, z] = \tfrac{1}{2}i(\lambda z^2 - 1/\lambda), \quad [L_3, z] = iz.$$

However, the introduction of i means that we must interpret z as a complex variable. In order to derive the form of this representation on a real manifold, we write $z = x + iy$ and identify real and imaginary parts. The representation so obtained can be shown to be equivalent to the canonical representation. Furthermore, the maximal subalgebra on which the representation vanishes is the subalgebra L_3 on which it is linear. This is in agreement with the general theory.

The above result suggests a method for constructing nonlinear representations for arbitrary noncompact semisimple Lie algebras, since every such algebra has a compact real form obtained by complexification [Ref. 11(a), Chap. 8, p. 310]. Thus, given an arbitrary nonlinear representation of a noncompact semisimple Lie algebra, complexification and separation into real and imaginary parts gives a nonlinear representation of the compact algebra which must be equivalent to the canonical representation. Thus, starting with the canonical representation for the compact form, a reversal of the above procedure must give all possible nonlinear representations of an arbitrary noncompact semisimple Lie algebra. Unfortunately, the reverse procedure is less easy to perform.

8. CONSTRUCTION OF INVARIANT LAGRANGIANS

The physical motivation behind the preceding work is the construction of Lagrangians which exhibit the supposed symmetries of nature and whose predictions may be compared with experiment. The preliminary step is therefore to construct a Lagrangian which is invariant under a desired symmetry, for example, $SU(2) \times SU(2)$. A subsequent step, which we shall not discuss here, is to break the symmetry in a well-defined way consistent with experiment, for example, by means of the PCAC hypothesis. It has been shown[4] that the results of calculations made on such Lagrangians in certain well-defined approximations (tree approximations) are in accord with results obtained from more laborious calculations based on current algebra. We also note that, although from the point of view of a symmetric Lagrangian all representations of the fields are equivalent, once symmetry breaking is introduced, different schemes may give different results [Ref. 3(d), p. 1757].

The construction of invariant Lagrangians is highly facilitated by the standard transformation property (6.12), satisfied in particular by the canonical baryon field, which ensures the following result: Any function, invariant under the linear subgroup, of fields with the transformation properties under the whole group of the standard baryon fields, is an invariant function under the whole group.[16] Thus, if $I(\psi)$ is the invariant function,

$$[X_{\tilde{\alpha}}, I(\psi)] = -I_{,r}R_{\tilde{\alpha}rs}\psi_s = 0,$$

which implies

$$[X_{\alpha}, I(\psi)] = -I_{,r}g_{\alpha\tilde{\alpha}}R_{\tilde{\alpha}rs}\psi_s = 0.$$

We also note that there can be no nontrivial invariant function of the ξ fields; this is a consequence of the transitive action of the group on the ξ component in the global form. Infinitesimally,

$$[X_{\alpha}, I(\xi)] = I_{,\beta}f_{\alpha\beta}(\xi) = 0 \Rightarrow I_{,\beta} = 0 \quad \text{or} \quad I = \text{const},$$

from the invertibility of $f_{\alpha\beta}(\xi)$.

A Lagrange function should in general depend upon the gradients of the fields with respect to the space–time coordinates. Such gradients transform linearly under the whole group, but need not fulfill the requirement of transforming in the standard way. The problem then arises of constructing analogous quantities, the so-called covariant derivatives, which do transform in the standard fashion. Thus, given the standard transformation equations

$$\begin{aligned} &[V_{\tilde{\alpha}}, \xi_{\beta}] = c_{\tilde{\alpha}\beta\gamma}\xi_{\gamma}; \quad [A_{\alpha}, \xi_{\beta}] = f_{\alpha\beta}(\xi), \\ &[V_{\tilde{\alpha}}, \psi] = -R_{\tilde{\alpha}}\psi; \quad [A_{\alpha}, \psi] = -g_{\alpha\tilde{\alpha}}(\xi)R_{\tilde{\alpha}}\psi, \end{aligned} \tag{8.1}$$

we require that the covariant derivatives obey

$$\begin{aligned} &[V_{\tilde{\alpha}}, D_{\mu}\xi_{\beta}] = c_{\tilde{\alpha}\beta\gamma}D_{\mu}\xi_{\gamma}, \\ &[A_{\alpha}, D_{\mu}\xi_{\beta}] = g_{\alpha\tilde{\alpha}}(\xi)c_{\tilde{\alpha}\beta\gamma}D_{\mu}\xi_{\gamma}, \\ &[V_{\tilde{\alpha}}, D_{\mu}\psi] = -R_{\tilde{\alpha}}D_{\mu}\psi, \\ &[A_{\alpha}, D_{\mu}\psi] = -g_{\alpha\tilde{\alpha}}(\xi)R_{\tilde{\alpha}}D_{\mu}\psi. \end{aligned} \tag{8.2}$$

It is easy to show that such covariant derivatives exist, providing $f_{\alpha\beta}(\xi)$ is invertible. For if we define

$$\begin{aligned} D_{\mu}\xi_{\beta} &= h_{\alpha\beta}(\xi)\partial_{\mu}\xi_{\alpha}, \\ D_{\mu}\psi &= \partial_{\mu}\psi + k_{\beta\alpha}(\xi)(D_{\mu}\xi_{\alpha})R_{\beta}\psi, \end{aligned} \tag{8.3}$$

then substituting Eq. (8.3) in Eqs. (8.1) and (8.2) ultimately leads to

$$\begin{aligned} (h_{\alpha\beta,\epsilon\iota} - h_{\alpha\beta,\iota\epsilon})f_{\gamma\epsilon}f_{\delta\iota} = 0, \\ (k_{\beta\alpha,\epsilon\iota} - k_{\beta\alpha,\iota\epsilon})f_{\gamma\epsilon}f_{\delta\iota} = 0, \end{aligned}$$

which are integrable provided $f_{\alpha\beta}$ is invertible. We also have that h and k are tensors under the subalgebra, and that h is invertible.

We may exhibit a general form of the covariant derivatives as follows. Given differential operators which obey

$$[V_{\tilde{\alpha}}, \mathcal{A}_{\beta}] = c_{\tilde{\alpha}\beta\gamma}\mathcal{A}_{\gamma}, \quad [A_{\alpha}, \mathcal{A}_{\beta}] = g_{\alpha\tilde{\alpha}}(\xi)c_{\tilde{\alpha}\beta\gamma}\mathcal{A}_{\gamma}, \tag{8.4}$$

we may then define covariant derivatives $D_{\mu}\xi$ and

$D_\mu\psi$ by

$$\partial_\mu\xi_\beta = [\mathcal{A}_\alpha, \xi_\beta]D_\mu\xi_\alpha,$$
$$D_\mu\psi = \partial_\mu\psi - [\mathcal{A}_\alpha, \psi]D_\mu\xi_\alpha, \tag{8.5}$$

which, given that $[\mathcal{A}_\alpha, \xi_\beta]$ is invertible, may be shown to have the correct transformation properties. Now given $f_{\alpha\beta}$ invertible, solutions to Eq. (8.4) necessarily exist, for it may be shown that we may write $\mathcal{A}_\alpha$ in the form

$$\mathcal{A}_\alpha = h^{-1}_{\beta\alpha}\frac{\partial}{\partial\xi_\beta} - k_{\bar\beta\alpha}R_{\bar\beta i j}\psi_j\frac{\partial}{\partial\psi_i}, \tag{8.6}$$

where $h_{\alpha\beta}$ and $k_{\beta\alpha}$ are defined by Eq. (8.3). However, the important fact is that we may further obtain solutions to Eq. (8.4) in the form

$$\mathcal{A}_\alpha = L_{\alpha\beta}(\xi)A_\beta + M_{\alpha\beta}(\xi)V_\beta, \tag{8.7}$$

where $L_{\alpha\beta}$ and $M_{\alpha\beta}$ are suitable functions of $\boldsymbol{\xi}$. Indeed, by comparison of Eqs. (8.6) and (8.7), we obtain

$$h^{-1}_{\beta\alpha} = L_{\alpha\gamma}f_{\gamma\beta} + M_{\alpha\bar\gamma}c_{\bar\gamma\beta\gamma}\xi_\gamma,$$
$$k_{\bar\gamma\alpha} = L_{\alpha\gamma}g_{\gamma\bar\gamma} + M_{\alpha\bar\gamma},$$

and it is easy to show that these equations have a solution for $L_{\alpha\beta}$ and $M_{\alpha\beta}$, given $f_{\alpha\beta}$ invertible.

The importance of this result is that we may go on to construct higher-order covariant derivatives transforming under G as prescribed by Eq. (8.2), but which involve several space–time differentiations. For example, $D^\mu D_\mu\psi$, which is a space–time invariant, takes the form

$$\begin{aligned} D^\mu D_\mu\psi &= \partial^\mu D_\mu\psi - [\mathcal{A}_\alpha, D_\mu\psi](D^\mu\xi_\alpha) \\ &= \partial^\mu\partial_\mu\psi - (\partial^\mu[\mathcal{A}_\beta, \psi])D_\mu\xi_\beta - [\mathcal{A}_\beta, \psi]\partial^\mu D_\mu\xi_\beta \\ &\quad - [\mathcal{A}_\alpha, \partial_\mu\psi](D^\mu\xi_\alpha) \\ &\quad + [\mathcal{A}_\alpha, [\mathcal{A}_\beta, \psi]](D^\mu\xi_\alpha)(D_\mu\xi_\beta). \end{aligned}$$

In the parity case, assuming the structure constants totally antisymmetric[cf. Ref. 11 (a), Chap. 8, pp. 310 and 311] and $g_{\alpha\bar\alpha}$ a constant, the solution takes the particularly simple form

$$L_{\alpha\beta} = \delta_{\alpha\beta}, \quad M_{\alpha\beta} = g_{\alpha\beta}, \tag{8.8}$$

and hence the covariant derivatives become

$$\partial_\mu\xi_\beta = (f_{\alpha\beta} + g_{\alpha\bar\alpha}c_{\bar\alpha\beta\gamma}\xi_\gamma)D_\mu\xi_\alpha, \quad D_\mu\psi = \partial_\mu\psi. \tag{8.9}$$

This generalizes a result of MacFarlane and Weisz[10] obtained in the special case when $G = K_n$ and $H = SU(n)$. As might be expected, the covariant derivatives for the ψ field are trivial, though it is perhaps worth noting that we may add an arbitrary multiple of the term

$$(f_{\beta\alpha} - g_{\beta\bar\alpha}c_{\bar\alpha\alpha\gamma}\xi_\gamma)^{-1}(\partial_\mu\xi_\beta)g_{\alpha\beta}R_\beta\psi$$

to $\partial_\mu\psi$ and still obtain a covariant derivative. Thus the choice of $D_\mu\psi$ is not unique, there being a corresponding arbitrariness in the choice of the Lagrangian.

It should be noted that in the constant case the expressions for $L_{\alpha\beta}$ and $M_{\alpha\beta}$ are simpler than the corresponding expression for $h_{\alpha\beta}$. Moreover, the former involve only the $g_{\alpha\beta}$ term. It might be anticipated from this that the general expressions will involve only $g_{\alpha\beta}$ and its first derivatives as in the construction of Christoffel symbols, though we have been unable to verify this conjecture.

In the case of the canonical representation, the covariant derivatives can be evaluated explicitly.[5,17] The solution takes the form

$$D_\mu\xi_\beta = F_{\alpha\beta}(\xi)\partial_\mu\xi_\alpha,$$
$$D_\mu\psi = \partial_\mu\psi + G_{\alpha\beta}(\xi)(\partial_\mu\xi_\alpha)R_\beta\psi,$$

where $F_{\alpha\beta}$ and $G_{\alpha\beta}$ are given by Eq. (5.9). In the parity case, this expression simplifies considerably to give

$$D_\mu\boldsymbol{\xi}\cdot\mathbf{A} = \Delta^{-1}\sinh\Delta\,\partial_\mu\boldsymbol{\xi}\cdot\mathbf{A},$$
$$D_\mu\psi = \partial_\mu\psi + K_\beta(\boldsymbol{\xi}, \partial_\mu\boldsymbol{\xi})R_\beta\psi,$$

where

$$K_\beta(\boldsymbol{\xi}, \partial_\mu\boldsymbol{\xi})V_\beta = \Delta^{-1}(\cosh\Delta - 1)(\partial_\mu\boldsymbol{\xi}\cdot\mathbf{A}).$$

There is a striking similarity between the expressions obtained in this and the $g_{\alpha\bar\alpha}$ constant case. Thus substitution from Eqs. (5.6), (5.11), (5.12), and (8.1) gives after a little manipulation

$$\partial_\mu\xi_\beta = (f_{\alpha\beta} - g_{\alpha\bar\alpha}c_{\bar\alpha\beta\gamma}\xi_\gamma)D_\mu\xi_\alpha,$$
$$D_\mu\psi = \partial_\mu\psi + g_{\alpha\bar\alpha}(D_\mu\xi_\alpha)R_{\bar\alpha}\psi,$$

an expression very similar to Eq. (8.9). This is apparently coincidental.

Having obtained the covariant derivatives, a Lagrange function invariant under G and space–time may be readily constructed. To do this,we merely combine covariant and contravariant components, the latter being obtained by replacing the representation $R_{\bar\alpha}$ by its adjoint representation and ∂_μ and ∂^μ. As we noted previously, the meson mass term cannot appear in such a Lagrangian, though it may be included as a symmetry-breaking term.[9,18]

Lagrangians constructed in the manner described above are essentially bilinear in the co- and contravariant derivatives. In addition, it is possible to write down multilinear invariants by making use of the following construction.

Suppose that, under the subgroup H, the vectors $y^{(k)}_\beta$ transform according to

$$[V_{\bar\alpha}, y^{(k)}_\beta] = c_{\bar\alpha\beta\gamma}y^{(k)}_\gamma.$$

Then the $C_{i\beta j}y_\beta^{(k)}$; $i, j = 1, 2, \cdots, n_G$, transform like tensors under H; and $I(n)$, given by

$$I(n) = c_{i_1\beta_1 i_2} y_{\beta_1}^{(1)} c_{i_2\beta_2 i_3} y_\beta^{(2)} \cdots c_{i_n\beta_n i_1} y_{\beta_n}^{(n)},$$

are invariant under H, for all $n \geq 2$. For $n = 2$, we obtain the familiar invariant $y_\beta^{(1)} y_\beta^{(2)}$, whilst for $n > 2$, either new invariants or functions of $I(2)$ appear, depending on the particular choice of G. When we replace y_β by the covariant derivatives $(D_\mu \xi_\beta)$, it follows from this that the resulting quantities are invariant under the whole group.

9. NONLINEARITY AND ZERO MASS

In the following we consider the transformation properties with respect to G of the underlying state space on which the field operators act. Particular emphasis is placed on the noninvariance of the vacuum which leads through the Goldstone boson theorem[19] to zero-mass particles.

We suppose that the action of G on the states is induced by a one–one onto map U_g chosen such that expectation values remain fixed. This leads in the usual way to the set $U_g : g \in G$ defining a unitary representation of G. Assumed invariance of the vacuum under G then gives

$$\langle 0 | (\exp \mathbf{a} \cdot \mathbf{x} - 1) x_i | 0 \rangle = 0, \qquad (9.1)$$

for all i. This is an identity in the group parameters $\mathbf{a}$ and hence Eq. (9.1) implies possibly infinitely many relationships between the vacuum expectation values of products of the field operators (n-point functions). In special cases, a direct contradiction[20] can be demonstrated, though this is less easy to show in general. However, for linear transformations no contradiction results as long as the vacuum expectation values of the fields vanish. Hence the vacuum may have the symmetry of the linear subgroup H and this leads as a consequence of the Goldstone boson theorem to the appearance of $(n_G - n_H)$ massless bosons, one for each degree of symmetry lost.[21] The precise relationship of this result to the fact, noted in Sec. 8, that the meson mass term must cause a symmetry breaking of the Lagrangian is of some interest and merits further study.

Further insight into the consequences of nonlinear transformations can be made with the aid of the following hypothesis. Suppose that, corresponding to the field operators x_i, there exist particle creation and annihilation operators $\hat{a}_i^*(\mathbf{k})$ and $\hat{a}_i(\mathbf{k})$ which satisfy the canonical commutation relation

$$[\hat{a}_i^*(\mathbf{k}), \hat{a}_j(\mathbf{k}')] = \delta_{ij}\delta(\mathbf{k} - \mathbf{k}'), \qquad (9.2)$$

and that the normalized sums $(2)^{-\frac{1}{2}}[\hat{a}_i^*(\mathbf{k}) + \hat{a}_i(\mathbf{k})]$ transform like the fields x_i under G. (In the free-field case, we should interpret these operators as Fourier transforms of the fields.) Then we may give the following explicit expression for the unitary operators U_g. This is

$$U_g \equiv U_a = \exp a_\mu \int \tfrac{1}{2}\{f_{\mu i}((2)^{-\frac{1}{2}}[\hat{\mathbf{a}}(\mathbf{k}) + \hat{\mathbf{a}}^*(\mathbf{k})]), (2)^{-\frac{1}{2}}[\hat{a}_i(\mathbf{k}) - \hat{a}_i^*(\mathbf{k})]\}_+ \, d\mathbf{k},$$

where the $\mathbf{a}$ define the group parameters of $g \in G$. U_g acts linearly on the states developed from the vacuum by the action of the creation operators $\hat{a}_i^*(\mathbf{k})$. Nonlinearity of the field transformation manifests itself as an action on the states which changes particle number. Moreover, given $f_{\mu i}$ in canonical form, it is easily shown that the vacuum is not invariant under the G/H cosets. Invariant states may be constructed by application of the averaging operator $\int_G U_g \, dg$ (where dg is the Haar measure) to an arbitrary state. Such states are in general linear combinations of states associated with differing particle number.

10. LINEARIZATION BY EMBEDDING

In Sec. 2 we considered the problem of linearizing the representation by transforming from one set of coordinates to another with the same number of components. Suppose we now permit the second set to have an arbitrary number of components. Then, in an essentially trivial fashion, we can always carry out a formal linearization. Thus, given a group representation specified by the analytic function $\mathbf{h}(\mathbf{x}, \mathbf{a})$ of the coordinates x_i and the group parameters a_μ, we expand it as power series in the x_i identifying each monomial as a distinct coordinate of the new space. That is, we write

$$y_{j_1 \cdots j_n} = x^{j_1} x^{j_2} \cdots x^{j_n}, \quad \text{for all } n = 1, 2, \cdots, j_i = 1, 2, \cdots.$$

Clearly $h(y)$ is a linear function and hence the group representation is linear in these coordinates. The y space is, in general, infinite dimensional with an infinite number of nonlinear relations between the individual coordinates which define a surface corresponding to the embedded manifold on which the group acts.

In certain special cases, a more judicious choice of coordinates may be made which permits the y space to be finite dimensional. This may be exemplified by a study of the nonlinear representations discussed by Chang and Gürsey.[3] In this, the transformation of the pion field (π_1, π_2, π_3) under the chiral group K_2 is determined by the 8×8 matrix function U with argument $i\gamma_5 \boldsymbol{\tau} \cdot \boldsymbol{\pi}$. Under isospin, the transformation of the

pion field is linear and this holds, irrespective of the functional form of U. On the other hand, the remaining part of the chiral transformation defined by

$$U \to U' = \exp(-\tfrac{1}{2}i\gamma_5\boldsymbol{\tau}\cdot\mathbf{a})U\exp(-\tfrac{1}{2}i\gamma_5\boldsymbol{\tau}\cdot\mathbf{a}) \tag{10.1}$$

is, in general, nonlinear. However, it is easy to see from Eq. (10.1) that if, in terms of some new fields, U is a linear function, then these new fields transform linearly. To find such fields, we first remark that the different functional forms for U given by Chang and Gürsey all lead to transformations on the pion fields which as representations of K_2 are equivalent. Hence it is sufficient to consider Model B described by these authors. In this, U takes the form

$$U = \left(\frac{1-f^2\pi^2}{1+f^2\pi^2}\right) + 2if\gamma_5\frac{\boldsymbol{\tau}\cdot\boldsymbol{\pi}}{(1+f^2\pi^2)},$$

where f is an arbitrary real constant. Let us make the change of variable defined by

$$\tilde{\pi}_i = \frac{2f\pi_i}{1+f^2\pi^2}, \quad i=1,2,3; \quad \tilde{\pi}_4 = \left(\frac{1-f^2\pi^2}{1+f^2\pi^2}\right). \tag{10.2}$$

It is easy to see that in terms of these new variables U is a linear function, and hence the new coordinates transform linearly. Moreover, Eq. (10.2) is just the stereographic projection of R^3 onto the surface of a sphere embedded in R^4, this surface being the representation manifold. Correspondingly, the functions [see, for example, Ref. 2, Eq. (2.21)] which define the infinitesimal generators of the transformation of the pion fields are linear in these new coordinates.

The physical significance of this result is that, by introduction of additional fields, it is always possible to replace nonlinear transformations by linear ones. We shall not dwell on the conceptual advantages or disadvantages inherent in such a procedure. It suffices to remark that a similar situation occurs in gravitation theory[22] and, as Fock[8] has shown in his study of the hydrogen atom, in ordinary quantum mechanics. Finally, in the example described above, the additional coordinate can be identified with the σ field of Gell-Mann and Lévy[2] introduced in their σ model for β decay.

ACKNOWLEDGMENTS

The authors would like to thank Professor Y. Ne'eman for encouragement, advice, and the hospitality of the Department of Physics of Tel Aviv University. They also acknowledge useful conversations with Professor L. P. Horwitz and Dr. J. T. Lewis. One of us A. I. S.) would also like to thank Professor C. A. Coulson for the hospitality of the Mathematical Institute, Oxford.

APPENDIX A

Let G be the additive group of the circle (the real line modulo 2π), and M the complex plane.[23] Then $T(a)$, defined by

$$z' \equiv T(a)z = f^{-1}(e^{ia}f(z)), \quad a\in[0,2\pi], \tag{A1}$$

where f and f^{-1} are analytic at $z=0$, is a nonlinear representation of G. Suppose further that

$$f(z) = z + O(z^2), \quad f^{-1}(z) = z + O(z^2)$$

(for example, choose f to be the sine function). Then $D(a)$ as defined by Eq. (2.7) is multiplication by $\exp ia$. Thus, applying Eq. (2.8), we obtain

$$\omega = \int_0^{2\pi} e^{-ia}f^{-1}(e^{ia}f(z))\,da/2\pi \tag{A2}$$

for the coordinates ω in which the representation is linear. To verify this, we set $b = (f(z)e^{ia})$ and substitute in (A2). This gives the contour integral

$$\omega = [f(z)/2\pi i]\int_C b^{-2}f^{-1}(b)\,db,$$

where C is a circle of radius $f(z)$ about $b=0$. The integrand has a simple pole at $b=0$; hence

$$\omega = f(z).$$

Substituting this result back into (A1), we obtain

$$\omega' = D(a)\omega = e^{ia}\omega,$$

which is a linear representation as required.

For noncompact groups the linearization procedure given in Sec. 2 may fail, even though it is possible to linearize the representation by a change of coordinates. This is shown in the following example.

Let G be the additive group of the real line and M the complex plane with the origin excluded. Then $T(a)$, defined by

$$T(a)z = z(1-iaz)^{-1}, \quad a\in(-\infty,\infty),$$

is a nonlinear representation of G. $D(a)$ as defined by Eq. (2.7) does not depend on a, being simply the identity representation. Hence it cannot be equivalent to $T(a)$ as equivalent representations must be in one-to-one correspondence. Correspondingly, the group integration (2.8) fails. A linearization can be effected through the transformation $\omega = \exp(-z^{-1})$. This gives the representation

$$T(a)\omega = (\exp ia)\omega,$$

with M' an infinitely sheeted Riemann surface. However, it should also be noted that the origin is excluded from M, and were it to be included, it would not then be possible to make the transformation analytic at this point.

APPENDIX B

In the following, we verify that

$$T(\mathbf{a})\mathbf{x} \equiv (\exp \mathbf{a} \cdot \mathbf{X})\mathbf{x}$$

with X_λ defined by (4.8) is a representation of the Lie group generated by the X_λ, and obtain the functions $\Xi_{\lambda\mu}(\mathbf{a})$ defined by (4.3). We start with two lemmas. In these K denotes as before the Lie algebra.

Lemma B1: Given $\mathbf{a}$, $\mathbf{b}$ vectors of suitable dimension and $X_\lambda \in K$, then

$$(\exp \mathbf{a} \cdot \mathbf{X})(\exp \mathbf{b} \cdot \mathbf{X}) = \exp \{\boldsymbol{\phi}(\mathbf{a}, \mathbf{b}) \cdot \mathbf{X}\},$$

where $\boldsymbol{\phi}$ is a vector-valued function of $\mathbf{a}$, $\mathbf{b}$ analytic at $\mathbf{a} = \mathbf{b} = \mathbf{0}$ (that is, at the identity $e \in G$).

Proof: This is an elementary consequence of the fact that given X, Y arbitrary in K, then

$$(\exp tX)(\exp tY) = \exp Z(t),$$

where $Z(t) \in K$ and $Z(t)$ is analytic in t at $t = 0$ [Ref. 11(c), Chap. V, p. 172].

Lemma B2: Given $\mathbf{h}(\mathbf{x})$ a vector-valued infinitely differentiable function on M, then

$$(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{h}(\mathbf{x}) = \mathbf{h}\{(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{x}\}.$$

Proof: We first note that if $f(t)$ is infinitely differentiable, then

$$\left\{\left(\exp \frac{d}{dt}\right) f(t)\right\} = f(t) + \frac{df(t)}{dt} + \cdots + \frac{1}{n!}\frac{d^n f(t)}{dt^n} + \cdots = f(t+1) = f\left\{\left(\exp \frac{d}{dt}\right) t\right\}. \tag{B1}$$

Now solve the set of differential equations (for which the integrability conditions are trivial)

$$a_\mu v_{i\mu}[\mathbf{x}(t)] = \frac{dx_i(t)}{dt} \tag{B2}$$

and substitute in Eq. (4.8). This gives

$$\mathbf{a} \cdot \mathbf{X} = \frac{d}{dt}. \tag{B3}$$

Now from Eq. (B2) by differentiation, it follows, since the $v_{i\mu}(\mathbf{x})$ are analytic, that the x_i are infinitely differentiable for all i. Hence $\mathbf{h}(\mathbf{x}(t))$ is infinitely differentiable in t. Then use of Eqs. (B1) and (B3) gives

$$(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{h}(\mathbf{x}(t)) = \mathbf{h}\{\mathbf{x}(\exp \mathbf{a} \cdot \mathbf{X}t)\} = \mathbf{h}\{(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{x}(t)\},$$

which proves the lemma.

The main result now follows easily. Define

$$(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{x} = \mathbf{h}(\mathbf{x}, \mathbf{a}). \tag{B4}$$

To show that this is a group representation we must verify Eq. (4.2). According to Lemma (B1), we have that

$$(\exp \mathbf{a} \cdot \mathbf{X} \exp \mathbf{b} \cdot \mathbf{X})\mathbf{x} = \{\exp \boldsymbol{\phi}(\mathbf{a}, \mathbf{b}) \cdot \mathbf{X}\}\mathbf{x} = \mathbf{h}[\mathbf{x}, \boldsymbol{\phi}(\mathbf{a}, \mathbf{b})]. \tag{B5}$$

On the other hand, from Lemma (B2),

$$(\exp \mathbf{a} \cdot \mathbf{X} \exp \mathbf{b} \cdot \mathbf{X})\mathbf{x} = \exp \mathbf{a} \cdot \mathbf{X}\{\mathbf{h}(\mathbf{x}, \mathbf{b})\} = \mathbf{h}\{(\exp \mathbf{a} \cdot \mathbf{X})\mathbf{x}, \mathbf{b}\} = \mathbf{h}[\mathbf{h}(\mathbf{x}, \mathbf{a}), \mathbf{b}],$$

which combined with Eq. (B5) proves the required result.

The functions $\Xi_{\lambda\mu}(\mathbf{a})$ defined by Eq. (4.3) can be obtained by differentiation of Eq. (B4). Thus, setting $\mathbf{h} = \mathbf{x}'$, we have

$$\frac{\partial x_i'}{\partial a_\mu} = \left\{\frac{\partial}{\partial a_\mu}(\exp \mathbf{a} \cdot \mathbf{X})\right\} x_i = \exp \mathbf{a} \cdot \mathbf{X}\left\{\frac{1 - e^{-\Delta_{\mathbf{a}\cdot\mathbf{x}}}}{\Delta_{\mathbf{a}\cdot\mathbf{x}}}\right\} X_\mu x_i \quad \text{from (5.5).}$$

Setting

$$\{(1 - e^{-\Delta_{\mathbf{a}\cdot\mathbf{x}}})/\Delta_{\mathbf{a}\cdot\mathbf{x}}\} X_\mu \equiv \Xi_{\lambda\mu}(\mathbf{a}),$$

we have

$$\frac{\partial x_i'}{\partial a_\mu} = \Xi_{\lambda\mu}(\mathbf{a})(\exp \mathbf{a} \cdot \mathbf{X}) X_\lambda x_i = \Xi_{\lambda\mu}(\mathbf{a})(\exp \mathbf{a} \cdot \mathbf{X}) v_{i\lambda}(\mathbf{x}) \quad \text{from Eq. (4.8)} = \Xi_{\lambda\mu}(\mathbf{a}) v_{i\lambda}(\mathbf{x}') \quad \text{from Lemma B2.}$$

The $\Xi^{-1}_{\lambda\mu}$ form a nonlinear representation of K linear only on the identity. It is the same as that obtained from the global procedure. Indeed it is just the infinitesimal form of the representation $g: \mathbf{a} \to \mathbf{a}'$ of G defined by right translation

$$(\exp \mathbf{a} \cdot \mathbf{X})g = \exp \mathbf{a}' \cdot \mathbf{X}.$$

This may be identified with Eq. (2.18).

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR grant number EOOAR-68-0010, through the European Office of Aerospace Research.

† Permanent address: Mathematical Institute, Oxford, England.
‡ Permanent address: Polytechnic Institute of Brooklyn, Brooklyn, New York 11201.

[1] S. Weinberg, Phys. Rev. Letters **18**, 1881 (1967).
[2] M. Gell-Mann and M. Lévy, Nuovo Cimento **16**, 705 (1960); J. Schwinger, Ann. Phys. (N.Y.) **2**, 407 (1957).

[3] (a) F. Gürsey, Nuovo Cimento **16**, 230 (1960); (b) J. Wess and B. Zumino, Phys. Rev. **163**, 1727 (1967); (c) L. S. Brown, Phys. Rev. **163**, 1802 (1967); (d) P. Chang and F. Gürsey, Phys. Rev. **164**, 1752 (1967); (e) Y. P. Yao, Phys. Rev. **171**, 1702 (1968).

[4] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969).

[5] C. G. Callan, S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2247 (1969).

[6] J. Schwinger, Phys. Rev. Letters **B24**, 473 (1967).

[7] D. B. Fairlie and K. Yoshida, Ann. Phys. (N.Y.) **46**, 326 (1968).

[8] V. Fock, Z. Physik **98**, 145 (1935); B. Kurşunoğlu, *Modern Quantum Theory* (Miller Freeman Publications, New York, 1962), Chap. XIII.

[9] S. Weinberg, Phys. Rev. **166**, 1568 (1968).

[10] A. J. MacFarlane and P. H. Weisz, Nuovo Cimento **559**, 853 (1968).

[11] General information on Lie groups and Lie algebras may be found in the following: (a) M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley Publ. Co., London, 1962); (b) P. M. Cohn, *Lie Groups* (Cambridge University Press, Cambridge, England, 1965); (c) N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., London, 1962); (d) S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., London, 1962).

[12] We use $T(g)\mathbf{x}$ to denote the function $\mathbf{h}(\mathbf{x}, \mathbf{a})$ when the parametrization of G is undefined. We also use $T(\mathbf{a})\mathbf{x}$ for $\mathbf{h}(\mathbf{x}, \mathbf{a})$.

[13] We are indebted to the referee for the remarks leading to this generalization.

[14] This point has been apparently overlooked in Refs. 4 and 5.

[15] A. Messiah, *Quantum Mechanics* (North-Holland Publ. Co., Amsterdam, 1965), Vol. I, Chap. VIII, pp. 339–40.

[16] The canonical transformation on the baryon fields performed by Chang and Gürsey in Ref. 3, Sec. 3, transforms their baryon field into standard form, thus ensuring the invariance of their baryon-mass term.

[17] In this, the error in Ref. 5 noted previously (Ref. 14) has been corrected.

[18] W. Sollfrey, Phys. Rev. **173**, 1805 (1968).

[19] J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962); D. Lurié, *Particles and Fields* (Interscience Publishers, Inc., New York, 1968), Chap. 10, pp. 472–78.

[20] P. B. Kantor and J. L. Pietenpol, Phys. Rev. Letters **21**, 241 (1968).

[21] T. W. B. Kibble, Phys. Rev. **155**, 1554 (1967).

[22] J. Rosen, Rev. Mod. Phys. **37**, 204 (1965).

[23] Although we choose M to be the complex plane, it is easy to see, by equating real and imaginary parts, that the analysis can be translated to a real two-dimensional analytic manifold.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Some Remarks on Classical Scalar Radiation from Point Sources

ROBERT G. CAWLEY
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland 20910

(Received 27 June 1969)

The radiation from a point source with proper time-dependent coupling strength to a real classical scalar field is worked out. The adiabatic switching limit reproduces the results of the physical case of constant coupling for the radiation. The problem of defining a "radiated" field is discussed.

1. INTRODUCTION

The radiation of a massive real Lorentz scalar field has recently been treated in detail for the case of a finite number of point sources having nonvanishing accelerations over a finite region of space–time.[1,2] The present paper extends this effort by working out the case where the source coupling to the field is time dependent. We apply the results to two problems, the first being the adiabatic switching of the source strength, and the second, whose purpose is to study the problem of localization of the source of the radiation field, is the sudden switching limit. We find that the adiabatic limit of the radiation from the source with time-dependent coupling reproduces the constant coupling results. The sudden limit is a representation of sources with "truncated world lines" and gives significant extra radiation from the time variation of the coupling, regardless of when the switching is performed.

It is also possible to apply our results to the problem of the classical theory of the massive Lorentz vector field, though we do not do this here.

Section 2 obtains the field of a single prescribed point source with τ-dependent coupling (superposition gives the many-particle result), Sec. 3 works out the radiation obtained for the example of a specific switching function (with compact support) and determines the adiabatic limit, Sec. 4 discusses the problem of localizing the source of the radiation, and Sec. 5 is a brief summary of results.

2. RADIATION FROM A POINT SOURCE OF TIME-VARYING STRENGTH

The Klein–Gordan equation for a real scalar field in the presence of an external source is[3]

$$(\Box + m^2)\varphi(x) = \rho(x), \tag{2.1}$$

and for $\rho(x)$ we consider the following:

$$\rho(x) = \int_{-\infty}^{+\infty} d\tau G(\tau)\delta^{(4)}(x - \xi(\tau)), \tag{2.2}$$

which represents a particle with world line *prescribed* by

$$x_\mu = \xi_\mu = \xi_\mu(\tau). \tag{2.3}$$

[3] (a) F. Gürsey, Nuovo Cimento **16**, 230 (1960); (b) J. Wess and B. Zumino, Phys. Rev. **163**, 1727 (1967); (c) L. S. Brown, Phys. Rev. **163**, 1802 (1967); (d) P. Chang and F. Gürsey, Phys. Rev. **164**, 1752 (1967); (e) Y. P. Yao, Phys. Rev. **171**, 1702 (1968).

[4] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969).

[5] C. G. Callan, S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2247 (1969).

[6] J. Schwinger, Phys. Rev. Letters **B24**, 473 (1967).

[7] D. B. Fairlie and K. Yoshida, Ann. Phys. (N.Y.) **46**, 326 (1968).

[8] V. Fock, Z. Physik **98**, 145 (1935); B. Kurşunoğlu, *Modern Quantum Theory* (Miller Freeman Publications, New York, 1962), Chap. XIII.

[9] S. Weinberg, Phys. Rev. **166**, 1568 (1968).

[10] A. J. MacFarlane and P. H. Weisz, Nuovo Cimento **559**, 853 (1968).

[11] General information on Lie groups and Lie algebras may be found in the following: (a) M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley Publ. Co., London, 1962); (b) P. M. Cohn, *Lie Groups* (Cambridge University Press, Cambridge, England, 1965); (c) N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., London, 1962); (d) S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., London, 1962).

[12] We use $T(g)\mathbf{x}$ to denote the function $\mathbf{h}(\mathbf{x}, \mathbf{a})$ when the parametrization of G is undefined. We also use $T(\mathbf{a})\mathbf{x}$ for $\mathbf{h}(\mathbf{x}, \mathbf{a})$.

[13] We are indebted to the referee for the remarks leading to this generalization.

[14] This point has been apparently overlooked in Refs. 4 and 5.

[15] A. Messiah, *Quantum Mechanics* (North-Holland Publ. Co., Amsterdam, 1965), Vol. I, Chap. VIII, pp. 339–40.

[16] The canonical transformation on the baryon fields performed by Chang and Gürsey in Ref. 3, Sec. 3, transforms their baryon field into standard form, thus ensuring the invariance of their baryon-mass term.

[17] In this, the error in Ref. 5 noted previously (Ref. 14) has been corrected.

[18] W. Sollfrey, Phys. Rev. **173**, 1805 (1968).

[19] J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962); D. Lurié, *Particles and Fields* (Interscience Publishers, Inc., New York, 1968), Chap. 10, pp. 472–78.

[20] P. B. Kantor and J. L. Pietenpol, Phys. Rev. Letters **21**, 241 (1968).

[21] T. W. B. Kibble, Phys. Rev. **155**, 1554 (1967).

[22] J. Rosen, Rev. Mod. Phys. **37**, 204 (1965).

[23] Although we choose M to be the complex plane, it is easy to see, by equating real and imaginary parts, that the analysis can be translated to a real two-dimensional analytic manifold.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Some Remarks on Classical Scalar Radiation from Point Sources

ROBERT G. CAWLEY
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland 20910

(Received 27 June 1969)

The radiation from a point source with proper time-dependent coupling strength to a real classical scalar field is worked out. The adiabatic switching limit reproduces the results of the physical case of constant coupling for the radiation. The problem of defining a "radiated" field is discussed.

1. INTRODUCTION

The radiation of a massive real Lorentz scalar field has recently been treated in detail for the case of a finite number of point sources having nonvanishing accelerations over a finite region of space–time.[1,2] The present paper extends this effort by working out the case where the source coupling to the field is time dependent. We apply the results to two problems, the first being the adiabatic switching of the source strength, and the second, whose purpose is to study the problem of localization of the source of the radiation field, is the sudden switching limit. We find that the adiabatic limit of the radiation from the source with time-dependent coupling reproduces the constant coupling results. The sudden limit is a representation of sources with "truncated world lines" and gives significant extra radiation from the time variation of the coupling, regardless of when the switching is performed.

It is also possible to apply our results to the problem of the classical theory of the massive Lorentz vector field, though we do not do this here.

Section 2 obtains the field of a single prescribed point source with τ-dependent coupling (superposition gives the many-particle result), Sec. 3 works out the radiation obtained for the example of a specific switching function (with compact support) and determines the adiabatic limit, Sec. 4 discusses the problem of localizing the source of the radiation, and Sec. 5 is a brief summary of results.

2. RADIATION FROM A POINT SOURCE OF TIME-VARYING STRENGTH

The Klein–Gordan equation for a real scalar field in the presence of an external source is[3]

$$(\Box + m^2)\varphi(x) = \rho(x), \tag{2.1}$$

and for $\rho(x)$ we consider the following:

$$\rho(x) = \int_{-\infty}^{+\infty} d\tau G(\tau)\delta^{(4)}(x - \xi(\tau)), \tag{2.2}$$

which represents a particle with world line *prescribed* by

$$x_\mu = \xi_\mu = \xi_\mu(\tau). \tag{2.3}$$

The world-line parameter τ we choose to be the future-favoring particle proper time,

$$u_\mu(\tau)^2 = 1, \quad u_0(\tau) = \frac{d\xi_0(\tau)}{d\tau} > 0. \tag{2.4}$$

We allow in Eq. (2.2) for the coupling strength of the particle G to the field $\varphi(x)$ to vary in a smooth way with the particle proper time. The case $G(\tau) = g =$ const was treated in Ref. 1 (hereafter called I).

To obtain the solution to Eq. (2.1) we use the retarded Green's function

$$\Delta_R(x) = [\theta(t)/2\pi][\delta(\lambda) - \tfrac{1}{2}m\lambda^{-\frac{1}{2}}\theta(\lambda)J_1(m\lambda^{\frac{1}{2}})], \tag{2.5}$$

where θ ($t = x_0$) is the unit step function and $\lambda = x^2$. The solution is

$$\varphi(x) = \varphi^{\text{in}}(x) + \int \rho(x')\Delta_R(x - x')\, d^4x', \tag{2.6}$$

where $\varphi^{\text{in}}(x)$ is a solution to the homogeneous variant of Eq. (2.1) satisfying appropriate initial conditions given for $\varphi(x)$ in the remote past.

We take $\varphi^{\text{in}}(x) = 0$ and proceed to simplify Eq. (2.6) as in I. Substitute Eqs. (2.2) and (2.5) to get

$$\varphi(x) = \frac{G_R}{4\pi(x - \xi_R)\cdot u_R} - \frac{m}{4\pi}\int_{-\infty}^{\tau_R(x)} d\tau G(\tau)\frac{J_1(m\lambda_\tau^{\frac{1}{2}})}{\lambda_\tau^{\frac{1}{2}}}, \tag{2.7}$$

with $\lambda_\tau = [x - \xi(\tau)]^2$, $G_R = G(\tau_R)$, and $\tau_R(x)$ the smaller of the two solutions to $\lambda_\tau = 0$; change variables to $\zeta = m\lambda_\tau^{\frac{1}{2}}$ with the help of

$$\frac{\partial\tau(x,\zeta)}{\partial\zeta} = -\zeta[m^2(x - \xi)\cdot u]^{-1},$$

to find

$$\varphi(x) = \frac{G_R}{4\pi(x - \xi_R)\cdot u_R} - \frac{1}{4\pi}\int_0^\infty d\zeta G(\tau)\frac{J_1(\zeta)}{(x - \xi(\tau))\cdot u(\tau)}; \tag{2.8}$$

and integrate by parts to get

$$\varphi(x) = \frac{-1}{4\pi m^2}\int_0^\infty d\zeta\zeta J_0(\zeta)\left(\frac{G[-1 + (x - \xi)\cdot w]}{[(x - \xi)\cdot u]^3} - \frac{dG/d\tau}{[(x - \xi)\cdot u]^2}\right) \tag{2.9}$$

$$= \frac{-1}{4\pi}\int_{-\infty}^{\tau_R(x)} d\tau J_0(m\lambda_\tau^{\frac{1}{2}})\left(\frac{G[-1 + (x - \xi)\cdot w]}{[(x - \xi)\cdot u]^2} - \frac{dG/d\tau}{(x - \xi)\cdot u}\right), \tag{2.10}$$

where $w_\mu = w_\mu(\tau) = d^2\xi_\mu(\tau)/d\tau^2$ is the 4-acceleration of the particle. We assume that w_μ vanishes unless $\tau_1 < \tau < \tau_2$. Expressions for $\varphi_{,\mu}(x)$ are also needed; we have

$$\begin{aligned}\varphi_{,\mu}(x) = \frac{-1}{4\pi m^2}\int_0^\infty d\zeta\zeta J_0(\zeta)\big(&G\{[(x - \xi)\cdot u]^2 w_\mu \\ &- [3(-1 + (x - \xi)\cdot w)(x - \xi)\cdot u]u_\mu \\ &+ [(x - \xi)\cdot w(x - \xi)\cdot u \\ &- 3(-1 + (x - \xi)\cdot w)^2](x - \xi)_\mu\} \\ &\times [(x - \xi)\cdot u]^{-5} \\ &+ \dot G\{[3(-1 + (x - \xi)\cdot w)](x - \xi)_\mu \\ &+ [2(x - \xi)\cdot u]u_\mu\}[(x - \xi)\cdot u]^{-4} \\ &- \ddot G(x - \xi)_\mu[(x - \xi)\cdot u]^{-3}\big),\end{aligned} \tag{2.11}$$

where the dot denotes differentiation with respect to τ. Useful in the derivation of this expression is the relation

$$\partial_\mu\tau(x, \zeta) = (x - \xi)_\mu/[(x - \xi)\cdot u]. \tag{2.12}$$

Equation (2.11) can be converted to a τ integral after the manner in which (2.10) is obtained from (2.9). Finally, it will be noted that all these results reproduce the corresponding constant $G(\tau)$ results of I.

We were concerned in I with the asymptotic behavior of the solutions (2.9), (2.10), and (2.11) for the case where $G(\tau)$ is constant, and where the source world line satisfied $w_\mu(\tau) \neq 0$ only for $\tau_1 < \tau < \tau_2$. We have retained the latter restriction in this paper, while for $G(\tau)$ we are interested in the following case:

$$\begin{aligned} G(\tau) &= 0, && \tau \leq \tau_e, \\ &= g(\tau), && \tau_e < \tau < \tau_l, \\ &= 0, && \tau_l \leq \tau, \end{aligned} \tag{2.13}$$

where $\tau_e < \tau_1$ and $\tau_l > \tau_2$. We also request of $G(\tau)$ that its derivative vanish at τ_e and τ_l.

Radiation from this source is defined essentially as in I. Enclose the portion of the particle's world line between $\xi_\mu(\tau_e)$ and $\xi_\mu(\tau_l)$ by a Minkowski cylinder Σ (Fig. 1) with axis along the time axis of an arbitrary observer frame and denote the radius of Σ by R. The cylinder extends into the infinite past and the infinite future. A Minkowski surface defines a Lorentz equivalent class of measurement procedures[2]: one of those defined by Σ is the following: measurement over all times of flux densities through a sphere of radius R. Thus, we define the radiated energy momentum into the φ field by

$$P_\mu^{\text{rad}} = \lim_{R\to\infty}\int_\Sigma d^3\sigma_\nu T_{\mu\nu}, \tag{2.14}$$

where $T_{\mu\nu}$ is the canonical stress–energy tensor of the

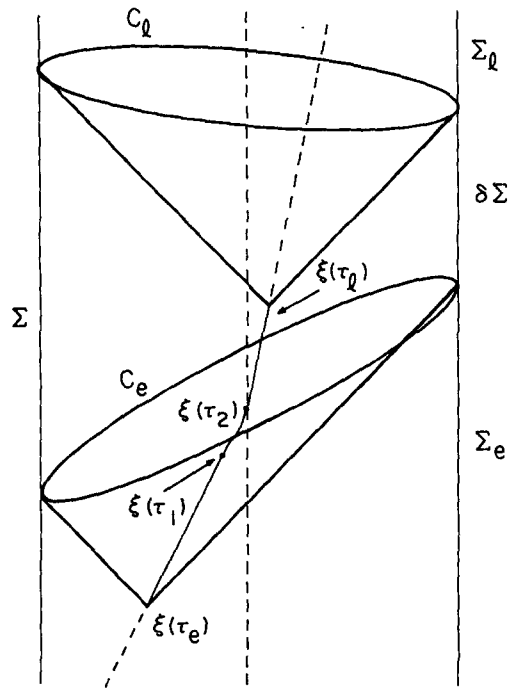


FIG. 1. World line of a particle accelerated over a finite interval $(\xi_0(\tau_1), (\xi_0(\tau_2))$ and coupled to a classical scalar field over a finite interval $(\xi_0(\tau_e), \xi_0(\tau_l)) \supset (\xi_0(\tau_1), \xi_0(\tau_2))$. The surface Σ represents the history of the observation sphere used in defining the radiation.

free φ field,

$$T_{\mu\nu} = \varphi_{,\mu}\varphi_{,\nu} - g_{\mu\nu}\mathfrak{L}, \tag{2.15}$$

$$\mathfrak{L} = \tfrac{1}{2}(\varphi_{,\alpha}\varphi_{,\alpha} - m^2\varphi^2). \tag{2.16}$$

The surface element $d^3\sigma_\nu$ is given by

$$d^3\sigma_\nu = n_\nu d^3\sigma, \tag{2.17}$$

where

$$n_\nu = (0, \hat{n}) = (0, -\hat{R}) \tag{2.18}$$

toward the time axis because Σ is a timelike surface. It can be seen that Eq. (2.18) singles out the μ-space indices of $T_{\mu\nu}$, and so Eq. (2.14) involves integration of flux densities in the observer frame of Fig. 1.

In order to determine the value of the limit in Eq. (2.14) we consider contributions to the integral from different regions; to accomplish this, we break up Σ into three parts as follows. Let the intersection of Σ and the forward light cone from $\xi_\mu(\tau_\alpha)$ be denoted by C_α, where α may be either of e or l (see Fig. 1). The portion below C_e we denote by Σ_e, that above C_l we call the "wake" and denote by Σ_l, while the region between C_e and C_l we call[4] the "surf" and denote by $\delta\Sigma$. Finally, choose origin of coordinates so that $\xi_\mu(\tau_2) = 0$.

It can be seen from Eqs. (2.10) and (2.11) that $\varphi(x) = 0$ on Σ_e so this portion makes no contribution to the radiation. For the problem studied in I it was found that the net contribution to the radiation from the surf region of the Σ integration vanished because $\varphi(x)$ and $\varphi_{,\mu}(x)$ were each $O(R^{-\frac{5}{4}})$ for large R and the surf has finite temporal extent. Scrutiny of Eqs. (3.9) and (3.10) of I shows that $\varphi(x)$, as given by Eqs. (2.9) and (2.10) of the present paper, is also $O(R^{-\frac{5}{4}})$; the same is true for $\varphi_{,\mu}(x)$. Hence, no net contribution to the radiation is provided in the present case by $\delta\Sigma$. The radiation is all received on Σ_l and signals from portions of the world line between τ_e and τ_l all contribute. Again the analysis of I goes through. One has only to replace Eq. (3.15) of I by

$$\int_0^\infty d\zeta \to \int_{\zeta_l}^{\zeta_e} = \int_{\zeta_l}^{\zeta_2} + \int_{\zeta_2}^{\zeta_1} + \int_{\zeta_1}^{\zeta_e}, \tag{2.19}$$

where $\zeta_\alpha \equiv m\lambda^{\frac{1}{2}}\tau_\alpha$, and then observe that Eq. (3.19) of I gives the same result as found for the case of constant G; a slight improvement of this result[2] is

$$\varphi(x) = O((t - R)^{-\frac{1}{4}}t^{-\frac{5}{4}}). \tag{2.20}$$

$\varphi_{,\mu}(x)$ has the same behavior. In this way we find the leading terms of $\varphi(x)$ and $\varphi_{,\mu}(x)$, over regions of Σ from which nonvanishing contributions to the radiation arise, to be

$$\varphi(x) \sim m^{-\frac{1}{2}}\lambda^{-\frac{3}{4}}\,\mathrm{Re}\,[\tilde{\varphi}(U)e^{-i\psi}] \tag{2.21a}$$

$$\varphi_{,\mu}(x) \sim m^{-\frac{1}{2}}\lambda^{-\frac{3}{4}}\,\mathrm{Re}\,[-imU_\mu\tilde{\varphi}(U)e^{-i\psi}], \tag{2.21b}$$

where $\lambda = x^2$, $U_\mu = \lambda^{-\frac{1}{2}}x_\mu$, and

$$\tilde{\varphi}(U) = \int_{\tau_e}^{\tau_l} d\tau e^{imU\cdot\xi(\tau)}\left(g(\tau)\frac{U\cdot w(\tau)}{[U\cdot u(\tau)]^2} - \dot{g}(\tau)\frac{1}{U\cdot u(\tau)}\right); \tag{2.22}$$

also $\psi(x) = m\lambda^{\frac{1}{2}} + \frac{3}{4}\pi$. Reference 2 shows how these polar forms are to be used for the evaluation of integrals of the type given in Eq. (2.14) and how a particle interpretation to the asymptotic field can be introduced. Indeed the only differences arising anywhere in the present problem come from the $\dot{g}(\tau)$ term of Eq. (2.22) so that all the results of I and II can be used without further modification. In particular the case of more than one source is handled by replacing $\tilde{\varphi}(U)$ in expressions for the radiation (energy, momentum, particle number) by a linear combination

$$\tilde{\varphi}(U) \to \sum_i \tilde{\varphi}^{(i)}(U) \tag{2.23}$$

of terms of the form of that on the right side of Eq. (2.22). Passage to the case of radiation from a distribution of point sources should be possible.[5]

3. DEFINITION OF RADIATED FIELD FROM AN ADIABATICALLY SWITCHED POINT SOURCE

We specialize the results of the preceding section to the choice of $g(\tau)$,

$$\begin{aligned} g(\tau) &= gh_e(\tau), \quad \tau_e < \tau \leq \tau_1, \\ &= g, \qquad \tau_1 \leq \tau \leq \tau_2, \\ &= gh_l(\tau), \quad \tau_2 \leq \tau < \tau_l, \end{aligned} \tag{3.1}$$

where for definiteness we take as switching functions

$$h_e(\tau) = \exp\left[-(\tau_1 - \tau)/(\tau - \tau_e)\right], \tag{3.2a}$$

$$h_l(\tau) = \exp\left[-(\tau - \tau_2)/(\tau_l - \tau)\right]. \tag{3.2b}$$

Remembering that $w = 0$ unless $\tau_1 < \tau < \tau_2$, we find that insertion into Eq. (2.25) results in

$$\tilde{\varphi}(U) = \tilde{\varphi}^{(0)}(U) + \delta\tilde{\varphi}(U), \tag{3.3}$$

where

$$\tilde{\varphi}^{(0)}(U) = (2\pi)^{-\frac{3}{2}} g \int_{\tau_1}^{\tau_2} d\tau e^{imU\cdot\xi} \frac{U \cdot w}{(U \cdot u)^2}, \tag{3.4}$$

which is the result for the case of a proper-time-independent, i.e., physical, source strength, and where

$$\delta\tilde{\varphi}(U) = \delta_e\tilde{\varphi}(U) + \delta_l\tilde{\varphi}(U), \tag{3.5}$$

with

$$\delta_e\tilde{\varphi}(U) = -(2\pi)^{-\frac{3}{2}} g \int_{\tau_e}^{\tau_1} d\tau \frac{e^{imU\cdot\xi}}{U \cdot u} \dot{h}_e(\tau) \tag{3.6a}$$

and

$$\delta_l\tilde{\varphi}(U) = -(2\pi)^{-\frac{3}{2}} g \int_{\tau_2}^{\tau_l} d\tau \frac{e^{imU\cdot\xi}}{U \cdot u} \dot{h}_l(\tau). \tag{3.6b}$$

Using the fact that $\xi(\tau)$ is linear in τ over the regions of integration in Eqs. (3.6), we find that a little algebra gives

$$\delta_e\tilde{\varphi}(U) = -(2\pi)^{-\frac{3}{2}} g \frac{e^{imU\cdot\xi_e}}{U \cdot u_1} \int_0^1 dx e^{i\lambda_e x} \left(\frac{e^{1-x^{-1}}}{x^2}\right), \tag{3.7a}$$

$$\delta_l\tilde{\varphi}(U) = +(2\pi)^{-\frac{3}{2}} g \frac{e^{imU\cdot\xi_l}}{U \cdot u_2} \int_0^1 dx e^{-i\lambda_l x} \left(\frac{e^{1-x^{-1}}}{x^2}\right), \tag{3.7b}$$

where $\xi_\alpha = \xi(\tau_\alpha)$, $u_\alpha = u(\tau_\alpha)$, with $\alpha = e$, 1, 2, or l, and where $\lambda_e = mU \cdot (\xi_1 - \xi_e)$ and $\lambda_l = mU \cdot (\xi_l - \xi_2)$. It can be seen that $\lambda_e \geq m(\tau_1 - \tau_l)$ and $\lambda_l \geq m(\tau_l - \tau_2)$ for arbitrary $u(\tau_1)$, $u(\tau_2)$, and U, since these are all timelike unit vectors.

We can show that Eqs. (3.7) imply that the $\delta_\alpha\tilde{\varphi}(U)$ are of order λ_α^{-1} when λ_α is large. Thus immediately we have

$$|\delta_\alpha\tilde{\varphi}(U)| \leq (2\pi)^{-\frac{3}{2}} g\, |I_\alpha|, \tag{3.8}$$

where

$$I_\alpha = \int_0^1 dx e^{i\lambda_\alpha x} \left(\frac{e^{1-x^{-1}}}{x^2}\right), \tag{3.9}$$

and I_α may be bounded by integrating by parts,

$$\begin{aligned} I_\alpha = (i\lambda_\alpha)^{-1} &\left(\frac{e^{i\lambda_\alpha x + 1 + x^{-1}}}{x^2}\right)\Bigg|_0^1 \\ &- (i\lambda_\alpha)^{-1} \int_0^1 dx e^{i\lambda_\alpha x} \frac{d}{dx}\left(\frac{e^{1-x^{-1}}}{x^2}\right). \end{aligned} \tag{3.10}$$

The integrated term is $(i\lambda_\alpha)^{-1} e^{i\lambda_\alpha}$ and the integral in the reflected term is bounded in absolute value by $6e$, so we have that $|I_\alpha| < (1 + 6e)\lambda_\alpha^{-1}$, and therefore by Eq. (3.8),

$$|\delta_\alpha\tilde{\varphi}(U)| = O(\lambda_\alpha^{-1}), \quad \lambda_\alpha \gg 1, \tag{3.11}$$

uniformly. If we take $m(\tau_1 - \tau_e) = m(\tau_l - \tau_2) \equiv \lambda \leq \lambda_\alpha$, we find, from Eqs. (3.3), (3.5), and (3.11), that

$$\tilde{\varphi}(U) = \tilde{\varphi}^{(0)}(U) + O(\lambda^{-1}) \tag{3.12}$$

when λ is large. All of the radiated quantities in the present example differ from those of the physical case of constant g by amounts of order $O(\lambda^{-1})$.

In a similar fashion one may show that the right sides of Eqs. (2.21) are of the form

$$\varphi(x) \sim \varphi^{(0)}(x) + (x^2)^{-\frac{1}{4}}(x \cdot v)^{-1} O(\lambda^{-1}), \tag{3.13a}$$

$$\varphi_{,\mu}(x) \sim \varphi^{(0)}_{,\mu}(x) + (x^2)^{-\frac{1}{4}}(x \cdot v)^{-1} U_\mu O(\lambda^{-1}), \tag{3.13b}$$

where v is a future-pointing unit timelike vector, and where $O(\lambda^{-1})$ represents a bound uniform in x.

Finally, it is interesting to note that the switching parameter, which must be dimensionless, is the product of source switching time, $\tau_l - \tau_2$ or $\tau_1 - \tau_e \equiv \Delta\tau$, and mass (range) parameter m for the radiated field, viz., $\lambda = m\Delta\tau$.

4. RADIATION SOURCE

The form of Eq. (2.22) suggests that for $G(\tau) =$ const the portion of the source world line between $\xi_\mu(\tau_1)$ and $\xi_\mu(\tau_2)$ is the source of the radiation. In the case of a mass-zero field it is possible to identify the portion of the source world history responsible for the radiation in just this fashion by means of the retardation condition, but when $m \neq 0$ this heuristic notion fails. The radiation from the source Eq. (2.2) with

$$G(\tau) = g = \text{const}$$

is not given correctly by the asymptotic form of the field from the truncated source,

$$\rho'(x) = g \int_{\tau_1}^{\tau_2} d\tau \delta^{(4)}(x - \xi(\tau)). \tag{4.1}$$

The source (4.1) may be regarded as one for which $g(\tau)$ is a square step, so two endpoint δ functions would show up in Eq. (2.22) from the $\dot{g}(\tau)$ term. That such terms are present may be seen more clearly from Eq. (2.8). The integration by parts leading to Eq. (2.9) involves two surface terms; when $\tau_1 < \tau_R < \tau_2$ the light-cone term is cancelled by the integrated term evaluated at τ_R, and the other integrated term does not contribute to the radiation, so again no contributions arise from the surf. In the wake, however, while $G_R = 0$ so that the light-cone term is absent from (2.8), both integrated terms are present and each contributes to the radiation. Finally, further verification is found by evaluating the right sides of Eqs. (3.6) in the sudden limit. In particular it should be noted that the extra terms do not "get smaller" if the truncation times $\xi_0(\tau_{e,l})$ tend to (plus/minus) infinity.

We conclude that the truncation of the source world line does not provide a localization of the source of the radiation. It cannot be in this sense, therefore, that the accelerated portion of the source world line may be regarded as the source of the radiation.

5. SUMMARY

We have found the fields, their asymptotic behavior, and the radiation for the case of a point source with proper time-dependent coupling strength when the source is accelerated over a finite time interval. The sudden switching limit reproduces the field and the radiation arising from a truncated source given as a proper time integral over just the accelerated portion of the world line; in addition to the physical part, endpoint contributions to the radiation also arise owing to the sudden switching, and thereby ruling out use of the truncated world line as a source for defining a "radiated field." On the other hand, the adiabatic switching limit gives back the physical result. It is interesting to note in the latter case that the switching parameter $\lambda = m\Delta\tau$, where $\Delta\tau$ is the switching time, indicating that the radiation due to the switching is larger for smaller m values, though for any specified m this has no effect on the adiabatic limit.

[1] R. G. Cawley and Egon Marx, Intern. J. Theoret. Phys. **1**, 153 (1968).

[2] R. G. Cawley, Ann. Phys. (N.Y.) **54**, 122 (1969). This is referred to in the text as II.

[3] Greek indices take all values (mod 4) and we use the modified summation convention with time-favoring metric. A comma is sometimes used to denote differentiation with respect to $x^\mu = \sum_\nu g^{\mu\nu}x_\nu$; thus

$$\varphi_{,\mu} = \partial_\mu\varphi = \left(\frac{\partial\varphi}{\partial x_0}, -\frac{\partial\varphi}{\partial x_k}\right).$$

[4] Note that this terminology differs slightly from that used in I.

[5] It is interesting to note some of the features of an extended (dispersing) source. We imagine a collection of sources confined to a finite region of space at $x_0 = 0$ and whose accelerations have all ceased by this time, with outgoing 4-velocities $u^{(i)}$, world lines $x_\mu = \xi_\mu^{(i)}(\tau)$, and coupling strengths $\tilde{g}(u^{(i)}, \tau)$. We build an extended source $\rho(x)$ formally by

$$\rho(x) \to \text{"lim"} \sum_i \int_{-\infty}^{+\infty} d\tau \tilde{g}(u^{(i)}, \tau)\delta^{(4)}(x - \xi^{(i)}(\tau))$$
$$= \int \frac{d^3\mathbf{u}}{2u_0} \int_{-\infty}^{+\infty} d\tau \tilde{G}(u, \tau)\delta^{(4)}(x - \xi(\mathbf{u}, \tau)).$$

This integral can be evaluated directly, and in particular, examined in the asymptotic region; the result is

$$\rho(x) = [\tilde{G}(U, \lambda^{\frac{1}{4}})/2\lambda^{\frac{3}{4}}][1 + O(\lambda^{-\frac{1}{4}})], \qquad \lambda \text{ large}, x_0 > 0.$$

The case of a point source takes this form formally with $\tilde{G}(U, \lambda^{\frac{1}{4}}) = g(\lambda^{\frac{1}{4}}) \cdot 2U_0\delta^{(3)}(\mathbf{U} - \mathbf{u}(\tau_2))$. The form above should be compared with the $\lambda^{-\frac{3}{4}}$ asymptotic behavior of the fields.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Two-Group Neutron Transport Theory in Spherical Geometry

T. W. SCHNATZ AND C. E. SIEWERT
Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina

(Received 23 May 1969)

A set of normal modes for the two-group steady-state neutron transport equation in spherical geometry is constructed. The singular eigenfunction-expansion technique is then used to develop a rigorous solution to the isotropically emitting spherical shell-source problem in an infinite medium.

I. INTRODUCTION

The singular eigenfunction-expansion technique introduced by Case has been used extensively in the areas of neutron transport theory and radiative transfer to construct rigorous solutions to a certain class of model problems.[1-5] This method has enjoyed particular success for energy-dependent problems,[5] for time-dependent theory,[6] for anisotropic scattering models,[3] for reactor-cell calculations,[7] and for several astrophysical applications.[4,8] Although Case's normal-mode expansion technique has been found suitable for a large number of applications, one of the major restrictions of the method is the difficulty with which the extension to nonplanar geometries is made.

Mitsis, by introducing a transform technique, solved the critical-sphere problem and he made an exhaustive study of the normal modes of the one-speed equation with spherical symmetry.[9] Leonard and Mullikin[10] and Erdmann and Siewert[11] also solved several problems in spherical one-speed theory. In the latter paper, two distinct approaches to spherical problems were employed: The first relied upon the spherical-to-plane geometry transformation for the density, and the second utilized more directly the normal modes of the equation for the angular density.

The N-group formulation discussed by Davison has been employed for investigating energy-dependent problems in neutron-transport theory.[12] This model also has been examined in light of the Case technique[13]; Leonard and Ferziger,[14] Siewert and Shieh,[15] and Yoshimura and Katsuragi[16] have made contributions to the theory of multigroup neutron transport, and Metcalf and Zweifel[17] have used this work to make numerical calculations for the Milne problem in two-group theory.

The purpose of the present paper is to blend the methods of Erdmann and Siewert[11] for spherical problems with the two-group analysis of Siewert and Shieh[15] in order to solve the isotropically emitting spherical-shell source problem for the two-group model in an infinite medium. In Sec. II the basic equations for this problem are given, and the normal modes of the two-group equation in spherical geometry are constructed, while Sec. III is devoted to the solution of the considered problem.

II. GENERAL ANALYSIS

We consider the Green's function associated with an isotropically emitting spherical-shell source in an infinite medium. Thus we seek a solution to the time-independent transport equation

$$\mu \frac{\partial}{\partial r} \mathbf{\Psi}(r_0; r, \mu) + \frac{(1-\mu^2)}{r} \frac{\partial}{\partial \mu} \mathbf{\Psi}(r_0; r, \mu) + \mathbf{\Sigma}\mathbf{\Psi}(r_0; r, \mu) = \mathbf{C}\int_{-1}^{1} \mathbf{\Psi}(r_0; r, \mu')\, d\mu' + \frac{\delta(r-r_0)}{8\pi r^2}\mathbf{Q}, \quad (1)$$

subject to the constraint that $\mathbf{\Psi}(r_0; r, \mu)$ must be bounded for all r, since the considered medium must be nonmultiplying. Here $\mathbf{\Psi}(r_0; r, \mu)$ is a vector whose two components represent the angular neutron fluxes in each of the energy groups, μ is the direction cosine of the propagating radiation, and r is the optical variable defined in terms of the smaller of the two total cross sections. Thus, the $\mathbf{\Sigma}$ matrix takes the form

$$\mathbf{\Sigma} = \begin{vmatrix} \sigma & 0 \\ 0 & 1 \end{vmatrix}, \quad \sigma > 1, \quad (2)$$

where σ is the ratio of the total cross section in the first group to the total cross section in the second group, and $\mathbf{C}$ is the transfer matrix with elements c_{ij}. In addition the components of $\mathbf{Q}$, q_1 and q_2, are used to indicate the intensities of the two group sources.

In the usual manner,[2] we need consider only the homogeneous version of Eq. (1); we thus replace the source term by the equivalent boundary condition

$$\mu[\mathbf{\Psi}(r_0; r_0^+, \mu) - \mathbf{\Psi}(r_0; r_0^-, \mu)] = (1/8\pi r_0^2)\mathbf{Q}. \quad (3)$$

We should like to construct the solution to this problem in a manner analogous to that used so successfully in plane geometry, i.e., Case's method of singular eigenfunction expansions.[1] First, a general

set of solutions, denoted as normal modes, to the homogeneous transport equation is determined. The desired solution is then written as a linear sum of these normal modes, and the arbitrary expansion coefficients in this sum are selected such that the boundary conditions of the problem are satisfied. This procedure is, of course, a classical technique; however, in contrast to problems in plane geometry where the necessary completeness theorems are usually available,[2] the solution here cannot be effected quite so readily.

We begin by constructing a set of normal modes for the equation

$$\left[\mu\frac{\partial}{\partial r}+\frac{(1-\mu^2)}{r}\frac{\partial}{\partial\mu}+\mathbf{\Sigma}\right]\mathbf{\Psi}(r,\mu) = \mathbf{C}\int_{-1}^{1}\mathbf{\Psi}(r,\mu')\,d\mu'. \quad (4)$$

Careful inspection of the similarities between the normal modes in one-speed transport theory for plane[2] and spherical[11] geometries suggests a form for the solutions here. In addition, the relationships between the normal modes for one-speed[2] and two-group[15] theory in plane geometry may also be used to advantage. It is therefore proposed that Eq. (4) has solutions of the form

$$\mathbf{\Psi}_\eta(r,\mu)=\sum_{m=0}^{\infty}[\tfrac{1}{2}(2m+1)]P_m(\mu)U_m(r,\eta)\mathbf{G}_m(\eta). \quad (5)$$

Here the Legendre polynomials are represented by $P_m(\mu)$, and

$$U_m(r,\eta)=A(\eta)k_m(r/\eta)+B(\eta)(-1)^m i_m(r/\eta), \quad (6)$$

with $A(\eta)$, $B(\eta)$, and, at this point, η being arbitrary. In addition,

$$i_m(x)=(\pi/2x)^{\frac{1}{2}}I_{m+\frac{1}{2}}(x) \quad (7a)$$

and

$$k_m(x)=(\pi/2x)^{\frac{1}{2}}K_{m+\frac{1}{2}}(x). \quad (7b)$$

Following Watson's notation, we have used $I_{m+\frac{1}{2}}(x)$ and $K_{m+\frac{1}{2}}(x)$ to denote the modified Bessel functions.[18]

If the proposed solution is substituted into Eq. (4), we observe that the G-vectors must be solutions of the recursion relation

$$(2m+1)\eta\mathbf{\Sigma}\mathbf{G}_m(\eta)=2\eta\mathbf{C}\mathbf{G}_0(\eta)\delta_{0,m}+(m+1)\mathbf{G}_{m+1}(\eta)+m\mathbf{G}_{m-1}(\eta),\quad m=0,1,2,\cdots. \quad (8)$$

In order to establish Eq. (8) the following expressions have been utilized[19]:

$$(2m+1)\mu P_m(\mu)=(m+1)P_{m+1}(\mu)+mP_{m-1}(\mu), \quad (9a)$$

$$(1-\mu^2)\frac{d}{d\mu}P_m(\mu)=(m+1)[\mu P_m(\mu)-P_{m+1}(\mu)], \quad (9b)$$

$$\frac{d}{dr}U_m(r,\eta)=-\frac{1}{\eta}U_{m-1}(r,\eta)-\frac{1}{r}(m+1)U_m(r,\eta), \quad (9c)$$

and

$$\eta(2m+1)U_m(r,\eta)=r[U_{m+1}(r,\eta)-U_{m-1}(r,\eta)]. \quad (9d)$$

Previous work by Siewert and Shieh[15] can now be used to find a set of G-vectors and thus to complete the justification of the solutions given by Eq. (5). In Ref. 15 an eigenvalue equation

$$(\eta\mathbf{\Sigma}-\mu\mathbf{I})\mathbf{F}(\eta,\mu)=\eta\mathbf{C}\int_{-1}^{1}\mathbf{F}(\eta,\mu')\,d\mu' \quad (10)$$

was encountered, and the eigenvalue spectrum and corresponding eigenvectors were established. If we multiply Eq. (10) by $P_m(\mu)$, integrate over μ from -1 to 1, and make the identification

$$\mathbf{G}_m(\eta)\triangleq\int_{-1}^{1}P_m(\mu)\mathbf{F}(\eta,\mu)\,d\mu, \quad (11)$$

we note that $\mathbf{G}_m(\eta)$ will be a solution to Eq. (8). Since $\mathbf{F}(\eta,\mu)$ is known for all acceptable values of η in the complex plane,[15] the G-vectors as given by Eq. (11) are determined; more explicitly, the discrete spectrum yields

$$\mathbf{G}_m(\eta_i)=\begin{vmatrix} 2c_{12}\eta_i Q_m(\sigma\eta_i) \\ 2\eta_i[c_{22}-2C\eta_i T(1/\sigma\eta_i)]Q_m(\eta_i)\end{vmatrix}, \quad (12)$$

where the η_i are the "positive" zeros of the dispersion function

$$\Omega(z)=1-2c_{11}zT(1/\sigma z)-2c_{22}zT(1/z)+4Cz^2T(1/z)T(1/\sigma z), \quad (13)$$

the degenerate spectrum $\eta\in(0,1/\sigma)$ yields

$$\mathbf{G}_{1,m}(\eta)=\begin{vmatrix} -c_{12}P_m(\sigma\eta) \\ 2CZ_m(\eta)+c_{11}P_m(\eta)\end{vmatrix},\quad \eta\in(0,1/\sigma), \quad (14a)$$

$$\mathbf{G}_{2,m}(\eta)=\begin{vmatrix} (2C/\sigma)Z_m(\sigma\eta)+c_{22}P_m(\sigma\eta) \\ -c_{21}P_m(\eta)\end{vmatrix},\quad \eta\in(0,1/\sigma), \quad (14b)$$

and the spectrum $\eta\in(1/\sigma,1)$ leads to

$$\mathbf{G}_{3,m}(\eta)=\begin{vmatrix} 2c_{12}\eta Q_m(\sigma\eta) \\ 2[c_{22}-2\eta CT(1/\sigma\eta)]Z_m(\eta)+P_m(\eta)[1-c_{11}\eta T(1/\sigma\eta)]\end{vmatrix},\quad \eta\in(1/\sigma,1). \quad (15)$$

Here we have used the notation $T(x) = \tanh^{-1} x$ and $C = \det \mathbf{C}$; the Legendre functions of the second kind are denoted by $Q_m(x)$, and the mth-order polynomials $Z_m(x)$,

$$Z_m(x) \triangleq \tfrac{1}{2}xP\int_{-1}^{1} \frac{P_m(\mu)}{x-\mu}\,d\mu - xT(x)P_m(x), \quad (16)$$

satisfy the same recursion formula, viz., Eq. (9a), as the Legendre polynomials; they begin differently, however,

$$Z_0(x) = 0, \quad Z_1(x) = -x, \quad \text{and} \quad Z_2(x) = -\tfrac{3}{2}x^2.$$

Now that a set of normal modes for Eq. (4) has been established, and before continuing to the final section where the considered problem is solved, several additional comments can be made: (a) Although the construction here of the normal modes has not been a formal derivation, these results should follow from an analysis similar to that used by Mitsis[9] for one-speed theory. (b) By no means have we proved that all solutions to Eq. (4) are given by these results; on the other hand, we do have solutions sufficiently general for the construction of the solution to the shell-source problem. (c) We have used the results for $\mathbf{F}(\eta, \mu)$ given by Siewert and Shieh[15] for the two-group model; however, the method employed here may also be used to extend the N-group theory of Yoshimura and Katsuragi[16] to spherical geometry.

III. SPHERICAL SHELL-SOURCE PROBLEM

We seek a bounded solution to Eq. (1) or, alternatively, a bounded solution to Eq. (4) subject to the "jump" boundary condition, Eq. (3). Since the Bessel functions $i_m(x)$ diverge as x increases without bound, and since the $k_m(x)$ behave similarly in the vicinity of the origin, we separate the desired solution in the usual manner[2]:

$$\Psi(r_0\colon r, \mu) = \sum_{m=0}^{\infty} [\tfrac{1}{2}(2m+1)]P_m(\mu)\mathbf{R}_m^{+}(r), \quad r > r_0, \quad (17a)$$

and

$$\Psi(r_0\colon r, \mu) = \sum_{m=0}^{\infty} [\tfrac{1}{2}(2m+1)]P_m(\mu)\mathbf{R}_m^{-}(r), \quad r < r_0, \quad (17b)$$

where

$$\mathbf{R}_m^{+}(r) \triangleq \sum_i A(\eta_i)\mathbf{G}_m(\eta_i)k_m(r/\eta_i) + \int_0^{1/\sigma} [A_1(\eta)\mathbf{G}_{1,m}(\eta) + A_2(\eta)\mathbf{G}_{2,m}(\eta)]k_m(r/\eta)\,d\eta + \int_{1/\sigma}^{1} A_3(\eta)\mathbf{G}_{3,m}(\eta)k_m(r/\eta)\,d\eta \quad (18a)$$

and

$$\mathbf{R}_m^{-}(\eta_i) \triangleq \sum_i B(\eta_i)\mathbf{G}_m(\eta_i)(-1)^m i_m(r/\eta_i) + \int_0^{1/\sigma} [B_1(\eta)\mathbf{G}_{1,m}(\eta) + B_2(\eta)\mathbf{G}_{2,m}(\eta)](-1)^m i_m(r/\eta)\,d\eta + \int_{1/\sigma}^{1} B_3(\eta)\mathbf{G}_{3,m}(\eta)(-1)^m i_m(r/\eta)\,d\eta. \quad (18b)$$

The solution given by Eqs. (17) clearly satisfies the homogeneous transport equation; there remains then only the necessity to constrain this solution to meet the condition given by Eq. (3) and thus to determine all of the unknown expansion coefficients, $A(\eta_i)$, $B(\eta_i)$, $A_\alpha(\eta)$, and $B_\alpha(\eta)$, $\alpha = 1, 2, 3$, appearing in the expression for $\Psi(r, \mu)$. We therefore substitute Eqs. (17) into Eq. (3) to find

$$\sum_{m=0}^{\infty} P_m(\mu)[(m+1)\mathbf{S}_{m+1}(r_0) + m\mathbf{S}_{m-1}(r_0)] = \mathbf{Q} \quad (19)$$

or, alternatively,

$$(m+1)\mathbf{S}_{m+1}(r_0) + m\mathbf{S}_{m-1}(r_0) = \mathbf{Q}\delta_{0,m}, \quad m = 0, 1, 2, \cdots, \quad (20)$$

where we have defined

$$\mathbf{S}_m(r_0) \triangleq 4\pi r_0^2[\mathbf{R}_m^{+}(r_0) - \mathbf{R}_m^{-}(r_0)]. \quad (21)$$

Noting that the neutron flux

$$\boldsymbol{\phi}(r_0\colon r) \triangleq \int_{-1}^{1} \Psi(r_0\colon r, \mu')\,d\mu' \quad (22)$$

is to be continuous across the surface $r = r_0$, we observe that $\mathbf{S}_0(r_0) = 0$, and thus Eq. (20) yields the following sufficiency conditions on the unknown vectors $\mathbf{S}_m(r_0)$:

$$\mathbf{S}_m(r_0) = 0, \quad \text{for } m \text{ even}, \quad (23a)$$

$$\mathbf{S}_1(r_0) = \mathbf{Q}, \quad (23b)$$

and

$$\mathbf{S}_m(r_0) = -\frac{2\cdot 4\cdot 6\cdot 8\cdots(m-1)}{3\cdot 5\cdot 7\cdot 9\cdots(m)}(-1)^{\frac{1}{2}(m+1)}\mathbf{Q}, \quad m = 3, 5, 7, \cdots. \quad (23c)$$

Equations (23) now represent the conditions from which we must extract the necessary results for all unknown expansion coefficients. Following the procedure used by Erdmann and Siewert,[11] we shall first show how to satisfy these conditions for the cases $m = 0$ and 1; further analysis will then reveal that these conditions are met for all m.

Referring to Eqs. (18), we note that $\mathbf{S}_0(r_0)$ can be made identically zero simply by insisting that

$$\frac{A(\eta_i)}{i_0(r_0/\eta_i)} = \frac{B(\eta_i)}{k_0(r_0/\eta_i)} \triangleq D(\eta_i) \tag{24a}$$

and

$$\frac{A_\alpha(\eta)}{i_0(r_0/\eta)} = \frac{B_\alpha(\eta)}{k_0(r_0/\eta)} \triangleq D_\alpha(\eta), \quad \eta \in (0, 1), \quad \alpha = 1, 2, 3. \tag{24b}$$

We now utilize Eqs. (18), (21), and (23) to write the condition on $\mathbf{S}_1(r_0)$ in the form

$$\begin{aligned}(2\pi^2)^{-1}\mathbf{Q} = &\sum_i D(\eta_i)\mathbf{G}_1(\eta_i)\eta_i^2 \\ &+ \int_0^{1/\sigma} [D_1(\eta)\mathbf{G}_{1,1}(\eta) + D_2(\eta)\mathbf{G}_{2,1}(\eta)]\eta^2\, d\eta \\ &+ \int_{1/\sigma}^1 D_3(\eta)\mathbf{G}_{3,1}(\eta)\eta^2\, d\eta,\end{aligned} \tag{25}$$

where we have made use of the property

$$\begin{aligned}i_0(x)k_m(x) - (-1)^m i_m(x)k_0(x) &\\ = \frac{\pi}{4x^2}\sum_{\alpha=0}^m \frac{1}{(2x)^\alpha} W_\alpha^m[1 - (-1)^{m+\alpha}],&\end{aligned} \tag{26a}$$

with[19]

$$W_\alpha^m = (m + \alpha)!/\alpha!\,(m - \alpha)!. \tag{26b}$$

The form of Eq. (25) is suggestive of a full-range expansion in terms of the eigenvectors $\mathbf{F}(\eta, \mu)$ of Eq. (10). We therefore use the results of Siewert and Shieh[15] and write

$$\begin{aligned}(2\pi^2)^{-1}\mathbf{Q} = \mu\Big\{&\sum_i D(\eta_i)[\mathbf{F}_{i+}(\mu) - \mathbf{F}_{i-}(\mu)]\eta_i^2 \\ &+ \int_0^{1/\sigma} D_1(\eta)[\mathbf{F}_1^{(1)}(\eta, \mu) - \mathbf{F}_1^{(1)}(-\eta, \mu)]\eta^2\, d\eta \\ &+ \int_0^{1/\sigma} D_2(\eta)[\mathbf{F}_2^{(1)}(\eta, \mu) - \mathbf{F}_2^{(1)}(-\eta, \mu)]\eta^2\, d\eta \\ &+ \int_{1/\sigma}^1 D_3(\eta)[\mathbf{F}^{(2)}(\eta, \mu) - \mathbf{F}^{(2)}(-\eta, \mu)]\eta^2\, d\eta\Big\}.\end{aligned} \tag{27}$$

The various **F**-vectors appearing in Eq. (27) are given explicitly in Ref. 15 [see Eqs. (6), (7), and (10)] and, for the sake of brevity, will not be repeated here. We note, however, that $\mathbf{F}(\eta, -\mu) = \mathbf{F}(-\eta, \mu)$, and thus the full-range completeness theorem[15] ensures that Eq. (27) has a solution. In addition the full-range orthogonality theorem[15] may be employed to obtain explicit results for the unknown coefficients in Eq. (27):

$$D(\eta_i) = [2\pi^2\eta_i^2 N(\eta_i)]^{-1}\tilde{\mathbf{G}}_0^\dagger(\eta_i)\mathbf{Q}, \tag{28a}$$

$$\begin{aligned}D_1(\eta) = &[2\pi^2\eta^2 N_1(\eta)]^{-1} \\ &\times [N_{22}(\eta)\tilde{\mathbf{G}}_{1,0}^\dagger(\eta) - N_{12}(\eta)\tilde{\mathbf{G}}_{2,0}^\dagger(\eta)]\mathbf{Q},\end{aligned} \tag{28b}$$

$$\begin{aligned}D_2(\eta) = &[2\pi^2\eta^2 N_1(\eta)]^{-1} \\ &\times [N_{11}(\eta)\tilde{\mathbf{G}}_{2,0}^\dagger(\eta) - N_{21}(\eta)\tilde{\mathbf{G}}_{1,0}^\dagger(\eta)]\mathbf{Q},\end{aligned} \tag{28c}$$

and

$$D_3(\eta) = [2\pi^2\eta^2 N_2(\eta)]^{-1}\tilde{\mathbf{G}}_{3,0}^\dagger(\eta)\mathbf{Q}. \tag{28d}$$

Here the superscript tilde denotes the transpose operation, the superscript dagger indicates an interchange of c_{ij} and c_{ji},

$$N_{11}(\eta) = \eta\{c_{12}c_{21} + [c_{11} - 2\eta CT(\eta)]^2 + \pi^2C^2\eta^2\}, \tag{29a}$$

$$N_{22}(\eta) = \eta\{c_{12}c_{21} + [c_{22} - 2\eta CT(\sigma\eta)]^2 + \pi^2C^2\eta^2\}, \tag{29b}$$

and

$$N_{ij}(\eta) = -c_{ji}\eta\{c_{11} + c_{22} - 2\eta C[T(\eta) + T(\sigma\eta)]\}, \quad \text{for } i \neq j. \tag{29c}$$

In addition,

$$N(\eta_i) = \eta_i^2\left[c_{22} - 2C\eta_i T\left(\frac{1}{\sigma\eta_i}\right)\right]\frac{d}{dz}\,\Omega(z)\Big|_{z=\eta_i}, \tag{30a}$$

$$\begin{aligned}N_1(\eta) = \eta^2C^2(&\{1 - 2\eta c_{11}T(\sigma\eta) - 2\eta c_{22}T(\eta) \\ &+ \eta^2C[4T(\eta)T(\sigma\eta) - \pi^2]\}^2 \\ &+ \pi^2\eta^2\{2C\eta[T(\eta) + T(\sigma\eta)] - c_{11} - c_{22}\}^2),\end{aligned} \tag{30b}$$

and

$$\begin{aligned}N_2(\eta) = \eta(&\{1 - 2\eta c_{11}T(1/\sigma\eta) - 2\eta c_{22}T(\eta) \\ &+ 4\eta^2CT(\eta)T(1/\sigma\eta)\}^2 \\ &+ \pi^2\eta^2\{c_{22} - 2\eta CT(1/\sigma\eta)\}^2).\end{aligned} \tag{30c}$$

If Eq. (27) is integrated over μ from -1 to 1, the resulting equation is identical with Eq. (25), and thus the expressions given by Eqs. (28) for the expansion coefficients $D(\eta_i)$ and $D_\alpha(\eta)$, $\alpha = 1, 2, 3$, coupled with Eqs. (24), ensure that the conditions on $\mathbf{S}_m(r_0)$ are satisfied for $m = 0$ and 1. It remains to be shown that these coefficients are correct for all m.

We begin the proof by considering the explicit expression

$$\mathbf{S}_m(r_0) = \tfrac{1}{2}\sum_{k=0}^m [W_k^m/(2r_0)^k][1 - (-1)^{m+k}]\mathbf{J}_k^m, \quad m = 0, 1, 2, \cdots, \tag{31a}$$

where we have used Eqs. (24) and (26) and defined

$$\mathbf{J}_k^m \triangleq 2\pi^2\Big\{\sum_i D(\eta_i)\mathbf{G}_m(\eta_i)\eta_i^{k+2} + \int_0^{1/\sigma}[D_1(\eta)\mathbf{G}_{1,m}(\eta) + D_2(\eta)\mathbf{G}_{2,m}(\eta)]\eta^{k+2}\,d\eta + \int_{1/\sigma}^1 D_3(\eta)\mathbf{G}_{3,m}(\eta)\eta^{k+2}\,d\eta\Big\}. \quad (31b)$$

In light of this expression the constraints on $\mathbf{S}_m(r_0)$ given by Eqs. (23) can be imposed by requiring

$$\mathbf{J}_k^m = \mathbf{0}, \quad 0 < k < m, \quad m + k = 3, 5, 7, \cdots, \quad (32a)$$

$$\mathbf{J}_0^1 = \mathbf{Q}, \quad (32b)$$

and

$$\mathbf{J}_0^m = -\frac{2\cdot 4\cdot 6\cdots(m-1)(-1)^{\frac{1}{2}(m+1)}}{3\cdot 5\cdot 7\cdots(m)}\mathbf{Q}, \quad m = 3, 5, 7, \cdots. \quad (32c)$$

If we multiply Eq. (27) by $\mu^{k-1}P_m(\mu)$ and integrate over μ from -1 to 1, we find

$$(4\pi^2)^{-1}\int_{-1}^1 \mu^{k-1}P_m(\mu)\,d\mu\mathbf{Q} = \sum_i D(\eta_i)\eta_i^2\int_{-1}^1 P_m(\mu)\mu^k\mathbf{F}_{i+}(\mu)\,d\mu + \int_0^{1/\sigma}\Big[D_1(\eta)\int_{-1}^1 P_m(\mu)\mu^k\mathbf{F}_1^{(1)}(\eta,\mu)\,d\mu + D_2(\eta)\int_{-1}^1 P_m(\mu)\mu^k\mathbf{F}_2^{(1)}(\eta,\mu)\Big]\eta^2\,d\eta + \int_{1/\sigma}^1 D_3(\eta)\int_{-1}^1 P_m(\mu)\mu^k\mathbf{F}^{(2)}(\eta,\mu)\,d\mu\eta^2\,d\eta, \quad m > 0, \quad m + k = 1, 3, 5, \cdots. \quad (33)$$

For $k = 0$ we observe that the right-hand side of Eq. (33) is $\mathbf{J}_0^m/2\pi^2$ and thus for all odd m we obtain

$$\mathbf{J}_0^m = \frac{1}{2}\int_{-1}^1 P_m(\mu)\frac{d\mu}{\mu}\mathbf{Q}, \quad m = 1, 3, 5, \cdots. \quad (34)$$

The integral in the above result is nonsingular for m odd, and, in fact, Eq. (34) establishes Eqs. (32b) and (32c).

If we now consider $0 < k < m$, the left-hand side of Eq. (33) is clearly zero. Furthermore, inspection of the eigenvectors in Ref. 15 reveals that we may use the expression

$$\frac{\mu^k}{a\xi - \mu} = -\mu^{k-1} - a\xi\mu^{k-2} - (a\xi)^2\mu^{k-3} - \cdots + \frac{(a\xi)^k}{(a\xi - \mu)}, \quad \text{for} \quad a = \sigma \text{ or } 1, \quad (35)$$

to write

$$\int_{-1}^1 P_m(\mu)\mu^k\mathbf{F}(\xi,\mu)\,d\mu = \int_{-1}^1 P_m(\mu)\begin{vmatrix} -\mu^{k-1} - (\sigma\xi)\mu^{k-2} - \cdots - (\sigma\xi)^{k-1} \\ -\mu^{k-1} - \xi\mu^{k-2} - \cdots - \xi^{k-1}\end{vmatrix}d\mu + \xi^k\boldsymbol{\Sigma}^k\int_{-1}^1 P_m(\mu)\mathbf{F}(\xi,\mu)\,d\mu, \quad \text{for} \quad \xi = \eta_i \quad \text{or} \quad \eta\in(0,1), \quad (36)$$

with $\boldsymbol{\Sigma}^k$ denoting the kth power of $\boldsymbol{\Sigma}$. For the values of k considered, Eq. (36) reduces to

$$\int_{-1}^1 P_m(\mu)\mu^k\mathbf{F}(\xi,\mu)\,d\mu = \xi^k\boldsymbol{\Sigma}^k\mathbf{G}_m(\xi), \quad \text{for} \quad \xi = \eta_i \quad \text{or} \quad \eta\in(0,1), \quad 0 < k < m, \quad m + k = 3, 5, 7, \cdots, \quad (37)$$

where $\mathbf{G}_m(\xi)$ is defined by Eq. (11). We note that Eq. (33) now may be written as

$$\boldsymbol{\Sigma}^k\mathbf{J}_k^m = \mathbf{0}, \quad 0 < k < m, \quad m + k = 3, 5, 7, \cdots. \quad (38)$$

Since $\boldsymbol{\Sigma}$ is a nonsingular matrix, Eq. (32a) follows directly, and the proof is complete.

Having successfully determined all the unknown expansion coefficients, we thus have a complete solution for the angular flux $\boldsymbol{\Psi}(r_0: r, \mu)$, and explicit expressions for the flux $\boldsymbol{\phi}(r_0: r)$ and the current,

$$\mathbf{j}(r_0: r) \triangleq \int_{-1}^1 \boldsymbol{\Psi}(r_0: r, \mu)\mu\,d\mu, \quad (39)$$

are immediately available:

$$\boldsymbol{\phi}(r_0: r) = (\pi/2rr_0)\Big\{\sum_i D(\eta_i)\mathbf{G}_0(\eta_i)\eta_i^2\sinh(r_0/\eta_i)e^{-r/\eta_i} + \int_0^{1/\sigma}[D_1(\eta)\mathbf{G}_{1,0}(\eta) + D_2(\eta)\mathbf{G}_{2,0}(\eta)]\sinh(r_0/\eta)e^{-r/\eta}\eta^2\,d\eta + \int_{1/\sigma}^1 D_3(\eta)\mathbf{G}_{3,0}(\eta)\sinh(r_0/\eta)e^{-r/\eta}\eta^2\,d\eta\Big\}, \quad r > r_0, \quad (40)$$

and

$$\mathbf{j}(r_0: r) = \frac{\pi(\boldsymbol{\Sigma} - 2\mathbf{C})}{2rr_0}\Big\{\sum_i D(\eta_i)\mathbf{G}_0(\eta_i)\eta_i^3 H\Big(\frac{r_0}{\eta_i}:\frac{r}{\eta_i}\Big) + \int_0^{1/\sigma}[D_1(\eta)\mathbf{G}_{1,0}(\eta) + D_2(\eta)\mathbf{G}_{2,0}(\eta)]H\Big(\frac{r_0}{\eta}:\frac{r}{\eta}\Big)\eta^3\,d\eta + \int_{1/\sigma}^1 D_3(\eta)\mathbf{G}_{3,0}(\eta)H\Big(\frac{r_0}{\eta}:\frac{r}{\eta}\Big)\eta^3\,d\eta\Big\}, \quad (41a)$$

where

$$H(x_0: x) \triangleq \sinh x_0 e^{-x}[(1/x) + 1], \qquad x > x_0, \tag{41b}$$

$$H(x_0: x) \triangleq \sinh x e^{-x_0}[(1/x) - \coth x], \quad x < x_0. \tag{41c}$$

We note that the result for $\phi(r_0: r)$ $r < r_0$ is obtained by interchanging r_0 and r in Eq. (40); in addition, we have used the relation

$$\mathbf{G}_1(\xi) = \xi(\mathbf{\Sigma} - 2\mathbf{C})\mathbf{G}_0(\xi), \quad \xi = \eta_i \quad \text{or} \quad \eta \in (0, 1), \tag{42}$$

to obtain Eq. (41a). Higher moments of $\Psi(r_0; r, \mu)$ may be obtained in a similar manner by integrating Eqs. (17), and finally, results for the two-group point-source problem are obtained by observing the limit as $r_0 \to 0$ in Eqs. (17a), (40), (41a), and (41b).

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation through grant GK-3072.

[1] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1967).
[3] N. J. McCormick and I. Kuščer, J. Math. Phys. **6**, 1939 (1965).
[4] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[5] J. Mika, Nucl. Sci. Eng. **22**, 235 (1965).
[6] I. Kuščer and P. F. Zweifel, J. Math. Phys. **6**, 1125 (1965).
[7] J. H. Ferziger and A. H. Robinson, Nucl. Sci. Eng. **21**, 382 (1965).
[8] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[9] G. J. Mitsis, Argonne National Laboratory Report ANL-6787, 1963.
[10] A. Leonard and T. W. Mullikin, Proc. Natl. Acad. Sci. (U.S.) **52**, 683 (1964).
[11] R. C. Erdmann and C. E. Siewert, J. Math. Phys. **9**, 81 (1968).
[12] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).
[13] R. Żelazny and A. Kuszell, Ann. Phys. (N.Y.) **16**, 81 (1961).
[14] A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. **26**, 181 (1966).
[15] C. E. Siewert and P. S. Shieh, J. Nucl. Energy **21**, 383 (1967).
[16] T. Yoshimura and S. Katsuragi, Nucl. Sci. Eng. **33**, 297 (1968).
[17] D. R. Metcalf and P. F. Zweifel, Nucl. Sci. Eng. **33**, 318 (1968).
[18] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1945).
[19] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Technique for Finding the Moment Equations of a Nonlinear Stochastic System*

N. G. F. SANCHO
Department of Mathematics, McGill University, Montreal, Quebec

(Received 2 June 1969)

A technique is described for deriving the moment equations of a nonlinear stochastic system with a random forcing term. The nonlinear term is then linearized by means of minimizing the mean-square error between nonlinear and linear terms. The traditional derivation of the Fokker–Planck equation for the conditional probability density, and hence the moment equations, is bypassed.

I. INTRODUCTION

The Fokker–Planck equation has been used extensively in the theoretical study of both linear and nonlinear systems driven by random excitation. In a recent paper by Morton and Corrsin,[1] the assumptions needed to derive the Fokker–Planck equation for a random forcing function which is not Gaussian white noise is given; and the experimental confirmation of the equation is shown for the moment equations in the steady state.

The purpose of this paper is to use a technique first described by Cumming[2] for deriving the differential equation governing the moment equations of a nonlinear stochastic system. A linearized form of the nonlinear term is derived by minimizing the mean-square error between the nonlinear and linear term as described by Bellman and Richardson.[3] A simpler form of equating the higher-order moments to zero, for certain initial conditions, is also given. This is a restatement of the Poincaré–Lyapunov stability theorem.[4]

The theory given in this paper could be generalized to an nth-order system with random coefficients.

In the derivation of these moment equations no guarantee is given for the preservation of moment inequalities. A closure and preservation technique derived by Bellman and Richardson[5] can then be applied.

where

$$H(x_0: x) \triangleq \sinh x_0 e^{-x}[(1/x) + 1], \qquad x > x_0, \tag{41b}$$

$$H(x_0: x) \triangleq \sinh x e^{-x_0}[(1/x) - \coth x], \quad x < x_0. \tag{41c}$$

We note that the result for $\phi(r_0: r)$ $r < r_0$ is obtained by interchanging r_0 and r in Eq. (40); in addition, we have used the relation

$$\mathbf{G}_1(\xi) = \xi(\mathbf{\Sigma} - 2\mathbf{C})\mathbf{G}_0(\xi), \quad \xi = \eta_i \quad \text{or} \quad \eta \in (0, 1), \tag{42}$$

to obtain Eq. (41a). Higher moments of $\Psi(r_0; r, \mu)$ may be obtained in a similar manner by integrating Eqs. (17), and finally, results for the two-group point-source problem are obtained by observing the limit as $r_0 \to 0$ in Eqs. (17a), (40), (41a), and (41b).

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation through grant GK-3072.

[1] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1967).
[3] N. J. McCormick and I. Kuščer, J. Math. Phys. **6**, 1939 (1965).
[4] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[5] J. Mika, Nucl. Sci. Eng. **22**, 235 (1965).
[6] I. Kuščer and P. F. Zweifel, J. Math. Phys. **6**, 1125 (1965).
[7] J. H. Ferziger and A. H. Robinson, Nucl. Sci. Eng. **21**, 382 (1965).
[8] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[9] G. J. Mitsis, Argonne National Laboratory Report ANL-6787, 1963.
[10] A. Leonard and T. W. Mullikin, Proc. Natl. Acad. Sci. (U.S.) **52**, 683 (1964).
[11] R. C. Erdmann and C. E. Siewert, J. Math. Phys. **9**, 81 (1968).
[12] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).
[13] R. Żelazny and A. Kuszell, Ann. Phys. (N.Y.) **16**, 81 (1961).
[14] A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. **26**, 181 (1966).
[15] C. E. Siewert and P. S. Shieh, J. Nucl. Energy **21**, 383 (1967).
[16] T. Yoshimura and S. Katsuragi, Nucl. Sci. Eng. **33**, 297 (1968).
[17] D. R. Metcalf and P. F. Zweifel, Nucl. Sci. Eng. **33**, 318 (1968).
[18] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1945).
[19] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Technique for Finding the Moment Equations of a Nonlinear Stochastic System*

N. G. F. SANCHO
Department of Mathematics, McGill University, Montreal, Quebec

(Received 2 June 1969)

A technique is described for deriving the moment equations of a nonlinear stochastic system with a random forcing term. The nonlinear term is then linearized by means of minimizing the mean-square error between nonlinear and linear terms. The traditional derivation of the Fokker–Planck equation for the conditional probability density, and hence the moment equations, is bypassed.

I. INTRODUCTION

The Fokker–Planck equation has been used extensively in the theoretical study of both linear and nonlinear systems driven by random excitation. In a recent paper by Morton and Corrsin,[1] the assumptions needed to derive the Fokker–Planck equation for a random forcing function which is not Gaussian white noise is given; and the experimental confirmation of the equation is shown for the moment equations in the steady state.

The purpose of this paper is to use a technique first described by Cumming[2] for deriving the differential equation governing the moment equations of a nonlinear stochastic system. A linearized form of the nonlinear term is derived by minimizing the mean-square error between the nonlinear and linear term as described by Bellman and Richardson.[3] A simpler form of equating the higher-order moments to zero, for certain initial conditions, is also given. This is a restatement of the Poincaré–Lyapunov stability theorem.[4]

The theory given in this paper could be generalized to an nth-order system with random coefficients.

In the derivation of these moment equations no guarantee is given for the preservation of moment inequalities. A closure and preservation technique derived by Bellman and Richardson[5] can then be applied.

II. APPROXIMATE CONDITIONS FOR A SYSTEM

We consider again the system given by Morton and Corrsin[1] of the form

$$\ddot{x}(t) + \alpha\dot{x}(t) + \psi(x(t)) = f(t), \tag{1}$$

where α is a constant, $\psi(x(t))$ is an arbitrary "smooth" function of $x(t)$, and f is a random forcing function which is not Gaussian white noise.

Let

$$\dot{x} \equiv s, \tag{2}$$

so that (1) can be written as

$$\dot{s} + \alpha s + \psi(x(t)) = f(t). \tag{3}$$

In using the Fokker–Planck equation, the derivation of $\langle \Delta x \mid x, s\rangle$, $\langle \Delta s \mid x, s\rangle$, $\langle (\Delta s)^2 \mid x, s\rangle$, etc., are needed; where Δ is a finite forward-increment operator over the time increment Δt, and $\langle - \mid x, s\rangle$ is the conditional expectation given x and s. Morton and Corrsin[1] have summarized the conditions of $f(t)$ which permit the conditional expectation as given above to be replaced by their Fokker–Planck approximation. Since these conditional expectations are needed in deriving the moment equations of the given system (1), we summarize the results again:

(1) The forcing function $f(t)$ is a stationary, normal random variable with zero average value and non-infinite integral scale.

(2) $f(t)$ is uncorrelated with some specific functions of $x(t)$.

(3) The largest statistically characteristic time of $f(t)$, $T_{\max}$, say, must be so much smaller than the smallest characteristic time of $x(t)$ and $\dot{x}(t)$, δ_x and δ_s, say, that there can exist a time θ which is very much larger than the former and very much smaller than both of the latter; i.e.,

$$\delta_x, \quad \delta_s \gg \theta \gg T_{\max}.$$

Let us now calculate the conditional expectation of the increments of the state variables. We have

$$\langle \Delta x \mid x, s\rangle = s\Delta t + O(\Delta t)^2, \tag{4}$$

$$\langle \Delta s \mid x, s\rangle = -[\alpha s + \psi]\Delta t + \left\langle \int_t^{t+\Delta t} f\, dt_1 \right\rangle + O(\Delta t)^2, \tag{5}$$

$$\langle (\Delta x)^2 \mid x, s\rangle = s^2(\Delta t)^2 + O(\Delta t)^2, \tag{6}$$

$$\langle \Delta x \Delta s \mid x, s\rangle = -s[\alpha s + \psi](\Delta t)^2 + s\left\langle \int_t^{t+\Delta t} f\, dt_1 \right\rangle \Delta t + O(\Delta t)^2, \tag{7}$$

$$\langle (\Delta s)^2 \mid x, s\rangle = [\alpha s + \psi]^2(\Delta t)^2 - 2[\alpha s + \psi]\left\langle \int_t^{t+\Delta t} f\, dt_1 \right\rangle \Delta t + \left\langle \left(\int_t^{t+\Delta t} f\, dt_1 \right)^2 \right\rangle + O(\Delta t)^2. \tag{8}$$

Let

$$\Lambda = \int_t^{t+\Delta t} \int_t^{t+\Delta t} \langle f(t_1) f(t_2)\rangle\, dt_1\, dt_2, \tag{9}$$

and introduce the autocorrelation function

$$R_f(\tau) \equiv \langle f(t) f(t+\tau)\rangle / \langle f^2\rangle. \tag{10}$$

Equation (9) can be written as

$$\Lambda(\Delta t) = \langle f^2\rangle \int_0^{\Delta t} \int_0^{\Delta t} R_f(t_1 - t_2)\, dt_1\, dt_2, \tag{11}$$

provided that $f(t)$ satisfies the three conditions given earlier.

We can now note that as Δt (or θ) $\to 0$, then

$$\Lambda \to \langle f^2\rangle (\Delta t)^2, \tag{12}$$

so that Eq. (8) can be written as

$$\langle (\Delta s)^2 \mid x, s\rangle = D\Delta t + O(\Delta t), \tag{13}$$

where $D \equiv \Lambda(\theta)\theta^{-1} \simeq \lim\ [\Lambda(\Delta t)/\Delta t]$ as $\Delta t \to \infty$. Furthermore, we can write for the ensemble average

$$\left\langle \int_t^{t+\Delta t} f\, dt_1 \right\rangle = \int_t^{t+\Delta t} \langle f\rangle\, dt_1 = 0. \tag{14}$$

III. MOMENT EQUATIONS OF THE STOCHASTIC SYSTEM

Consider now $B(x, s; t)$, an arbitrary function of $x(t)$ and $s(t)$, whose partial derivatives B_x, B_s, B_{xs}, etc., are continuous and bounded on any interval of $x(t)$ and $s(t)$. We derive a differential equation for the expected value of B, $\langle B\rangle$. Using Δ again as a finite forward-increment operator over the time increment Δt, we have, using Taylor's series,

$$\Delta B = B_x\Delta x + B_s\Delta s + \tfrac{1}{2}B_{xx}(\Delta x)^2 + \tfrac{1}{2}B_{ss}(\Delta s)^2 + B_{xs}(\Delta x \Delta s) + O(\Delta x, \Delta s). \tag{15}$$

Taking the conditional expectation of (15) given x and s, we have

$$\langle \Delta B \mid x, s\rangle = B_x s\Delta t - B_s(\alpha s + \psi)\Delta t + \tfrac{1}{2}B_{ss}D\Delta t + O(\Delta t)^2. \tag{16}$$

Now, taking the expected value of (16), we have as $\langle\langle \Delta B \mid x, s\rangle\rangle = \langle \Delta B\rangle$,

$$\langle \Delta B\rangle = \langle B_x s\rangle \Delta t - \langle B_s(\alpha s + \psi)\rangle \Delta t + \tfrac{1}{2}\langle B_{ss}D\rangle \Delta t + O(\Delta t)^2. \tag{17}$$

Dividing through by Δt and taking the limit as $\Delta t \to 0$, and interchanging $\langle\ \rangle$ and d operators on the left-hand side, we obtain the ordinary differential equation

$$\frac{d\langle B\rangle}{dt} = \langle B_x s\rangle - \langle B_s(\alpha s + \psi)\rangle + \tfrac{1}{2}\langle B_{ss} D\rangle. \quad (18)$$

The moment equations of the system can be found by substituting $B = x, s, x^2, xs, s^2$, etc., respectively.

Example: Let $B = x^2, s^2, xs$; then the moment equations can be written as:

$$\frac{d\langle x^2\rangle}{dt} = 2\langle xs\rangle, \quad (19)$$

$$\frac{d\langle s^2\rangle}{dt} = -2\langle s(\alpha s + \psi)\rangle + D, \quad (20)$$

and

$$\frac{d\langle xs\rangle}{dt} = \langle s^2\rangle - \alpha\langle xs\rangle - \langle x\psi\rangle. \quad (21)$$

It is easy to see that if ψ is linear then moments of order two or less are obtained on the right-hand side of (20) and (21), but if ψ is nonlinear, say $x + \beta x^3$, then moments of order four are obtained on the right-hand side. The solution of the moment equations could therefore not be solved when ψ is nonlinear.

IV. LINEARIZATION

A simple form of linearization as given by Bellman and Richardson[3] is to use the Poincaré–Lyapunov stability theorem[4]; i.e., given a vector differential equation of the form

$$\frac{dz}{dt} = Az + f(z), \quad z(0) = c, \quad (22)$$

then for $\|c\|$ sufficiently small and A negative definite, the system is stable for $f(z)$ nonlinear and $t \to \infty$. The coefficients of z; z_0, z_1, etc., may be thought of as the order of moments of system.

Hence if we wish to find, say, first-order moments of the system we use (18) to find the higher-moment equations of the system and ignore the higher-order terms on the right-hand side. For example, the first-order moments of the system are written, using Eq. (18) and $\psi = x + \beta x^3$, as

$$\frac{d\langle x\rangle}{dt} = \langle s\rangle, \quad \langle x(0)\rangle = c_1, \quad (23)$$

$$\frac{d\langle s\rangle}{dt} = -\alpha\langle s\rangle - \langle x\rangle - \beta\langle x^3\rangle, \quad \langle s(0)\rangle = c_2, \quad (24)$$

$$\frac{d\langle x^2\rangle}{dt} = 2\langle xs\rangle, \quad \langle x^2(0)\rangle = c_3, \quad (25)$$

$$\frac{d\langle s^2\rangle}{dt} = -2\alpha\langle s^2\rangle - 2\langle xs\rangle + D, \quad \langle s^2(0)\rangle = c_4, \quad (26)$$

$$\frac{d\langle xs\rangle}{dt} = \langle s^2\rangle - \alpha\langle xs\rangle - \langle x^2\rangle, \quad \langle x(0)s(0)\rangle = c_5, \quad (27)$$

$$\frac{d\langle x^3\rangle}{dt} = 3\langle x^2 s\rangle, \quad \langle x^3(0)\rangle = c_6, \quad (28)$$

$$\frac{d\langle x^2 s\rangle}{dt} = 2\langle xs^2\rangle - \alpha\langle x^2 s\rangle - \langle x^3\rangle, \quad \langle x^2(0)s(0)\rangle = c_7, \quad (29)$$

$$\frac{d\langle xs^2\rangle}{dt} = \langle s^3\rangle - 2\alpha\langle s^2 x\rangle - 2\langle sx^2\rangle + D\langle x\rangle, \quad \langle x(0)s^2(0)\rangle = c_8, \quad (30)$$

and

$$\frac{d\langle s^3\rangle}{dt} = -3\alpha\langle s^3\rangle - 3\langle xs^2\rangle + 3D\langle s\rangle, \quad \langle s^3(0)\rangle = c_9, \quad (31)$$

where we have ignored moments of order four or higher on the right-hand side, i.e., if the initial values of the various moments given are sufficiently small. The equations as given above are nonlinear and in a form which can be solved.

A better form of linearization as given by Bellman and Richardson[3] is to linearize the nonlinear term, and minimize the mean-square error between the nonlinear and linear term. If $\psi(x(t)) = x + \beta x^3$, where β is a constant, then put

$$x^3 \simeq ax, \quad (32)$$

where a is time independent, or taking the expected value

$$\langle x^3\rangle \simeq a\langle x\rangle. \quad (33)$$

Using a mean-square norm as a measure of approximation, we choose to minimize the quantity

$$\int_0^\infty (\langle x^3\rangle - a\langle x\rangle)^2\, dt. \quad (34)$$

The a is readily determined by the condition

$$\int_0^\infty \langle x^3\rangle\langle x\rangle\, dt = a\int_0^\infty \langle x\rangle^2\, dt. \quad (35)$$

The $\langle x^3\rangle$ and $\langle x\rangle$ term are again derived from Eq. (18) after the system is linearized. A more detailed calculation of the a term for a first-order nonlinear stochastic equation is given by Sancho.[6]

V. CLOSURE AND PRESERVATION OF MOMENT PROPERTIES

In deriving the approximate moment equations for a nonlinear stochastic system we have not shown that

moment properties are preserved over an interval of time. Certain moment inequalities must be fulfilled; for example, we must have

$$\langle x^2\rangle \geq \langle x\rangle^2. \tag{36}$$

A simple closure technique given by Bellman[5] is applied. We summarize the results as follows:

Let Y be the solution of the linear-matrix differential equation

$$Y' = BY + YB^T, \quad Y(0) = c, \tag{37}$$

where B is an arbitrary real matrix and B^T denotes the transpose. If c is positive definite, then $Y(t)$ is positive definite for $t > 0$. The proof follows from the representation

$$Y(t) = e^{Bt}ce^{B^Tt}. \tag{38}$$

Consider the first two moments $\langle x\rangle$ and $\langle x^2\rangle$ as derived from (18) and the associated matrix system

$$\begin{pmatrix}\langle x^0\rangle & \langle x\rangle\\ \langle x\rangle & \langle x^2\rangle\end{pmatrix}' = \begin{pmatrix}0 & \langle s\rangle\\ \langle s\rangle & -\alpha\langle s\rangle - \langle x\rangle - \beta\langle x^3\rangle\end{pmatrix}, \tag{39}$$

where $\langle x^0\rangle = 1$ and $\langle x^3\rangle$ given by the solution of moment equation after it has been linearized one way or another. Let

$$B = \begin{pmatrix}b_1 & b_2\\ b_3 & b_4\end{pmatrix}, \tag{40}$$

where b_1, b_2, b_3, b_4 are real parameters to be determined, and the associated matrix system

$$\begin{pmatrix}u_0 & u_1\\ u_1 & u_2\end{pmatrix}' = B\begin{pmatrix}u_0 & u_1\\ u_1 & u_2\end{pmatrix} + \begin{pmatrix}u_0 & u_1\\ u_1 & u_2\end{pmatrix}B^T. \tag{41}$$

We wish to determine the elements b_1, b_2, b_3, b_4 so that

$$\left\| \begin{pmatrix}0 & \langle s\rangle\\ \langle s\rangle & -\alpha\langle s\rangle - \langle x\rangle - \beta\langle x^3\rangle\end{pmatrix} - B\begin{pmatrix}\langle x^0\rangle & \langle x\rangle\\ \langle x\rangle & \langle x^2\rangle\end{pmatrix} - \begin{pmatrix}\langle x^0\rangle & \langle x\rangle\\ \langle x\rangle & \langle x^2\rangle\end{pmatrix}B^T \right\|^2 \tag{42}$$

is a minimum, where we have

$$\|Y\|^2 = \int_0^\infty tr(YY^T)\,dt = \int_0^\infty \left(\sum_{i,j} y_{ij}^2\right) dt. \tag{43}$$

If $\langle x^0\rangle = 1$, then in (41) we can take $b_1 = b_2 = 0$. The resulting equations of (41) are

$$\begin{aligned} u_1' &= b_4u_1, \quad u_1(0) = k_1,\\ u_2' &= 2b_4u_2, \quad u_2(0) = k_2. \end{aligned} \tag{44}$$

The results of (37) ensures that the matrix

$$\begin{pmatrix}u_0 & u_1\\ u_1 & u_2\end{pmatrix} \tag{45}$$

obtained by (41) is positive definite.

The minimizing procedure in (42) leads to the value b_4, after the solution of $\langle x\rangle$, $\langle s\rangle$, $\langle x^2\rangle$, and $\langle x^3\rangle$ has been substituted in. The solution of (44) gives

$$u_1 = k_1e^{b_4t} \quad \text{and} \quad u_2 = k_2e^{2b_4t}. \tag{46}$$

Since $(u_1^2/u_2) = (k_1^2/k_2)$, the moment inequality is clearly preserved.

* This research was completed when the author was attending the Summer Research Institute of the Canadian Mathematical Congress, Queen's University, Kingston, Ontario.

[1] J. B. Morton and S. Corrsin, J. Math. Phys. **10**, 361 (1969).
[2] I. G. Cumming, Intern. J. Control **5**, 85 (1967).
[3] R. Bellman and J. M. Richardson, Quart. Appl. Math. **20**, 333 (1963).
[4] R. Bellman, *Stability Theory of Differential Equations* (McGraw-Hill Book Co., New York, 1959).
[5] R. Bellman and J. M. Richardson, J. Math. Anal. Appl. **23**, 639 (1968).
[6] N. G. F. Sancho, "On the Approximate Moment Equations of a Nonlinear, Stochastic Differential Equation," J. Math. Anal. Appl. (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On the Solution of a Generalized Wiener–Hopf Equation

R. MITTRA AND S. W. LEE
Department of Electrical Engineering, University of Illinois, Urbana, Ill. 61801

(Received 30 August 1968)

This paper deals with the generalized Wiener–Hopf equation

$$G(\alpha)X_+(\alpha) + H(\alpha)X_+(-\alpha) - Y_-(\alpha) + \Psi^{(i)}(\alpha) = 0, \quad \tau_1 < \tau < \tau_2,$$

where $\alpha\ (= \sigma + i\tau)$ is a complete variable, $G(\alpha)$, $H(\alpha)$, and $\Psi^{(i)}(\alpha)$ are known functions, and $X_+(\alpha)$ and $Y_-(\alpha)$ are unknowns, analytic in upper and lower half-planes, respectively, as indicated by their respective subscripts. This type of equation arises in a class of boundary-value problems in electromagnetic theory, the geometries of which may be described as modified Wiener–Hopf type. The method of approach, which is fundamentally different than those currently available in the literature, is based on a pairing of singularities in the complex α plane. This leads to a functional equation which is exactly solvable in its asymptotic form. The knowledge of this solution permits one to employ one of several rapidly converging numerical procedures available in the literature for a more accurate solution. Two examples illustrating the application of the procedure are included in the paper.

1. INTRODUCTION

In electromagnetic theory, the Wiener–Hopf equation

$$G(\alpha)\bar{X}_+(\alpha) - \bar{Y}_-(\alpha) + \Psi^{(i)}(\alpha) = 0, \quad \tau_1 < \tau < \tau_2 \tag{1.1}$$

arises in a class of boundary value problems involving junctions of semi-infinite geometries. A number of problems belonging to this category are illustrated in Fig. 1 (for a more complete list of problems, see Noble, Ref. 1). The various functions appearing in (1.1) typically play the following roles: $G(\alpha)$ is associated with the Fourier transform of the Green's function, with $\alpha = \sigma + i\tau$ as the transform (complex) variable; $\Psi^{(i)}(\alpha)$ corresponds to the transform of the incident field; $\bar{X}_+(\alpha)$ and $\bar{Y}_-(\alpha)$ represent transforms of unknown electric and magnetic fields in semi-infinite ranges. The subscripts "+" and "−" indicate the regions of regularity of the corresponding functions in the complex α plane. For instance, $\bar{X}_+(\alpha)$ is regular in the upper half-plane defined by $\tau > \tau_1$, and $\bar{Y}_-(\alpha)$ in the lower half-plane $\tau < \tau_2$. The problem is to simultaneously solve for the two unknowns $\bar{X}_+(\alpha)$ and $\bar{Y}_-(\alpha)$ by an application of the Wiener–Hopf procedure.[1] The central step in this procedure involves the transformation of (1.1) into the form

$$\bar{F}_+(\alpha) = \bar{F}_-(\alpha), \quad \tau_1 < \tau < \tau_2. \tag{1.2}$$

This is done by factorizing $G(\alpha)$ into a product $G_+(\alpha)G_-(\alpha)$, followed by some rearrangements. Note that the function on the lhs of (1.2) is regular in the upper half α plane, while the function on the rhs is regular in the lower half α plane. Since these two half-planes overlap in $\tau_1 < \tau < \tau_2$, both sides of (1.2) must be equal to an entire function $P(\alpha)$ by analytic continuation, and then (1.2) is valid for all α in the complex plane. The function $P(\alpha)$ cannot be uniquely determined without imposing an additional constraint on the asymptotic behaviors of $\bar{X}_+(\alpha)$ and $\bar{Y}_-(\alpha)$ for large $|\alpha|$. The asymptotic behavior is based on an application of the edge condition, which requires

$$\bar{X}_+(\alpha) \sim \alpha^{-\bar{\nu}}, \quad \bar{Y}_-(\alpha) \sim \alpha^{-\bar{\mu}} \tag{1.3}$$

as $|\alpha| \to \infty$ in their respective half-planes. Here $\bar{\nu}$ and $\bar{\mu}$ are determinable from the known behavior of the fields in the neighborhood of the edge. Once $P(\alpha)$ is known, $\bar{X}_+(\alpha)$, and $\bar{Y}_-(\alpha)$ are derived from $\bar{F}_+(\alpha)$, and $\bar{F}_-(\alpha)$, respectively, thus completing the solution.

The present paper is concerned with an extended form of (1.1), viz.,

$$G(\alpha)X_+(\alpha) + H(\alpha)X_+(-\alpha) - Y_-(\alpha) + \Psi^{(i)}(\alpha) = 0, \quad \tau_1 < \tau < \tau_2, \tag{1.4}$$

where $H(\alpha)$ is a known function. As might be conjectured, the range of applicability of (1.4) is considerably wider than the conventional Wiener–Hopf equation (1.1). In general, (1.4) arises in the formulation of boundary value problems belonging to a class of modified Wiener–Hopf geometries, in which one of the semi-infinite subregions is modified. A number of modified geometries, which correspond to the problems in Fig. 1 and which may be formulated in terms of (1.4), are sketched in Fig. 2. These and similar problems in this category are of wide practical interest, as evidenced by a large number of publications that have dealt with them in recent years. (A partial list is given in Refs. 2–6.)

To date, the most important method for attacking (1.4) has been provided by Jones[2] for the special case when $H(\alpha)$ *is a meromorphic function* with infinitely

many simple poles at $\alpha = \{-i\gamma_n\}$ with $\mathrm{Re}\,\gamma_n > 0$ in the lower half α plane. He transforms (1.4) into an infinite set of linear simultaneous equations, which take the form

$$G_+(i\gamma_p)A_p + \sum_{n=1}^{\infty} \frac{c_n}{i(\gamma_p + \gamma_n)G_+(i\gamma_n)} = -\phi_+^{(i)}(i\gamma_n), \quad \text{for } p = 1, 2, 3, \cdots, \quad (1.5)$$

where $A_n = X_+(i\gamma_n) =$ unknowns, $c_n =$ residue of $H(\alpha)$ at $\alpha = -i\gamma_n$,

$$G(\alpha) = G_+(\alpha)G_-(\alpha) = G_+(\alpha)G_+(-\alpha),$$

$$[\Psi^{(i)}(\alpha)/G_-(\alpha)] = \phi_+^{(i)}(\alpha) + \phi_-^{(i)}(\alpha).$$

The unknown function $X_+(\alpha)$ is obtained by substituting A_n from the solution of (1.5). This gives

$$X_+(\alpha) = \frac{-1}{G_+(\alpha)}\left(\phi^{(i)}(\alpha) + \sum_{n=1}^{\infty} \frac{c_n}{\alpha + i\gamma_n} \frac{A_n}{G_+(i\gamma_n)}\right). \quad (1.6)$$

The original problem is thus reduced to that of solving the matrix equation (1.5), which is numerically inverted after truncation to a finite size. Unfortunately, it is difficult to establish the convergence of the above procedure[1] and, in addition, it is not possible to establish the satisfaction of the edge condition.

The method to be presented in this paper for attacking (1.5) is different from that of Jones. It is based on an examination of the singularities of the functions appearing in (1.5) in the complex α plane. By invoking the fact that (1.5) is valid for all α, the singularities of various functions are paired such that they are canceled exactly. This procedure leads to a functional equation, which is exactly solvable in its asymptotic form. This provides a built-in check for the convergence of the procedure, guarantees the satisfaction of the edge condition, and permits the extraction of a rapidly converging and accurate numerical solution to the problem.

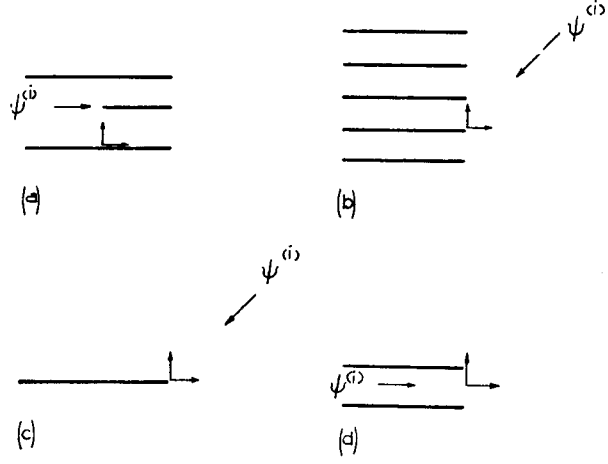


FIG. 1. Typical Wiener–Hopf problems.

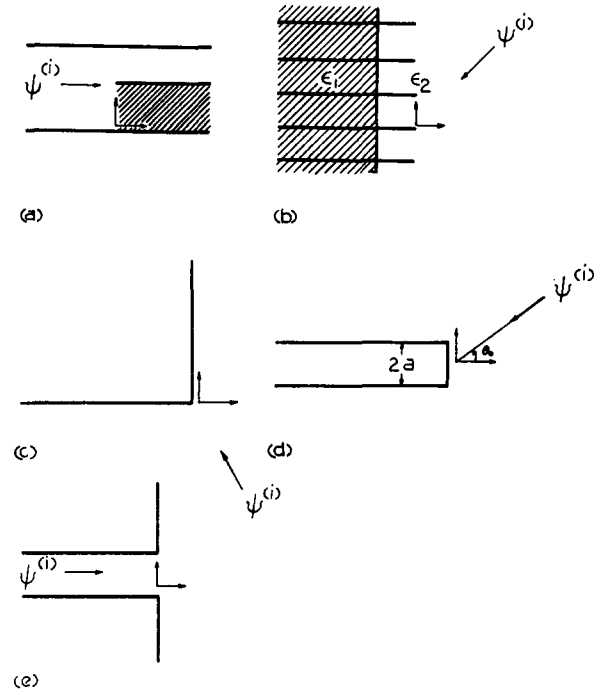


FIG. 2. Modified Wiener–Hopf problems.

2. DERIVATION OF FUNCTIONAL EQUATION

Before proceeding with the discussion of the solution of the modified equation

$$F(\alpha) = G(\alpha)X_+(\alpha) + H(\alpha)X_+(-\alpha) - Y_-(\alpha) + \Psi^{(i)}(\alpha) = 0, \quad \tau_1 < \tau < \tau_2, \quad (2.1)$$

it is desirable to categorize the various situations that arise in the investigation of modified Wiener–Hopf problems. Table I presents a list of typical situations, together with the characteristics of pole-zero and branch singularities of $G(\alpha)$ and $H(\alpha)$ in (3.1). It is useful to point out that:

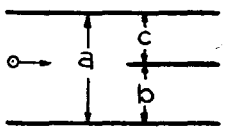
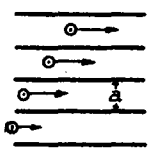
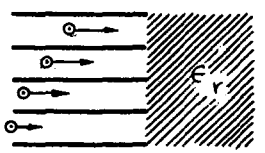
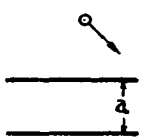
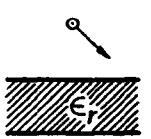
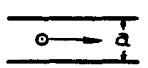
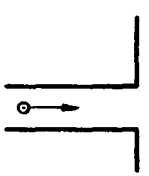
(i) The singularities of $G(\alpha)$ are those which correspond to the spectral eigenvalues in the *unmodified* geometries;

(ii) the singularities of $H(\alpha)$ are also present in $G(\alpha)$; however, the spectrum of singularities of $G(\alpha)$ may be broader than that of $H(\alpha)$;

(iii) $H(\alpha)$ contains only pole singularities if the modified region is a closed one (periodic structures may be considered as closed regions with periodic boundary conditions), while $H(\alpha)$ contains a branch singularity if the modified region is an open one.

The known function $\Psi^{(i)}(\alpha)$ in (2.1) is related to the incident field. Since, in principle, an arbitrary source

Table I. Classification of Modified Wiener-Hopf Geometries.

Group		Basic Wiener-Hopf Geometry	Modified Wiener-Hopf Geometry	G(α) Zeros in l.h.p.	G(α) Poles in l.h.p.	G(α) Branch in l.h.p.	H(α) Poles in l.h.p.	H(α) Branch in l.h.p.
I	A			$\alpha = \{-i\gamma_{na}\}$	$\alpha = \begin{matrix}\{-i\gamma_{nb}\} \\ \{-i\gamma_{nc}\}\end{matrix}$	None	$\alpha = \{-i\gamma_{nb}\}$	None
	B			$\alpha = \{-i\gamma_{na}\}$	$\alpha = \{-i\beta_{na}\}$	None	$\alpha = \{-i\beta_{na}\}$	None
	C			None	$\alpha = \{-i\gamma_{na}\}$	$\sqrt{\alpha-k}$	$\alpha = \{-i\gamma_{na}\}$	None
II				None	$\alpha = \{-i\gamma_{na}\}$	$\sqrt{\alpha-k}$	None	$\sqrt{\alpha-k}$

$$\gamma_{nd} = [(n\pi/d)^2 - k^2]^{1/2}\ ,\ \gamma'_{nd} = [(n\pi/d)^2 - \epsilon_r k^2]^{1/2},\ \beta_{nd} = [(\frac{2n\pi + u}{d})^2 - k^2]^{1/2}$$

is expressible in terms of a superposition of spectral waves, the incident field in Wiener–Hopf problems is generally taken to be a spectral wave, resulting in the following form:

$$\Psi^{(i)}(\alpha) = \Psi_0/(\alpha + k^{(i)}), \tag{2.2}$$

where, for definiteness, we assume, without loss of generality, that $\operatorname{Im} k^{(i)} > 0$ and $\alpha = k^{(i)}$ is not a singularity of $G(\alpha)$.

The unknown $X_+(\alpha)$ and $Y_-(\alpha)$ in (2.1) are the transforms of the electric or magnetic fields over a semi-infinite aperture. Their asymptotic behaviors are determined by the edge conditions, which require

$$X_+(\alpha) \sim \alpha^{-\nu}, \quad Y_-(\alpha) \sim \alpha^{-\mu}, \quad \text{as} \quad |\alpha| \to \infty, \tag{2.3}$$

where ν and μ are two numbers known *a priori*.

In attacking (2.1), we observe first of all that $F(\alpha)$ is regular and identically zero in the strip $\tau_1 < \tau < \tau_2$. Let $F(\alpha) = F_+(\alpha) + F_-(\alpha)$ where $F_+(\alpha)$ is regular for $\tau > \tau_1$ and $F_-(\alpha)$ is regular for $\tau < \tau_2$. Then, by using the well-known methods[1] for relating $F_+(\alpha)$ and $F_-(\alpha)$ to $F(\alpha)$, it is readily shown that $F_+(\alpha) = -F_-(\alpha)$ for all α, i.e., $F(\alpha)$ is identically zero everywhere in the complex α plane. Thus, it is no longer necessary to restrict (2.1) in the strip $\tau_2 < \tau < \tau_1$. The above fact allows one to derive a functional equation for $X_+(\alpha)$ as we will now show.

The derivation of the functional equation is based on the observation that, if $F(\alpha)$ is identically zero for all α, then the singularities of the individual functions in (2.1) must pair up in a manner so as to cancel each other, thereby assuring that the total contribution is indeed zero everywhere. First consider Group I.C in Table I where the common singularities of $G(\alpha)$ and $H(\alpha)$ in the lower half-plane of α are at simple poles $\alpha = -i\gamma_{na} = -i[(n\pi/a)^2 - k^2]^{\frac{1}{2}}$, for $n = 1, 2, \cdots$. Multiplying (2.1) by $(\alpha + i\gamma_{pa})$ and letting $\alpha \to (-i\gamma_{pa})$, we derive the desired functional equation

$$X_+(-i\gamma_{pa}) \operatorname{Res} G(-i\gamma_{pa}) + X_+(+i\gamma_{pa}) \operatorname{Res} H(-i\gamma_{pa}) = 0, \quad \text{for} \quad p = 1, 2, 3, \cdots, \tag{2.4}$$

where the symbol $\operatorname{Res} G(-i\gamma_{pa})$ implies the residue of $G(\alpha)$ at $\alpha = -i\gamma_{pa}$. The last two terms $Y_-(\alpha)$ and $\Psi^{(i)}(\alpha)$ in (2.1) do not contribute to (2.4) since they are regular at $\alpha = -i\gamma_{pa}$. For the special case $H(\alpha) \to 0$, (2.4) reduces to

$$X_+(-i\gamma_{pa}) \operatorname{Res} G(-i\gamma_{pa}) = 0, \quad p = 1, 2, 3, \cdots. \tag{2.5}$$

Thus, in this event, $X_+(\alpha)$ has simple (in electromagnetic-wave problems, the zeros are simple if there are no degenerate modes) zeros at $\alpha = -i\gamma_{pa}$, a fact that can be verified by applying the conventional Wiener–Hopf technique. The functional equations for problems in Group I.A and I.B take the same form as (2.4), except with γ_{na} replaced by γ_{nb} or β_{na}.

For problems in Group II, we note that $G(\alpha)$ and $H(\alpha)$ have a common branch singularity at $\alpha = -k$ in the lower half α plane. Let this branch cut be designated by Σ, shown in Fig. 3. Then, evaluating $F(\alpha)$ in (2.1) at points α^+ and α^- located on two sides of Σ, respectively, and taking the difference between the two expressions, we arrive at the functional equation

$$G(\alpha^+)X_+(\alpha^+) - G(\alpha^-)X_+(\alpha^-) + [H(\alpha^+) - H(\alpha^-)]X_+(-\alpha) = 0. \tag{2.6}$$

Once again, if $H(\alpha) \to 0$ or if it is regular along Σ, the above equation reduces to

$$G(\alpha^+)X_+(\alpha^+) - G(\alpha^-)X_+(\alpha^-) = 0, \quad \text{with} \quad \alpha \in \Sigma. \tag{2.7}$$

Equation (2.7), as well as (2.6), indicates that, if $G(\alpha)$ has a branch singularity at $\alpha = -k$, so must $X_+(\alpha)$, and at an identical location.

The solutions of the functional equation in (2.6) is more involved than the case where $H(\alpha)$ is meromorphic. In the present paper, we consider the solution of (2.4), and we will present the solution of (2.6) in a future communication.

3. SOLUTION FOR GROUPS I

We now discuss a method of solution of (2.4) which requires, as a first step, an examination of the singularities of $X_+(\alpha)$ in the lower half α plane. The mathematical considerations for the three cases in Group I are very similar. For definiteness, let us concentrate on Group I.C. For this case, $G(\alpha)$ has poles at $\alpha = -i\gamma_{na}$ and, in addition, a branch point at $\alpha = -k$ in the lower half α plane. Denote

$$G(\alpha) = [G_-(\alpha)]\left[G_+^{(b)}(\alpha)\left(e^{\chi(\alpha)} \prod_{n=1}^{\infty} \times \left[1 + \frac{\alpha}{i\gamma_{na}}\right] e^{i\alpha a/n\pi}\right)^{-1}\right], \tag{3.1}$$

where

$$\chi(\alpha) = i(\alpha a/\pi)[1 - c - \ln(\alpha a/i\pi)],$$
$$c = \text{Euler const} = 0.57721.$$

Here $G_-(\alpha)$ is regular in the lower α plane and $G_+^{(b)}(\alpha)$ is regular everywhere except for the branch cut Σ in the lower α plane. Furthermore, both $G_-(\alpha)$ and $G_+^{(b)}(\alpha)$ have algebraic behavior as $|\alpha| \to \infty$. Observe that, if $F(\alpha)$ in (2.1) is to be identically zero for all α, then the

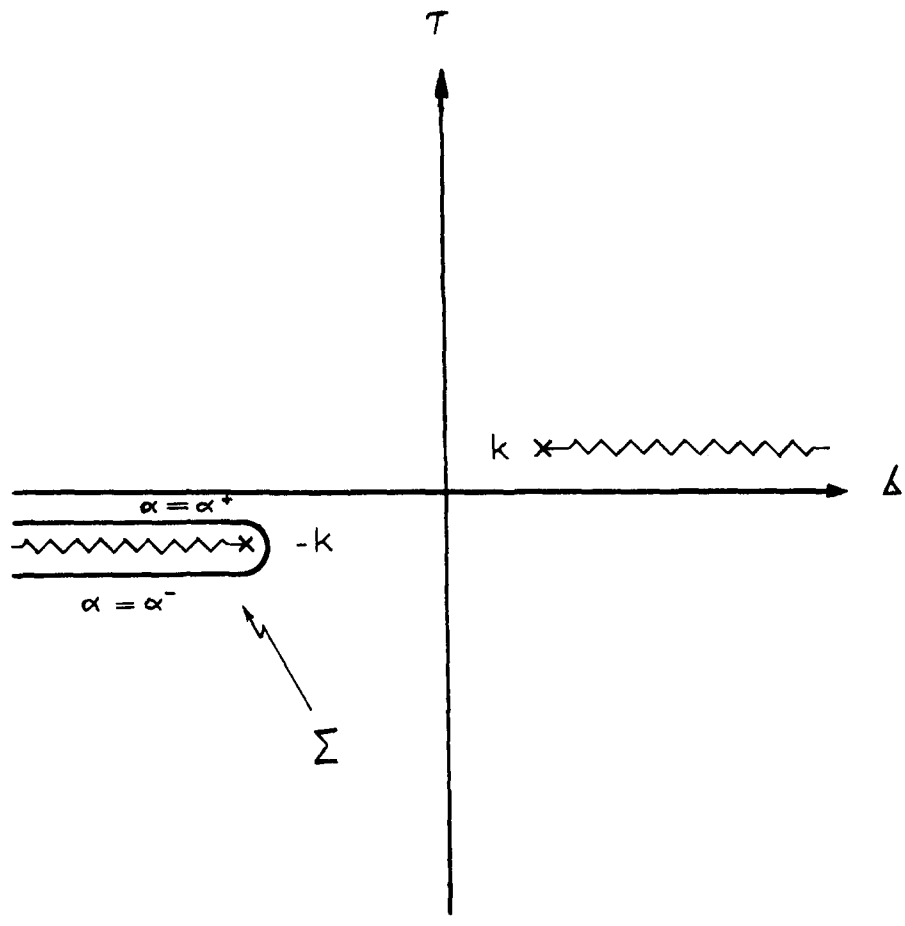


FIG. 3. Branch cuts in complex α plane.

following must be true:

(i) $X_+(\alpha)$ must contain a factor $[1/G^{(b)}(\alpha)]$;

(ii) $X_+(\alpha)$ must have a simple pole at $\alpha = -k^{(i)}$ so that

$$G(-k^{(i)}) \operatorname{Res} X_+(-k^{(i)}) + \Psi_0 = 0; \quad (3.2)$$

(iii) $X_+(\alpha)$ can have no singularities other than those specified by (i) and (ii).

We may, therefore, write

$$X_+(\alpha) = \frac{Q(\alpha)}{(\alpha + k^{(i)})G_+^{(b)}(\alpha)}, \quad (3.3)$$

where $Q(\alpha)$ is an entire function, yet to be determined. To this end, consider first of all the special case $H(\alpha) \to 0$, whence the modified equation in (2.1) reduces to the standard Wiener–Hopf Eq. (1.1). Let the solution corresponding to this limiting case be denoted by $\bar{X}_+(\alpha)$. Then, in view of (2.5), $\bar{X}_+(\alpha)$ will have simple zeros at $\alpha = -i\gamma_{pa}$. Hence, $\bar{X}_+(\alpha)$ may be represented as

$$\bar{X}_+(\alpha) = \bar{Q}_0(\alpha)\left[e^{\chi(\alpha)} \prod_{p=1}^{\infty} \left(1 + \frac{\alpha}{i\gamma_{pa}}\right) e^{i\alpha a/p\pi}\right] \times [(\alpha + k^{(i)})G_+^{(b)}(\alpha)]^{-1}. \quad (3.4)$$

At this point, we recall that the edge conditions stated in (1.3) required $\bar{X}_+(\alpha)$ to behave algebraically at infinity. This, in turn, requires that $\bar{Q}_0(\alpha)$ in (3.4) be, at most, a finite polynomial. Specifically, when the source term is given [in general, $Q_0(\alpha)$ is a polynomial of degree n, where n is the number of incident spectral waves] by (2.2), $Q_0(\alpha)$ turns out to be a constant, say, $\bar{Q}_0$. This constant is determined from (3.8), which yields, with the help of (3.1),

$$\bar{Q}_0(\alpha) = \bar{Q}_0 = -\Psi_0/G_-(-k^{(i)}). \quad (3.5)$$

Insertion of (3.5) into (3.4) completes the solution of the standard Wiener–Hopf equation in (1.1). This result, of course, checks after slight rearrangement with that obtainable through the conventional Wiener–Hopf method.[1] Let us now return to (2.1) and the expression of $X_+(\alpha)$ in (3.3). The set of equations in (2.4) can be satisfied by properly specifying the infinite zeros of $X_+(\alpha)$, denoted by $\alpha = -i\Gamma_{na}$. Even though the Γ_{na} must be regarded as unknowns at this stage, we will show that the asymptotic behavior of Γ_{na} is given by

$$\Gamma_{na} \sim (n\pi/a) + \Delta, \quad \text{as} \quad n \to \infty, \quad (3.6)$$

where Δ is a constant and $\Delta < (\pi/a)$. To this end, we rewrite (3.3) as

$$X_+(\alpha) = Q_0(\alpha)\bar{Q}_0\left[e^{\chi(\alpha)}\prod_{p=1}^{\infty}\left(1+\frac{\alpha}{i\Gamma_{pa}}\right)e^{i\alpha a/p\pi}\right] \times [(\alpha + k^{(i)})G_+^{(b)}(\alpha)]^{-1} = \left(Q_0(\alpha)\prod_{p=1}^{\infty}\frac{1+\alpha/i\Gamma_{pa}}{1+\alpha/i\gamma_{pa}}\right)\bar{X}_+(\alpha), \quad (3.7)$$

where $Q_0(\alpha)$ is a polynomial of degree q, an integer. Using (1.3), we can derive the asymptotic behavior as

$$X_+(\alpha) \sim \alpha^{q-[(\Delta a/\pi)+\tilde{\nu}]}, \quad \text{as} \quad |\alpha| \to \infty, \quad \tau > 0. \quad (3.8)$$

Invoking the edge condition given in (2.3), we derive

$$\nu = [(\Delta a/\pi) + \tilde{\nu}] - q. \quad (3.9)$$

The edge condition imposes the bound $(\nu - \tilde{\nu}) < 1$. Hence, we have, from (3.9),

$$\Delta = (\pi/a)(\nu - \tilde{\nu}), \quad (3.10a)$$

$$q = 0. \quad (3.10b)$$

Thus, the entire function $Q_0(\alpha)$ in (3.7) again turns out to be a constant Q_0. This constant is determinable from the condition in (3.2), and the result is

$$Q_0(\alpha) = Q_0\prod_{p=1}^{\infty}\frac{1-k^{(i)}/\gamma_{pa}}{1-k^{(i)}/\Gamma_{pa}}. \quad (3.11)$$

Inserting (3.11) into (3.7), we have the solution of $X_+(\alpha)$ provided that the Γ_{pa} are known. Thus, the original problem has been reduced to that of determining the set of zeros of $X_+(\alpha)$.

To this end, substitute (3.1), (3.7), and (3.11) into (2.4). This gives, after some manipulations,

$$\Delta_{pa} = \frac{i\Gamma_{pa}}{\gamma_{pa}}\frac{\operatorname{Res} H(-i\gamma_{pa})}{G_-(-i\gamma_{pa})G_+^{(b)}(+i\gamma_{pa})}\frac{k^{(i)}-i\gamma_{pa}}{k^{(i)}+i\gamma_{pa}}e^{\chi(i\gamma_{pa})} \times \prod_{n=1}^{\infty}(1+\gamma_{pa}/\Gamma_{na})e^{-\gamma_{pa}a/n\pi}\prod_{n=1}^{\infty}{}^{(p)}\frac{1-\gamma_{pa}/\gamma_{na}}{1-\gamma_{pa}/\Gamma_{na}}, \quad \text{for} \quad p = 1, 2, 3, \cdots, \quad (3.12)$$

where $\Delta_{pa} = \Gamma_{pa} - \gamma_{pa}$. Recall that, from the edge condition, we have shown in (3.10a) that

$$\Delta_{pa} \sim \Delta = (\pi/a)(\nu - \tilde{\nu}), \quad \text{as} \quad p \to \infty. \quad (3.13)$$

In a large number of practical problems, the values of Δ_{pa} approach the asymptotic value in an extremely rapid fashion.[7] Therefore, one needs only to solve for the first few Δ_{pa} from (3.12), while the rest is replaced by the asymptotic value Δ. The first few Δ_{pa} can be solved numerically by iterations or direct matrix inversions without difficulty. Detailed discussions and examples for these procedures have appeared recently in connection with the "modified residue calculus method" [7–12] and will not be repeated here.

It is important also to point out that, in a given physical problem, the result in (3.13) can also be obtained by asymptotically expanding (3.12) for large p.[7] Thus, (3.13) is indeed consistent with (3.12), as one should expect.

We next discuss two examples which will illustrate the application of the procedure described above.

4. DIFFRACTION BY A THICK HALF-PLANE

The problem of diffraction of a plane wave by a conducting thick half-plane [Fig. 2(d)] is a very important problem in the theory of diffraction; yet a satisfactory solution to this problem is yet to be derived. The derivation of the Wiener–Hopf equation can be found in Jones,[2] and Noble (Ref. 1, pp. 187–189). Writing this equation in the notations of (2.1), we may obtain

$$G(\alpha) = (\gamma e^{-\gamma a}\cosh \gamma a)^{-1}, \quad (4.1a)$$

$$H(\alpha) = (\tanh \gamma a)/\gamma, \quad (4.1b)$$

$$\Psi^{(i)}(\alpha) = \psi_0/(\alpha - k\cos\theta_0), \quad (4.1c)$$

$$\tau_1 = k_2\cos\theta_0, \quad \tau_2 = k_2, \quad k = k_1 + ik_2, \quad (4.1d)$$

$$\gamma = (\alpha^2 - k^2)^{\frac{1}{2}}, \quad \text{and} \quad \gamma \to +\operatorname{Re}\alpha \text{ as } \alpha \to \infty \text{ along real axis.} \quad (4.1e)$$

Note that the common singularities of $G(\alpha)$ and $H(\alpha)$ are simple poles at $\alpha = \alpha = \pm i\gamma_{na}$; hence, the Wiener–Hopf equation is of type I.C.

Following the procedure outlined in Sec. 3, it is not difficult to show that, corresponding to (3.7),

$$X_+(\alpha) = \left(Q_0\prod_{m=1}^{\infty}\frac{1+\alpha/i\Gamma_{2m-1}}{1+\alpha/i\gamma_{2m-1}}\right)\bar{X}_+(\alpha), \quad (4.2)$$

where

$$\bar{X}_+(\alpha) = \psi_0[(\alpha - k\cos\theta_0)G_-(k\cos\theta_0)G_+(\alpha)]^{-1},$$

$$G_+(\alpha)G_-(\alpha) = G_+(\alpha)G_-(-\alpha) = (\gamma e^{-\gamma a}\cosh\gamma a)^{-1}.$$

Thus, the problem reduces to that of determining the set of zeros Γ_{ma} in (5.2).

Recall that the edge condition in the case of a thick half-plane requires $X_+(\alpha) \sim \alpha^{-\frac{1}{2}}$ as $|\alpha| \to \infty$, whereas for a thin half-plane, $\bar{X}_+(\alpha) \sim \alpha^{-\frac{1}{2}}$ as $|\alpha| \to \infty$ in the upper half-plane. From (3.13), we have

$$\Delta_{na} = \Gamma_{ma} - \gamma_{ma} \sim \Delta = \pi/6a, \quad \text{as} \quad m \to \infty. \quad (4.3)$$

The values of Γ_{ma} approach the asymptotic value very rapidly, provided that ka is not too large. Therefore, it is necessary to dermine only the first few zeros of $X_+(\alpha)$.

To this end, we need to solve for the first few equations in the functional equations, which in the

present problem are

$$\Delta_{pa} = \frac{\Gamma_{pa}}{\gamma_{pa}(-\gamma_{pa}a)[G_+(i\gamma_{pa})]^2} \times \frac{k^{(i)} - i\gamma_{pa}}{k^{(i)} - i\gamma_{pa}} \prod_{m=1}^{\infty} \frac{1 + \gamma_{pa}/\Gamma_{ma}}{1 + \gamma_{pa}/\gamma_{ma}} \times \prod_{\substack{m=1 \\ \text{odd}}}^{\infty}{}^{(p)} \frac{1 - \gamma_{pa}/\gamma_{na}}{1 - \gamma_{pa}/\Gamma_{na}}, \quad \text{for} \quad p = 1, 3, 5, \cdots. \tag{4.4}$$

The explicit form of $G_+(\alpha)$ can be found in Noble (Ref. 1, p. 103). The first few Γ_m can readily be computed by using one of the methods discussed in Refs. 10 and 11.

A very similar problem to the diffraction by a thick half-plane is that of diffraction by a solid semi-infinite cylinder. The latter problem was recently attacked by the authors[9] using the method developed in here.

5. SCATTERING BY DIELECTRIC-LOADED BIFURCATED WAVEGUIDE

The Wiener–Hopf problems are typically formulated in terms of Jones' method (by "Jones' method," we mean the method described in Ref. 1, Chap. 2 and Ref. 1, p. 181, for the thick half-plane problem). However, the method is not readily extensible to problems involving two different homogeneous media. In this section, we illustrate an extension of Green's function method of formulation for a structure in which the medium has a discontinuity in the longitudinal direction (the direction where the Fourier transform is taken). The formulations of problems involving two transverse semi-infinite media are much easier (see, for example, Ref. 13). It is believed that such a formulation of the modified Wiener–Hopf equation has not appeared elsewhere.

For the sake of illustration, we will choose the problem of a dielectric-loaded parallel-plate bifurcated waveguide, the geometry of which is shown in Fig. 4. The incident wave is given by

$$E_y^{(i)} = \phi^{(i)} = \sin(\pi/a)x e^{-\gamma_{1a}z}, \quad \text{for all } z, \tag{5.1}$$

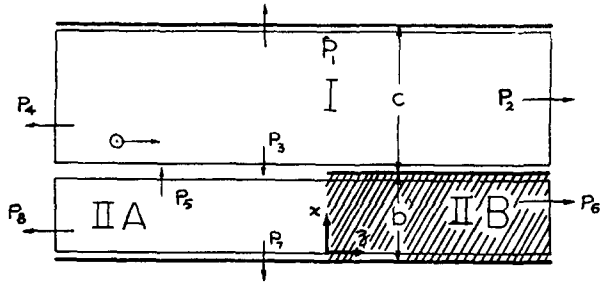


FIG. 4. Dielectric-loaded bifurcated waveguide.

where $\gamma_{1a} = [(\pi/a)^2 - k^2]^{\frac{1}{2}} = -i[k^2 - (\pi/a)^2]^{\frac{1}{2}}$, and $k = \omega[\mu_0\epsilon_0]^{\frac{1}{2}}$. First, let us consider the region I defined by $b < x < a$, where the total field ϕ is expressed as the sum of the incident field $\phi^{(i)}$ and the scattered field $\phi^{(s)}$. Introduce the Green's function via the equation

$$(\nabla^2 + k^2)g^{(1)}(x, x_0; z - z_0) = -\delta(x - x_0)\delta(z - z_0), \tag{5.2}$$

where

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

with the boundary conditions

$$g^{(1)} = 0, \quad \text{at} \quad x = b+ \quad \text{and} \quad x = a, \tag{5.3a}$$

$$g^{(1)} \text{ satisfies radiation condition as } |z| \to \infty. \tag{5.3b}$$

By conventional procedures, it can be shown that

$$\phi^{(s)}(x, z) = -\int_{-\infty}^{0} \left(\phi^{(s)} \frac{\partial g^{(1)}}{\partial x_0}\right)_{x_0=b+} dz_0 + \int_{0}^{\infty} \left(\phi^{(i)} \frac{\partial g^{(1)}}{\partial x_0}\right)_{x_0=b+} dz_0, \quad \text{for} \quad a > x > b, \quad \text{all } z, \tag{5.4}$$

where

$$g^{(1)}(x, x_0; z - z_0) = \frac{1}{c}\prod_{m=1}^{\infty} \frac{1}{\gamma_{mc}} \sin\frac{m\pi}{c}(x_0 - b)\sin\frac{m\pi}{c}(x - b)e^{-\gamma_{mc}|z-z_0|},$$

$$\gamma_{mc} = [(m\pi/c)^2 - k^2]^{\frac{1}{2}}.$$

The situation in region II defined by $b > x > 0$ is more complicated because of the presence of the dielectric medium. Note that the total field $\phi = \phi^{(i)} + \phi^{(s)}$ satisfies

$$(\nabla^2 + K(z)k^2)\phi = 0, \quad \text{for} \quad b > x > 0, \tag{5.5}$$

where $K(z) = 1$, for $z < 0$, and $K(z) = \epsilon_r$, for $z > 0$, and is discontinuous at $z = 0$. The boundary conditions for ϕ are:

(i) $\phi = 0$ at $x = 0$, all z, and $x = b-$, $z > 0$; (5.6a)

(ii) $\phi^{(s)}$ satisfies the radiation conditions as $|z| \to \infty$; (5.6b)

(iii) ϕ and $\partial\phi/\partial z$ are continuous at $0 < x < b$, $z = 0$. (5.6c)

We introduce a Green's function $g^{(2)}(x, x_0, z - z_0)$, satisfying

$$[\nabla^2 + K(z)k^2]g^{(2)} = -\delta(x - x_0)\delta(z - z_0), \quad \text{for} \quad 0 < x < b, \tag{5.7}$$

where the δ source is located in region II.A, i.e., $z_0 < 0$. The boundary conditions for $g^{(2)}$ are:

(i) $g^{(2)} = 0$, at $x = 0$, and $x = b-$; (5.7a)

(ii) $g^{(2)}$ satisfies radiation condition as $|z| \to \infty$; (5.7b)

(iii) $g^{(2)}$ and $\partial g^{(2)}/\partial z$ is continuous at $0 < x < b$, $z = 0$. (5.7c)

It is not difficult to show such a Green's function in the following explicit form:

$$g^{(2)}(x, x_0; z - z_0) = g^{(3)}(x, x_0; z - z_0), \quad \text{for} \quad z < 0,$$
$$= g^{(4)}(x, x_0; z - z_0), \quad \text{for} \quad z > 0, \tag{5.8a}$$

where

$$g^{(3)}(x, x_0; z - z_0) = \frac{1}{b}\prod_{m=1}^{\infty}\frac{1}{\gamma_{mb}} \sin\frac{m\pi}{b}x_0 \sin\frac{m\pi}{b} \times (e^{-\gamma_{mb}|z-z_0|} + R_m e^{-\gamma_{mb}|z+z_0|}), \tag{5.8b}$$

$$R_m = (\gamma_{mb} - \gamma'_{mb})/(\gamma_{mb} + \gamma'_{mb}),$$
$$\gamma_{mb} = [(m\pi/b)^2 - k^2]^{\frac{1}{2}}, \quad \gamma'_{mb} = [(m\pi/b)^2 - \epsilon_r k^2]^{\frac{1}{2}}. \tag{5.8d}$$

The form of $g^{(4)}$ is of no direct interest and is omitted here.

Applying the Green's identity to functions $\phi(x, z)$ and $g^{(2)}(x, z)$ and $g^{(2)}(x - x_0; z - z_0)$ in region II, one has after some manipulations

$$-\phi^{(s)}(x, z) = \int_{-\infty}^{0}\left(\phi^{(s)}\frac{\partial g^{(3)}}{\partial x_0}\right)_{x_0=b-} dz_0 + q(x, z), \quad \text{for} \quad 0 < x < b, \quad z < 0, \tag{5.9a}$$

where

$$q(x, z) = \phi^{(i)}(x, z) + \int_{-\infty}^{0}\left(\phi^{(i)}\frac{\partial g^{(3)}}{\partial x_0}\right)_{x_0=b-} dz_0 + \int_{P_s}\left(\phi^{(i)}\frac{\partial g^{(3)}}{\partial z_0} - g^{(3)}\frac{\partial \phi^{(i)}}{\partial z_0}\right) dx_0. \tag{5.9b}$$

The next step is to express $q(x, z)$ in terms of the field over the aperture $x = b$, $z < 0$. This is detailed below.

Realize first that $\phi^{(i)}$ satisfies the homogeneous wave equation

$$(\nabla^2 + k^2)\psi^{(i)} = 0. \tag{5.10}$$

In order to derive an alternate form of the integral

$$\int_{P_s}\left(\phi^{(i)}\frac{\partial g^{(3)}}{\partial z_0} - g^{(3)}\frac{\partial \phi^{(i)}}{\partial z_0}\right) dx_0$$

appearing in (5.9b), it is desired to have the equation obtained by applying the operator $(\nabla^2 + k^2)$ to $g^{(3)}$. From (5.8b), it can be shown that, for all z,

$$(\nabla^2 + k^2)g^{(3)} = -\delta(x - x_0)\delta(z - z_0) - \delta(z + z_0) \times \left(\frac{2}{b}\prod_{m=1}^{\infty} R_m \sin\frac{m\pi}{b}x_0 \sin\frac{mm\pi}{b}x\right). \tag{5.11}$$

An application of Green's identity to $\phi^{(i)}$ and $g^{(3)}$ in region II gives

$$q(x, z) = -l(x, z) - \int_0^{\infty}\left(\phi^{(i)}\frac{\partial g^{(3)}}{\partial x_0}\right)_{x_0=b-} dz_0, \quad \text{for} \quad z < 0, \tag{5.12a}$$

where

$$l(x, -z) = \int_0^b\left(\frac{2}{b}\prod_{m=1}^{\infty} R_m \sin\frac{m\pi}{b}x_0 \sin\frac{m\pi}{b}x\right) \times \phi^{(i)}(x, -z)\, dx_0. \tag{5.12b}$$

Substituting (5.12a) into (5.9a), one has

$$-\phi^{(s)}(x, z) = \int_{-\infty}^{0}\left(\phi^{(s)}\frac{\partial g^{(3)}}{\partial x_0}\right)_{x_0=b-} dz_0 - \int_0^{\infty}\left(\phi^{(i)}\frac{\partial g^{(3)}}{\partial x_0}\right)_{x_0=b-} dz_0 - l(x, -z), \quad \text{for} \quad 0 < x < b, \quad z < 0, \tag{5.13}$$

which is the desired expression for the scattered field in region II.A.

Now, apply the boundary conditions

$$\phi^{(s)}(x = b+, z) = \phi^{(s)}(x = b-, z), \quad \text{for} \quad z < 0, \tag{5.14a}$$

$$\left.\frac{\partial \phi^{(s)}}{\partial x}\right|_{x=b+} = \left.\frac{\partial \phi^{(s)}}{\partial x}\right|_{x=b-}, \quad \text{for} \quad z < 0, \tag{5.14b}$$

to (5.4) and (5.13). There results

$$\int_{-\infty}^{0}\phi^{(s)}(b, z_0)[g(|z - z_0|) + h(|z + z_0|)]\, dz_0, \int_0^{\infty}\phi^{(i)}(b, z_0)[g(|z - z_0|) + h(|z + z_0|)]\, dz_0 - l'(-z), \quad \text{for} \quad z < 0, \tag{5.15}$$

where

$$g(|z - z_0|) = \left[\frac{\partial^2}{\partial x \partial x_0}\left(g^{(1)}(x, x_0; z - z_0) + \frac{1}{b}\sum_{m=1}^{\infty}\frac{1}{\gamma_{mb}} \sin\frac{m\pi}{b}x_0 \sin\frac{m\pi}{b} e^{-\gamma_{mb}|z-z_0|}\right)\right]_{\substack{x=b \\ x_0=b}},$$

$$h(|z + z_0|) = \left[\frac{\partial^2}{\partial x \partial x_0}\left(\frac{1}{b}\sum_{m=1}^{\infty}\frac{R_m}{\gamma_{mb}} \sin\frac{m\pi}{b}x_0 \times \sin\frac{m\pi}{b} x e^{+\gamma_{mb}|z+z_0|}\right)\right]_{\substack{x=b \\ x_0=b}},$$

$$l'(-z) = \left[\frac{\partial}{\partial x} l(x, -z)\right]_{x=b}.$$

Equation (5.15) is an integral equation of semi-infinite range for the aperture $\phi^{(s)}(b, z)$, $z < 0$.

The transformation of (5.15) into a Wiener–Hopf equation in a Fourier transformation is quite routine. Omitting the details, the final form of the transformed equation is

$$G(\alpha)Z_-(\alpha) + H(\alpha)Z_-(-\alpha)Y_+(\alpha) = G(\alpha)\Phi_+^{(i)}(\alpha) - H(\alpha)\Phi_+^{(i)}(-\alpha) - L(\alpha), \quad |\tau| < \operatorname{Re}\gamma_{1a}, \tag{5.16}$$

where

$$G(\alpha) = (\sinh \gamma a)/(\sinh \gamma c \sinh \gamma b),$$

$$H(\alpha) = \frac{1}{b}\sum_{m=1}^{\infty} \frac{1}{i\gamma_{mb}}\left(\frac{m\pi}{b}\right)^2 R_m\left(\frac{1}{a + i\gamma_{mb}} - \frac{1}{a - i\gamma_{mb}}\right),$$

$$\Phi_+^{(i)}(\alpha) = i \sin(\pi/a)b/(\alpha + i\gamma_{1a}),$$

$$Z_-(\alpha) = \int_{-\infty}^{0} \phi^{(s)}(b, z)e^{i\alpha z}\, dz,$$

$$Y_+(\alpha) = \text{an unknown function},$$

$$L_-(\alpha) = \frac{1}{2\pi}\,\frac{-2i \sin \pi b/a}{b(\alpha - i\gamma_{1a})} \sum_{m=1}^{\infty} \rho_m \frac{(m\pi/b)^2}{\gamma_{1a}^2 - \gamma_{mb}^2}.$$

Note that (5.16) is in a slightly different form from (2.1). However, the method for solution given in Sec. 3 is readily adapted to this case as we now show. Recognize as a first step that

$$[Z_-(\alpha) - \Phi_+^{(i)}(\alpha)]$$

in (5.16) plays the same role as $X_+(\alpha)$ in (2.1). Then, it is not difficult to show that

$$X(\alpha) = Z_-(\alpha) - \Phi_+^{(i)}(\alpha) = -\frac{\Phi_+^{(i)}(\alpha)G_-(-\gamma_{1a})}{G_-(\alpha)} \prod_{n=1}^{\infty} \frac{1 + \alpha/i\gamma_{nb}}{1 + \alpha/i\Gamma_{nb}}, \tag{5.17}$$

where $\Gamma_{nb} \sim \gamma_{nb}$ as $n \to \infty$. The first few Γ_{nb} are determined from the functional equation

$$X(i\gamma_{mb}) \operatorname{Res} G(i\gamma_{mb}) - X(-i\gamma_{mb}) \operatorname{Res} H(i\gamma_{mb}) = 0, \quad \text{for } m = 1, 2, 3, \cdots. \tag{5.18}$$

It is interesting to mention that, for this particular closed-region problem, the solution to the scattered field can also be formulated by the conventional mode matching procedure through the normal mode representations. This procedure leads to an infinite set of linear equations. By applying the modified residue calculus method,[7] the set of equations is then reduced exactly to the functional equation in (5.18). Thus, although the initial forms of the equations in the above two methods are radically different, the functional equation takes an identical form in both of these cases. Extensive numerical computations related to (5.18) can be found in Ref. 7.

6. CONCLUSION

This paper has presented a method for attacking the modified Wiener–Hopf equation (1.1) when $H(\alpha)$ is meromorphic. The solution is reduced to the finding of an infinite zeros Γ_{ma} from a set of linear equations. The asymptotic value of Γ_{ma} can be determined explicitly from the edge condition and, in many practical examples, Γ_{ma} approaches the asymptotic value in an extremely rapid fashion. Therefore, we need to determine only the first few Γ_{ma} for an accurate solution to (1.1).

The method for attacking (1.1) when $H(\alpha)$ has branch singularities will be appearing in a future communication.

[1] B. Noble, *Methods Based on the Wiener–Hopf Technique for the Solution of Partial Differential Equations* (Pergamon Press, Inc., New York, 1958).
[2] D. S. Jones, Proc. Roy. Soc. (London) **A217**, 153 (1953).
[3] D. S. Jones, Phil. Trans. Roy. Soc. **A247**, 499 (1955).
[4] W. E. Williams, "Step Discontinuities in Waveguides," Report No. EM-77, Institute of Mathematical Sciences, New York University, New York, 1955.
[5] V. M. Papadopoulos, Quart. J. Mech. Appl. Math. **10**, 191 (1957).
[6] S. W. Lee, IEEE Trans. MTT **15**, 364 (1967).
[7] R. Mittra, S. W. Lee, and G. F. VanBlaricum, Jr., Intern. J. of Eng. Sci. **6**, 295 (1968).
[8] S. W. Lee and R. Mittra, IEEE Trans. Antennas & Propagation **16**, 513 (1968).
[9] S. W. Lee and R. Mitra, Can. J. of Physics **47**, 1959 (1969).
[10] R. Mittra and S. W. Lee, Alta Frequenza, Special Issue, **38**, 291 (1969).
[11] G. F. Van Blaricum, Jr., and R. Mittra, IEEE Trans. MTT, No. 6, 310 (1969).
[12] R. Mittra and T. Itoh, IEEE Trans. MTT, No. 6, 319 (1969).
[13] C. P. Bates and R. Mittra, Radio Sci. **3**, 251 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Colorings of a Hexagonal Lattice

R. J. BAXTER
Mathematics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 13 May 1969)

The number of ways W^L of coloring the bonds of a hexagonal lattice of L sites (L large) with three colors so that no adjacent bonds are colored alike is calculated exactly, giving $W = 1.20872\cdots$. This is equivalent to counting the number of 4-colorings of the faces of the lattice and can also be regarded as a multiple-dimer problem. If one introduces activities corresponding to certain vertex configurations, then the system is found to have an infinite-order phase transition between two ordered states.

1. INTRODUCTION

Amongst the few exactly soluble 2-dimensional lattice models in statistical mechanics are the ice,[1] F model,[2,3] and KDP models.[4] Together with the dimer problems,[5] they have the property that each bond of the lattice can be in one of two states, and that (with suitable definitions) the number of bonds in each state is the same in every row of the lattice. The partition function is found to be given by equations similar to those encountered with the 1-dimensional Heisenberg chain[6] and 1-dimensional system of bosons with δ-function interaction.[7]

This paper is concerned with a model which has properties similar to those described above, except that a bond can be in one of three states, namely, counting the number of ways the bonds of a plane hexagonal lattice can be colored with three colors A, B, and C so that no two bonds at a vertex have the same color. The three types of colored bonds can be regarded as different species of dimers, with the rule that no two dimers of the same type can overlap. Alternatively, it is shown in the Appendix that the problem is also equivalent to counting the number of ways of coloring the faces of the hexagonal lattice with four colors so that no two adjacent faces are colored alike.

One can obtain a rough approximation, which can be shown to be a lower bound,[8] for the number of colorings by simple combinatorial arguments. Suppose the lattice has L sites, and hence $3L/2$ bonds (L large). Then each bond can have one of three possible colors, but of the 3^5 colorings of the five bonds at two neighboring vertices, only 12 are allowed. Neglecting any further correlations, this suggests that the number of colorings Z_L is given approximately by

$$Z_L \simeq 3^{3L/2}(12/3^5)^{L/2} = (\tfrac{4}{3})^{L/2}. \tag{1.1}$$

Since Z_L is expected to be proportional to W^L, where W is some constant, when L is large, it follows that

$$W^2 \simeq \tfrac{4}{3}. \tag{1.2}$$

Regarding colors B and C as dimers, and a bond colored A as empty, an upper bound on Z_L is obtained by counting the number of ways of independently placing two sets of close-packed dimers on the lattice. This is just the product of the number of ways of placing each of the sets separately, each of which is proportional to $\exp(LG/2\pi)$, where[9]

$$G = \tfrac{1}{2}\sqrt{3}\{1^{-2} + 2^{-2} - 4^{-2} - 5^{-2} + 7^{-2} + 8^{-2} - 10^{-2} - 11^{-2} + \cdots\} \tag{1.3}$$
$$= 1.014938\cdots.$$

Thus $Z_L < \exp(LG/\pi)$, and hence

$$W^2 < \exp(2G/\pi) = 1.90814\cdots. \tag{1.4}$$

In this paper, W is obtained exactly and is found to be given by a coupled pair of integral equations similar to those that arise in the double Heisenberg chain[10] and a 1-dimensional system of spin-$\frac{1}{2}$ fermions with δ-function interaction.[11] It is found that

$$W^2 = \frac{2^2}{1.3}\,\frac{5^2}{4.6}\,\frac{8^2}{7.9}\cdots$$
$$= 1.46099\cdots. \tag{1.5}$$

In addition, one can generalize the problem by associating weights, or activities, with the various configurations of colors at a vertex. In particular, suppose the colors A, B, and C to be ordered cyclically, so that B follows A, C follows B, and A follows C, and divide the vertices into two types a and b as shown in Fig. 1. Then one can associate an activity z_1 with a-vertices (b-vertices) at which the colors of the surrounding bonds go anticlockwise (clockwise) round the vertex, and an activity z_2 when they go in the opposite direction. The partition function Z_L is then

$$Z_L = W^L = \sum z_1^m z_2^n g(m, n), \tag{1.6}$$

where the summation is over all nonnegative integral values of m and n, such that $m + n = L$, and $g(m, n)$ is the number of allowed ways of coloring the lattice so that m vertices have weight z_1 and n have weight z_2.

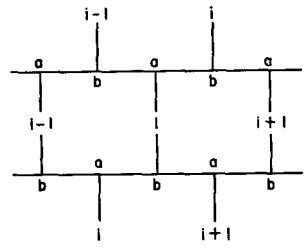


FIG. 1. The labeling of the vertical bonds in successive rows of the lattice and the division of vertices into types a and b.

Clearly Z_L must be of the form $z_1^L f(z_2/z_1)$, so one can without loss of generality set

$$z_1 = z, \quad z_2 = z^{-1}. \tag{1.7}$$

When $z \gg 1$, this system assumes an ordered configuration in which the vertical bonds are colored cyclically from left to right, and conversely when $z \ll 1$. When $z = 1$, the model reduces to the original coloring problem and it is found that, at this point, W, regarded as a function of z, has an essential singularity similar to that occurring at the infinite-order phase transition of the F model.

2. THE TRANSFER MATRIX

Suppose the hexagonal lattice to be drawn as in Fig. 1, with M rows, each containing N vertical bonds, and suppose that the lattice is wound on a cylinder so that the Nth bond of a row precedes the first.

As each row interacts only with the rows immediately above and below, when M is large, the partition function Z is given by

$$Z \sim \lambda^M, \tag{2.1}$$

where λ is the largest eigenvalue of the transfer matrix $\mathbf{V}$. If Φ and Φ' are the colorings of two successive rows of vertical bonds, then $\mathbf{V}$ has elements $V(\Phi, \Phi')$ which are zero if Φ cannot follow Φ', or else equal to the product of the activities of the intermediate vertices. Denoting the elements of the eigenvector of $\mathbf{V}$ as $x(\Phi)$, we see that λ is given by the equation

$$\lambda x(\Phi) = \sum_{\Phi'} V(\Phi, \Phi') x(\Phi). \tag{2.2}$$

To determine the elements $V(\Phi, \Phi')$, first draw lines through all B- and C-colored vertical bonds and all A-colored horizontal bonds. Then these lines cannot begin or end at an interior vertex and must propagate from one row to the next. Thus the number of vertical bonds not covered by a line, namely A-bonds, is the same in each row. Similarly, so are the numbers of B- and C-bonds.

The transfer matrix is therefore made up of a series of diagonal blocks relating colorings Φ and Φ' containing the same number of bonds of each color. One can therefore look at a particular diagonal block and suppose that in each row of vertical bonds there are n lines, of which r are colored C, i.e., there are N—n, n—r, r bonds colored A, B, and C, respectively. A row-coloring Φ can then be prescribed by giving the positions $i_1, \cdots, i_n$ of the lines, and by specifying that lines $p_1, \cdots, p_r$, at positions $i_{p_1}, \cdots, i_{p_r}$, are colored C. Thus the elements $x(\Phi)$ of the eigenvector can be written explicitly as $x(p_1, \cdots, p_r \mid i_1, \cdots, i_n)$.

Given such a prescription for the coloring Φ of a row, one has to determine the allowed colorings Φ' of the row below. First note that the lines cannot travel along more than one horizontal bond between the rows, so, using the labeling of Fig. 1, a vertical line at i must correspond to a line at i or $i+1$ in the row below. Such possible shifts of position can conveniently be expressed by operators $\mathbf{S}_1, \cdots, \mathbf{S}_n$, defined by

$$\mathbf{S}_m x(p_1, \cdots, p_r \mid i_1, \cdots, i_n) = x(p_1, \cdots, p_r \mid i_1, \cdots, i_{m-1}, i_m + 1, i_{m+1}, \cdots, i_n). \tag{2.3}$$

$x(p_1, \cdots, p_r \mid i_1, \cdots, i_n)$ can then be regarded as the element $i_1, \cdots, i_n$ of a vector $\mathbf{x}(p_1, \cdots, p_r)$.

Consider now two C-colored lines p and p' (at i_p and $i_{p'}$) in the upper row, with a number of B-colored lines in between. By considering the two possible sequences of alternate B, C colorings of the intermediate horizontal bonds, one finds the following rules for going from the upper to the lower row:

(i) At most one intermediate line q $(p < q < p')$ can change to color C. When this happens, a line between p and q must move from position i to $i+1$, while the positions of lines between q and p' are unchanged.

(ii) If no intermediate lines change color, then either they all remain at the same position, or else they all move from i to $i+1$. In the former case line p moves to $i_p + 1$ without changing color, while in the latter, line p' stays at $i_{p'}$ without changing color.

Thus in any event there must be a line q, where $p \leq q \leq p'$, which is colored C in the lower row. Figure 2 shows the case (i), where in general lines p and p', as well as q, change color. These lines can each either stay at their original position i, or move to $i+1$.

All vertices between and on lines occur in pairs of activities z and z^{-1}, except when a line changes color. The contribution of such a line to the activity product can be determined as the product of the activities of the vertices between the two surrounding lines. These

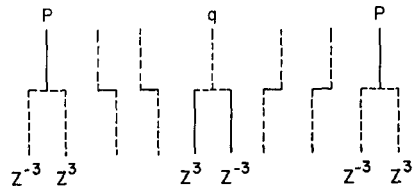


FIG. 2. Typical propagation of lines from upper row to lower. Solid vertical lines represent C-colored bonds, the broken one B-colored bonds. Lines p, q, and p' change color and may move either to the right (from bond i to bond $i+1$) or to the left (bond i to bond i). The contribution to the activity product of each possibility is shown under the corresponding move.

contributions are shown in Fig. 2 under the allowed moves of the line.

The possible moves of the lines $p, p+1, \cdots, p'-1$ shown in Fig. 2 can therefore be obtained, with correct vertex weights, by applying the operators

$$\mathbf{L}(p, q, p') = (z^{-3} + z^{3}\mathbf{S}_p)\mathbf{S}_{p+1} \cdots \mathbf{S}_{q-1}(z^{3} + z^{-3}\mathbf{S}_q) \tag{2.4a}$$

for $p < q < p'$. Similarly, the special cases mentioned in rule (ii), corresponding to $q = p$ and $q = p'$, correspond to operators

$$\mathbf{L}(p, p, p') = \mathbf{S}_p, \tag{2.4b}$$

$$\mathbf{L}(p, p', p') = (z^{-3} + z^{3}\mathbf{S}_p)\mathbf{S}_{p+1} \cdots \mathbf{S}_{p'-1}(z^{-3} + z^{3}\mathbf{S}_{p'})^{-1}, \tag{2.4c}$$

where the inverse operator in (2.4c) serves to cancel the corresponding term arising from the lines to the right of p'.

Ordering the lines so that

$$i_1 < i_2 < \cdots < i_n, \tag{2.5}$$

$$p_1 < p_2 < \cdots < p_r, \tag{2.6}$$

and considering successive pairs of C-colored lines, it follows that the transfer matrix equation (2.2) can be written explicitly as

$$\mathbf{x}(p_1, \cdots, p_r) = \sum_{q_1=p_1}^{p_2} \sum_{q_2=p_2}^{p_3} \cdots \sum_{q_r=p_r}^{p_{r+1}} \mathbf{L}_1 \cdots \mathbf{L}_r \mathbf{x}(q_1, \cdots, q_r), \tag{2.7}$$

where the summand is to be replaced by zero if any two q's are equal,

$$\mathbf{L}_j = \mathbf{L}(p_j, q_j, p_{j+1}), \tag{2.8}$$

and the cyclic conventions $p_{r+1} = p_1 + n$,

$$\mathbf{x}(p_1, \cdots, p_r) = \mathbf{x}(p_r - n, p_1, \cdots, p_{r-1}), \tag{2.9}$$

$$x(p_1, \cdots, p_r \mid i_1, \cdots, i_n) = x(p_1 + 1, \cdots, p_r + 1 \mid i_n - N, i_1, \cdots, i_{n-1}) \tag{2.10}$$

have been adopted.

One exceptional case must be considered, namely, when two lines m and $m+1$ are originally next to one another. The right-hand side of (2.7) then contains two types of terms where they coincide, arising from one of the q's being m or $m+1$. Such terms should be set equal to zero, which is equivalent to requiring that the analytic form of the eigenvector be such that

$$z^{-3}\mathbf{x}(q_1, \cdots, q_r) + z^{3}\mathbf{x}(q_1, \cdots, q_{j-1}, q_j + 1, q_{j+1}, \cdots, q_r) = 0 \tag{2.11}$$

when $i_{q_j} = i_{q_j+1}$.

3. THE FORM OF THE EIGENVECTOR

A. Dependence on $p_1, \cdots, p_r$

The problem now is to solve the eigenvalue equation (2.7), with boundary conditions (2.9)–(2.11). As a first step, ignore conditions (2.10) and (2.11), and look for general solutions of (2.7) and (2.9). Then it is natural to consider solutions which are eigenvectors of $\mathbf{S}_1, \cdots, \mathbf{S}_n$, namely,

$$x(p_1, \cdots, p_r \mid i_1, \cdots, i_n) = s_1^{i_1} s_2^{i_2} \cdots s_n^{i_n} X(p_1, \cdots, p_r). \tag{3.1}$$

Substituting this expression into (2.7), each $\mathbf{S}_m$ can be replaced by s_m, and the equation becomes formally similar to that arising in the F model.[2] In fact, it can be solved by a similar ansatz, namely,

$$X(p_1, \cdots, p_r) = \sum P_{1\ldots r} u_1(p_1) u_2(p_2) \cdots u_r(p_r), \tag{3.2}$$

where $P_{1\ldots r}$ and $u_m(p)$ are short-hand notations such that

$$P_{1\ldots r} \equiv P(\mu_1, \mu_2, \cdots, \mu_r), \tag{3.3}$$

$$u_m(p) \equiv u(\mu_m, p), \tag{3.4}$$

$\mu_1, \cdots, \mu_r$ is some permutation of the numbers $1, \cdots, r$, and the summation in (3.2) is over all such permutations.

Substituting the forms (3.1) and (3.2) of the eigenvector into (2.7), it is found convenient to define the quantities

$$\kappa = z^{6} + z^{-6}, \tag{3.5}$$

$$v(\mu, p) = s_1 s_2 \cdots s_p u(\mu, p)/(z^{-3} + z^{3}s_p), \tag{3.6}$$

$$f(\mu, p) = s_p^{-1} v(\mu, p) + \sum_{q=1}^{p-1} [s_q + s_q^{-1} + \kappa] v(\mu, q), \tag{3.7}$$

$$g(\mu, p) = f(\mu, p) + \kappa v(\mu, p). \tag{3.8}$$

Dividing both sides by various common factors, Eq. (2.7) then becomes

$$\lambda \sum P_{1\ldots r} v_1(p_1) \cdots v_r(p_r) = \sum P_{1\ldots r} \prod_{j=1}^{r} \{f_j(p_{j+1}) - g_j(p_j)\} \tag{3.9}$$

[using short-hand notations analogous to (3.4)], except that the right-hand side of this equation includes spurious contributions arising from $q_{k-1} = p_k = q_k$ in (2.7) for $k = 1, \cdots, r$. Such contributions must be subtracted, giving correction terms to the right-hand side of (3.9) of type

$$-\sum P_{1\cdots r} v_{k-1}(p_k) v_k(p_k) \prod_j \{f_j(p_{j+1}) - g_j(p_j)\}, \quad (3.10)$$

for $k = 1, \cdots, r$, where the product in the summand is over all values of j from 1 to r, *other than* $k-1$ and k.

Equation (3.9) can be solved by requiring that $g(\mu, p)$ and $v(\mu, p)$ satisfy a relation of the form

$$g(\mu, p) = t_\mu v(\mu, p) - c_\mu, \quad (3.11)$$

where t_μ and c_μ are adjustable parameters. Taking the difference of (3.7) for two successive values of p and using (3.8), it follows that

$$v(\mu, p) = c_\mu s_1 \cdots s_p w_{\mu,1} w_{\mu,2} \cdots w_{\mu,p}/(s_p + t_\mu), \quad (3.12)$$

where

$$w_{\mu,p} = (s_p + t_\mu)/(s_p t_\mu - \kappa s_p + 1). \quad (3.13)$$

From (3.1) and (3.2), the cyclic convention (2.9) implies that the coefficients $P_{1,\cdots,r}$ satisfy the relation

$$u_1(p+n) P_{2,3,\cdots,r,1} = u_1(p) P_{1,2,\cdots,r} \quad (3.14)$$

for all positive p. Substituting the form for $u(\mu, p)$ given by (3.6) and (3.12), and interpreting s_{p+n} as s_p, the terms dependent on p in (3.14) cancel, leaving

$$\left\{\prod_{q=1}^{n} w_{\mu_1,q}\right\} P_{2,3,\cdots,r,1} = P_{1,2,\cdots,r}. \quad (3.15)$$

Expanding the product on the rhs of (3.9) and using (3.8) and (3.11), it is found that the first and last terms are proportional to that on the lhs [after relabeling $\mu_1, \cdots, \mu_r$ in the first term and using (3.15)]. Thus these terms can be made to cancel by requiring that λ be given by

$$\lambda = s_1 \cdots s_n \prod_{\mu=1}^{r} (t_\mu - \kappa) + (-)^r t_1 \cdots t_r. \quad (3.16)$$

All other terms in the expansion of the product are of the same type as those occurring in the correction terms such as (3.10). On examination, it is found that these can be made to cancel one another by requiring that

$$\sum [1 - \kappa t_{\mu_k} + t_{\mu_{k-1}} t_{\mu_k}] P_{1,\cdots,r} = 0, \quad (3.17)$$

where, for a given permutation $\mu_1, \cdots, \mu_r$, the summation in (3.17) is over two terms, one given by this permutation and the other by interchanging μ_{k-1} and μ_k.

Equations (3.15) and (3.17) define the coefficients $P_{1\cdots r}$. They have a nontrivial solution if and only if

$$w_{\mu,1} w_{\mu,2} \cdots w_{\mu,n} = -\prod_{\nu=1}^{r} \left\{-\frac{1 - \kappa t_\mu + t_\mu t_\nu}{1 - \kappa t_\nu + t_\mu t_\nu}\right\} \quad (3.18)$$

for $\mu = 1, \cdots, r$.

B. Dependence on $i_1, \cdots, i_n$

As yet no attention has been paid to conditions (2.10) and (2.11). To do so, note that for given $s_1, \cdots, s_n$ Eqs. (3.13) and (3.18) define the parameters $t_1, \cdots, t_r$, and that the eigenvalue λ can then be obtained from (3.16). Further, $t_1, \cdots, t_r$ and λ are then *symmetric* functions of $s_1, \cdots, s_n$.

It follows that there are $n!$ expressions of the type (3.1) which satisfy Eqs. (2.7) and (2.9), corresponding to the various permutations of $s_1, \cdots, s_n$. This suggests that the complete solution may be of the form

$$x(p_1, \cdots, p_r \mid i_1, \cdots, i_n) = \sum_{\{\alpha\}} Q_{\alpha_1, \cdots, \alpha_n} x_{\{\alpha\}}(p_1, \cdots, p_r \mid i_1, \cdots, i_n), \quad (3.19)$$

where $\alpha_1, \cdots, \alpha_n$ is some permutation of the numbers $1, \cdots, n$, the summation is over all such permutations, and $x_{\{\alpha\}}$ is obtained from the above Eqs. (3.1)–(3.13) for x by replacing $s_1, \cdots, s_n$ by $s_{\alpha_1}, \cdots, \alpha_n$.

By using this form for the eigenvector, (2.10) is found to imply that

$$s_{\alpha_1}^N Q_{\alpha_2, \cdots, \alpha_n, \alpha_1} = \left\{\prod_{\mu=1}^{r} w_{\mu,\alpha_1}\right\} Q_{\alpha_1, \cdots, \alpha_n}, \quad (3.20)$$

while after some cancellations (2.11) gives

$$\sum [1 + \kappa s_{\alpha_{p+1}} + s_{\alpha_p} s_{\alpha_{p+1}}] Q_{\alpha_1, \cdots, \alpha_n} = 0, \quad (3.21)$$

the summation being over the two interchanges of α_p and α_{p+1}. Equations (3.20) and (3.21) thus play a similar role for the permutations $\alpha_1, \cdots, \alpha_n$ to that of (3.15) and (3.17) for $\mu_1, \cdots, \mu_r$. They have a solution if and only if

$$s_\alpha^{-N} w_{1,\alpha} w_{2,\alpha} \cdots w_{r,\alpha} = -\prod_{\beta=1}^{n} \left\{-\frac{1 + \kappa s_\beta + s_\alpha s_\beta}{1 + \kappa s_\alpha + s_\alpha s_\beta}\right\} \quad (3.22)$$

for $\alpha = 1, \cdots, n$.

All the conditions on the eigenvector have now been satisfied. It follows that (3.13), (3.18), and (3.22) define $s_1, \cdots, s_n$ and $t_1, \cdots, t_r$. The eigenvalue λ can then be calculated from (3.16) and the partition function from (2.1).

4. SOLUTION IN THE THERMODYNAMIC LIMIT

Equations (3.18) and (3.22) form a coupled set of equations similar to those arising in certain 1-dimensional problems,[10,11] and can be solved in a similar manner.

First make the transformations

$$s_\alpha = [1 - z^{-6}\exp(i\theta_\alpha)]/[\exp(i\theta_\alpha) - z^{-6}], \quad (4.1)$$

$$t_\mu = [z^6 - z^{-6}\exp(i\phi_\mu)]/[1 - \exp(i\phi_\mu)] \quad (4.2)$$

(where now $i^2 = -1$). Taking the logarithms of both sides of (3.18) and (3.22), and using the definitions (3.5) and (3.13), the equations become

$$\sum_\beta F_1(\theta_\beta - \phi_\mu) = \tfrac{1}{2}\pi(r + 1 - 2\mu) - \sum_\nu F_2(\phi_\mu - \phi_\nu), \quad (4.3)$$

$$NF_1(\theta_\alpha) + \sum_\nu F_1(\theta_\alpha - \phi_\nu) = -\tfrac{1}{2}\pi(n + 1 - 2\alpha) + \sum_\beta F_2(\theta_\alpha - \theta_\beta), \quad (4.4)$$

for $\mu = 1, \cdots, r$ and $\alpha = 1, \cdots, n$, respectively, the summations over ν and β also being over these ranges. The functions $F_m(\theta)$ are defined by

$$F_m(\theta) = \tan^{-1}\left\{\frac{z^{6m} + 1}{z^{6m} - 1}\tan\tfrac{1}{2}\theta\right\}, \quad (4.5)$$

$$-\tfrac{1}{2}\pi < F_m(\theta) \leq \tfrac{1}{2}\pi. \quad (4.6)$$

The first terms on the right-hand sides of (4.3) and (4.4) arise from particular choices of the branch of the logarithm function. When $z > 1$, these choices ensure that the θ's and ϕ's are real and satisfy the relations

$$\theta_1 < \theta_2 < \cdots < \theta_n, \quad \theta_{n+1-\alpha} = -\theta_\alpha,$$
$$\phi_1 < \phi_2 < \cdots < \phi_r, \quad \phi_{r+1-\mu} = -\phi_\mu. \quad (4.7)$$

When r, n, and N become large, their ratios remaining constant, the spacing between successive θ's and ϕ's becomes small, while $\theta_1 = -\theta_n$ and $\phi_1 = -\phi_r$ tend to certain limiting values U and V. In the limit the θ's and ϕ's are continuously distributed over the ranges $(-U, U)$ and $(-V, V)$, respectively.

If $N\rho(\theta)\,d\theta$ is the number of θ_α's between θ and $\theta + d\theta$, and $N\sigma(\phi)\,d\phi$ is the number of ϕ_μ's between ϕ and $\phi + d\phi$, then Eqs. (4.3) and (4.4) become

$$\int_{-U}^{U} d\theta'\rho(\theta')F_\theta(\theta' - \phi) = -\pi\int_{-V}^{\phi} d\phi'\sigma(\phi') - \int_{-V}^{V} d\phi'\sigma(\phi')F_2(\phi - \phi'), \quad (4.8a)$$

$$F_1(\theta) + \int_{-V}^{V} d\phi'\sigma(\phi')F_1(\theta - \phi') = \pi\int_{-U}^{\theta} d\theta'\rho(\theta') + \int_{-U}^{U} d\theta'\rho(\theta')F_2(\theta - \theta'), \quad (4.8b)$$

and U and V can be regarded as determined by the relations

$$n/N = \int_{-U}^{U} d\theta\rho(\theta), \quad (4.9a)$$

$$r/N = \int_{-V}^{V} d\phi\sigma(\phi). \quad (4.9b)$$

Applying the same transformations and limiting procedures to the eigenvalue equation (3.16), it is found that

$$\frac{1}{N}\log\frac{\lambda}{2} = \frac{1}{2}\int_{-V}^{V} d\phi\sigma(\phi)\log\left\{\frac{\tau + \tau^{-1} - 2\cos\phi}{2 - 2\cos\phi}\right\}, \quad (4.10)$$

where

$$\tau = z^{12}. \quad (4.11)$$

When $U = V = \pi$, the Eqs. (4.8) can be solved by differentiating and Fourier-analyzing. This gives

$$\rho(\theta) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{im\theta}\frac{(\tau^{m/2} + \tau^{-m/2})}{(\tau^m + 1 + \tau^{-m})}, \quad (4.12a)$$

$$\sigma(\phi) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\frac{e^{im\phi}}{(\tau^m + 1 + \tau^{-m})}, \quad (4.12b)$$

and hence from (4.9),

$$n = \tfrac{2}{3}N, \quad r = \tfrac{1}{3}N. \quad (4.13)$$

Thus in this case there are equal numbers of A-, B-, and C-colored bonds in each row of the lattice. Intuitively, we expect that this corresponds to the maximum eigenvalue λ. Substituting the form (4.12b) of $\sigma(\phi)$ into (4.10) and performing the ϕ integration, it is found that

$$\frac{1}{N}\log\frac{\lambda}{2} = \frac{1}{6}\log\tau + \sum_{m=1}^{\infty}\frac{1}{m}\frac{\tau^{-m}(1 - \tau^{-m})^2}{1 - \tau^{-3m}}. \quad (4.14)$$

Taylor-expanding the summand in this equation and summing each term with respect to m, it follows that

$$W^2 = Z^{1/MN} = \lambda^{1/N} = z^2\prod_{p=1}^{\infty}\frac{(1 - \tau^{1-3p})^2}{(1 - \tau^{2-3p})(1 - \tau^{-3p})} \quad (4.15)$$

when $z > 1$.

5. CONCLUSIONS

Equation (4.15), together with the definition (4.11), expresses the partition function Z in terms of the activity z when $z > 1$. When $z < 1$, it should be replaced by z^{-1} in these equations, since it is apparent from the symmetry of the problem that the partition function is unchanged by such a transformation.

When $z \gg 1$, the system assumes an ordered state in which the colors go anticlockwise round each a

vertex, and clockwise around each b vertex. Thus $W \sim z$, as is indeed given by (4.15).

When $z \to 1$, the problem reduces to the original unweighted coloring problem, and the expression (4.15) tends to the appropriate limit, namely,

$$W^2 = \prod_{p=1}^{\infty} \frac{(3p-1)^2}{(3p-2)3p}. \tag{5.1}$$

Further, W is found to have a Taylor expansion about $z = 1$ which is formally invariant with respect to the transformation of replacing z by z^{-1}. However, it is not analytic at this point. Rather, it has an essential singularity. Indeed, it appears from (4.15) that $W(z)$ is a lacunary function, having a solid wall of singularities on the unit circle. Thus at $z = 1$ the system undergoes an infinite-order phase transition similar to that of the F model,[2,3] except that in this case the transition is between two ordered states, rather than from a disordered to an ordered state.

6. ACKNOWLEDGMENT

The author is grateful to Professor Elliott Lieb for many useful comments, in particular for pointing out the equivalences discussed in the Appendix.

APPENDIX

Lenard (Ref. 1) has pointed out that the square-lattice ice model is equivalent to a coloring problem of the faces of the lattice. Similarly, the three-color bond problem discussed in this paper ($z = 1$) is equivalent to counting the number of ways of coloring the faces of the hexagonal lattice with four colors a, b, c, and d so that no adjacent faces are colored alike.

To see this, suppose that some allowed coloring of the faces has been made and regard the colors as ordered cyclically so that b follows a, c follows b, d follows c, and a follows d. Now imagine an observer standing on a face f of the lattice, looking across a bond to an adjacent face. If the adjacent color cyclically follows or precedes that of f, place an arrow on the intermediate bond pointing to the right or left, respectively, or else leave the bond empty. Do this to all bonds. Then the configuration of arrows is specified uniquely and can be seen to be such that at any vertex there are two and only two arrows, either both pointing in or both pointing out.

Further, once one face of the lattice has been colored, to any configuration of arrows on bonds satisfying this condition at each vertex there corresponds an allowed coloring of the faces. Allowing for the four possible colorings of the initial face, it follows that if

the number of allowed arrow configurations $= Z$,

then

the number of allowed 4-colorings of the faces $= 4Z$.

Now consider the arrow configurations. Following successive arrows round the lattice, one traces out a series of nonintersecting closed polygons such that each vertex lies on a polygon. Further, each polygon can be in one of two states, one being obtained from the other by reversing the arrows along it. Thus,

$$Z = \sum_m 2^m g(m),$$

where $g(m)$ is the number of ways of covering the lattice with m polygons. However, this expression for Z is precisely the number of ways of coloring the bonds of the lattice with three colors A, B, and C, since successive B, C bonds also form closed polygons which cover the lattice and can be in two states, corresponding to interchanging B and C. Thus

$Z =$ number of allowed three colorings of the bonds,

and the required equivalence is established.

[1] E. H. Lieb, Phys. Rev. **162**, 162 (1967).
[2] E. H. Lieb, Phys. Rev. Letters **18**, 1046 (1967).
[3] R. J. Baxter, J. Math. Phys. **10**, 1211 (1969).
[4] E. H. Lieb, Phys. Rev. Letters **19**, 108 (1967). C. P. Yang, *ibid.*, **19**, 586 (1967).
[5] P. W. Kasteleyn, J. Math. Phys. **4**, 287 (1963).
[6] C. N. Yang and C. P. Yang, Phys. Rev. **150**, 321 (1966).
[7] E. H. Lieb and W. Linger, Phys. Rev. **130**, 1605 (1963).
[8] The approximation can also be obtained as a variational approximation for the maximum eigenvalue λ of the transfer matrix (2.2) by trying $x(\Phi) = \text{const}$. Thus it must be a lower bound.
[9] See Ref. 5 for a general discussion. Alternatively, the dimer problem can be solved by drawing lines through empty vertical bonds and full horizontal bonds, and then applying reasoning similar to, but simpler than, that of this paper.
[10] E. H. Lieb and F. Y. Wu, Phys. Rev. Letters **C20**, 1445 (1968).
[11] C. N. Yang, Phys. Rev. Letters **23**, 1312 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Concavity of Magnetization of an Ising Ferromagnet in a Positive External Field

ROBERT B. GRIFFITHS*
Institute for Theoretical Physics,† State University of New York at Stony Brook, Stony Brook, New York 11790

C. A. HURST
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia 5001

AND

S. SHERMAN‡
Department of Mathematics, Indiana University, Bloomington, Indiana 47401
and
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia 5001

(Received 23 June 1969)

An inequality for correlation funtions in an Ising model with purely ferromagnetic interactions between pairs of spins is established and used to show that the magnetization in such a model is a concave function of external field H for $H > 0$. The concavity of magnetization, which holds not only for spin-$\frac{1}{2}$ but also for arbitrary-spin Ising ferromagnets, provides a basis for certain thermodynamic inequalities near the ferromagnetic critical point, including one involving the "high temperature" indices α and γ.

I. INTRODUCTION

That the magnetization M should be a concave function of magnetic field H for $H \geq 0$ (and a convex function for $H \leq 0$, since M is an odd function of H) is one of a series of plausible conjectures useful in analyzing the behavior of ferromagnets near their critical points. In particular, it is important for certain thermodynamic inequalities for critical point exponents, including an inequality involving the "high temperature" exponents α and γ.[1] We wish to present a proof of this conjecture in the case of an Ising model with purely ferromagnetic interactions (of arbitrary range) between pairs of spins. The proof is obtained, as in earlier work of an analogous nature,[2,3] by establishing a particular type of inequality for correlation functions. Whereas the desire to prove the concavity of M motivated a consideration of the inequality in question [see (2.8) and (2.10)], the latter is of interest in its own right and may possibly have other applications. Our proof employs the methods of Ref. 3 (hereafter referred to as KS), Sec. 9. However, the reader should note that the Hamiltonian (2.9) for which the present results are obtained is much less general than that employed in KS. While the proof is carried out for a spin-$\frac{1}{2}$ Ising model (each Ising "spin" can have two possible values), the inequality (2.8), and consequently the result on concavity of magnetization, is easily extended to Ising models with arbitrary magnitude of spin.

A statement of the problem and some notation is found in Sec. II together with a generalization to Ising models with spin greater than $\frac{1}{2}$. In Sec. III, the correlation function inequality is transformed into an equivalent problem in the theory of graphs, and the proof is completed with the help of a theorem on graphs in Sec. IV. The application of concavity in magnetization to inequalities for critical point exponents is discussed briefly in Sec. V, with particular reference to 2- and 3-dimensional Ising ferromagnets.

II. NOTATION AND PRINCIPAL THEOREM

Consider an Ising model with spins σ_i, $i = 1, 2 \cdots n$, which take the values ± 1, and a Hamiltonian

$$\mathcal{H} = -\sum_{1 \leq i < j \leq n} \sum J_{ij}\sigma_i\sigma_j - H\sum_{i=1}^{n} \sigma_i, \tag{2.1}$$

where H is the external magnetic field and we assume that

$$J_{ij} \geq 0, \tag{2.2}$$

that is, all exchange interactions are ferromagnetic. Let

$$s = \sigma_1 + \sigma_2 + \cdots + \sigma_n \tag{2.3}$$

and let

$$M = \langle s \rangle / n \tag{2.4}$$

be the average magnetization per spin. Here angular brackets denote a thermal average. It follows from general convexity considerations[4] [for which (2.2) is not necessary] that

$$\frac{\partial M}{\partial H} \geq 0 \tag{2.5}$$

for all H. We wish to establish the following:

Theorem 1:

$$\frac{\partial^2 M}{\partial H^2} \leq 0, \quad \text{for all} \quad H \geq 0, \tag{2.6}$$

that is, the magnetization is a concave function of

positive external field. This amounts to showing [KS, relation (7.8)] that

$$\langle s^3\rangle - 3\langle s\rangle\langle s^2\rangle + 2\langle s\rangle^3 \leq 0, \tag{2.7}$$

which is in turn implied by

$$\langle\sigma_i\sigma_j\sigma_k\rangle - \langle\sigma_i\rangle\langle\sigma_j\sigma_k\rangle - \langle\sigma_j\rangle\langle\sigma_i\sigma_k\rangle - \langle\sigma_k\rangle\langle\sigma_i\sigma_j\rangle + 2\langle\sigma_i\rangle\langle\sigma_j\rangle\langle\sigma_k\rangle \leq 0, \tag{2.8}$$

for arbitrary choice of i, j, k (not necessarily distinct). A completely equivalent method for treating the problem (see Ref. 2) is to introduce a "ghost" spin σ_0, with the magnetic field acting on a spin σ_i replaced with a corresponding exchange interaction $J_{0i} = H$, and correlation functions involving an odd number of spin operators modified by inserting σ_0, e.g., $\langle\sigma_i\rangle$ becomes $\langle\sigma_i\sigma_0\rangle$. The new Hamiltonian is simply

$$\mathcal{H} = -\sum_{0\leq i<j\leq n}\sum J_{ij}\sigma_i\sigma_j, \tag{2.9}$$

and Theorem 1 is an immediate consequence of

Theorem 2: Provided all $J_{ij} \geq 0$ in (2.9),

$$\langle\sigma_i\sigma_j\sigma_k\sigma_l\rangle - \langle\sigma_i\sigma_l\rangle\langle\sigma_j\sigma_k\rangle - \langle\sigma_j\sigma_l\rangle\langle\sigma_i\sigma_k\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_i\sigma_j\rangle + 2\langle\sigma_i\sigma_l\rangle\langle\sigma_j\sigma_l\rangle\langle\sigma_k\sigma_l\rangle \leq 0. \tag{2.10}$$

Note that, if $l = 0$ so that σ_l is the "ghost spin," (2.10) is equivalent to (2.8). But, in fact, (2.10) is more general, and we could interpret σ_i as the ghost spin so that

$$\langle\sigma_j\sigma_k\sigma_l\rangle - \langle\sigma_l\rangle\langle\sigma_j\sigma_k\rangle - \langle\sigma_k\rangle\langle\sigma_j\sigma_l\rangle - \langle\sigma_j\rangle\langle\sigma_k\sigma_l\rangle + 2\langle\sigma_l\rangle\langle\sigma_j\sigma_l\rangle\langle\sigma_k\sigma_l\rangle \leq 0 \tag{2.11}$$

is valid for any spins j, k, l in the presence of a positive magnetic field $H \geq 0$.

Before starting the proof of Theorem 2, let us consider the generalization of Theorem 1 and the inequality (2.8) in the case of Ising models with spin $\frac{1}{2}p$, p an integer ≥ 1, in which σ in (2.1), (2.3), etc., is replaced by S, a variable which takes on the values $-p, -p+2, -p+4, \cdots, p$. It is easily shown that, with σ replaced by S,

$$\langle S_iS_jS_k\rangle - \langle S_i\rangle\langle S_jS_k\rangle - \langle S_j\rangle\langle S_iS_k\rangle - \langle S_k\rangle\langle S_iS_j\rangle + 2\langle S_i\rangle\langle S_j\rangle\langle S_k\rangle \leq 0 \tag{2.12}$$

is a consequence of (2.8) itself. For this we write

$$S_i = \sum_{i'=1}^{p}\sigma_{ii'} \tag{2.13}$$

(and similar expressions for S_j and S_k), where $\sigma_{ii'} = \pm 1$. The system described with spins S_i and ferromagnetic bonds is for many purposes identical to an "analog" system[5] in which only the $\sigma_{ii'}$, $\sigma_{jj'}$, and $\sigma_{kk'}$ occur with suitable ferromagnetic interactions. In particular, the correlations appearing in (2.12) may be evaluated by replacing each S with the corresponding σ's according to (2.13). The sum of all the resulting terms corresponding to a *particular* choice for i', j', and k' is nonpositive by (2.8), and consequently (2.12) is itself nonpositive. The same argument cannot be used to generalize (2.10) to the case of Ising models with arbitrary spin because σ_l plays a special role and cannot simply be replaced by S_l.

III. GRAPHICAL FORMULATION OF THEOREM

Since the temperature (always assumed positive) enters in an inconsequential way in our arguments, we shall henceforth suppose $\beta = (kT)^{-1} = 1$. Using the result $\sigma_i^2 = 1$ for any i, we can write (see KS, Secs. 9, 11)

$$\exp\left(\sum_{i<j}J_{ij}\sigma_i\sigma_j\right) = \sum_{A\subset N}\pi_A\sigma^A, \tag{3.1}$$

where N is the set $\{0, 1, 2, \cdots, n\}$ and

$$\sigma^A = \prod_{i\in A}\sigma_i. \tag{3.2}$$

In particular, when A is the null set ϕ, $\sigma^\phi = 1$, and the corresponding π_ϕ is a multiple of the partition function,

$$2^{n+1}\pi_\phi = Z = \sum_{\sigma_i=\pm 1}\exp\left(\sum_{i<j}J_{ij}\sigma_i\sigma_j\right), \tag{3.3}$$

while, in general,

$$2^{n+1}\pi_A = Z\langle\sigma^A\rangle. \tag{3.4}$$

Consequently, (2.10) amounts to showing that, for any four spins 0, 1, 2, 3 [corresponding to l, i, j, k in (2.10), respectively],

$$\pi_\phi^2\pi_{0123} - \pi_\phi(\pi_{01}\pi_{23} + \pi_{02}\pi_{13} + \pi_{03}\pi_{12}) + 2\pi_{01}\pi_{02}\pi_{03} \leq 0. \tag{3.5}$$

In the case we are considering, in which only pair interactions occur in the Hamiltonian, a graphical interpretation of the π_A is very convenient. With vertices i and j from the set N, a graph ν is completely specified by giving the number of edges $\nu(ij)$ connecting i and j.[6] We shall denote by $\delta\nu$ the set of odd vertices in ν, vertices which are end points for an odd number of edges in ν.

For each graph ν, define the weight

$$W(\nu) = \prod_{i<j}[(J_{ij})^{\nu(ij)}/\nu(ij)!]. \tag{3.6}$$

We can then write [KS, Sec. 9]

$$\pi_A = \sum_{\nu:\delta\nu=A}W(\nu). \tag{3.7}$$

Note that, in the case $A = \phi$, the sum includes the null graph (no edges) with weight 1.

Consider next a typical term appearing in (3.5):

$$\pi_{01}\pi_{02}\pi_{03} = \sum_{\nu_{01}}\sum_{\nu_{02}}\sum_{\nu_{03}} W(\nu_{01})W(\nu_{02})W(\nu_{03}), \quad (3.8)$$

where

$$\delta\nu_{0j} = \{0j\}, \quad (3.9)$$

and we shall in general use the convention that, if a graph ν has subscripts, these are the elements of the set $\delta\nu$. It is rather natural to associate with each summand in (3.8) a single graph μ, with

$$\mu(ij) = \nu_{01}(ij) + \nu_{02}(ij) + \nu_{03}(ij), \quad (3.10)$$

which permits us to rewrite (3.8) as

$$\pi_{01}\pi_{02}\pi_{03} = \sum_{\mu} D_\mu(\pi_{01}\pi_{02}\pi_{03})W(\mu). \quad (3.11)$$

Here, with $\mu!$ an abbreviation for the product over all pairs $i<j$ of $\mu(ij)!$, and $\nu_{01}!$, $\nu_{02}!$, $\nu_{03}!$ similarly defined,

$$D_\mu(\pi_{01}\pi_{02}\pi_{03}) = \sum_{\nu_{01}}\sum_{\nu_{02}}\sum_{\nu_{03}} \frac{\mu!}{\nu_{01}!\,\nu_{02}!\,\nu_{03}!}, \quad (3.12)$$

and this last sum is over ordered triples $(\nu_{01}, \nu_{02}, \nu_{03})$ satisfying (3.9) and (3.10). If for a particular μ no such triple exists, $D_\mu = 0$.

Of course, an analogous procedure may be used for the other terms in (3.5). Thus, in order to establish (3.5)—note that all $W(\mu)$ are positive—it suffices to show that for every μ

$$2D_\mu(\pi_{01}\pi_{02}\pi_{03}) \leq D_\mu(\pi_\phi\pi_{01}\pi_{23}) + D_\mu(\pi_\phi\pi_{02}\pi_{13}) + D_\mu(\pi_\phi\pi_{03}\pi_{12}) - D_\mu(\pi_\phi^2\pi_{0123}). \quad (3.13)$$

Hitherto, we have treated the graphs as having indistinguishable edges in the sense that two graphs ν and ν' are the same if $\nu(ij) = \nu'(ij)$ for every pair (ij). However, the combinatorial factor D_μ has a particularly simple interpretation if we change our point of view and suppose that the graph μ has *labeled edges* which can be regarded as distinguishable. *From this point on, we shall define* a graph μ to be a set of (distinguishable) edges. Two graphs μ and ν will be considered identical only if they contain precisely the same edges, and not simply the same number of edges connecting any pair of vertices. The vertices of a graph μ, $V(\mu)$, is the set of all end points of the edges in the set μ.[6] When b is a vertex, "μ contains b" means $b \in V(\mu)$. We shall be concerned with *decompositions* (ν, ν', ν'') of a graph μ into mutually disjoint (no edges in common) subgraphs

$$\mu = \nu + \nu' + \nu'' \quad (3.14)$$

(we shall use $+$ in place of $\cup$ for union of disjoint graphs), with the convention that (ν, ν', ν'') and $(\omega, \omega', \omega'')$ are the same decomposition only if $\nu = \omega$, $\nu' = \omega'$, and $\nu'' = \omega''$.

With these definitions, we can interpret

$$D_\mu(\pi_{01}\,\pi_{02}\,\pi_{03})$$

as the total number of decompositions of the graph μ (with labeled edges) of the form $(\nu_{01}, \nu_{02}, \nu_{03})$, the summand in (3.12) being precisely the correct combinatorial factor to go from a system of unlabeled to one with labeled edges. Similarly, $D_\mu(\pi_\phi^2\,\pi_{0123})$ is the number of decompositions of μ in the form $(\nu_\phi, \nu'_\phi, \nu_{0123})$, where note that ν_ϕ or ν'_ϕ (or both) may be the null graph with no edges at all.

Another concept we shall need is that of the symmetric difference of two graphs. The symmetric difference of two sets A and B, written $A\Delta B$, is the set of all elements which appear in A or in B, but not in both. Similarly, the graph $\rho = \nu\Delta\nu'$ is the set of (labeled) edges which occur in either ν or ν', but not in both. The relation

$$\delta(\nu\Delta\nu') = (\delta\nu)\Delta(\delta\nu') \quad (3.15)$$

states, in words, that the odd vertices of $\nu\Delta\nu'$ are those which are either even in ν and odd in ν', or odd in ν and even in ν'. We use this result in establishing the following:

Lemma 1: Let ω be a graph, and a and b two of its vertices. Define

$$F(\omega) = \{\nu : \nu \doteq \omega,\ \delta\nu = \phi\}, \quad (3.16)$$

$$E(\omega; a, b) = \{\nu : \nu \subset \omega,\ \delta\nu = \{ab\}\}. \quad (3.17)$$

Provided $E(\omega; a, b)$ is not empty, the following equality holds:

$$\#[E(\omega; a, b)] = \#[F(\omega)], \quad (3.18)$$

where $\#[\ ]$ stands for the number of elements in the set $[\ ]$. The proof is obtained by choosing some element $\pi \in E(\omega; a, b)$ and noting that, for fixed π,

$$\nu \to \nu' = \pi\Delta\nu \quad (3.19)$$

is a one-to-one mapping of $E(\omega; a, b)$ into $F(\omega)$ and vice versa. The result may also be stated as follows: The subsets of μ are an Abelian group under symmetric difference, with the null graph as the identity. Clearly, $F(\omega)$ is a subgroup and $E(\omega; a, b)$ a coset.

Note that $E(\omega; a, b)$ is empty if and only if the vertices a and b lie in different connected components of the graph ω, a simple consequence of the fact that any connected component of a graph must have an

even number of odd vertices.[7] We shall use ω $(a = b)$ as an abbreviation for saying that a and b are in the same connected component of ω, and ω $(a \neq b)$ for the contrary.

Lemma 1 may be used to simplify (3.13). All decompositions below refer to the graph μ with $\delta\mu = \{0123\}$. First let us establish the result

$$\#\{(\nu_\phi, \nu'_\phi, \nu_{0123}):(\mu - \nu_\phi)\,(0 = 3)\} = \#\{(\nu_\phi, \nu_{03}, \nu_{12})\}. \quad (3.20)$$

We shall in fact show that this holds for each choice of ν_ϕ. Consider the case $(\mu - \nu_\phi)\,(0 \neq 3)$. The left side of (3.20) is zero, but the right side also vanishes because $\mu - \nu_\phi$ contains no subgraph ν_{03} whose only odd vertices are 0 and 3. When, on the other hand, $(\mu - \nu_\phi)\,(0 = 3)$, we apply Lemma 1 with ω replaced by $\mu - \nu_\phi$ and $a = 0$, $b = 3$. For a fixed ν_ϕ, the number of decompositions $(\nu_\phi, \nu'_\phi, \nu_{0123})$ is simply the number of ways of choosing a subgraph ν'_ϕ of $\mu - \nu_\phi$, since this choice uniquely determines ν_{0123}. Consequently, the left side of (3.20) is $\#[F(\mu - \nu_\phi)]$, while, by an analogous argument, the right side is $\#[E(\mu - \nu_\phi; 0, 3)]$. On the basis of (3.20) and the definition of D_μ, we have

$$\begin{aligned} D_\mu(\pi_\phi^2\pi_{0123}) &= \#\{(\nu_\phi, \nu'_\phi, \nu_{0123})\} \\ &= \#\{(\nu_\phi, \nu'_\phi, \nu_{0123}):(\mu - \nu_\phi)\,(0 = 3)\} \\ &\quad + \#\{(\nu_\phi, \nu'_\phi, \nu_{0123}):(\mu - \nu_\phi)\,(0 \neq 3)\} \\ &= D_\mu(\pi_\phi\pi_{03}\pi_{12}) \\ &\quad + \#\{(\nu_\phi, \nu'_\phi, \nu_{0123}):(\mu - \nu_\phi)\,(0 \neq 3)\}. \end{aligned} \quad (3.21)$$

Note that $\mu - \nu_\phi$ has, like μ, four odd vertices, and in the case $(\mu - \nu_\phi)\,(0 \neq 3)$ they must occur in pairs in different connected components. The two mutually exclusive possibilities are $(\mu - \nu_\phi)\,(0 = 1)$ and $(\mu - \nu_\phi)\,(0 = 2)$. Applying Lemma 1 to the graph $\mu - \nu_\phi$ for these two cases, with $a = 0$ and $b = 1$ and 2, respectively, we obtain

$$\begin{aligned} &\#\{(\nu_\phi, \nu'_\phi, \nu_{0123}):(\mu - \nu_\phi)\,(0 \neq 3)\} \\ &\quad = \#\{(\nu_\phi, \nu_{01}, \nu_{23}):(\mu - \nu_\phi)\,(0 \neq 3)\} \\ &\qquad + \#\{(\nu_\phi, \nu_{02}, \nu_{13}):(\mu - \nu_\phi)\,(0 \neq 3)\}. \end{aligned} \quad (3.22)$$

In addition, we have

$$\begin{aligned} &D_\mu(\pi_\phi\pi_{01}\pi_{23}) \\ &\quad = \#\{(\nu_\phi, \nu_{01}, \nu_{23}):(\mu - \nu_\phi)\,(0 \neq 3)\} \\ &\qquad + \#\{(\nu_\phi, \nu_{01}, \nu_{23}):(\mu - \nu_\phi)\,(0 = 3)\} \end{aligned} \quad (3.23)$$

and the analogous equation where 1 and 2 are everywhere interchanged; combining these results with (3.13), (3.21), and (3.22), we have

$$\begin{aligned} &2D_\mu(\pi_{01}\pi_{02}\pi_{03}) \\ &\quad \leq \#\{(\nu_\phi, \nu_{01}, \nu_{23}):(\mu - \nu_\phi)\,(0 = 3)\} \\ &\qquad + \#\{(\nu_\phi, \nu_{02}, \nu_{13}):(\mu - \nu_\phi)\,(0 = 3)\}. \end{aligned} \quad (3.24)$$

The application of Lemma 1 twice to the graph $\mu - \nu_\phi$ shows that the two terms on the right side of (3.24) are equal, and its application to the graph $\mu - \nu_{02}$ shows that

$$D_\mu(\pi_{01}\pi_{02}\pi_{03}) = \#\{(\nu_\phi, \nu_{02}, \nu_{13}):(\mu - \nu_{02})\,(0 = 3)\}. \quad (3.25)$$

Hence, we have (noting that $\mu - \nu_{02} = \nu_\phi + \nu_{13}$, etc.)

$$\begin{aligned} &\#\{(\nu_\phi, \nu_{02}, \nu_{13}):(\nu_\phi + \nu_{13})\,(0 = 3)\} \\ &\quad \leq \#\{(\nu_\phi, \nu_{02}, \nu_{13}):(\nu_{02} + \nu_{13})\,(0 = 3)\} \end{aligned} \quad (3.26)$$

as an inequality completely equivalent to (3.13).

IV. A THEOREM ON GRAPHS

We shall show that the inequality (3.26) is satisfied separately for every choice of $\nu_{13} \subset \mu$. The result follows, in fact, from the following theorem:

Theorem 3: Let the graph μ (with labeled edges) be the sum of two disjoint (no edges in common) subgraphs ω and ρ. Suppose that $a \neq b$ are among the vertices of ω, and w (which could be a or b) is one of the vertices of ρ. Define

$$C = \{\nu:\nu \subset \omega,\ (\nu + \rho)\,(a = w) \text{ or } (\nu + \rho)\,(b = w) \text{ or both}\} \quad (4.1)$$

and let E and F be defined by (3.16) and (3.17). Then, if $E(\omega; a, b)$ is not empty, the following inequality is valid:

$$\#[F(\omega) \cap C] \leq \#[E(\omega; a, b) \cap C]. \quad (4.2)$$

Before carrying out the proof of this theorem, let us see how it can be used to establish (3.26). The ρ in the theorem corresponds to ν_{13} and the ω to $\mu - \nu_{13}$, while $a = 0$, $b = 2$, and $w = 3$. We then interpret the right side of (3.26) as the number of ways of choosing a subgraph ν_{02} in $\mu - \nu_{13}$ such that 0 and 3 are connected in $(\nu_{02} + \nu_{13})$; the graph $\nu_\phi = \mu - \nu_{13} - \nu_{02}$ is determined uniquely when ν_{02} is known. Similarly, the left side of (3.26) is the number of ways of choosing ν_ϕ in $\mu - \nu_{13}$ with a similar connectivity condition. The correspondence between (3.26) and (4.2) should now be plain except for one minor point. In (4.2) we allow the possibility $(\nu + \rho)\,(2 = 3)$ in place of $(\nu + \rho)\,(0 = 3)$. However, since 0 and 2 are connected in ν_{02}, $(\nu_{02} + \nu_{13})\,(0 = 3)$ implies $(\nu_{02} + \nu_{13})\,(2 = 3)$.

Therefore, the right-hand member in (4.2), with the above interpretation, is the same as the right-hand member in (3.26), while the left-hand member in (4.2) is in general larger than the corresponding term in (3.26). This still permits us, of course, to infer (3.26) from (4.2).

The foregoing remarks show that Theorem 3 has (3.26) as one of its consequences when the right side of (3.26) is not zero [this corresponds to the stipulation preceding (4.2)]. If we restrict ourselves to μ for which $\delta\mu = \{0123\}$ and to $\nu_{13} \subset \mu$ [otherwise both sides of (3.26) vanish] and note that a possible choice for ν_ϕ is the null subgraph, it is clear that the right side of (3.26) will vanish only if 0 and 3 are in separate connected components of μ, a situation in which the left side of (3.26) must also, obviously, vanish.

Proof of Theorem 3: Define

$$G = \{\nu : \nu \subset \omega,\ (\nu + \rho)\ (a \neq w),\ (\nu + \rho)\ (b \neq w)\} = \{\nu : \nu \subset \omega\} - C \tag{4.3a}$$

and note that

$$\begin{aligned}\#[E(\omega; a, b) \cap G] + \#[E(\omega; a, b) \cap C] &= \#[E(\omega; a, b)] = \#[F(\omega)] \\ &= \#[F(\omega) \cap G] + \#[F(\omega) \cap C],\end{aligned} \tag{4.3b}$$

where the second equality is a consequence of Lemma 1, together with the hypothesis that $E(\omega; a, b)$ is not empty. Consequently, (4.2) is equivalent to

$$\#[E(\omega; a, b) \cap G] \leq \#[F(\omega) \cap G]. \tag{4.4}$$

Given any $\nu \in G$, let τ be the connected component of $\nu + \rho$ which contains w; of course, neither a nor b belong to $V(\tau)$. Let $\bar{\mu}$ be the set of edges of μ for which neither end point belongs to $V(\tau)$, and let $\bar{\omega}$ be $\omega \cap \bar{\mu}$. Evidently, we can write

$$\nu = \lambda + \kappa, \tag{4.5}$$

where

$$\lambda = \nu \cap \tau, \quad \kappa = \nu \cap \bar{\mu}; \tag{4.6}$$

note that κ is a subgraph of $\bar{\omega}$. Let $G(\lambda)$ denote the subset of graphs in G which give rise to a specific λ. We shall prove (4.4) by showing that for every λ

$$\#[E(\omega; a, b) \cap G(\lambda)] \leq \#[F(\omega) \cap G(\lambda)] \tag{4.7}$$

and then summing over λ. Consider a λ for which the left side of (4.7) is nonzero, that is, there is some $\nu \in E(\omega; a, b) \cap G(\lambda)$. Since $\delta\nu = \{ab\}$, neither a nor b can be elements in $V(\tau)$, and since $V(\tau)$ and $V(\bar{\mu})$ have no elements in common, we conclude that $\delta\lambda = \phi$ and $\delta\kappa = \{ab\}$. Hence, κ is a member of $E(\bar{\omega}; a, b)$. Indeed, for *any* $\kappa \in E(\bar{\omega}; a, b)$, $\lambda + \kappa$ belongs to $E(\omega; a, b) \cap G(\lambda)$. Consequently the left side of (4.7) is equal to $\#[E(\bar{\omega}; a, b)]$. Since

$$\lambda \in F(\omega) \cap G(\lambda),$$

the right side of (4.7) is nonzero and an analogous argument shows that it must equal $\#[F(\bar{\omega})]$. Lemma 1 then implies that the left and right sides of (4.7) are equal. On the other hand, if λ is such that the left side of (4.7) vanishes, it is still possible for the right side to be positive, leading to an inequality. This observation completes the proof of Theorem 3.

V. THERMODYNAMIC INEQUALITIES FOR CRITICAL POINT INDICES

Certain inequalities based on thermodynamic arguments have been used to check the consistency of critical point indices derived from experimental data for fluids and magnetic systems or estimated on the basis of power series expansions. One assumes that thermodynamic variables of interest vary as some power law near the critical point; for example, on the critical isentrope $S = S_c$ (the subscript c denotes the value of a quantity at the critical point),

$$H \sim |M|^{\delta_s} \operatorname{sgn}(M), \tag{5.1a}$$

$$T - T_c \sim |M|^{\sigma}, \tag{5.1b}$$

whereas on the critical isotherm $T = T_c$, the variation of entropy is given by

$$S_c - S \sim |M|^{\zeta+1}. \tag{5.1c}$$

The other indices we shall employ, α, α', β, γ, γ', and δ, are defined in the usual way.[8,9]

Thermodynamic inequalities must be based on some hypotheses regarding the behavior of thermodynamic functions; the following three are of interest for the present discussion:

I. The thermodynamic free energy $F(H, T)$ is a concave function of the variables H and T together.

II. At constant $H > 0$, $M(H, T)$ is a nonincreasing function of T.

III. At constant T and for $H > 0$, $M(H, T)$ is a concave function of H.

Of these hypotheses, the first is of considerable generality, and can be shown to hold for most model magnetic systems (Ising, Heisenberg, ferro- or antiferromagnetic) because the Hamiltonian depends linearly on the field.[4] Among its consequences is the result that M must be monotone nondecreasing in H at fixed T. On the other hand, II and III would not be expected to hold except in ferromagnets and have been definitely established (in this and a preceding paper) only for Ising ferromagnets (of arbitrary spin[5]). It is also plausible that they should hold, for example, in Heisenberg ferromagnets, and the same series

expansion techniques used to estimate critical indices can also be employed to check the hypotheses. (We remark that additional inequalities[10] for Ising ferromagnets may be derived using the well-known result that zeros of the partition function lie on the unit circle in the complex $z = \exp(-\beta H)$ plane; we shall not discuss these here.)

Table I[11] lists thermodynamic inequalities for the indices mentioned above which have been established on the basis of Hypotheses I–III. (For details, see Ref. 1.) Note in particular that Hypothesis III appears to be necessary in order to obtain an inequality involving the "high temperature" ($T > T_c$) exponents α and γ. The inequalities in the table are independent of one another except that 1 may be obtained by combining 2 and 3. However, 1 can be established on the basis of considerably weaker assumptions than are needed for 3.

Of these inequalities 6 is exceptional in that it allows for two possibilities:

$$\text{(i)} \qquad \delta = \delta_s, \tag{5.2a}$$

$$\text{(ii)} \qquad \delta > \delta_s \geq \zeta + \sigma. \tag{5.2b}$$

In case (ii) one obtains, with the help of 4 and 7, the result

$$(1 - \alpha)(2 - \alpha')\delta_s \geq (2 - \alpha)(1 - \alpha')\delta + (\alpha - \alpha'), \tag{5.3}$$

which for $\alpha \geq \alpha'$ (both quantities are assumed nonnegative and cannot exceed 1) implies $\delta_s \geq \delta$, contrary to (5.2b). We conclude that case (ii) is only compatible with $\alpha < \alpha'$, and $\alpha \geq \alpha'$ implies that δ and δ_s are identical.

TABLE I. Inequalities for critical exponents.[a]

Inequality	Hypotheses[b]
1. $2\beta + \gamma' \geq 2 - \alpha'$	
2. $\beta(\delta + 1) \geq 2 - \alpha'$	
3. $\gamma' \geq \beta(\delta - 1)$	II, III
4. $(2 - \alpha')\zeta + 1 \geq (1 - \alpha')\delta$	II
5. $\delta \geq \delta_s$	
6. $\delta_s \geq \min\{\delta, \zeta + \sigma\}$	II
7. $(2 - \alpha)\sigma \geq \delta_s + 1$	II
8. $\gamma(\delta_s + 1) \geq (2 - \alpha)(\delta_s - 1)$	II, III

[a] Taken from Table II of Ref. 1.
[b] Hypothesis I (convexity) is understood in all cases.

To illustrate the inequalities, consider first the 2-dimensional Ising ferromagnet with nearest-neighbor interactions for which $\alpha = \alpha' = 0$ and $\beta = \frac{1}{8}$ are exact results.[8] We have, therefore, $\delta_s = \delta \geq 15$ by 2 and $\gamma' \geq \frac{7}{4}$, $\gamma \geq \frac{7}{4}$ by 3 and 8, respectively. In addition, there is independent evidence (see Ref. 8), which cannot as yet be regarded as entirely rigorous, that γ and γ' are both exactly $\frac{7}{4}$. If this is the case, then, from either 3 or 8, it follows that δ must be exactly 15.

For 3-dimensional Ising ferromagnets, no results are known exactly, but it is believed at present that[8] $\alpha = \frac{1}{8}$ and $\gamma = \frac{5}{4}$. Inserting these values in 8 yields $\delta_s \leq 5$ and, if we admit the possibility that α may be as large as 0.14 and γ as large as 1.253 (error estimates from Ref. 8), the maximum value of δ_s is still only 5.13. Therefore, if $\alpha' \leq \alpha$ (in accord with the estimate of $\frac{1}{61}$ for α'), we can conclude that $\delta = \delta_s$ is unlikely to be as large as 5.2 estimated from series expansions, and a value of 5, which accords better with scaling ideas, would seem preferable. It is, however, not impossible that α' exceeds α, especially since the former seems rather difficult to estimate accurately. In any case, it is evident that the thermodynamic inequalities pose nontrivial constraints on the choice of acceptable indices.

* Permanent address: Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.
† Supported in part by the Atomic Energy Commission.
‡ Supported by ARGC Contract 66/15220 and NSF GP 7946.
[1] R. B. Griffiths, J. Chem. Phys. **43**, 1958 (1965).
[2] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).
[3] D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).
[4] R. B. Griffiths, J. Math. Phys. **5**, 1215 (1964).
[5] R. B. Griffiths, J. Math. Phys. **10**, 1559 (1969).
[6] It is convenient to regard N as the vertex set of the graph ν. This is to be distinguished from the set of nonisolated vertices, vertices which are end points of one or more edges, for which we use the symbol $V(\nu)$.
[7] O. Ore, *Theory of Graphs* (American Mathematical Society, Providence, R.I., 1962), p. 24.
[8] Results on critical indices are summarized in the review article by M. E. Fisher, Rept. Progr. Phys. **30**, 615 (1967), and a convenient table giving definitions and theoretical values for various models will be found at the end of the article.
[9] We use the definition of α and α' given in Appendix A of Ref. 1, which results in the value 0 whenever the heat capacity has a finite limit at the critical point. In particular, *negative* values of α and α' are strictly excluded when dealing with the inequalities in Table I.
[10] G. A. Baker, Jr., Phys. Rev. Letters **20**, 990 (1968).
[11] G. A. Baker, Jr., and D. S. Gaunt [Phys. Rev. **155**, 545 (1967)] claim to have derived the third inequality in Table I with the opposite sign, $\gamma' \leq \beta(\delta - 1)$. Their derivation, however, involves an interchange of limits which they fail to justify and which in the case of a "homogeneous" equation of state [R. B. Griffiths, Phys. Rev. **158**, 176 (1967)] is, in general, incorrect.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Generalized Anharmonic Oscillator

CARL M. BENDER*†
Department of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 2 June 1969)

The generalized anharmonic oscillator is defined by the Hamiltonian H_N, which in the coordinate space representation is given by $H_N = -d^2/dx^2 + \frac{1}{4}x^2 + g(\frac{1}{2}x^2)^N$. The analytic properties of the energy levels of H_N as functions of complex coupling g are derived and described. Zeroth-order WKB techniques are used in the mathematical analysis. For all N, the results are qualitatively similar to those for the ordinary anharmonic oscillator in which $N = 2$ and, thus, the results are model independent for this wide class of models. The limiting case $N \to \infty$ is solved exactly without using WKB techniques. The exact solution agrees with the WKB solution to zeroth order. This agreement is most impressive and testifies to the accuracy and utility of WKB methods.

I. INTRODUCTION

In a recent paper[1] (hereafter referred to as BW), Bender and Wu used WKB techniques to elucidate the analytic structure of the energy levels of the anharmonic oscillator as a function of complex coupling constant λ. The anharmonic oscillator is described by the Hamiltonian

$$H = \tfrac{1}{2}\dot{\varphi}^2 + \tfrac{1}{2}m^2\varphi^2 + \lambda\varphi^4, \tag{1}$$

where

$$[\varphi, \dot{\varphi}] = i. \tag{2}$$

In BW it was shown that the energy levels for real λ, $E_0(\lambda)$, $E_1(\lambda)$, $\cdots$, are branches of a complex function $E(\lambda)$. $E(\lambda)$ has an infinite sequence of square-root-type branch point singularities where level crossing occurs. This sequence of branch points is located in the complex λ plane near $\arg \lambda = \pm 270°$ and has a limit point at the origin. These results explain the divergence of the Feynman perturbation series which is a power series in λ.[2]

In this paper we raise the following question: Do the unexpected and exciting results of BW depend on the $\lambda\varphi^4$ interaction or are they model independent? Specifically, if we generalize the interaction to $g\varphi^{2N}$, does $E(g)$ in this model still exhibit the same general analytic structure as $E(\lambda)$? The answer to this question is *yes*, indicating that the phenomena brought to light in BW are more fundamental and far reaching than one might have originally expected.

To analyze the $g\varphi^{2N}$ interaction, we have greatly extended the WKB techniques described in BW. Although we will rely somewhat on those results of BW which are easily extendible to the $g\varphi^{2N}$ case, we will strongly emphasize the powerful new techniques we have developed.

The generalized anharmonic oscillator Hamiltonian is

$$H_N = \tfrac{1}{2}\dot{\varphi}^2 + \tfrac{1}{2}m^2\varphi^2 + g\varphi^{2N}, \tag{3}$$

where φ obeys the commutation relation in Eq. (2). H_N in Eq. (3) reduces to the Hamiltonian analyzed in BW [Eq. (1)] when $N = 2$.

We transform to the coordinate space representation by substituting

$$\varphi = 2^{-\frac{1}{2}}x \tag{4}$$

and

$$\dot{\varphi} = -i2^{\frac{1}{2}}\frac{d}{dx}. \tag{5}$$

We also let $m = 1$ without loss of generality.

The energy eigenvalue equation in coordinate space is

$$\left\{\frac{-d^2}{dx^2} + \frac{x^2}{4} + g\frac{x^{2N}}{2^N} - E(g)\right\}\Phi(x) = 0, \tag{6a}$$

where

$$\lim_{x\to\pm\infty}\Phi(x) = 0. \tag{6b}$$

Equations (6) and (7) define the analytic properties of $E(g)$, and it is the purpose of this paper to study these equations. In Sec. II we set up the mathematical framework for doing the WKB analysis. Next, we present zeroth-order WKB calculations of the wavefunction in coordinate space for N small (Sec. III) and for N large (Sec. IV). In Sec. V we use the WKB wavefunction $\Phi(x)$ to determine the qualitative analytic properties of $E(g)$. In Sec. VI we consider the limit as $N \to \infty$.

II. MATHEMATICAL FRAMEWORK FOR WKB ANALYSIS

A. The Analytic Continuation of $E(g)$

The coordinate representation [Eq. (6)] implies that, for large real $|x|$ and real g,

$$\Phi(x) \sim \exp\left[\frac{-|x|^{N+1}g^{\frac{1}{2}}}{(N+1)2^{\frac{1}{2}N}}\right]. \tag{7}$$

Hence, the boundary condition in Eq. (6b) holds for complex $|x| \to \infty$ in the sector $|\arg x| < \pi/2(N+1)$, as well as on the real x axis.

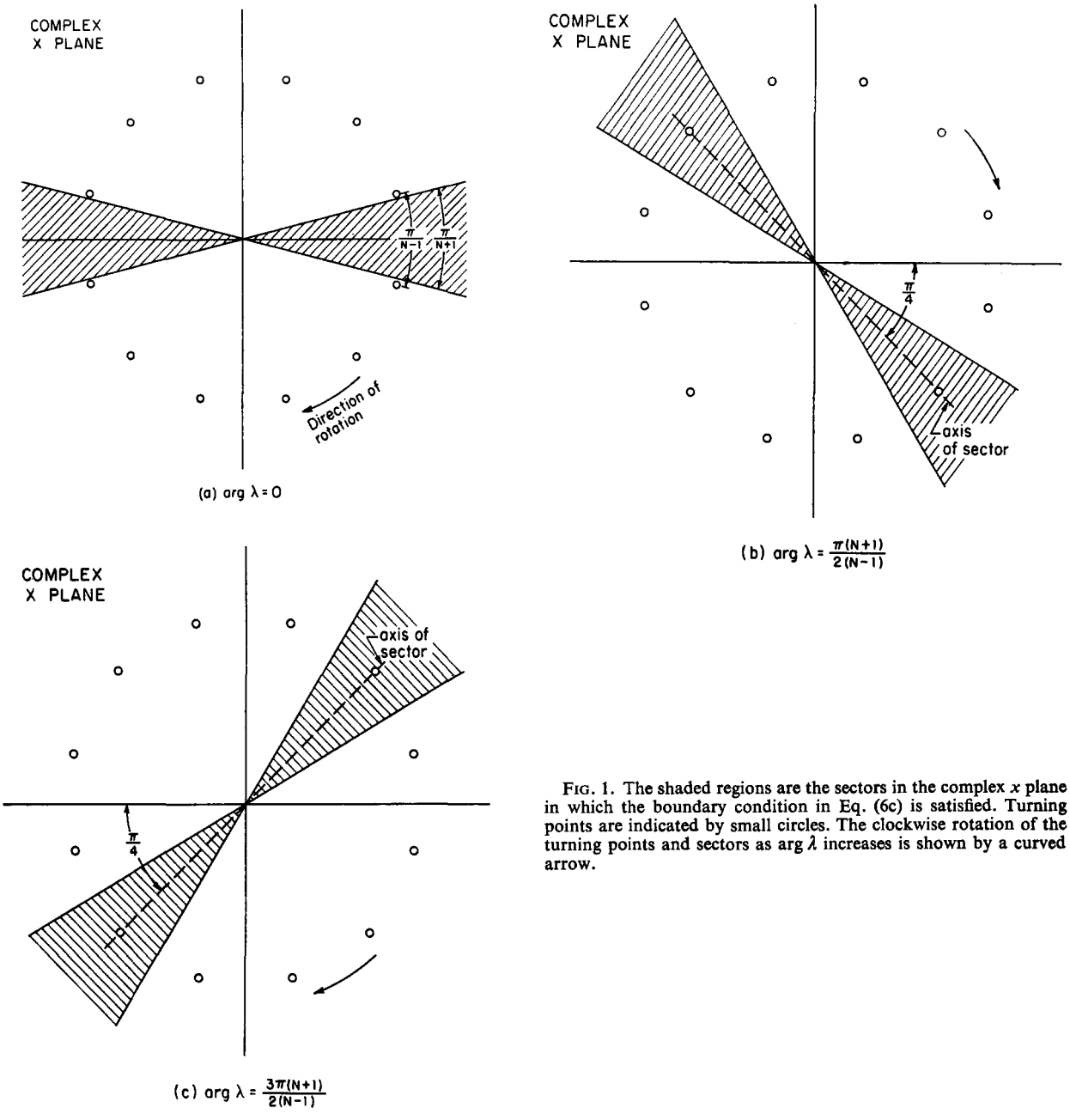


FIG. 1. The shaded regions are the sectors in the complex x plane in which the boundary condition in Eq. (6c) is satisfied. Turning points are indicated by small circles. The clockwise rotation of the turning points and sectors as arg λ increases is shown by a curved arrow.

Equation (6) allows us to continue $E(g)$ into the complex g plane. For complex g, we *define* $E(g)$ by Eq. (6a) and by the general boundary condition

$$|\arg(\pm x) + [2(N+1)]^{-1}\arg g| < \pi[2(N+1)]^{-1}. \tag{6c}$$

Given N and arg g, Eq. (6c) defines two sectors in the x plane centered about

$$\arg x = -[2(N+1)]^{-1}\arg g$$

and

$$\arg x = -[2(N+1)]^{-1}\arg g + \pi, \tag{8a}$$

each having an angular opening of

$$\theta = \pi/(N+1). \tag{8b}$$

These sectors are labeled on Fig. 1.

The analytic continuation $E(g)$ is defined for x in these sectors.

B. The Turning Points

The turning points for the zeroth-order WKB solution to Eq. (6a) are the solutions of

$$\tfrac{1}{4}x^2 + 2^{-N}gx^{2N} - E = 0. \tag{9}$$

It is convenient to let

$$g = \lambda^{N-1}. \tag{10}$$

Then, as in BW, we let $|\lambda|$ be small:

$$|\lambda| \ll 1. \tag{11a}$$

Moreover, we only consider low-lying energy levels:

$$|E| = O(1). \tag{11b}$$

Using the approximations in Eqs. (11), we solve Eq. (9) to zeroth order in λ. There are $2N$ turning points. Two of them lie near the origin at a radial distance of order 1:

$$x \sim \pm(2E)^{\frac{1}{2}}. \tag{12}$$

The rest of the turning points lie on a circle of radius $|\lambda|^{-\frac{1}{2}}$ centered about the origin:

$$x \sim \lambda^{-\frac{1}{2}} e^{\pi i/2(N-1)} 2^{N-2/2(N-1)} e^{\pi i m/N-1}, \quad m = 0, 1, \cdots, 2N-3. \tag{13}$$

The turning points in Eq. (13) are indicated by small circles on Fig. 1.

C. $\Phi(x)$ Near the Origin

As in BW, we approximate the wavefunction $\Phi(x)$ for x restricted to a sector in Eq. (8) by breaking up each sector into four regions. When x is near the origin (that is, when it is in region A), we approximate Eq. (6a) by

$$\left(-\frac{d^2}{dx^2} + \tfrac{1}{4}x^2 - E\right)\Phi(x) = 0. \tag{14}$$

Two independent solutions to this equation are the parabolic cylinder functions $D_{E-\frac{1}{2}}(\pm x)$ (see BW). For even- or odd-parity wavefunctions

$$\Phi_A(x) = C[D_{E-\frac{1}{2}}(x) \pm D_{E-\frac{1}{2}}(-x)]. \tag{15}$$

To zeroth order in λ, this solution is a good approximation to the wavefunction Φ when $|x| \ll |\lambda|^{-\frac{1}{2}}$. But, $|x|$ can range as large as, say, $|\lambda|^{-\frac{1}{4}}$. Thus, we can disregard the turning points in Eq. (12), even if they lie in the sector defined by Eq. (8).

D. Rotation of the Turning Points in Eq. (13)

When the turning points in Eq. (13) lie in the sector [Eq. (8)], they may *not* be disregarded. When *do* they lie in the sector?

When $\arg\lambda$ is near 0, the turning points lie outside and on either side of the sector [see Fig. 1(a)]. Thus, there are no obstructions and the parabolic cylinder function of Eq. (15) asymptotically connects to Eq. (7). This connection gives the same condition on E as in BW:

$$E = 2n + \tfrac{1}{2}, \quad n = 0, 1, 2, \cdots, \text{ for even-parity wavefunctions;}$$
$$E = 2n + \tfrac{3}{2}, \quad n = 0, 1, 2, \cdots, \text{ for odd-parity wavefunctions.} \tag{16}$$

As $\arg\lambda$ increases from 0, both the sectors and the turning points rotate clockwise [see Fig. 1(a)]. The turning points rotate slightly faster than the sectors. Thus, a turning point eventually enters the sector and blocks the path of the above connection, making Eq. (16) invalid. The turning point lies in the center of the sector when

$$\arg\lambda = \pi(N+1)/2(N-1), \quad \arg x = -\tfrac{1}{4}\pi, \tag{17}$$

as is shown in Fig. 1(b). This situation is analyzed in Secs. III and IV.

As $\arg\lambda$ continues to increase, the turning point leaves the sector and eventually another one enters. This second turning point is in the center of the sector when

$$\arg\lambda = 3\pi(N+1)/2(N-1), \quad \arg x = -\tfrac{3}{4}\pi. \tag{18}$$

This case is symmetrical to that in Eq. (17) [see Fig. 1(c)] and need not be treated separately.

When $\arg\lambda$ reaches $2\pi(N+1)/(N-1)$, parity invariance makes this configuration of turning points indistinguishable from that in Fig. 1(a). Since we recover Fig. 1(a) when $\arg g = 2\pi(N+1)$, $E(g)$ has an $(N+1)$-root branch-point singularity about the origin in the g plane. This is not an important feature of $E(g)$ and can be removed in perturbation theory by expanding in powers of g^{N+1}. This would not affect the divergence of perturbation theory which results from the singularities described in Sec. V.

E. The Configuration in Fig. 1(c)

It is useful to introduce new notation:

$$\rho = \lambda \exp[-i\pi(N+1)/2(N-1)], \tag{19a}$$
$$r = xe^{\frac{1}{4}i\pi} 2^{-(N-2)/2(N-1)}, \tag{19b}$$
$$\epsilon = iE(\lambda) 2^{N/(N-1)}, \tag{19c}$$
$$T = 2^{-1/(N-1)}. \tag{19d}$$

Equation (19) simplifies Eq. (6a) to

$$O = \left[\frac{d^2}{dr^2} + T^2(r^2 - \epsilon - \rho^{N-1} r^{2N})\right]\Phi(r). \tag{20}$$

Following BW, we express the locations of the two relevant turning points [see Fig. 1(c)],[3] which we now call r_0 and r_1:

$$r_0 \sim \epsilon^{\frac{1}{2}} = O(1) \tag{21a}$$

and

$$r_1 \sim \rho^{-\frac{1}{2}}. \tag{21b}$$

In terms of r_0 and r_1, in region A, $0 \leq |r| \ll |r_1|$; in region B, $|r_0| \ll |r| \ll |r_1|$; in region C, $|r| \lesssim |r_1|$; in region D, $|r| \sim |r_1|$.

III. WKB SOLUTION FOR "SMALL" N

In this section we find $\Phi(x)$ to zeroth order in λ, assuming N is "small." That is,

$$|\rho|\, N \ll 1. \tag{22}$$

In the next section we treat the case where N is "large." That is, $|\rho|\, N$ is not small compared to one, but rather[4]

$$N^{-1} \leq |\rho| \ll 1. \tag{23}$$

We approximate Φ in regions A, B, C, and D as follows:

Region A

Near the origin, Φ is given by Eq. (15).

Regions B and C

In these regions, Φ is given by the zeroth-order WKB solution (see BW)

$$\Phi_{\mathrm{WKB}} = (-\epsilon + r^2 - \rho^{N-1}r^{2N})^{-\frac{1}{4}} \times \left\{C_2 \exp\left[iT\int_{r_0}^{r}(-\epsilon + r^2 - \rho^{N-1}r^{2N})^{\frac{1}{2}}\,dr\right] + C_3 \exp\left[-iT\int_{r}^{r}(-\epsilon + r^2 - \rho^{N-1}r^{2N})^{\frac{1}{2}}\,dr\right]\right\}. \tag{24}$$

Region D

Near the turning point r_1, Eq. (20) reduces to the Airy equation

$$\left[\frac{d^2}{dR^2} + T^2(2N-2)\rho^{-\frac{1}{2}}R\right]\Phi_D(R) = 0, \tag{25}$$

after we substitute

$$R = r_1 - r. \tag{26}$$

It is essential in the derivation of Eq. (25) that N be "small." Otherwise, we could not have truncated the binomial expansion for $(1 - R/r_1)^{2N-2}$. When N is "large," the substitution in Eq. (26) is useless.

The solution to Eq. (25) is the Airy function

$$\Phi_D(R) = D(\tfrac{1}{3}y)^{\frac{1}{2}}K_{\frac{1}{3}}(2y^{\frac{3}{2}}), \tag{27}$$

where

$$y = -[9\sqrt{\rho}/T^2(2N-2)]^{-\frac{1}{3}}R. \tag{28}$$

Observe that, since $\Phi_D(R) \to 0$ asymptotically for $|r| \gg |r_1|$, the boundary condition in Eq. (6c) is already incorporated in Eq. (27).

Asymptotic Connection of Φ Across Regions A and B

Following BW, we asymptotically connect the even-parity case of Eq. (15) in region A to Eq. (24) in region B. The leading term in the asymptotic expansion of Eq. (15) is

$$\Phi_A(r) \sim C_1 r^{-\frac{1}{2}}\left(\frac{\exp[\frac{1}{4}i(2r^2T - \epsilon T\log r^2T - \frac{1}{2}\pi)]}{\Gamma(\frac{1}{4} - \frac{1}{4}i\epsilon T)} + \frac{\exp[-\frac{1}{4}i(2r^2T - \epsilon T\log r^2T - \frac{1}{2}\pi)]}{\Gamma(\frac{1}{4} + \frac{1}{4}i\epsilon T)}\right). \tag{29}$$

In region B, Eq. (24) behaves asymptotically as

$$\Phi_B(r) \sim r^{-\frac{1}{2}}\{C_2 \exp[\tfrac{1}{4}iT(2r^2 - \epsilon - \epsilon\log 4r^2/\epsilon)] + C_3 \exp[-\tfrac{1}{4}iT(2r^2 - \epsilon - \epsilon\log 4r^2/\epsilon)]\}. \tag{30}$$

Joining Eqs. (29) and (30) determines the ratio of C_2 to C_3:

$$\frac{C_2}{C_3} = \frac{\Gamma(\frac{1}{4} + \frac{1}{4}i\epsilon T)}{\Gamma(\frac{1}{4} - \frac{1}{4}i\epsilon T)} \times \exp\left(\tfrac{1}{2}iT\epsilon\log\frac{4}{T\epsilon} + \tfrac{1}{2}iT\epsilon - \tfrac{1}{4}i\pi\right). \tag{31}$$

The derivation of Eq. (31) is valid for both "large" and "small" N.

Asymptotic Connection of Φ Across Regions C and D

To find the asymptotic expansion of Eq. (24) in region C, we approximate the integral

$$\int_{r_0}^{r} dr(-\epsilon + r^2 - \rho^{N-1}r^{2N})^{\frac{1}{2}}$$

by decomposing it into two parts which we denote by A and $B(r)$:

$$A = \int_{r_0}^{r_1} dr(-\epsilon + r^2 - \rho^{N-1}r^{2N})^{\frac{1}{2}}, \tag{32}$$

$$B(r) = -\int_{r}^{r_1} dr(-\epsilon + r^2 - \rho^{N-1}r^{2N})^{\frac{1}{2}}. \tag{33}$$

We evaluate $B(r)$ by substituting Eq. (26) and by terminating binomial expansions. In terms of the new variable R,

$$B(R) \sim -r_1\int_0^R (2N-2)^{\frac{1}{2}}\rho^{\frac{1}{4}}R^{\frac{1}{2}} = -\rho^{-\frac{1}{4}}R^{\frac{3}{2}}\tfrac{2}{3}(2N-2)^{\frac{1}{2}}. \tag{34}$$

The evaluation of A in Eq. (32) is much more difficult. (In BW, where $N = 2$, A was given exactly in terms of elliptic functions. When $N > 2$, this method is no longer applicable.) Let $\bar{r} = (r_0r_1)^{\frac{1}{2}}$. Then,

$$A = A_1 + A_2, \tag{35}$$

where

$$A_1 = \int_{r_0}^{\bar{r}} dr(-\epsilon + r^2 - \rho^{N-1} r^{2N})^{\frac{1}{2}} \tag{36}$$

and

$$A_2 = \int_{\bar{r}}^{r_1} dr(-\epsilon + r^2 - \rho^{N-1} r^{2N})^{\frac{1}{2}}. \tag{37}$$

We approximate A_1 by

$$A_1 \sim \int_{r_0}^{\bar{r}} dr(r^2 - r_0^2)^{\frac{1}{2}} = \tfrac{1}{2}\bar{r}^2 - \tfrac{1}{2} r_0^2 \log(2\bar{r}/r_0) - \tfrac{1}{4} r_0^2. \tag{38}$$

Correspondingly, we approximate A_2 by substituting $x = r/r_1$ and by noting that[3] $r_1^{2N-2} \sim \rho^{1-N}(1 - \rho\epsilon)$:

$$\begin{aligned} A_2 &\sim (r_1 \rho^{\frac{1}{2}})^{N-1} r_1^2 \int_{\bar{r}/r_1}^1 x\, dx(1 - x^{2N-2})^{\frac{1}{2}} \\ &\qquad \times \left(1 - \frac{\epsilon}{r_1^2 x^2} \frac{1 - x^2}{1 - x^{2N-2}}\right)^{\frac{1}{2}} \\ &\sim (1 - \tfrac{1}{2}\rho\epsilon)\left[r_1^2 \int_{\bar{r}/r_1}^1 x\, dx(1 - x^{2N-2})^{\frac{1}{2}} \right. \\ &\qquad \left. - \tfrac{1}{2}\epsilon \int_{\bar{r}/r_1}^1 \frac{dx(1 - x^2)}{x(1 - x^{2N-2})^{\frac{1}{2}}} \right]. \end{aligned} \tag{39}$$

We have terminated the square-root expansion in Eq. (39) because we are only interested in approximations to zeroth order in λ.

To do the first integral in Eq. (39), we note that

$$r_1^2 \int_0^{\bar{r}/r_1} x\, dx(1 - x^{2N-2})^{\frac{1}{2}} \sim \tfrac{1}{2}\bar{r}^2.$$

Thus, letting $t = x^{2N-2}$,

$$\begin{aligned} r_1^2 &\int_{\bar{r}/r_1}^1 x\, dx(1 - x^{2N-2})^{\frac{1}{2}} \\ &\sim -\tfrac{1}{2}\bar{r}^2 + \frac{r_1^2}{2N-2} \int_0^1 t^{[1/(N-1)-1]}(1 - t)^{(\frac{3}{2}-1)}\, dt \\ &= -\tfrac{1}{2}\bar{r}^2 + \frac{r_1^2}{2N-2} \frac{\Gamma((N-1)^{-1})\Gamma(\frac{3}{2})}{\Gamma(\frac{3}{2} + (N-1)^{-1})} \\ &= \tfrac{1}{2}\bar{r}^2 + \frac{r_1^2}{N+1} \frac{\Gamma((N-1)^{-1}))\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2} + (N-1)^{-1})}. \end{aligned} \tag{40}$$

The substitution $t = x^{2N-2}$ is also needed in the second integral in Eq. (39), which is done as follows:

$$\begin{aligned} -\tfrac{1}{2}\epsilon &\int_{\bar{r}/r_1}^1 \frac{dx}{x} \frac{(1 - x^2)}{(1 - x^{2N-2})^{\frac{1}{2}}} \\ &= -\tfrac{1}{2}\epsilon \int_{(\bar{r}/r_1)^{2N-2}}^1 \frac{dt}{(2N-2)t} \frac{1 - t^{1/(N-1)}}{(1 - t)^{\frac{1}{2}}} \\ &\sim -\tfrac{1}{2}\epsilon \int_{(\bar{r}/r_1)^{2N-2}}^1 \frac{dt}{(2N-2)t(1 - t)^{\frac{1}{2}}} \\ &\quad + \tfrac{1}{2}\epsilon \int_0^1 \frac{dt}{(2N-2)} t^{[1/(N-1)-1]}(1 - t)^{\frac{1}{2}-1} \\ &\sim -\frac{\epsilon}{2(2N-2)} \log\left[4\left(\frac{r_1}{\bar{r}}\right)^{2N-2}\right] \\ &\quad + \frac{\epsilon\Gamma(\frac{1}{2})\Gamma((N-1)^{-1})}{2(2N-2)\Gamma(\frac{1}{2} + (N-1)^{-1})} \\ &= -\tfrac{1}{2}\epsilon \log\frac{r_1}{T\bar{r}} + \frac{\epsilon}{(2N-2)} \frac{\Gamma(\frac{3}{2})\Gamma((N-1)^{-1})}{\Gamma(\frac{1}{2} + (N-1)^{-1})}. \end{aligned} \tag{41}$$

Combining Eqs. (35), (39), (40), and (41) and using[3] $r_1^2 \sim \rho^{-1} - \epsilon(N-1)^{-1}$ and $r_0^2 \sim \epsilon$ gives

$$\begin{aligned} A = &-\tfrac{1}{4}\epsilon + \tfrac{1}{4}\epsilon \log \tfrac{1}{4} T^2 \rho\epsilon \\ &+ [\rho(N+1)]^{-1} \frac{\Gamma((N-1)^{-1})\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2} + (N-1)^{-1})}. \end{aligned} \tag{42}$$

Observe that, as one would expect, all dependence on $\bar{r}$ in Eq. (42) has cancelled.

The procedure we used to derive Eq. (42) is perfectly general as long as N is "small" [see Eq. (22)]. It may be used to evaluate A to any order in ρ; in each order the dependence on $\bar{r}$ cancels.

This completes the computation of the asymptotic expansion of Eq. (24) in region C. The result is

$$\begin{aligned} \Phi_C \sim R^{-\frac{1}{4}} \{ &C_2 \exp[iTA + iTB(r)] \\ &+ C_3 \exp[-iTA - iTB(r)] \}, \end{aligned} \tag{43}$$

where A and $B(r)$ are given in Eqs. (42) and (34).

Finding the asymptotic expansion of Φ_D in Eqs. (27) and (28) is straightforward and follows BW. The result is

$$\begin{aligned} \Phi_D \sim C_4 R^{-\frac{1}{4}} \{ &\exp[\tfrac{2}{3} i\rho^{-\frac{1}{4}} T(2N-2)^{\frac{1}{2}} R^{\frac{3}{2}} - \tfrac{1}{4} i\pi] \\ &+ \exp[-\tfrac{2}{3} i\rho^{-\frac{1}{4}} T(2N-2)^{\frac{1}{2}} R^{\frac{3}{2}} + \tfrac{1}{4} i\pi] \}. \end{aligned} \tag{44}$$

Joining Eqs. (43) and (44) gives a second determination of the ratio of C_2 to C_3:

$$\begin{aligned} \frac{C_2}{C_3} = \exp\Bigg[&\tfrac{1}{2} i\pi + \tfrac{1}{2} iT\epsilon - \tfrac{1}{2} iT\epsilon \log \tfrac{1}{4} T^2 \rho\epsilon \\ &- \frac{2iT}{(N+1)\rho} \frac{\Gamma((N-1)^{-1})\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2} + (N-1)^{-1})} \Bigg]. \end{aligned} \tag{45}$$

We re-emphasize that Eq. (45) is valid only when N is "small" [see Eq. (22)]. Otherwise, the evaluation of A in Eq. (32) and the determination of the differential equation (25) are meaningless.

Finally, we combine Eqs. (31) and (45). This gives

an implicit relation between E and ρ:

$$\frac{\Gamma(\frac{1}{4}+\frac{1}{2}E)}{\Gamma(\frac{1}{4}-\frac{1}{2}E)} = \exp\left[\tfrac{5}{4}\pi i - E\log T\rho + \frac{i}{(N+1)\rho}\frac{T\Gamma((N-1)^{-1})\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}+(N-1)^{-1})}\right]. \quad (46a)$$

Equation (46) is the result we have sought. It is the even-parity generalization[5] of Eq. (9a) in BW for "small" $N \geq 2$. The analogous equation for odd parity [corresponding to Eq. (9b) in BW] is

$$\frac{\Gamma(\frac{3}{4}+\frac{1}{2}E)}{\Gamma(\frac{3}{4}-\frac{1}{2}E)} = \exp\left[-\tfrac{5}{4}\pi i - E\log T\rho + \frac{i}{(N+1)\rho}\frac{T\Gamma((N-1)^{-1})\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}+(N-1)^{-1})}\right]. \quad (46b)$$

The implications of Eq. (46) are discussed in Sec. V.

IV. WKB SOLUTION FOR "LARGE" N

The asymptotic connection of Φ across regions A and B in Eq. (31) is valid for any N. However, the connection across regions C and D in Eq. (45) is only valid when N is "small." This is because the "small-N" approximation to $r^{2N} = (r_1 - R)^{2N}$, which involves the termination of a binomial expansion, becomes less accurate as N increases relative to r_1 and wrong when $N > r_1$. In this section, we rederive the "large-N" analog of Eq. (45).[6]

When r is near r_1 [that is, when R in Eq. (26) is small], we define

$$t \equiv 2RN/r_1 = [(r_1 - r)/r_1]2N. \quad (47)$$

This new variable is useful because it enables us to approximate the expression $-\epsilon + r^2 - \rho^{N-1}r^{2N}$ as follows:

Since r_1 is a root of $-\epsilon + r^2 - \rho^{N-1}r^{2N}$, we derive the identity

$$-\epsilon + r^2 - \rho^{N-1}r^{2N} = -\epsilon + r_1^2(1 - t/2N)^2 - (r_1^2 - \epsilon)(1 - t/2N)^{2N}. \quad (48)$$

But N is large and $t/2N$ is small. Thus,

$$-\epsilon + r^2 - \rho^{N-1}r^{2N} \sim (r_1^2 - \epsilon)(1 - e^{-t}). \quad (49)$$

We rely heavily on Eq. (49) to derive the asymptotic expansions of Φ in regions C and D.

Asymptotic Expansion of Φ in Region C

To compute the asymptotic behavior of Φ_C we must recalculate A in Eq. (32) and B in Eq. (33). By substituting Eq. (49), B becomes

$$B(t) \sim -\frac{r_1}{2N}(r_1^2 - \epsilon)^{\frac{1}{2}}\int_0^t dx\,(1 - e^{-x})^{\frac{1}{2}} = -\frac{r_1}{2N}(r_1^2 - \epsilon)^{\frac{1}{2}}\{2\log[1 + (1 - e^{-t})^{\frac{1}{2}}] + t - 2(1 - e^{-t})^{\frac{1}{2}}\}. \quad (50)$$

But, $t \gg 1$. Therefore,

$$B(t) \sim -r_1(2N)^{-1}(r_1^2 - \epsilon)^{\frac{1}{2}}(t + 2\log 2 - 2) \quad (51)$$

$$\sim -(2N\rho)^{-1}(1 - \tfrac{1}{2}\rho\epsilon)(t + 2\log 2 - 2). \quad (52)$$

In going from Eq. (51) to Eq. (52), we used the solution in Footnote 3.

To evaluate A when N is large, we let $r = x\rho^{-\frac{1}{2}}$ in Eq. (32). Then,

$$A \sim \rho^{-1}\int_{(\rho\epsilon)^{\frac{1}{2}}}^{1-\rho\epsilon/2N} dx(-\rho\epsilon + x^2 - x^{2N})^{\frac{1}{2}}. \quad (53)$$

The smallness of x^{2N} in the range of integration leads to an approximation of the integral in Eq. (53):

$$A \sim \rho^{-1}\int_{(\rho\epsilon)^{\frac{1}{2}}}^{1-\rho\epsilon/2N} dx(x^2 - \rho\epsilon)^{\frac{1}{2}} - (2\rho)^{-1}\int_{(\rho\epsilon)^{\frac{1}{2}}}^{1-\rho\epsilon/2N} dx\,\frac{x^{2N}}{(x^2 - \rho\epsilon)^{\frac{1}{2}}}. \quad (54)$$

The first term in Eq. (54) is a standard integral. The second term is evaluated by observing that $\rho\epsilon$ is small compared to x^2 near $x = 1$, which is the region of greatest contribution to the integral. Hence, the second term is approximately $-(4\rho N)^{-1}$. The final result of computing Eq. (54) is

$$A \sim \tfrac{1}{4}\epsilon\log\tfrac{1}{4}\rho\epsilon + \frac{2N-1}{4\rho N} - \frac{N\epsilon}{4(N-1)} \sim \tfrac{1}{4}\epsilon\log\tfrac{1}{4}\rho\epsilon + (2\rho)^{-1} - \tfrac{1}{4}\epsilon. \quad (55)$$

Observe that Eq. (55) could also be obtained by letting $N \to \infty$ in Eq. (42) (also, see Footnote 5).

Combining Eqs. (52) and (55) gives[6]

$$\Phi_C(t) \sim C_2\exp\{iT[(-t/2N\rho)(1 - \tfrac{1}{2}\rho\epsilon) - \tfrac{1}{4}\epsilon + \tfrac{1}{4}\epsilon\log\tfrac{1}{4}\rho\epsilon]\} + C_3\exp\{-iT[-(t/2N\rho) \times (1 - \tfrac{1}{2}\rho\epsilon) - \tfrac{1}{4}\epsilon + (2\rho)^{-1} + \tfrac{1}{4}\epsilon\log\tfrac{1}{4}\rho\epsilon]\}. \quad (56)$$

Asymptotic Expansion of Φ in Region D

Using Eqs. (47) and (49) to rewrite Eq. (20) in region D gives

$$\left\{\frac{d^2}{dt^2} + \left[\frac{T}{2N\rho}(1 - \tfrac{1}{2}\rho\epsilon)\right]^2(1 - e^{-t})\right\}\Phi_D(t) = 0. \quad (57)$$

Because we are considering N to be "large," Eq. (57) is completely different from its corresponding "small-N" equation (25).

To transform Eq. (57) into a more familiar form, we substitute

$$S = 2Fe^{-\frac{1}{2}t} \tag{58a}$$

and

$$F = (T/2N\rho)(1 - \tfrac{1}{2}\rho\epsilon), \tag{58b}$$

which yield

$$\left[S^2\frac{d^2}{dS^2} + S\frac{d}{dS} + (4F^2 - S^2)\right]\Phi_D(S) = 0. \tag{59}$$

Equation (59) is a modified Bessel equation (an ordinary Bessel equation with S replaced by iS), where $\nu^2 = -4F^2$. The solution to Eq. (59) is[7]

$$\Phi_D(t) = (\text{Const})K_{2Fi}(2Fe^{-\frac{1}{2}t}). \tag{60}$$

The argument of the modified Bessel function in Eq. (60) is very small. Therefore, the asymptotic behavior of $\Phi_D(t)$ is found by expanding Eq. (60) in a Taylor series about zero. To do this, we combine the following two formulas[8]:

$$K_\nu(z) = \tfrac{1}{2}\pi(\sin\nu\pi)^{-1}[I_{-\nu}(z) - I_\nu(z)], \tag{61a}$$

$$I_\nu(z) = \frac{(\frac{1}{2}z)^\nu e^{-z}}{\Gamma(\nu+1)}\,{}_1F_1(\nu + \tfrac{1}{2}; 2\nu + 1; 2z), \tag{61b}$$

$${}_1F_1(\nu + \tfrac{1}{2}; 2\nu + 1; 2z) = 1 + z + O(z^2) \tag{61c}$$

After some manipulation we obtain the asymptotic behavior of Φ in region D:

$$\Phi_D \sim (\text{const})\left\{\exp\left[\frac{-iT}{2N\rho}(1 - \tfrac{1}{2}\rho\epsilon) + \frac{iT}{N\rho}\log\frac{T}{2N\rho} + \frac{i\gamma T}{N\rho}\right] - \exp\left[\frac{itT}{2N\rho}(1 - \tfrac{1}{2}\rho\epsilon) - \frac{iT}{N\rho}\log\frac{T}{2N\rho} - \frac{i\gamma T}{N\rho}\right]\right\}. \tag{62}$$

Connection of Regions C and D

Joining Eq. (56) to Eq. (62) gives the connection formula

$$\frac{C_2}{C_3} = \exp\left[iT\left(\frac{2}{N\rho}\log\frac{T}{2N\rho} + \tfrac{1}{2}\epsilon - \rho^{-1} - \tfrac{1}{2}\epsilon\log\tfrac{1}{4}\rho\epsilon + \frac{2\gamma}{N\rho}\right) + i\pi\right]. \tag{63}$$

Equation (63) is the result of asymptotically joining Φ across regions C and D and is the "large-N" analog of Eq. (45).

Finally, we combine Eqs. (63) and (31) and get

$$\frac{\Gamma(\frac{1}{4} + \frac{1}{2}E)}{\Gamma(\frac{1}{4} - \frac{1}{2}E)} = \exp\left[-E\log\frac{\rho}{T} + \tfrac{3}{4}i\pi + \frac{iT}{\rho} - \frac{2iT}{N\rho}\log\frac{T}{2N\rho} - \frac{2\gamma iT}{N\rho}\right]. \tag{64a}$$

For odd parity, the corresponding result is

$$\frac{\Gamma(\frac{3}{4} + \frac{1}{2}E)}{\Gamma(\frac{3}{4} - \frac{1}{2}E)} = \exp\left[-E\log\frac{\rho}{T} - \tfrac{3}{4}i\pi + \frac{iT}{\rho} - \frac{2iT}{N\rho}\log\frac{T}{2N\rho} - \frac{2\gamma iT}{N\rho}\right]. \tag{64b}$$

Equation (64) is the "large-N" analog of Eq. (46). (γ is Euler's constant.)

V. QUALITATIVE ANALYTIC PROPERTIES OF $E(\rho)$

Following the analytic procedures of BW, we ascertain that, for arbitrary (but fixed) N: (a) *the only singularities of $E(\rho)$ in the ρ plane are square-root-type branch-point singularities* (except at the origin); (b) *a given value of $\rho \neq 0$ is a branch point if and only if*

$$\int\Phi^2(x)\,dx = 0. \tag{65}$$

Approximating $\Phi(x)$ in Eq. (65) by the zeroth-order WKB wavefunction determined in Sec. III gives

$$0 = -\psi(E + \tfrac{1}{2}) + \tfrac{1}{2}\pi\cot(\tfrac{1}{4} - \tfrac{1}{2}E)\pi + \log 2/T\rho, \tag{66a}$$

for even-parity wavefunctions, and

$$0 = -\psi(E + \tfrac{1}{2}) + \tfrac{1}{2}\pi\cot(\tfrac{3}{4} - \tfrac{1}{2}E)\pi + \log 2/T\rho, \tag{66b}$$

for odd-parity wavefunctions, where ψ is the logarithmic derivative of the Γ function.

Solving Eqs. (46)[9] and (66) simultaneously gives the approximate location in the ρ plane of the branch points of $E(\rho)$. These branch points are at

$$E = 2m + \tfrac{1}{2} + \eta,$$

$$\eta^{-1} = \log\frac{4\Gamma(\frac{1}{2})(M)(N+1)\Gamma(\frac{1}{2} + (N-1)^{-1})}{T^2\Gamma((N-1)^{-1})} + \gamma - \sum_{k=1}^{2m}k^{-1},$$

$$\rho^{-1} = \frac{2(N+1)\Gamma(\frac{1}{2} + (N-1)^{-1})}{T\Gamma((N-1)^{-1})}\left\{(M + \tfrac{1}{2}m - \tfrac{1}{8})\pi^{\frac{1}{2}} - \frac{i}{2\pi^{\frac{1}{2}}}\left[\log(2m)! + \tfrac{1}{2}\log\tfrac{1}{2}\pi - \log\log M - 1 - (2m + \tfrac{1}{2})\times\log\frac{4\Gamma(\frac{1}{2})(M)(N+1)\Gamma(\frac{1}{2} + (N-1)^{-1})}{T^2\Gamma((N-1)^{-1})}\right]\right\}, \tag{67a}$$

for even parity, and

$$E = 2m + \tfrac{3}{2} + \eta,$$

$$\eta^{-1} = \log \frac{4\Gamma(\frac{1}{2})(M+1)(N+1)\Gamma(\frac{1}{2}+(N-1)^{-1})}{T^2\Gamma((N-1)^{-1})} + \gamma - \sum_{k=1}^{2m+1} k^{-1},$$

$$\rho^{-1} = \frac{2(N+1)\Gamma(\frac{1}{2}+(N-1)^{-1})}{T\Gamma((N-1)^{-1})}\left\{[M+1+\tfrac{1}{2}m+\tfrac{1}{8}]\pi^{\frac{1}{2}} - \frac{i}{2\pi^{\frac{1}{2}}}\left[\log(2m+1)! + \tfrac{1}{2}\log\tfrac{1}{2}\pi - \log\log M - 1 - (2m+\tfrac{3}{2}) \times \log\frac{4\Gamma(\frac{1}{2})(M+1)(N+1)\Gamma(\frac{1}{2}+(N-1)^{-1})}{T^2\Gamma((N-1)^{-1})}\right]\right\}, \quad (67b)$$

for odd parity. In Eq. (67), γ is Euler's constant, M is a large positive integer, and m is a small positive integer.

Equation (67) is the main result of this paper and shows that, for any N, there is an infinite sequence of branch points approaching the origin. Further analysis (which is not given here because it has been published elsewhere[1]) reveals that $E(\lambda)$ has an infinitely-sheeted Riemann surface with the mth energy level corresponding to the mth Riemann sheet. Level crossing occurs where consecutive sheets are joined, namely at the branch points.

None of these qualitative analytic properties depend on the choice of N. (As N varies, the most that happens is that the locations of the branch points shift slightly.) Hence, these results are *model independent* for a large class of models. This fact strengthens the conjecture made in BW that these discoveries are characteristic of more complicated and realistic field theories and encourages us to continue these investigations. It is our hope that a deeper knowledge of WKB theory may be obtained by applying these methods to a sequence of models of increasing complexity and structure and that this, in turn, might lead to a greater understanding of perturbation theory, in particular, and field theory, in general.

VI. THE $N \to \infty$ LIMIT

The $N \to \infty$ case of the generalized anharmonic oscillator may be solved exactly. Comparing the solution with the "large-N" WKB result of Sec. IV is a stiff test of the mathematical methods we have used in this paper.

A. The WKB $N \to \infty$ Result

Letting $N \to \infty$ in Eq. (64) gives

$$\frac{\Gamma(\frac{1}{4}+\frac{1}{2}E)}{\Gamma(\frac{1}{4}-\frac{1}{2}E)} = \exp\left[\tfrac{3}{4}\pi i + \frac{i}{\rho} - E\log\rho\right], \quad (68a)$$

for even parity, and

$$\frac{\Gamma(\frac{3}{4}+\frac{1}{2}E)}{\Gamma(\frac{3}{4}-\frac{1}{2}E)} = \exp\left[-\tfrac{3}{4}\pi i + \frac{i}{\rho} - E\log\rho\right], \quad (68b)$$

for odd parity.

Observe that the limit as $N \to \infty$ of the "small-N" WKB result in Eq. (46) exists. However, this limiting equation differs slightly from Eq. (68); to wit, $\pm\frac{3}{4}\pi i$ is replaced by $\pm\frac{5}{4}\pi i$. Equation (68) is the correct equation because the limit as $N \to \infty$ is uniform (the choice of ρ does not depend on N). More precisely, when N is large, the slope of the potential near the turning point at r_1 (in region D) becomes extremely steep. [In the derivation of Eq. (64), as opposed to the derivation of Eq. (46), we took great care to approximate the steep part of the potential.] The effect of this steepness is illustrated when we take the limit as $N \to \infty$ and allow the walls of the potential to become infinitely steep. The potential then becomes a square well with a parabolic floor. It is well known that the WKB connection formula across a vertical wall differs by a phase factor of $e^{\frac{1}{2}i\pi}$ from the connection formula across a nonvertical wall. Hence, we would expect that the correct limiting equation would differ from the incorrect one by a total phase of $e^{\frac{1}{2}i\pi}$ (a factor of $e^{\frac{1}{4}i\pi}$ for each wall of the potential well). This *is* what we observe.

B. The Exact Solution

Substituting $g = \lambda^{N-1} = (i\rho)^{N-1}$ into Eq. (6a) gives

$$\left[\frac{-d^2}{dx^2} + \tfrac{1}{4}x^2 + \frac{1}{i\rho}(\tfrac{1}{2}i\rho x^2)^N - E\right]\Phi(x) = 0. \quad (69)$$

If we let $N \to \infty$, keeping $\rho \neq 0$, Eq. (69) becomes

$$\left[-\frac{d^2}{dx^2} + \tfrac{1}{4}x^2 - E\right]\Phi(x) = 0 \quad (70a)$$

with the associated boundary condition

$$\Phi\left[\pm\left(\frac{2}{i\rho}\right)^{\frac{1}{2}}\right] = 0. \quad (70b)$$

Equation (70a) is a parabolic-cylinder-function equation, and its solution is

$$\Phi(x) = (\text{Const})[D_{E-\frac{1}{2}}(x) \pm D_{E-\frac{1}{2}}(-x)], \quad (71)$$

for even or odd parity.

The energy eigenvalue equation follows from combining Eqs. (70b) and (71):

$$D_{E-\frac{1}{2}}[(2/i\rho)^{\frac{1}{2}}] \pm D_{E-\frac{1}{2}}[-(2/i\rho)^{\frac{1}{2}}] = 0, \qquad (72)$$

for even or odd parity.

Equation (72) is an *exact* implicit relation between E and ρ.

C. Comparison of Eqs. (68) and (72)

To rewrite Eq. (72) in a form amenable for comparison with Eq. (68), we observe that the arguments of the parabolic cylinder functions in Eq. (72) are large and thus expand $D_{E-\frac{1}{2}}$ into its asymptotic series. We give the full asymptotic series for $D_\nu(z)$ below[10]:

$$D_\nu(z) \sim z^\nu e^{-\frac{1}{4}z^2} \sum_{n=0}^{\infty} \frac{(-\frac{1}{2}\nu)_n(\frac{1}{2} - \frac{1}{2}\nu)_n}{n!\,(-\frac{1}{2}z^2)^n}, \qquad (73a)$$

for $-\frac{3}{4}\pi < \arg z < \frac{3}{4}\pi$, and

$$D_\nu(z) \sim z^\nu e^{-\frac{1}{4}z^2} \sum_{n=0}^{\infty} \frac{(-\frac{1}{2}\nu)_n(\frac{1}{2} - \frac{1}{2}\nu)_n}{n!\,(-\frac{1}{2}z^2)^n} - \frac{(2\pi)^{\frac{1}{2}}}{\Gamma(-\nu)} e^{\nu\pi i} z^{-(\nu+1)} e^{\frac{1}{4}z^2} \sum_{n=0}^{\infty} \frac{(1 + \frac{1}{2}\nu)_n(\frac{1}{2} + \frac{1}{2}\nu)_n}{n!\,(\frac{1}{2}z^2)^n}, \qquad (73b)$$

for $\frac{1}{4}\pi < \arg z < \frac{5}{4}\pi$.

Combining Eqs. (72) and (73) and doing some heavy manipulation gives

$$\frac{\Gamma(\frac{1}{4} + \frac{1}{2}E)}{\Gamma(\frac{1}{4} - \frac{1}{2}E)} = e^{\frac{3}{4}\pi i} e^{i/\rho} e^{-E\log\rho} g(\rho), \qquad (74a)$$

for even parity, and

$$\frac{\Gamma(\frac{3}{4} + \frac{1}{2}E)}{\Gamma(\frac{3}{4} - \frac{1}{2}E)} = e^{-\frac{3}{4}\pi i} e^{i/\rho} e^{-E\log\rho} g(\rho), \qquad (74b)$$

for odd parity, where

$$g(\rho) = \frac{\Gamma(\frac{1}{2} + E)\sum_{n=0}^{\infty} \frac{\Gamma(2n + \frac{1}{2} - E)}{n!}(-\frac{1}{4}i\rho)^n}{\Gamma(\frac{1}{2} - E)\sum_{n=0}^{\infty} \frac{\Gamma(2n + \frac{1}{2} + E)}{n!}(\frac{1}{4}i\rho)^n}. \qquad (74c)$$

Recall that Eq. (68) is valid to zeroth order in ρ. We note that to zeroth order in ρ, $g(\rho) = 1$.[11] Thus, Eqs. (68) and (72) agree. This agreement is most impressive and strongly attests to the usefulness and power of WKB techniques as an analytical tool in mathematical physics.

ACKNOWLEDGMENT

The author gives warm thanks to Professor T. T. Wu for many productive discussions.

* Supported by a National Science Foundation Predoctoral Fellowship

† Present address: Institute for Advanced Study, Princeton, N.J. 08540.

[1] C. M. Bender and T. T. Wu, Phys. Rev. Letters **21**, 406 (1968). The results of this paper are greatly expanded by Bender and Wu, Phys. Rev. **184**, 1231 (1969).

[2] The divergence of perturbation theory becomes more rapid as N increases.

[3] r_0 and r_1 on Fig. 1(c) are those roots of the term in parentheses in Eq. (20) which are real when ρ and ϵ are real. We can solve for these roots *exactly*:

$$r_0 = \epsilon^{\frac{1}{2}}\Big[1 + \omega + \omega^2\frac{4N-1}{2!} + \omega^3\frac{(6N-1)(6N-3)}{3!} + \omega^4\frac{(8N-1)(8N-3)(8N-5)}{4!} + \cdots\Big],$$

where $\omega = \frac{1}{2}\rho^{N-1}\epsilon^{N-1}$ and

$$r_1 = \rho^{-\frac{1}{2}}\Big[1 - \eta - \eta^2\frac{(2N+1)}{2!} - \eta^3\frac{(2N+3)(4N+1)}{3!} - \eta^4\frac{(2N+5)(4N+3)(6N+1)}{4!} - \cdots\Big],$$

with $\eta = \rho\epsilon/2(N-1)$.

Note that we have avoided the turning point at r_0 by approximating Φ by parabolic cylinder functions in region A [see Eq. (15)]. Nevertheless, as in BW, r_0 will be a useful quantity in Secs. III and IV. Note also that $r_0 \ll r_1$.

[4] For N both "small" [Eq. (22)] and "large" [Eq. (23)] the two power series in the previous footnote converge.

[5] Let $f(N) = \Gamma\Gamma((N-1)^{-1})\Gamma(\frac{1}{2})\Gamma^{-1}(\frac{1}{2} + (N-1)^{-1})$. Then, $f(2) = 1$, correctly reducing Eq. (46) to Eq. (9) in BW. Note also that $f(N)$ can be rewritten as

$$f(N) = (N-1)\Gamma^2\Big(1 + \frac{1}{N-1}\Big)2^{1/(N-1)}\Gamma^{-1}(1 + 2/(N-1)).$$

Then, using the Taylor series expansion about $z = 1$ of $\Gamma(z)$, we observe that as N increases, a good approximation to $f(N)$ is $(N-1)2^{1/(N-1)}$.

[6] In this section, we consider $N \gg 1$. We thus neglect terms of order one compared to N.

[7] In Eq. (58) [as in Eq. (27)], there is only one free constant. This is because the boundary condition [Eq. (6c)], which applies when $r \gg r_0$ (that is, when t is large and negative), only admits this one Bessel function.

[8] A. Erdélyi, W. Magnus, F. Oberhettenger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Co., New York, 1953), Vol. 2, Chap. VII.

[9] We use the "small-N" equation (46) because, given N, we are free to choose $|\rho|$ sufficiently small.

[10] The parabolic cylinder functions are discussed in Ref. 8, Chap. VIII. Equation (73b) is given incorrectly there and has been corrected in E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, London, 1962).

[11] $g(\rho)$ in Eq. (74c) is the ratio of two asymptotic (divergent) series. The functions that these series represent are convergent integrals, to wit,

$$g(\rho) = \frac{\Gamma(\frac{1}{2} + E)\int_0^\infty dt\, e^{-t}e^{-\frac{1}{2}t^2\rho}\, t^{-\frac{1}{2}-E}}{\Gamma(\frac{1}{2} - E)\int_0^\infty ds e^{-s}e^{\frac{1}{2}s^2\rho}\, s^{-\frac{1}{2}+E}}.$$

$g(\rho)$ can be extended into a single well-defined power series in ρ by formally inverting and multiplying the power series in Eq. (74c). This power series is also asymptotic (divergent).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On the Inverse Scattering Problem at Fixed Energy for Potentials Having Nonvanishing First Moments

JOSEPH R. COX
Department of Physics, Florida Atlantic University, Boca Raton, Florida
AND
KERRIS W. THOMPSON
Department of Mathematics and Department of Physics, Florida Atlantic University, Boca Raton, Florida

(Received 2 June 1969)

It has been shown previously by Newton that his solution (and its extension by Sabatier) of the problem of finding a central potential from a knowledge of all phase shifts at fixed energy yields a series whose expansion coefficients converge slowly unless the first moment of the potential vanishes. In particular, any truncation of the series after a finite number of terms necessarily results in potentials which have vanishing first moments. In this paper we propose a new, but formally somewhat similar, series for the potential which, for such truncations, does not suffer from this physically rather severe restriction. The series also furnishes new exact solutions of the Schrödinger equation at fixed energy. A closed-form expression for the scattering amplitude is obtained for a specific example. The problem of constructing the new series from the phase shifts is not discussed.

1. INTRODUCTION

The "inverse scattering problem at fixed energy" is the problem of finding a nonrelativistic central potential which will, for one energy, result in a prescribed set of phase shifts of all angular momenta. A general method of solving this problem has been given by Newton[1] and extended by Sabatier[2] and Newton.[3,4] However, because of the complexity of the method, considerable work remains to be done before it can be effectively utilized in the actual calculation of potentials. For example, quite apart from the problems associated with the lack of uniqueness of the potential which is to be determined by a given set of phase shifts[1-3,5] and which are present for whatever method one develops, one encounters the following practical difficulty in the employment of the method: In order to obtain a potential which has a nonvanishing first moment (which, of course, includes many potentials of physical interest), it is necessary to sum infinite series which depend on the interparticle distance r and which are not uniformly convergent in r.[3] Any attempt to truncate this series and thus approximate it by a finite sum results in a potential whose first moment vanishes.

The principle aim of this paper is to show that a different series than the one employed by Newton and generalized by Sabatier can be constructed which does in general yield, in the case of finite sums, potentials with nonvanishing first moments, and which thus lends hope that rapid convergence can be achieved in many instances for such potentials in the infinite-sum case. The series also generates new exact solutions to the Schrödinger equation at fixed energy. In this paper we will discuss finite sums only; however, as we will illustrate, these sums in general correspond to some of the infinite sums in the Newton–Sabatier formalism. The difficult problem of constructing the new series from the phase shifts is not discussed here, although its solution would of course be necessary before any complete formulation of the solution of the inverse problem alternative to the Newton–Sabatier formulation could be carried out.

In Sec. 2 some details of the derivation of Newton's series for the potential are reviewed and the resulting convergence difficulty of potentials with nonzero first moments is briefly discussed. In Sec. 3 a different but formally similar series is investigated which leads to potentials which, as is later demonstrated in Sec. 5, have nonzero first moments. While the initial steps are very similar, the analysis in the final stages is seen to be quite different. In Sec. 4 some illustrative examples easily obtainable from the new series are compared with corresponding examples obtained via the Newton–Sabatier series. For the cases considered, closed-form expressions for the scattering amplitudes are obtained by means of contour integrations in the complex angular momentum plane. Finally, In Sec. 5 the convergence properties of the Newton–Sabatier potentials and those of Sec. 3 are contrasted; in addition, the representation of a specific potential from Sec. 3 by a Newton–Sabatier series is discussed.

Since the energy is always fixed, we will choose the normalization $E = k^2 = 1$.

2. BRIEF REVIEW OF NEWTON'S METHOD

In his first paper[1] Newton begins with a given function $f(r, r')$, defined by the infinite series

$$f(r, r') = \sum_l c_l u_l(r) u_l(r'), \tag{2.1}$$

where the sum runs over the positive integers and zero,

c_l are real coefficients, and $u_l(r) \equiv rj_l(r)$ are the regular Ricatti–Bessel functions. Introducing the operator

$$D_0(r) \equiv r^2\left(\frac{\partial^2}{\partial r^2} + 1\right), \tag{2.2}$$

it is observed that $f(r, r')$ satisfies the partial differential equation

$$D_0(r)f(r, r') = D_0(r')f(r, r') \tag{2.3a}$$

and the boundary conditions

$$f(0, r') = f(r, 0) = 0. \tag{2.3b}$$

Next, a function $K(r, r')$ is defined by the equation

$$K(r, r') = f(r, r') - \int_0^r dr'' r''^{-2} K(r, r'')f(r'', r'), \tag{2.4}$$

which is an analog to the Gel'fand–Levitan equation.[6] Newton shows that, subject to the weak assumption that the c_l's in (2.1) increase at most by a fixed power of l for large l, Eq. (2.4) has a unique solution. Also, $K(r, r')$ satisfies the partial differential equation

$$\mathfrak{D}(r)K(r, r') = D_0(r')K(r, r') \tag{2.5a}$$

with the boundary condition

$$K(r, 0) = 0, \tag{2.5b}$$

where

$$\mathfrak{D}(r) \equiv D_0(r) - r^2\mathfrak{V}(r), \tag{2.6}$$

and

$$\mathfrak{V}(r) \equiv -2r^{-1}\frac{d}{dr}[r^{-1}K(r, r)]. \tag{2.7}$$

By means of Eqs. (2.5)–(2.7), it is then shown that the regular solution of the Schrödinger equation

$$\mathfrak{D}(r)\phi_l(r) = l(l + 1)\phi_l(r) \tag{2.8a}$$

satisfying the boundary condition

$$\lim_{r\to 0}[\phi_l(r)/u_l(r)] = 1 \tag{2.8b}$$

is given by

$$\phi_l(r) = u_l(r) - \int_0^r dr' r'^{-2} K(r, r')u_l(r') \tag{2.9}$$

and that $K(r, r')$ has the expansion

$$K(r, r') = \sum_l c_l\phi_l(r)u_l(r'), \tag{2.10}$$

where, just as in (2.1), the sum runs over the positive integers and zero.

By expressing the asymptotic behavior for large r of $\phi_l(r)$ in the form

$$\phi_l(r) \sim |f_l| \sin(r - \tfrac{1}{2}\pi l + \delta_l), \tag{2.11}$$

where $f_l = |f_l| e^{-i\delta_l}$ is the Jost function and δ_l is the phase shift, Newton obtains from (2.9) and (2.10) the "angular momentum dispersion relation"

$$f_l = 1 + \tfrac{1}{2}i\sum_{l'}\frac{e^{i\pi(l'-l)} - 1}{(l' - l)(l' + l + 1)}c_{l'}f_{l'}. \tag{2.12a}$$

From the real and imaginary parts of (2.12) multiplied by $e^{i\delta_l}$, he then obtains

$$\sin\delta_l = \sum_{l'}\Gamma^{(1)}_{ll'}c_{l'}|f_{l'}| \tag{2.13a}$$

and

$$|f_l| = \cos\delta_l + \sum_{l'}\Gamma^{(2)}_{ll'}c_{l'}|f_{l'}|, \tag{2.13b}$$

where $\Gamma^{(1)}_{ll'}$ and $\Gamma^{(2)}_{ll'}$ are functions of known form of δ_l, $\delta_{l'}$, l, and l'.

The manner in which the inverse problem is to be solved is then as follows: For a given set of phase shifts δ_l, the infinite set of equations (2.13a) yields the products $c_{l'}|f_{l'}|$, and these, when inserted in (2.13b), yield the $|f_l|$ and hence the c_l's. The c_l determine the function $f(r, r')$ according to Eq. (2.1) and thus, via equations (2.4) and (2.7), the potential $\mathfrak{V}(r)$.

The procedure is not unique,[1–3,5] and thus the investigation of those classes of potentials, all of which give rise to the same set of phase shifts, is of considerable interest. Sabatier[2] made the important observation that such classes of potentials can be greatly enlarged by including in the sum in (2.1) nonintegral values of l. Subsequently, Newton[3] showed that the additional angular momenta that appear in such a generalized expansion (2.1) or, equivalently, by Eqs. (2.7) and (2.10), in the expansion of the potential, are directly related to the singularities in l of f_l. Thus, if the sum in (2.10) is now replaced by a sum over a set Λ, so that by (2.7) we have

$$-\tfrac{1}{2}r\int_0^r dr'\mathfrak{V}(r')r' = \sum_{\lambda\in\Lambda}c_{\lambda-\frac{1}{2}}\phi_{\lambda-\frac{1}{2}}(r)u_{\lambda-\frac{1}{2}}(r), \tag{2.14}$$

then the set Λ consists of the set of points λ, where $f(-\lambda) \equiv f_{-\lambda-\frac{1}{2}}$ is singular, plus the set of all positive integers.[3]

The generalization of (2.12a) corresponding to (2.14) which gives the correct "dynamical interpolation for nonintegral l values is[3]

$$f(\lambda) = 1 + \tfrac{1}{2}i\sum_{\lambda'\in\Lambda}\frac{e^{i\pi(\lambda'-\lambda)} - 1}{\lambda'^2 - \lambda^2}c_{\lambda-\frac{1}{2}}f(\lambda'), \tag{2.12b}$$

from which it incidently follows that[3]

$$f(-n)/f(n) = 1 + (\pi/2n)c_{n-\frac{1}{2}}, \quad n = 1, 2, \cdots. \tag{2.15}$$

Finally, Newton[3] demonstrates that if the potential $\mathfrak{V}(r)$ in (2.14) has a nonvanishing first moment, i.e.,

$$\int_0^\infty r\mathfrak{V}(r)\,dr \neq 0, \tag{2.16}$$

then the series (2.14) must, unfortunately, converge very slowly, especially for large r [as is, in fact, readily evident from (2.14)], and that the c_l's cannot be absolutely summable. Also, any attempt to approximate (2.1) or (2.14) by a finite sum always results in a potential with vanishing first moment.

3. FINITE SERIES WHICH GIVE RISE TO POTENTIALS WITH NONVANISHING FIRST MOMENTS

Let S and T be two disjoint finite sets of N distinct real numbers chosen from the interval $(-\frac{1}{2}, \infty)$. Instead of Eq. (2.1) we begin with the function $g(r, r')$, defined by

$$g(r, r') = \sum_{l\in S} \gamma_l u_l(r_<) v_l(r_>), \tag{3.1a}$$

where γ_l are real coefficients, and $u_l(r)$ and $v_l(r)$ are, respectively, the regular Ricatti–Bessel functions $(\frac{1}{2}\pi r)^{\frac{1}{2}} J_{l+\frac{1}{2}}(r)$ [just as in Eq. (2.1)] and the irregular spherical Weber–Schläfli functions $(\frac{1}{2}\pi r)^{\frac{1}{2}} Y_{l+\frac{1}{2}}(r)$.[7] A more general starting point would be a function which is the sum of $g(r, r')$ of (3.1a) and $f(r, r')$ of (2.1). Since, however, it will turn out that, to have potentials with nonvanishing first moments with finite sums, it suffices to have (3.1a), we will, in the interest of clarity of presentation, omit terms of type (2.1).

The function $g(r, r')$ satisfies the partial differential equation

$$D_0(r)g(r, r') = D_0(r')g(r, r') \tag{3.2a}$$

and

$$g(0, r') = g(r, 0) = 0, \tag{3.2b}$$

just as $f(r, r')$ satisfies (2.3a) and (2.3b). Note that, in order to give meaning to (3.2a) at $r = r'$, it is necessary to associate the appropriate distribution functions with the second derivative $\partial^2 g(r, r')/\partial r^2$. The correctness of (3.2a) at $r = r'$ follows from the fact that (3.1a) is a sum of Green's functions. Alternatively, (3.1a) can be written in the form

$$g(r, r') \equiv A(r, r')\theta(r - r') + A(r', r)\theta(r' - r), \tag{3.1b}$$

where $A(r, r')$ is smooth at $r = r'$ and $\theta(x)$ is the step function. Upon substituting this expression into (3.2a), it is found that the resulting distribution functions $\delta(r - r')$ and $\delta'(r - r')$ are multiplied by functions which have, respectively, first- and second-order zeros at $r = r'$ and can accordingly be replaced by zero.

In analogy with (2.4) we define uniquely a function $L(r, r')$ by

$$L(r, r') = g(r, r') - \int_0^r dr'' r''^{-2} L(r, r'')g(r'', r'), \quad r \geq r', \tag{3.3a}$$

and

$$\lim_{r\to 0} r^{-\frac{1}{2}} |L(r', r)| = O(1), \tag{3.3b}$$

from which it follows, upon differentiating and integrating twice by parts and using (3.1b) and (3.2), that $L(r, r')$ satisfies the partial differential equation

$$D(r)L(r, r') = D_0(r')L(r, r'), \tag{3.4}$$

where

$$D(r) \equiv D_0(r) - r^2 V(r), \tag{3.5}$$

and

$$V(r) \equiv -\frac{2}{r}\frac{d}{dr}\left(\frac{L(r, r)}{r}\right). \tag{3.6}$$

The steps leading from (3.2) and (3.3) to (3.4)–(3.6) are identical to those of Newton in going from (2.3) and (2.4) to (2.5)–(2.7),[1] except that singularities arising from differentiation of (3.1b) now appear. However, it is precisely the inclusion of these distribution functions in the meaning of $\partial^2 g(r, r')/\partial r^2$ at $r = r'$ that allows one to integrate the expression

$$\int_0^r L(r, r'') \frac{\partial^2}{\partial r''^2} g(r'', r')\, dr''$$

twice by parts by the usual formula without special regard to the point $r'' = r'$. This can be checked with the aid of (3.1b) by breaking up the range of integration into two parts.

Next, in analogy with (2.9), we introduce the function

$$\bar{\phi}_l(r) = u_l(r) - \int_0^r dr' r'^{-2} L(r, r')u_l(r'), \quad l > -\tfrac{1}{2}, \tag{3.7}$$

from which it readily follows, upon use of (3.4), that $\bar{\phi}_l(r)$ satisfies the Schrödinger equation

$$D(r)\bar{\phi}_l(r) = l(l + 1)\bar{\phi}_l(r) \tag{3.8}$$

with potential $V(r)$ given by (3.6). The boundary conditions satisfied by $\bar{\phi}_l(r)$ will be found after we have solved (3.3).

In order to solve (3.3) we make the ansatz

$$L(r, r') = \sum_{L\in T} A_L(r)u_L(r'), \tag{3.9}$$

where, in addition to the $A_L(r)$, the set T [which, we recall, must by definition be disjoint with S of (3.1a)] is to be determined. Substituting (3.9) in (3.3) and using (3.1a), we obtain

$$\sum_{L\in T} A_L(r)u_L(r') = \sum_{l\in S} \gamma_l v_l(r)u_l(r') + \sum_{L\in T,\, l\in S} \frac{\gamma_l A_L(r)}{L(L + 1) - l(l + 1)} \times \{u_l(r')W[u_L(r), v_l(r)] - u_L(r')\}, \tag{3.10}$$

where $W[\alpha(x), \beta(x)] \equiv \alpha(x)\beta'(x) - \alpha'(x)\beta(x)$ is the Wronskian and where we have used the identity

$$\frac{d}{dr} W[u_l(r), v_L(r)] = [L(L+1) - l(l+1)]\frac{u_l(r)v_L(r)}{r^2}. \tag{3.11}$$

Since the functions $u_l(r')$, $l \in S \cup T$, are linearly independent, it follows from (3.10) that

$$\sum_{l \in S} [l(l+1) - L(L+1)]^{-1}\gamma_l = 1, \quad L \in T, \tag{3.12}$$

and

$$\sum_{L \in T} A_L(r) \frac{W[u_L(r), v_l(r)]}{l(l+1) - L(L+1)} = v_l(r), \quad l \in S. \tag{3.13}$$

Instead of determining the set T via (3.12) from a given set S and coefficients γ_l, we can equally well treat the γ_l as unknowns to be determined from (3.12) by a choice of the set T, as well as the set S. It is shown in Appendix A that if this is done, then (3.12) determines the γ_l uniquely. Thus, we shall assume from now on that the sets S and T are the given quantities; Eq. (3.13) then determines the $A_L(r)$ and hence, by (3.6), (3.7), and (3.9), the potential $V(r)$ and the solution $\bar{\phi}_l(r)$ of the Schrödinger equation.

In order to investigate the small-r behavior of $A_L(r)$ we note that, for small r,

$$\begin{aligned} u_l(r) &= [\pi^{\frac{1}{2}}/2^{l+1}\Gamma(l+\tfrac{3}{2})]r^{l+1} + O(r^{l+3}), \\ v_l(r) &= -[2^l\Gamma(l+\tfrac{1}{2})/\pi^{\frac{1}{2}}]r^{-l} + O(r^{-l+\epsilon}), \end{aligned} \tag{3.14}$$

where ϵ is such that $0 < \epsilon < \min(2, 2l+1)$.[8] Substituting (3.14) into (3.13) and rearranging, we have

$$\sum_{L \in T} \frac{r^L A_L(r)\sqrt{\pi}}{2^{L+1}\Gamma(L+\frac{3}{2})} [(L-l)^{-1} + O(r^\epsilon)] = 1 + O(r^\epsilon), \quad l \in S. \tag{3.15}$$

It is shown in Appendix A that a matrix with elements $(L-l)^{-1}$ is nonsingular. It follows that a matrix with elements $(L-l)^{-1} + O(r^\epsilon)$ is also nonsingular in some neighborhood of the origin. In the immediate vicinity of the origin it must have an inverse which differs from the inverse of the matrix with elements $(L-l)^{-1}$ by a matrix with elements of order r^ϵ. It then follows that, by inverting the set of equations (3.15), we have

$$r^L A_L(r) = a_L + O(r^\epsilon), \tag{3.16}$$

where a_L is a constant. We will show below that $a_L \neq 0$.

Consideration of the small-r behavior of (3.13) yields, upon use of (3.14) and (3.16),

$$\sum_{L \in T} b_L (L-l)^{-1} = 1, \quad l \in S, \tag{3.17}$$

where

$$b_L = a_L \pi^{\frac{1}{2}}[2^{L+1}\Gamma(L+\tfrac{3}{2})]^{-1}. \tag{3.18}$$

In order to find a_L, we introduce the function

$$u(l) \equiv 1 + \sum_{L \in T} b_L (l-L)^{-1}, \tag{3.19}$$

where l is arbitrary. Since (3.17) implies that

$$\mu(l) = 0, \quad l \in S, \tag{3.20}$$

and the definition (3.19) implies that

$$\mu(l) \to 1, \quad \text{as} \quad l \to \infty, \tag{3.21}$$

it follows that (3.19) can be written in the form

$$u(l) = \prod_{M \in S} (l-M) \Big/ \prod_{L \in T} (l-L). \tag{3.22}$$

According to (3.19), the residue of $\mu(l)$ at $l = L \in T$ is b_L, so from (3.18) and (3.22) we have

$$a_L = 2^{L+1}\Gamma(L+\tfrac{3}{2})\pi^{-\frac{1}{2}} \prod_{M \in S} (L-M) \Big/ \prod_{\substack{K \in T \\ K \neq L}} (L-K), \quad L \in T. \tag{3.23}$$

Since S and T are disjoint, it is apparent from (3.23) that, as was stated above, $a_L \neq 0$. Incidently, it then follows from (3.6), (3.9), (3.14), and (3.16) that the potential can be no more singular than r^{-1} at the origin.

Substituting (3.9) into (3.7) and using an identity similar to (3.11), we obtain

$$\bar{\phi}_l(r) = u_l(r) + \sum_{L \in T} A_L(r) \frac{W[u_L(r), u_l(r)]}{L(L+1) - l(l+1)}. \tag{3.24}$$

From (3.14), (3.16), and (3.23) it then follows that the behavior for small r of $\bar{\phi}_l(r)$ is

$$\bar{\phi}_l(r) = \tau(l) r^{l+1} + O(r^{l+2}), \tag{3.25}$$

where

$$\tau(l) = \pi^{\frac{1}{2}}[2^l\Gamma(l+\tfrac{3}{2})]^{-1} \times \left(1 - \sum_{L \in T} (L+l+1)^{-1} \prod_{M \in S} (L-M) \Big/ \prod_{\substack{K \in T \\ K \neq L}} (L-K)\right). \tag{3.26}$$

Thus, $\bar{\phi}_l(r)$ is a regular solution of (3.8), but differs from the regular solution $\phi_l(r)$ of (2.9) by an l-dependent factor. Comparing (3.26) with (3.19) and using (3.17), (3.18), and (3.23), we have

$$\tau(l) = \pi^{\frac{1}{2}}[2^l\Gamma(l+\tfrac{3}{2})]^{-1}\mu(-l-1), \tag{3.27a}$$

or, in virtue of (3.22),

$$\tau(l) = \pi^{\frac{1}{2}}[2^l\Gamma(l+\tfrac{3}{2})]^{-1} \times \prod_{M\in S}(l+M+1)\Big/\prod_{L\in T}(l+L+1). \quad (3.27b)$$

Consequently, the poles and zeros of $\tau(l)$ are all in the region $l < -\frac{1}{2}$.

Substituting the large-r asymptotic forms[7]

$$u_l(r) = \sin(r - \tfrac{1}{2}\pi l) + O(r^{-1}),$$
$$v_l(r) = -\cos(r - \tfrac{1}{2}\pi l) + O(r^{-1}) \quad (3.28)$$

in (3.13), we see that the behavior for large r of $A_L(r)$ is of the form

$$A_L(r) = B_L \cos(r - \tfrac{1}{2}\pi\beta_L) + O(r^{-1}). \quad (3.29)$$

According to (3.28), (3.29), and (3.13), B_L and β_L are given by

$$\sum_{L\in T} B_L \frac{\cos(r - \pi\tfrac{1}{2}\beta_L)\cos[\tfrac{1}{2}\pi(L-l)]}{L(L+1) - l(l+1)} = \cos(r - \tfrac{1}{2}\pi l), \quad l \in S, \quad (3.30)$$

or, since (3.30) is an identity in r,

$$\sum_{L\in T} \xi_L \frac{\cos[\pi(L-l)/2]}{L(L+1) - l(l+1)} = e^{+\frac{1}{2}i\pi l}, \quad l \in S, \quad (3.31)$$

where

$$\xi_L \equiv B_L e^{+\frac{1}{2}i\pi\beta_L}.$$

At this point a difficulty can arise. It may turn out that (3.31) has no (finite) solution ξ_L, $L \in T$. If, for example, S consisted only of even integers and the set T only of odd integers, then the left-hand side of (3.31) would vanish, whereas the right-hand side would not. This state of affairs might be due to a choice of S and T which would lead to a potential with a singularity in r on the positive real axis or perhaps a potential which does not have a physically acceptable form for large r. A somewhat similar difficulty also arises in connection with the Newton–Sabatier potentials.[2] In order to avoid such problems, we will henceforth restrict S and T by requiring that (3.31) have a (unique) solution.

Substituting the asymptotic forms (3.28) and (3.29) in (3.24), we have

$$\bar{\phi}_l(r) = \sin(r - \tfrac{1}{2}\pi l) + \sum_{L\in T} B_L \times \frac{\cos(r - \tfrac{1}{2}\pi\beta_L)\sin[\tfrac{1}{2}\pi(l-L)]}{L(L+1) - l(l+1)} + O\left(\frac{1}{r}\right), \quad (3.32)$$

where B_L and β_L are given by (3.30) or (3.31). Comparing (3.32) with (2.11) and noting that $\bar{\phi}_l(r) = \mu(-l-1)\phi_l(r)$, where $\mu(l)$ is given by (3.22), we obtain for the Jost function $f_l = |f_l|\, e^{-i\delta_l}$ the expression

$$f_l = \prod_{L\in T}(l+L+1)\Big/\prod_{M\in S}(l+M+1) \times \left(1 + i\sum_{L\in T}\xi_L e^{-i\pi l/2}\frac{\sin[\tfrac{1}{2}\pi(L-l)]}{L(L+1) - l(l+1)}\right). \quad (3.33)$$

Thus, the Jost function is given explicitly by (3.31) and (3.33), once a choice of the sets S and T has been made. The phase shifts are simply the negatives of the arguments of (3.33) for integer l. The potential giving rise to the phase shifts can be calculated explicitly from (3.13), (3.9), and (3.6), as can the wavefunction $\bar{\phi}_l(r)$ from (3.13) and (3.24). It will be shown in Sec. 5 that all such potentials have in general nonvanishing first moments.

4. ONE-TERM EXAMPLES

In the special case that the sum in Newton's series (2.1) reduces to a single term, the resulting scattering amplitude can be found in exact closed form.[2] We will show in this section that the corresponding single-term case for our series (3.1a) also yields an exact (but different) closed-form scattering amplitude. The method which we use is different and perhaps more general than that originally used by Sabatier.[2] One advantage of the method used below is that Sabatier's result for one term in (2.1) corresponding to integer l is readily extended to include noninteger l values.

When the right-hand sides of (2.1) and (3.1a) reduce to a single term and we relax the restriction in (2.1) that L be zero or a positive integer, we have, respectively,

$$f(r, r') = c_L u_L(r) u_L(r') \quad (4.1)$$

and

$$g(r, r') = \gamma_M u_M(r_<) v_M(r_>), \quad (4.2)$$

where L and M are real numbers satisfying $L > -\frac{1}{2}$ and $M > -\frac{1}{2}$. We will now compare the results which follow from (4.1) with those which follow from (4.2).

For the input (4.1) Eq. (2.12b) reduces to

$$f_l = 1 + i\left(\frac{e^{i\pi(L-l)} - 1}{2(L-l)(L+l+1)}\right)c_L f_L. \quad (4.3)$$

[Recall that $f(l+\frac{1}{2}) \equiv f_l$.] Letting $l \to L$ in (4.3) yields

$$[1 + \pi c_L(4L+2)^{-1}]f_L = 1, \quad (4.4)$$

which, together with (4.3), implies that

$$f_l = 1 + ia[e^{i\pi(L-l)} - 1][2(L-l)(L+l+1)]^{-1}, \quad (4.5)$$

where

$$a \equiv c_L[1 + \pi c_L(4L + 2)^{-1}]^{-1}. \tag{4.6}$$

The partial-wave amplitude $a_l = (1/2i)[e^{2i\delta_l} - 1]$ is thus given by

$$a_l = \frac{a[1 - \cos \pi(L - l)]}{2(L - l)(L + l + 1) + ia[e^{i\pi(L-l)} - 1]}. \tag{4.7}$$

Thus, for integer values of l, (4.7) yields

$$a_l = \begin{cases} -a[\cos \pi L - 1]\{2(L - l)(L + l + 1) \\ \qquad + ia[e^{i\pi L} - 1]\}^{-1}, & l \text{ even}, \\ +a[\cos \pi L + 1]\{2(L - l)(L + l + 1) \\ \qquad - ia[e^{i\pi L} + 1]\}^{-1}, & l \text{ odd}. \end{cases} \tag{4.8}$$

The scattering amplitude $f(\theta)$ is given by

$$f(\theta) = \sum_{l=0}^{\infty} (2l + 1)a_l P_l(\cos \theta). \tag{4.9}$$

It is convenient to separate (4.9) into two parts and write

$$f(\theta) = f_1(\theta) + f_2(\theta), \tag{4.10}$$

where

$$f_1(\theta) = \sum_{\substack{l=0 \\ (l \text{ even})}}^{\infty} (2l + 1)a_l P_l(\cos \theta) \tag{4.11a}$$

and

$$f_2(\theta) = \sum_{\substack{l=1 \\ (l \text{ odd})}}^{\infty} (2l + 1)a_l P_l(\cos \theta), \tag{4.11b}$$

the sums in (4.11a) and (4.11b) being over only even l and odd l, respectively. When a_l is of the form (4.8), $f_1(\theta)$ is a sum of type (B1) and $f_2(\theta)$ is a sum of type (B2) of Appendix B. It follows from the results of this appendix that

$$f_1(\theta) = \frac{-a\pi}{4 \cos \pi b_1}(1 - \cos \pi L) \times [P_{b_1-\frac{1}{2}}(\cos \theta) + P_{b_1-\frac{1}{2}}(-\cos \theta)] \tag{4.12}$$

and that

$$f_2(\theta) = \frac{+a\pi}{4 \cos \pi b_2}(1 + \cos \pi L) \times [P_{b_2-\frac{1}{2}}(\cos \theta) - P_{b_2-\frac{1}{2}}(-\cos \theta)], \tag{4.13}$$

where

$$b_1^2 = L(L + 1) + \tfrac{1}{2}ia[e^{i\pi L} - 1] + \tfrac{1}{4} \tag{4.14}$$

and

$$b_2^2 = L(L + 1) - \tfrac{1}{2}ia[e^{i\pi L} + 1] + \tfrac{1}{4}. \tag{4.15}$$

Although the validity of (4.12) and (4.13) is not dependent on the choice of sign of the real parts of b_1 and b_2, we may, for definiteness, require that Re $b_1 > 0$ and Re $b_2 > 0$.

The special choices of L an even integer or L an odd integer imply $f_1(\theta) \equiv 0$ or $f_2(\theta) \equiv 0$, respectively, and so we have

$$\begin{aligned} f(\theta) &= \frac{+a\pi}{2 \cos \pi b_2}[P_{b_2-\frac{1}{2}}(\cos \theta) \\ &\quad - P_{b_2-\frac{1}{2}}(-\cos \theta)], \quad L \text{ even}, \\ &= \frac{-a\pi}{2 \cos \pi b_1}[P_{b_1-\frac{1}{2}}(\cos \theta) \\ &\quad + P_{b_1-\frac{1}{2}}(-\cos \theta)], \quad L \text{ odd}, \end{aligned} \tag{4.16}$$

where b_1 and b_2 now reduce to

$$b_1^2 = (L + \tfrac{1}{2})^2 - ia, \quad L \text{ odd}, \tag{4.17}$$

and

$$b_2^2 = (L + \tfrac{1}{2})^2 - ia, \quad L \text{ even}. \tag{4.18}$$

The special cases (4.16) have been previously derived by Sabatier.[2,9]

The Jost function corresponding to (4.2) is given by (3.33); it is

$$f_l = \left(\frac{l + L + 1}{l + M + 1}\right) \times \left(1 + i\xi \frac{e^{-\frac{1}{2}i\pi l} \sin [\frac{1}{2}\pi(L - l)]}{(L - l)(L + l + 1)}\right), \tag{4.19}$$

where, from (3.31),

$$\xi \cos [\tfrac{1}{2}\pi(L - M)] = e^{\frac{1}{2}i\pi M}(L - M)(L + M + 1), \quad |L - M| \neq \text{odd integer}. \tag{4.20}$$

Also, according to (3.12),

$$\gamma_M = (M - L)(L + M + 1), \tag{4.21}$$

so (4.21) defines γ_M for given numbers M and L. Eliminating ξ from (4.19) and (4.20), we obtain

$$f_l = [(l + M + 1)(L - l)]^{-1}[(l + L + 1)(L - l) + ibe^{+\frac{1}{2}i\pi(M-l)} \sin [\tfrac{1}{2}\pi(L - l)]], \tag{4.22}$$

where

$$b \equiv (L - M)(L + M + 1)/\cos [\tfrac{1}{2}\pi(L - M)]. \tag{4.23}$$

The partial-wave amplitude $\tilde{a}_l$ obtained from (4.22) is

$$\tilde{a}_l = \frac{b \cos [\frac{1}{2}\pi(l - M)] \sin [\frac{1}{2}\pi(l - L)]}{(L + l + 1)(L - l) - ibe^{\frac{1}{2}i\pi(M-l)} \sin [\frac{1}{2}\pi(l - L)]}, \tag{4.24}$$

which, for integer l, becomes

$$\begin{aligned} \tilde{a}_l &= \frac{-b \cos [\frac{1}{2}\pi M] \sin [\frac{1}{2}\pi L]}{L(L + 1) + ibe^{\frac{1}{2}i\pi M} \sin [\frac{1}{2}\pi L] - l(l + 1)}, \quad l \text{ even}, \\ &= \frac{+b \sin [\frac{1}{2}\pi M] \cos [\frac{1}{2}\pi L]}{L(L + 1) - be^{\frac{1}{2}i\pi M} \cos [\frac{1}{2}\pi L] - l(l + 1)}, \quad l \text{ odd}. \end{aligned} \tag{4.25}$$

The scattering amplitude $g(\theta)$ corresponding to (4.25) is then

$$g(\theta) = \sum_{l=0}^{\infty} (2l + 1)\tilde{a}_l P_l(\cos\theta). \tag{4.26}$$

Comparing the structure of (4.25) with that of (4.8), we see immediately that (4.26) may be summed by the same method that was used to sum (4.9); the result is

$$g(\theta) = g_1(\theta) + g_2(\theta), \tag{4.27}$$

where, from (B12), we have

$$g_1(\theta) = \frac{\pi b}{2\cos(\pi c_1)} \cos(\tfrac{1}{2}\pi M)\sin(\tfrac{1}{2}\pi L) \times [P_{c_1-\frac{1}{2}}(\cos\theta) - P_{c_1-\frac{1}{2}}(-\cos\theta)], \tag{4.28}$$

and, from (B13),

$$g_2(\theta) = \frac{\pi b}{2\cos(\pi c_2)} \sin(\tfrac{1}{2}\pi M)\cos(\tfrac{1}{2}\pi L) \times [P_{c_2-\frac{1}{2}}(\cos\theta) - P_{c_2-\frac{1}{2}}(-\cos\theta)], \tag{4.29}$$

and where c_1 and c_2 are given (to within a sign) by

$$c_1^2 = L(L + 1) + ibe^{\frac{1}{2}i\pi M}\sin(\tfrac{1}{2}\pi L) + \tfrac{1}{4} \tag{4.30}$$

and

$$c_2^2 = L(L + 1) - be^{\frac{1}{2}i\pi M}\cos(\tfrac{1}{2}\pi L) + \tfrac{1}{4}. \tag{4.31}$$

Special care is necessary in interpreting (3.28) and (3.29) when L is chosen to be an integer. In this case, $c_1 = \pm(L + \frac{1}{2})$ if L is even and $c_2 = \pm(L + \frac{1}{2})$ if L is odd; as a result, (4.28) and (4.29) become indeterminate for L even and L odd, respectively, and it is necessary to return to (4.24) to evaluate a_L when L is integer. We find

$$\tilde{a}_L = -\tfrac{1}{2}\pi b\cos[\tfrac{1}{2}\pi(L - M)] \times [2L + 1 + \tfrac{1}{2}i\pi be^{i\pi(M-L)}]^{-1}. \tag{4.32}$$

This is the only term which contributes to the indeterminate expression (4.28) or (4.29). Hence, for L an integer we have, in place of (4.28) and (4.29),

$$g_1(\theta) = -\pi b\cos[\tfrac{1}{2}\pi(L - M)]P_L(\cos\theta) \times \left[1 + \frac{i\pi b}{4L + 2} e^{\frac{1}{2}i\pi(M-L)}\right]^{-1}, \quad L \text{ even}, \tag{4.33}$$

and

$$g_2(\theta) = -\pi b\cos[\tfrac{1}{2}\pi(L - M)]P_L(\cos\theta) \times \left[1 + \frac{i\pi b}{4L + 2} e^{\frac{1}{2}i\pi(M-L)}\right]^{-1}, \quad L \text{ odd}. \tag{4.34}$$

We note from (4.28) and (4.29) that there is also a simplification if M is chosen to be an integer. For noninteger L, integer M implies $g_2(\theta) \equiv 0$ if M is even and $g_1(\theta) \equiv 0$ if M is odd. According to (4.33) and (4.34), this statement is also true for integer L, since in (4.20) we are assuming that $|L - M|$ cannot equal an odd integer.

The scattering amplitudes (4.10) and (4.28) which result from (4.1) and (4.2), respectively, are very similar in structure; each is of the general form

$$E_1[P_{\kappa_1}(\cos\theta) - P_{\kappa_1}(-\cos\theta)] + E_2[P_{\kappa_2}(\cos\theta) + P_{\kappa_2}(-\cos\theta)], \tag{4.35}$$

where E_1, E_2, κ_1, and κ_2 are complex constants. Although, as we will show in the next section, the potential which gives rise to (4.1) is different than the potential which gives rise to (4.2), the similar structure (4.35) of $f(\theta)$ and $g(\theta)$ suggests that for some choices of parameters they might be identical. We will now demonstrate that this can never be the case.

Suppose that Eqs. (4.12)–(4.15) with L replaced by the symbol K are identical with Eqs. (4.28)–(4.31), respectively. Then we must have

$$a(1 - \cos\pi K) = -2b\cos\tfrac{1}{2}\pi M\sin\tfrac{1}{2}\pi L, \tag{4.36}$$

$$a(1 + \cos\pi K) = 2b\sin\tfrac{1}{2}\pi M\cos\tfrac{1}{2}\pi L, \tag{4.37}$$

$$K(K + 1) + \tfrac{1}{2}ia(e^{i\pi K} - 1) = L(L + 1) + ibe^{\frac{1}{2}i\pi M}\sin\tfrac{1}{2}\pi L, \tag{4.38}$$

and

$$K(K + 1) - \tfrac{1}{2}ia(e^{i\pi K} + 1) = L(L + 1) - be^{\frac{1}{2}i\pi M}\cos\tfrac{1}{2}\pi L, \tag{4.39}$$

where b is given by (4.23) and is thus determined once L and M are given, and, according to (4.6) with L replaced by K, a is a real, but otherwise free, parameter which depends on c_K. From (4.36) and (4.37) it follows that

$$a = b\sin\tfrac{1}{2}\pi(M - L) \tag{4.40}$$

and from (4.38) and (4.39) it follows that

$$iae^{i\pi K} = be^{i\frac{1}{2}\pi(L+M)}. \tag{4.41}$$

Thus, from (4.41) we conclude that $|a| = |b|$, and hence from (4.40) that $|M - L|$ is an odd integer. However, this violates (4.20). Thus, $f(\theta)$ and $g(\theta)$ as given by (4.10) and (4.28) can never be identical.

5. THE FIRST MOMENTS OF THE POTENTIALS OF SEC. 3

As was mentioned in Sec. 2 and as is evident from (2.14) in the limit as $r \to \infty$, when (2.16) holds, then the sum on the right-hand side of (2.16) must converge very slowly, and indeed if the sum is finite, then

the potential $\mathcal{U}(r)$ must necessarily have zero first moment. Let us compare this latter result with the corresponding result for a potential $V(r)$ resulting from the finite sum (3.1a) of Sec. 3. From (3.6) and (3.9) we find, with the aid of (3.14), (3.16), (3.23), (3.28), and (3.29), that

$$\int_0^\infty rV(r)\,dr = 2\sum_{L\in T}\left[\prod_{M\in S}(L-M)\Big/\prod_{\substack{K\in T\\K\neq L}}(L-K)\right]. \tag{5.1}$$

Since sets S and T are disjoint, each term in the sum on the right-hand side of (5.1) is nonzero; and since S and T are otherwise arbitrary, the sum itself will in general be nonzero. Hence, in contrast to the potentials given by (2.16) which must necessarily for finite sums have vanishing first moments, the potentials given by finite sum (3.1a) do not in general have vanishing first moments. In particular, the one-term example resulting from (4.2) gives a potential with only one term in the sum on the right-hand side of (5.1); thus, this potential always has a nonvanishing first moment. [It follows, incidently, from this that the potentials which give rise to $f(\theta)$ and $g(\theta)$ of Sec. 4 can never be the same, since they have, respectively, vanishing and nonvanishing first moments.]

Let us suppose that f_l as given by (3.33) arises from a potential such that the right-hand side of (5.1) is nonzero. Then, if this Jost function f_l satisfies Newton's dispersion relation (2.12b), it follows that the sum on the right-hand side of (2.12b) and hence on the right-hand side of (2.14) must be an infinite sum. Thus, finite sums of the type (3.1a) in general correspond to infinite sums of the type (2.1), where the summation is now over the infinite set Λ.

We will now show that f_l as given by (3.33) always does, in fact, satisfy the dispersion relation (2.12b). Newton[3] has shown that in order for a Jost function to satisfy (2.12b) it is sufficient that the functions $g_1(\lambda^2)$ and $g_2(\lambda^2)$, defined by

$$g_1(\lambda^2) \equiv \tfrac{1}{2}(1 - i\cot\pi\lambda)f(\lambda) + \tfrac{1}{2}(1 + i\cot\pi\lambda)f(-\lambda) \tag{5.2}$$

and

$$g_2(\lambda^2) \equiv \tfrac{1}{2}i[f(\lambda) - f(-\lambda)]/\sin\pi\lambda \tag{5.3}$$

[where, as in Sec. 2, $f(\lambda) \equiv f_{\lambda-\frac{1}{2}}$], possess the Mittag-Leffler expansions

$$g_1(\lambda^2) = 1 + \sum_\Lambda \frac{a_{\lambda'}}{\lambda'^2 - \lambda^2} \tag{5.4}$$

and

$$g_2(\lambda^2) = \sum_\Lambda \frac{b_{\lambda'}}{\lambda'^2 - \lambda^2}, \tag{5.5}$$

and are such that

$$g_1(\lambda^2) \to 1$$

and

$$g_2(\lambda^2) \to 0, \tag{5.6}$$

as $\operatorname{Im}\lambda \to \pm\infty$. [The set Λ in (5.4) and (5.5) is the same as in Sec. 2.]

The properties (5.6) follow from (5.2) and (5.3) if $f(\lambda) \to 1$ as $\operatorname{Im}\lambda \to -\infty$ and $e^{i\pi\lambda}f(\lambda)$ is bounded by an inverse power of $\operatorname{Im}\lambda$ as $\operatorname{Im}\lambda \to +\infty$. That these conditions are fulfilled by (3.33) is readily checked; the inverse power of $\operatorname{Im}\lambda$ is 2. In order to establish the existence of the expansions (5.4) and (5.5), we first show that

$$g_1(\lambda^2) \to 1,$$
$$g_2(\lambda^2) \to 0, \tag{5.7}$$

for $|\lambda| \to \infty$ and $\arg\lambda \neq 0, \pi$. The denominator $\sin\pi\lambda$ in (5.3) dominates over $f(\lambda)$ and $f(-\lambda)$ as given by (3.33) for $|\lambda| \to \infty$, $\arg\lambda \neq 0, \pi$; this establishes (5.7) for the function $g_2(\lambda^2)$. Defining

$$K(\lambda) \equiv (1 + i\cot\pi\lambda)f(-\lambda) = (ie^{-i\pi\lambda}/\sin\pi\lambda)f(-\lambda), \tag{5.8}$$

so that by (5.2) we have

$$g_1(\lambda^2) = \tfrac{1}{2}[K(\lambda) + K(-\lambda)], \tag{5.9}$$

we find, upon comparing (5.8) and (3.33), that

$$K(\lambda) \to 2 \quad \text{as} \quad |\lambda| \to \infty, \quad 0 < \arg\lambda < \pi, \tag{5.10}$$

and

$$K(\lambda) \to 0 \quad \text{as} \quad |\lambda| \to \infty, \quad -\pi < \arg\lambda < 0. \tag{5.11}$$

Substitution of (5.1) and (5.11) in (5.9) then establishes (5.7) for the function $g_1(\lambda^2)$. Next, we note that, according to (3.33), $f(\lambda)$ is analytic except for a finite number of poles; consequently, from (5.2) and (5.3) it follows that $g_1(\lambda^2)$ and $g_2(\lambda^2)$ are meromorphic functions of λ, and that they have poles only at the points $\lambda = \pm n$, $n = 0, \pm1, \pm2, \cdots$, augmented by the finite set of points for which $f(-\lambda)$ and $f(+\lambda)$ have poles. It then follows in the standard manner,[10] upon use of (5.7), that $g_1(\lambda^2)$ and $g_2(\lambda^2)$ possess the expansions (5.4) and (5.5), where the set Λ consists of the positive integers plus the (positive) points at which $f(-\lambda)$ becomes singular, in agreement with the discussion in Sec. 2. Consequently, (3.33) can always be represented by a dispersion relation of the form (2.12b).

Finally, we consider for the special case M integer the representation of the one-term example of Sec. 4, which results in the Jost function (4.22) in terms of the Newton–Sabatier formalism of Sec. 2. First, we note that (4.22) has no singularities. Thus, the set Λ of Eqs.

(2.14) and (2.12b) consists only of the positive integer values of $\lambda = l + \frac{1}{2}$ or, equivalently, of odd half-integer values of l. Thus, the c_l in (2.14) or (2.12b) are nonzero only for odd half-integer values of l. However, by using (2.15), these values of c_l for (4.22) can be calculated explicitly. When this is done, it is found that there are an infinite number of them and that for large values of l they have the behavior

$$c_l = -(4/\pi)(L - M) + O(l^{-1}). \tag{5.12}$$

Since $L \neq M$, we see that the c_l do not vanish for large l, but only tend to a constant. Thus, for this example, which has a potential with nonvanishing first moment, the description in terms of the formalism of Sec. 3 requires a one-term series, whereas the description in terms of the formalism of Sec. 2 requires a (slowly converging) infinite series.[11]

ACKNOWLEDGMENTS

It is a pleasure to thank Professor R. G. Newton for a helpful discussion and Professor J. M. Freeman for a critical reading of part of the manuscript.

APPENDIX A

We show here that, with S and T defined as in the first paragraph of Sec. 3, a matrix with elements $[L(L+1) - l(l+1)]^{-1}$ or a matrix with elements $(L - l)^{-1}$, where $L \in T$ and $l \in S$, is nonsingular. From this result it follows, in particular, that the coefficients γ_l are uniquely determined by (3.12). We note that if $L, K \in S \cup T$, and $L(L+1) = K(K+1)$, then $L = K$. (The possibility $L = -K - 1$ is excluded by the requirements $L > -\frac{1}{2}$ and $K > -\frac{1}{2}$.) With this result it follows that matrices of the above forms are examples of a general class of matrices which we show below are nonsingular.

If a_i and b_i, $i = 1, \cdots, n$, are $2n$ distinct real or complex numbers, then the $n \times n$ matrix A with elements

$$A_{ij} = (a_i - b_j)^{-1} \tag{A1}$$

is nonsingular. This result can be obtained by induction on n. For $n = 1$ the result is true. Suppose it is true for $n = N - 1 > 1$; then it is true for $n = N$. To show this we expand the determinant of A about the first row, which yields

$$|A| = \sum_{i=1}^{n} \frac{C_i}{(a_1 - b_i)}, \tag{A2}$$

where the C_i are nonzero constants, being determinants of $(N-1) \times (N-1)$ matrices of the form (A1), which are nonzero by assumption.

Introducing the function

$$F(z) = \sum_{i=1}^{n} \frac{C_i}{(z - b_i)}, \tag{A3}$$

where $F(z)$ is the determinant of the matrix A with a_1 replaced by z, and then setting $z = a_i$, $i = 2, \cdots, n$, we observe that the first and ith rows of A are equal, and hence

$$F(a_i) = 0, \quad i = 2, \cdots, n. \tag{A4}$$

Since from (A3) we also have

$$F(z) \to 0, \quad \text{as} \quad |z| \to \infty, \tag{A5}$$

it follows from (A3), (A4), and (A5) that

$$F(z) = C \prod_{i=2}^{n} (z - a_i) \Big/ \prod_{i=1}^{n} (z - b_i), \tag{A6}$$

where C is a nonzero constant. Hence, from (A6) it follows that

$$|A| = F(a_1) \neq 0, \tag{A7}$$

which is the required result.

APPENDIX B

We wish to sum the two series

$$\sum_{l \text{ even}} (2l + 1) a_l P_l(x), \quad l \geq 0, \tag{B1}$$

and

$$\sum_{l \text{ odd}} (2l + 1) a_l P_l(x), \quad l \geq 0, \tag{B2}$$

where a_l has the form

$$a_l = [b^2 - (l + \tfrac{1}{2})^2]^{-1} = [b^2 - \tfrac{1}{4} - l(l+1)]^{-1}. \tag{B3}$$

This function is invariant under the replacement

$$l \to -l - 1. \tag{B4}$$

Adding (B1) and (B2) and using (B3), we have

$$\sum_{l=0}^{\infty} (2l + 1)[b^2 - (l + \tfrac{1}{2})^2]^{-1} P_l(x). \tag{B5}$$

Making the replacement (B4) in (B5) and using the relation $P_l(x) = P_{-l-1}(x)$, we have, in place of (B5),

$$-\sum_{l=-\infty}^{-1} (2l + 1)[b^2 - (l + \tfrac{1}{2})^2]^{-1} P_l(x). \tag{B6}$$

For l a positive integer or zero, $P_l(x) = (-)^l P_l(-x)$, and for l a negative integer, $P_l(x) = (-)^{1-l} P_l(-x)$. Hence, from the equality of the expressions (B5) and (B6) we have

$$\sum_{l=0}^{\infty} (2l + 1) a_l P_l(x) = \tfrac{1}{2} \sum_{-\infty}^{\infty} (2l + 1) a_l (-)^l P_l(-x). \tag{B7}$$

The sum on the left-hand side of (B7) can be replaced by a contour integral by using the properties of $\sin \pi l$.

Thus, we have

$$\sum_{l=0}^{\infty}(2l+1)a_lP_l(x)=\frac{1}{4i}\int_C(2l+1)\frac{a_l}{\sin\pi l}P_l(-x)\,dl, \tag{B8}$$

where C is a contour encircling the zeros of $\sin\pi l$ but excluding the poles of a_l.

The contour C can be replaced by two contours by adding two infinite semicircles to the integral; thus, we have two clockwise contours—one in the upper and one in the lower half l-plane. We can check that the contributions from the infinite semicircles are zero by considering the asymptotic expansion in l[12]:

$$P_l(\cos\theta)=\frac{\Gamma(l+1)}{\Gamma(l+\frac{3}{2})}\left(\frac{2}{\pi\sin\theta}\right)^{\frac{1}{2}}\times\{\cos[(l+\tfrac{1}{2})\theta-\tfrac{1}{4}\pi]+O(l^{-1})\}, \tag{B9}$$

where $\epsilon<\theta<\pi-\epsilon$ and $\epsilon>0$.

Comparing (B9) with (B8) and using (B3), we see that the integrand of (B8) tends to zero, as $|l|$ tends to infinity, sufficiently rapidly to guarantee zero contribution from the infinite semicircular contour integrals that we have added to (B8). Hence we have

$$\sum_{l=0}^{\infty}(2l+1)a_lP_l(x)=\tfrac{1}{2}\pi\left(\frac{P_{b-\frac{1}{2}}(-x)}{\sin(b-\frac{1}{2})\pi}-\frac{P_{-b-\frac{1}{2}}(-x)}{\sin(b+\frac{1}{2})\pi}\right)$$

or

$$\sum_{l=0}^{\infty}(2l+1)a_lP_l(x)=-\frac{\pi}{\cos\pi b}P_{b-\frac{1}{2}}(-x). \tag{B10}$$

Replacing x by $-x$, we also have

$$\sum_{l=0}^{\infty}(2l+1)a_l(-)^lP_l(x)=-\frac{\pi}{\cos\pi b}P_{b-\frac{1}{2}}(x). \tag{B11}$$

Thus, addition and subtraction (B10) and (B11) yields

$$\sum_{l\,\text{even}}(2l+1)a_lP_l(x)=-\frac{\pi}{2\cos\pi b}[P_{b-\frac{1}{2}}(x)+P_{b-\frac{1}{2}}(-x)] \tag{B12}$$

and

$$\sum_{l\,\text{odd}}(2l+1)a_lP_l(x)=\frac{\pi}{2\cos\pi b}[P_{b-\frac{1}{2}}(x)-P_{b-\frac{1}{2}}(-x)], \tag{B13}$$

where b is the parameter introduced in the definition (B3) of a_l.

[1] R. G. Newton, J. Math. Phys. **3**, 75 (1962).
[2] P. C. Sabatier, J. Math. Phys. **7**, 1515, 2079 (1966); **8**, 905, 1957 (1967); **9**, 1241 (1968).
[3] R. G. Newton, J. Math. Phys. **8**, 1566 (1967).
[4] Earlier references which contain work related to the same problem include J. A. Wheeler, Phys. Rev. **99**, 630 (1955); T. Regge, Nuovo Cimento **14**, 951 (1959); A. Martin and G. Y. Targonski, Nuovo Cimento **20**, 1182 (1951).
[5] P. J. Redmond, J. Math. Phys. **5**, 1547 (1964).
[6] See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Book Co., Inc., New York, 1966).
[7] G. N. Watson, *Theory of Bessel Functions* (Cambridge Univ. Press, Cambridge, England, 1944), 2nd ed.
[8] (3.14) follows from the definitions $v_l(z)\equiv(\pi z/2)^{\frac{1}{2}}Y_{l+\frac{1}{2}}(z)$ and $Y_\nu(z)\equiv[J_\nu(z)\cos\pi\nu-J_{-\nu}(z)]/\sin\pi\nu$, together with the asymptotic formula $Y_\nu(z)=-z^{-\nu}2(\nu-1)!/\pi+O(z^\nu\log z)+O(z^{-\nu+2})$.
[9] Equation (4.16) for L odd corrects an over-all sign mistake in Sabatier's result, which is given on p. 2091 of the second-cited paper of Ref. 2.
[10] See, for example, E. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge Univ. Press, Cambridge, England, 1952).
[11] The convergence properties of the series (2.1) have been investigated by Sabatier (see Ref. 2).
[12] *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. I, Sec. 3.9.1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Note on the Uniqueness of the Solution of an Equation of Interest in the Inverse Scattering Problem

JOSEPH R. COX
Department of Physics, Florida Atlantic University, Boca Raton, Florida 33432
AND
KERRIS W. THOMPSON
Department of Mathematics and Department of Physics, Florida Atlantic University, Boca Raton, Florida 33432

(Received 4 September 1969)

It is proved that the solutions recently obtained by the authors [J. Math. Phys. **11**, 805 (1970)] of the Regge–Newton integral equation (of interest in connection with the inverse scattering problem at fixed energy) are, for a given kernel, inhomogeneity, and boundary condition, uniquely determined.

I. INTRODUCTION

A general method of determining a central potential from a knowledge of all phase shifts has been developed by Newton[1] and extended by Sabatier[2] and Newton.[3] In order to obtain the potential by this method, it is generally necessary to sum an infinite series. Unfortunately, this series always converges slowly except when the first moment of the potential vanishes.[3] In particular, any truncation of this series after a finite number of terms necessarily results in a potential with a vanishing first moment.

As a possible first step in the circumvention of such difficulties, we have introduced in a recent paper[4] a different series than the one used by Newton and generalized by Sabatier, which has the advantage that it does, in general, yield potentials with nonvanishing first moments for finite sums. However, we omitted in our paper any discussion of the uniqueness of the solution of the Regge–Newton integral equation (the analog in the fixed-energy case of the Gel'fand–Levitan equation[5]) which results when our proposed series is used as input for the theory. Since the uniqueness of the solution of this equation is essential if any method based on it is to work, and since neither of the two methods[1,5] of proving the uniqueness of the analogous solution for the Newton–Sabatier series are applicable, we construct this uniqueness proof in the present paper.

II. UNIQUENESS PROOF

Let S and T denote two disjoint finite sets, each containing N distinct real numbers chosen from the interval $(-\frac{1}{2}, \infty)$. Define the function $g(r, r')$ by

$$g(r, r') = \sum_{M\in S} \gamma_M u_M(r_<) v_M(r_>), \tag{1}$$

where the γ_M satisfy

$$\sum_{M\in S} [M(M+1) - L(L+1)]^{-1}\gamma_M = 1, \quad L \in T, \tag{2}$$

and $u_M(r)$ and $v_M(r)$ are the regular Riccatti–Bessel functions $(\frac{1}{2}\pi r)^{\frac{1}{2}} J_{M+\frac{1}{2}}(r)$ and the irregular spherical Weber–Schläfli functions $(\frac{1}{2}\pi r)^{\frac{1}{2}} Y_{M+\frac{1}{2}}(r)$,[6] respectively. Let $L(r, r')$ be any solution of

$$L(r, r') = g(r, r') - \int_0^r dr'' r''^{-2} L(r, r'') g(r'', r'), \quad r' < r, \tag{3}$$

which has the small-r' behavior

$$r'^{-\frac{1}{2}} L(r, r') = O(1). \tag{4}$$

The problem is to show that $L(r, r')$ is uniquely determined by Eqs. (1)–(4).

Introducing the function

$$\rho[l(l+1)] = 1 + \sum_{M\in S} [l(l+1) - M(M+1)]^{-1}\gamma_M \tag{5}$$

and comparing (5) with (2), we have

$$\rho[l(l+1)] = 0, \quad l \in T \quad \text{or} \quad -l-1 \in T.$$

Since also $\rho[l(l+1)] \to 1$ as $|l(l+1)| \to \infty$,

$$\rho[l(l+1)] = \prod_{L\in T} [l(l+1) - L(L+1)] \times \prod_{M\in S} [l(l+1) - M(M+1)]$$

and, hence, upon comparison with (5) we conclude that

$$\gamma_M = \prod_{L\in T} [M(M+1) - L(L+1)] \Big/ \prod_{\substack{M'\in S \\ M'\neq M}} [M(M+1) - M'(M'+1)], \quad M \in S.$$

Thus, for sets S and T given, the coefficients γ_M, $M \in S$, are uniquely determined and are nonzero.

Suppose there were two different solutions of (3), each satisfying (4). Then their difference $\Omega(r, r')$ would satisfy the homogeneous version of (3),

$$\Omega(r, r') = -\int_0^r dr'' r''^{-2} \Omega(r, r'') g(r'', r'), \tag{6}$$

and also, from (4),

$$r'^{-\frac{1}{2}}\Omega(r, r') = O(1). \tag{7}$$

Define $\bar{\Omega}_M(r, r')$ by

$$\bar{\Omega}_M(r, r') \equiv -\int_0^r dr'' r''^{-2}\Omega(r, r'')g_M(r'', r'), \tag{8}$$

where

$$g_M(r, r') \equiv \gamma_M u_M(r_<)v_M(r_>). \tag{9}$$

It then follows from (1), (6), (8), and (9) that

$$\sum_{M\in S} \bar{\Omega}_M(r, r') = \Omega(r, r'). \tag{10}$$

Applying the operator

$$D_0(r') \equiv r'^2(\partial^2/\partial r'^2 + 1) \tag{11}$$

to (8) and noting that

$$[D_0(r') - M(M+1)]g_M(r'', r') = \gamma_M r'^2\delta(r' - r''), \tag{12}$$

we obtain

$$[D_0(r') - M(M+1)]\bar{\Omega}_M(r, r') + \gamma_M\Omega(r, r') = 0. \tag{13}$$

Letting $\gamma_M^{-\frac{1}{2}}$ denote one of the square roots of γ_M^{-1} (recall that $\gamma_M \neq 0$, $M \in S$) and defining

$$\Omega_M(r, r') \equiv \gamma_M^{-\frac{1}{2}}\bar{\Omega}_M(r, r'), \quad M \in S, \tag{14}$$

Eq. (13) becomes, with the aid of (10) and (14),

$$[D_0(r') - M(M+1)]\Omega_M(r, r') + \gamma_M^{\frac{1}{2}}\sum_{M'\in S}\gamma_{M'}^{\frac{1}{2}}\Omega_{M'}(r, r') = 0, \quad M \in S, \tag{15}$$

or, in matrix notation,

$$[D_0(r') - \Lambda + aa^{\mathrm{T}}]\Omega = 0, \tag{16}$$

where Ω and a denote, respectively, column vectors with components $\Omega_M(r, r')$ and $\gamma_M^{\frac{1}{2}}$, and Λ denotes the diagonal matrix with diagonal elements $M(M+1)$.

In order to decouple Eqs. (15), let us consider the eigenvalue problem

$$(\Lambda - aa^{\mathrm{T}})x = \lambda x \tag{17}$$

or

$$(\Lambda - \lambda I)x = (a^{\mathrm{T}}x)a. \tag{18}$$

We first show that $\lambda = M(M+1)$, $M \in S$, is not an eigenvalue of $\Lambda - aa^{\mathrm{T}}$. If it were, then the Mth component of (18),

$$[M(M+1) - \lambda]x_M = (a^{\mathrm{T}}x)\gamma_M^{\frac{1}{2}},$$

would imply that $a^{\mathrm{T}}x = 0$ and hence (18) would become

$$[\Lambda - M(M+1)]x = 0, \quad M \in S,$$

which would have as its solution a vector x with components

$$x_{M'} = x_M\delta_{MM'}, \tag{19}$$

so that

$$0 = a^{\mathrm{T}}x = \sum_{M'\in S}\gamma_{M'}^{\frac{1}{2}}x_{M'} = \gamma_M^{\frac{1}{2}}x_M, \quad M \in S. \tag{20}$$

However, (19) and (20) imply that $x \equiv 0$. Hence, $\lambda = M(M+1)$, $M \in S$, is not an eigenvalue of $\Lambda - aa^{\mathrm{T}}$. From this fact it follows that $\Lambda - \lambda I$ is nonsingular and thus from (18) that

$$x = (a^{\mathrm{T}}x)(\Lambda - \lambda I)^{-1}a. \tag{21}$$

Multiplying (21) on the left by a^{T}, we have, since $a^{\mathrm{T}}x \neq 0$,

$$a^{\mathrm{T}}(\Lambda - \lambda I)^{-1}a = 1$$

or

$$\sum_{M\in S}[M(M+1) - \lambda]^{-1}\gamma_M = 1,$$

from which we conclude, upon comparison with (2), that the eigenvalues of $\Lambda - aa^{\mathrm{T}}$ are distinct, and that they are given by

$$\lambda = L(L+1), \quad L \in T.$$

Thus, $\Lambda - aa^{\mathrm{T}}$ can be transformed to the diagonal form

$$D = P^{-1}(\Lambda - aa^{\mathrm{T}})P, \tag{22}$$

where the diagonal matrix D has diagonal elements $L(L+1)$, $L \in T$.

Introducing

$$\Omega' \equiv P^{-1}\Omega \tag{23}$$

with components $\Omega'_L(r, r')$ and using (22), Eq. (16) becomes

$$[D_0(r') - D]\Omega' = 0$$

or, equivalently,

$$[D_0(r') - L(L+1)]\Omega'_L(r, r') = 0, \quad L \in T. \tag{24}$$

Since $u_L(r)$ and $v_L(r)$, as introduced in (1), are linearly independent solutions of

$$[D_0(r') - L(L+1)]y_L(r') = 0,$$

the most general solution of (24) must be of the form

$$\Omega'_L(r, r') = A'_L(r)u_L(r') + B'_L(r)v_L(r'). \tag{25}$$

Hence, according to (10), (14), (23), and (25), $\Omega(r, r')$ is of the form

$$\Omega(r, r') = \sum_{L\in T}[A_L(r)u_L(r') + B_L(r)v_L(r')]. \tag{26}$$

The small-r behavior of $u_L(r)$ and $v_L(r)$ is[6]

$$u_L(r) = [\pi^{\frac{1}{2}}/2^{L+1}\Gamma(L + \tfrac{3}{2})]r^{L+1} + O(r^{L+3}),$$
$$v_L(r) = -[2^L\Gamma(L + \tfrac{1}{2})/\pi^{\frac{1}{2}}]r^{-L} + O(r^{-L+\epsilon}), \tag{27}$$

where ϵ is such that $0 < \epsilon < \min(2, 2L + 1)$. Thus, since by assumption $\frac{1}{2} < L < \infty$, it follows from (7), (26), and (27) that $B_L(r) \equiv 0$, $L \in T$. Equation (26), therefore, reduces to

$$\Omega(r, r') = \sum_{L \in T} A_L(r) u_L(r'). \tag{28}$$

Substituting (28) in (6) and using (1) and the identity

$$\frac{d}{dr} W[u_M(r), v_L(r)] = [L(L+1) - M(M+1)] u_M(r) v_L(r)/r^2,$$

where $W[u_M(r), v_L(r)] \equiv u_M(r)v_L'(r) - u_M'(r)v_L(r)$ is the Wronskian, we obtain

$$\sum_{L \in T} C_{ML}(r) A_L(r) = 0, \quad M \in S, \tag{29}$$

where

$$C_{ML}(r) \equiv W[u_L(r), v_M(r)] \times [M(M+1) - L(L+1)]^{-1}. \tag{30}$$

It follows from properties of Bessel functions that $\det C(r)$, where $C(r)$ is the $N \times N$ matrix with elements $C_{ML}(r)$, $M \in S$, $L \in T$, is an analytic function of r in some region containing the positive real r axis $(0 < r < \infty)$ and, hence, can have only isolated zeros on $(0, \infty)$, unless, of course, $\det C(r) \equiv 0$ on $(0, \infty)$. However, in our previous paper[4] we have shown that $\det C(r) \neq 0$ in some neighborhood of $r = 0$, so the latter possibility is excluded. Since, therefore, $\det C(r)$ can only have isolated zeros on $(0, \infty)$, the only (continuous) solution of (29) is

$$A_L(r) \equiv 0, \quad L \in T. \tag{31}$$

Therefore, according to (28) and (31), Eqs. (6) and (7) only have the trivial solution $\Omega(r, r') \equiv 0$. Therefore, $L(r, r')$ is uniquely determined by Eqs. (1)–(4).

Finally, we remark that mere vanishing of $L(r, r')$ at $r' = 0$ does not assure the uniqueness of $L(r, r')$. Indeed, if (4) is replaced by the weaker requirement

$$L(r, 0) = 0, \tag{4'}$$

and, for example, one chooses the sets S and T in such a way that each contains only one member, M and L, respectively, and $M > L$, $L \in (-\frac{1}{2}, 0)$, $M \in (-\frac{1}{2}, \infty)$, so that (1) reduces to

$$g(r, r') = \gamma_M u_M(r_<) v_M(r_>),$$

where

$$\gamma_M = (M - L)(M + L + 1),$$

then it is readily verified that (3) and (4') have the family of solutions

$$L(r, r') = \{W[u_L(r), v_M(r)]\}^{-1} \gamma_M v_M(r) u_L(r') + f(r)\{W[v_L(r), v_M(r)] u_L(r') - W[u_1(r), v_M(r)] v_L(r')\},$$

where $f(r)$ is completely arbitrary.

ACKNOWLEDGMENT

We thank Professor R. G. Newton for pointing out to us the necessity of constructing the above uniqueness proof.

[1] R. G. Newton, J. Math. Phys. **3**, 75 (1962).
[2] P. C. Sabatier, J. Math. Phys. **7**, 1515, 2079 (1955); **8**, 905, 1957 (1967); **9**, 1241 (1968).
[3] R. G. Newton, J. Math. Phys. **8**, 1566 (1967).
[4] J. R. Cox and K. W. Thompson, J. Math. Phys. **11**, 805 (1970) (preceding paper).
[5] See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Book Company, Inc., New York, 1966).
[6] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1944), 2nd ed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Curvature Collineations in Empty Space-Times

C. D. COLLINSON
Department of Applied Mathematics, University of Hull, Hull, England

(Received 20 June 1969)

It is shown that the only curvature collineations admitted by an empty space-time, not of Petrov type *N*, are conformal motions. The curvature collineations admitted by the plane-fronted gravitational waves are found.

1. INTRODUCTION

A Riemannian space is said to admit a curvature collineation if there exists an infinitesimal transformation $\bar{x}^i = x^i + \epsilon\xi^i(x)$ for which $\pounds R^i_{jkl} = 0$, where $\pounds$ denotes the Lie derivative with respect to the vector ξ^i and R^i_{jkl} is the curvature tensor of the space. Katzin, Levine, and Davis[1] have recently investigated such collineations and have derived several results of geometrical and physical interest. The application of such collineations to the space-times of general relativity is, however, limited by the following theorem:

Theorem: The only curvature collineations admitted by an empty space-time, not of Petrov type *N*, are conformal motions.

This theorem is proved in Sec. 2. The curvature collineations admitted by the plane-fronted gravitational waves are found in Sec. 3. As a result, it is seen that empty space-times of Petrov type *N* do admit curvature collineations other than conformal motions.

2. PROOF OF THE THEOREM

Katzin, Levine, and Davis show that a necessary condition for a curvature collineation is

$$h_{ia}R^a_{jkl} + h_{ja}R^a_{ikl} = 0, \tag{2.1}$$

where

$$h_{ij} = \xi_{i;j} + \xi_{j;i}. \tag{2.2}$$

In order to exploit Eq. (2.1) for the space-times of general relativity, the null tetrad[2] components of the equation are written out explicitly. The resulting equations are then solved for empty space-times, the algebra being simplified for each Petrov type by choosing the usual canonical forms for the tetrad components of the Weyl tensor. The calculations are straightforward and it is found that, if the space-time is not of Petrov type *N*,

$$h_{ij} = \tfrac{1}{2}\phi g_{ij}, \tag{2.3}$$

where g_{ij} is the metric tensor of the space-time. Equation (2.3) defines a conformal motion, and so the theorem stated in the introduction is proved. For space-times of Petrov type *N*, it is found that

$$h_{ij} = \tfrac{1}{2}\phi g_{ij} + \alpha l_i l_j, \tag{2.4}$$

where l_i is the principal vector of the Weyl tensor satisfying

$$C^i_{jkl}l_i = 0. \tag{2.5}$$

The question now arises as to whether or not empty space-times of Petrov type *N* can admit curvature collineations other than conformal motions. To answer this question, the curvature collineations admitted by the plane-fronted gravitational waves are found (these space-times are of Petrov type *N*).

3. THE PLANE-FRONTED GRAVITATIONAL WAVES

The metric of the plane-fronted gravitational waves can be written, in terms of two real coordinates ρ and σ, and a complex coordinate z, in the form

$$ds^2 = 2\,d\rho\,d\sigma - 2H\,d\sigma^2 - 2\,dz\,d\bar{z},$$

where

$$\frac{\partial H}{\partial \rho} = \frac{\partial^2 H}{\partial z \partial \bar{z}} = 0.$$

The vectors[3]

$$l^i = \delta^i_1, \quad n^i = \delta^i_2 + H\delta^i_1, \quad m^i = \delta^i_0,$$

where $(x^1, x^2, x^0, \bar{x}^0) \equiv (\rho, \sigma, z, \bar{z})$ form a null tetrad, the intrinsic derivatives associated with the tetrad being

$$D\phi = l^i\phi_{,i} = \frac{\partial \phi}{\partial \rho},$$

$$\Delta\phi = n^i\phi_{,i} = \frac{\partial \phi}{\partial \sigma} + H\frac{\partial \phi}{\partial \rho},$$

and

$$\delta\phi = m^i\phi_{,i} = \frac{\partial \phi}{\partial \bar{z}}.$$

The only nonzero spin coefficient is

$$\nu = -\frac{\partial H}{\partial z}, \tag{3.1}$$

while the only nonzero tetrad component of the Weyl tensor is

$$\psi_4 = -\frac{\partial^2 H}{\partial z^2}. \tag{3.2}$$

Equation (2.4) can be written as

$$\pounds g_{ij} = \tfrac{1}{2}\phi g_{ij} + \alpha l_i l_j. \tag{3.3}$$

In empty space the equation defining a curvature collineation can be written $\pounds C^i_{jkl} = 0$ or, using Eqs. (2.5) and (3.3),

$$\pounds C_{ijkl} = \tfrac{1}{2}\phi C_{ijkl}. \tag{3.4}$$

The null tetrad components of Eq. (3.4) and of Eq. (3.3) (with $\alpha = 0$) have been given in a previous paper.[4] Substituting (3.1) and (3.2) into these equations yields (with $\alpha \neq 0$)

$$D\xi_1 = 0, \tag{3.5}$$

$$\Delta\xi_2 = -\frac{\partial H}{\partial z}\xi_3 - \frac{\partial H}{\partial \bar{z}}\bar{\xi}_3 + \tfrac{1}{2}\alpha, \tag{3.6}$$

$$\delta\xi_3 = 0, \tag{3.7}$$

$$\Delta\xi_1 + D\xi_2 - \tfrac{1}{2}\phi = 0, \tag{3.8}$$

$$\delta\xi_1 + D\xi_3 = 0, \tag{3.9}$$

$$\delta\xi_2 + \Delta\xi_3 = -\frac{\partial H}{d\bar{z}}\xi_1, \tag{3.10}$$

$$\delta\bar{\xi}_3 + \delta\bar{\xi}_3 + \tfrac{1}{2}\phi = 0, \tag{3.11}$$

$$D\xi_3 = 0, \tag{3.12}$$

$$-\frac{\partial^3 H}{\partial\sigma\partial z^2}\xi_1 + \frac{\partial^3 H}{\partial z^3}\xi_3 - 2\frac{\partial^2 H}{\partial z^2}(\Delta\xi_1 - \bar{\delta}\bar{\xi}_3 - \phi/4) = 0. \tag{3.13}$$

Equations (3.5) and (3.7)–(3.12) can be integrated to give

$$\xi_1 = \xi_1(\sigma),$$

$$\xi_2 = -[a + \bar{a} + \dot{\xi}_1]\rho - z\bar{z}\dot{a} - \dot{b}\bar{z} - \dot{\bar{b}}z - H\xi_1 + c(\sigma),$$

$$\xi_3 = a(\sigma)z + b(\sigma),$$

and

$$\tfrac{1}{2}\phi = -(a + \bar{a}),$$

where $a - \bar{a} = ik_1$ (const). Equation (3.13) is manipulated using the separation of variables technique. Several cases arise. Only one case is discussed here, namely, that corresponding to $\partial^3 H/\partial z^3 = 0$ (this is a plane gravitational wave). In this case, the coordinate system can be chosen so that

$$H = \alpha(\sigma)z^2 + \bar{\alpha}(\sigma)\bar{z}^2.$$

Imposing the further condition $(\alpha\bar{\alpha})^{-\frac{1}{4}}[\log \alpha/\bar{\alpha}] \neq$ const, Eq. (3.13) then yields $\xi_1 = 0$, $a = \bar{a}$, so that the null tetrad component of the most general curvature collineation are

$$\xi^1 = 0,$$

$$\xi_2 = -2a\rho - \dot{a}z\bar{z} - \dot{b}\bar{z} - \dot{\bar{b}}z + c(\sigma),$$

and

$$\xi_3 = a(\sigma)z + b(\sigma).$$

Substituting these into Eq. (3.6) yields

$$\tfrac{1}{2}\alpha = -2\dot{a}\rho - \ddot{a}z\bar{z} - \bar{z}(\ddot{b} - 2\bar{\alpha}\dot{\bar{b}}) - z(\ddot{\bar{b}} - 2\alpha b) + \dot{c}.$$

For a conformal motion, α is zero. Since α does not vanish identically, it follows that not all the curvature collineations are conformal motions. The conformal motions admitted by the plane gravitational waves considered here can be found by putting $\alpha = 0$. It is then clear that

$$a = k_1, \quad c = k_2, \quad \text{and} \quad \ddot{b} - 2\bar{\alpha}\dot{\bar{b}} = 0,$$

where k_1 and k_2 are constants. The equation $\ddot{b} - 2\bar{\alpha}\dot{\bar{b}} = 0$ admits four independent solutions, and so there is a six-parameter group of conformal motions. The conformal motions with $k_1 = 0$ form a five-parameter group of motions (the existence of this group is well known). The conformal motion with $k_2 = b = 0$ is a homothetic motion (since ϕ is a constant).

[1] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).
[2] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[3] C. D. Collinson, J. Math. Phys. **9**, 403 (1968).
[4] C. D. Collinson and D. C. French, J. Math. Phys. **8**, 701 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

General Relativistic Thin-Sandwich Theorem*

ARTHUR KOMAR
Belfer Graduate School of Science, Yeshiva University, New York, New York 10033

(Received 6 June 1969)

The configuration variables for the gravitational field g_{mn} are assigned arbitrarily on two infinitesimally neighboring spacelike hypersurfaces. We then investigate the extent to which a solution of the vacuum Einstein field equations can be found consistent with the given assignment. A local approach, employing Dirac's Hamiltonian formalism, reveals that solutions can be found locally which are nonunique and highly unstable.

I. INTRODUCTION

In the Feynman history integral approach to quantization, it is essential that the independent classical configuration variables can be assigned arbitrarily on two spacelike hypersurfaces and that a unique classical trajectory can be interpolated to such an assignment. With a view toward understanding the quantum theory of gravitation, it was proposed[1] to investigate the corresponding problem in the general theory of relativity. In view of the complexity of the general problem, it was found necessary to introduce the simplification that the two spacelike surfaces were infinitesimally neighboring. Within that more limited context, a variational approach seemed to indicate that in general with reasonable boundary conditions at infinity there were unique solutions.[1] However, the precise structure of the equations was not exhibited or analyzed. A by-product of such a result would be the impossibility of constructing general-relativistic observables exclusively from the configuration variables.

The purpose of this note is to present an alternative attack on this "thin-sandwich" problem which is completely local. The virtue of such an approach is relative clarity of the ensuing expressions resulting in a corresponding simplicity of the analysis. We shall determine the extent of arbitrariness of local solutions when they exist. Of course, we will not be able to treat questions of uniqueness in the large via our approach. Recent work of Bergmann[2] attacks this same problem from a point of view complementary to ours, namely, by considering arbitrary variations of the configuration variables away from a given classical trajectory. With the ensuing linearization of the equations, the global problem may become more amenable.

II. HAMILTONIAN TREATMENT

Introducing the canonical field variables g_{mn} and p^{mn}, we can write the Einstein field equations in canonical form by defining the four Hamiltonian constraints[3]

$$H_s \equiv p^{mn}g_{mn,s} - 2(p^{mn}g_{ms})_{,n} = 0, \tag{2.1}$$

$$H_L \equiv |g_{ab}|^{-\frac{1}{2}}[p^{mn}p_{mn} - \tfrac{1}{2}(p^m_m)^2] + |g_{ab}|^{+\frac{1}{2}}\,{}^3R = 0. \tag{2.2}$$

With the above definitions, an arbitrary infinitesimal coordinate transformation

$$X^{\mu'} = X^\mu + \xi^\mu, \tag{2.3}$$

described by the descriptor

$$\xi^\mu = \delta^\mu_s \xi^s + l^\mu \xi^L, \tag{2.4}$$

(where l^μ is the unit normal to the spacelike surface on which the canonical variables are defined) is generated by the Hamiltonian

$$H = -\int (\xi^s H_s + \xi^L H_L)\, d_3X. \tag{2.5}$$

We shall require the fundamental existence theorem[4]: To every solution $g_{mn}(X^s)$ and $p^{mn}(X^s)$ of the constraint equations (2.1) and (2.2), there exists a solution of the vacuum Einstein field equations, $G_{\mu\nu} = 0$, unique up to a diffeomorphism generated by H, which preserves the initial values of g_{mn} and p^{mn}.

Let us assume that, rather than assigning $g_{mn}(X^s)$ and $p^{mn}(X^s)$ on some initial surface, we are given instead $g_{mn}(X^s)$ and $\delta g_{mn}(X^s) \equiv h_{mn}(X^s)$. The "thin-sandwich problem" can now be posed: Determine $p^{mn}(X^s)$, which satisfies the constraint equations (2.1) and (2.2) in such a way that $h_{mn}(X^s)$ can be obtained by commuting g_{mn} with H of Eq. (2.5). Thus, we require

$$h_{mn} = [g_{mn}, H]. \tag{2.6}$$

Employing the explicit expressions (2.1), (2.2), and (2.5), we readily find

$$h_{mn} = -\xi_{m;n} - \xi_{n;m} - 2\,|g_{ab}|^{-\frac{1}{2}}\,\xi_L(p_{mn} - \tfrac{1}{2}g_{mn}p^s_s) \tag{2.7}$$

(the semicolon denotes covariant differentiation with respect to the spatial metric g_{mn}). Solving Eq. (2.7) for p_{mn}, we find

$$p_{mn} = (|g_{ab}|^{\frac{1}{2}}/2\xi_L)[-h_{mn} - \xi_{m;n} - \xi_{n;m} + g_{mn}(h^s_s + 2\xi^s{}_{;s})]. \quad (2.8)$$

Substituting this expression into the four constraint equations (2.1) and (2.2), we obtain four equations for the four as yet undetermined functions, ξ_s and ξ_L. Each solution ξ_s and ξ_L of the constraint equations will provide, via Eq. (2.8), a solution of the constraint equations in terms of canonical variables g_{mn} and p^{mn} and, thus, via the fundamental existence theorem, a unique Ricci flat Riemannian manifold, consistent with our given "thin sandwich."

The fourth constraint, Eq. (2.2), being a quadratic algebraic equation in p_{mn}, can be solved immediately for ξ_L^2, yielding

$$\xi_L^2 = (4 \cdot {}^3R)^{-1}[(h^s_s)^2 + 4h^r_r\xi^s{}_{;s} + 4(\xi^s{}_{;s})^2 - h^{rs}h_{rs} - 4h^{rs}\xi_{r;s} - 2\xi^{r;s}\xi_{r;s} - 2\xi^{r;s}\xi_{s;r}]. \quad (2.9)$$

If we substitute this expression into the three remaining constraint equations, we find an extremely complicated set of second-order nonlinear equations for the three functions ξ^s. Although the details of the equations are not terribly illuminating, we shall write them down should the interested reader wish to compare them with the corresponding relations of Bergmann's linearized approach[2]:

$$\begin{aligned}
&[2\xi^m{}_{;m}{}^s - \xi^{m;s}{}_m - \xi^{s;m}{}_m + h^{m;s}_m - h^{sm}{}_{;m}] \\
&\quad \times [4(\xi^m{}_{;m})^2 - 2\xi^{m;n}(\xi_{m;n} + \xi_{n;m}) - 4h^{mn}\xi_{m;n} \\
&\quad + 4h^m_m\xi^n{}_{;m} + (h^m_m)^2 - h^{mn}h_{mn}] \\
&\quad - \tfrac{1}{2}\{8\xi^p{}_{;p}\xi^q{}_{;qm} - 4\xi^{p;q}{}_m(\xi_{p;q} + \xi_{q;p}) + 4h^q_q\xi^p{}_{p;m} \\
&\quad - 4h^{pq}\xi_{p;qm} + 4h^p{}_{p;m}\xi^q{}_{;q} - 4h^{pq}{}_{;m}\xi_{p;q} + 2h^p_p h^q{}_{q;m} \\
&\quad - 2h^{pq}h_{pq;m} - ({}^3R_{,m}/{}^3R)[4(\xi^p{}_{;p})^2 \\
&\quad - 2\xi^{p;q}(\xi_{p;q} + \xi_{q;p}) - 4h^{pq}\xi_{p;q} \\
&\quad + 4h^p_p\xi^q{}_{;q} + (h^p_p)^2 - h^{pq}h_{pq}]\} \\
&\quad \times [g^{ms}(h^n_n + 2\xi^n{}_{;n}) - \xi^{s;m} - \xi^{m;s} - h^{ms}] = 0.
\end{aligned} \quad (2.10)$$

III. CONCLUSION

An examination of Eq. (2.9) reveals that if the initial configuration variables g_{mn} are assigned so that ${}^3R = 0$, one cannot solve Eq. (2.2) for ξ_L. If we ignore this possibility for a moment, another pathology arises when the right-hand side of Eq. (2.9) becomes negative, thereby leading to imaginary ξ_L. In general, we may avoid this pathology by selecting $\xi^s(X^2, X^3)$ and $(\partial\xi^s/\partial\xi^1)(X^2, X^3)$ arbitrarily on an $X^1 =$ constant surface, in such a way that the right-hand side of Eq. (2.9) is initially positive and such that the three equations (2.10) can be solved for $\partial^2\xi^s/\partial X^{12}$ in terms of the given functions. We can thus assert that solutions do exist at least for some finite distance off the initial surface. However, in view of our freedom to assign six arbitrary functions of two variables, the solutions are by no means unique. In addition, they are unstable to small changes in the Cauchy data (contrary to the indications of Ref. 1).

For the case where ${}^3R = 0$, we must retain ξ_L in Eq. (2.1) and treat Eq. (2.2) as a first-order differential equation exclusively involving ξ^s. An investigation of the resulting equations reveals a situation quite analogous to the case of nonvanishing ${}^3R$. Namely, for a variety of situations, the six functions $\xi_L(X^2, X^3)$, $\xi^s(X^2, X^3)$, $(\partial\xi^2/\partial X^1)(X^2, X^3)$, $(\partial\xi^3/\partial X^1) \times (X^2, X^3)$ can be assigned arbitrarily (subject to inequalities) on the initial surface $X^1 =$ const and uniquely determine an unstable solution for some distance off the surface.

We conclude that, in a local region of a spacelike hypersurface, solutions to the "thin-sandwich problem" can be found which are neither unique nor stable. A particularly puzzling feature of the equations obtained is that, when one examines the leading terms (highest derivatives) for the purpose of ascertaining the behavior of the linearized problem, one finds that there are assignments of g_{mn} and h_{mn} for which the resulting equations for ξ^s and ξ_L may become mutually incompatible. It is not clear why the linearized problem should have such a radically different structure and of what significance it may be.

* This work is supported in part by the United States Air Force under Grant No. AF-AFOSR 68-1524.

[1] R. F. Baierlein, P. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962).
[2] P. G. Bergmann, *Cincinnati Sesquicentennial*, Louis Witten, Ed. (to be published).
[3] P. A. M. Dirac, Phys. Rev. **114**, 924 (1959).
[4] Y. Foures-Bruhat, Acta Met. **88**, 141 (1952).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Summability Methods in Perturbation Theory

M. REEKEN
Battelle Institute, Advanced Studies Center, 1227 Carouge-Geneva, Switzerland

(Received 4 September 1969)

The Mittag-Leffler summability method is applied to operator-valued analytic functions and a corresponding procedure for perturbation theory is derived, which has a bigger region of convergence. This region is explicitly described.

1. INTRODUCTION

Let H_0 be a closed linear operator acting in a Banach space and let V be another closed linear operator acting in this space, which is relatively bounded with respect to H_0, with relative bound equal to zero.

We will be interested in the resolvent operator $R_\lambda(z) = (H_0 + \lambda V - z)^{-1}$ defined for z not in the spectrum of $H_\lambda = H_0 + \lambda V$ (here λ is a complex parameter). Let $\Sigma'(H_0)$ be a finite system of eigenvalues of H_0, separated from the rest of the spectrum of H_0 by a closed curve Θ encircling $\Sigma'(H_0)$. Then, there is a convex region Λ of the complex plane, containing the origin, such that for all $\lambda \in \Lambda$, the spectrum $\Sigma(H_\lambda)$ of $H_0 + \lambda V$ is likewise separated by Θ into a part $\Sigma'(H_\lambda)$ and a remainder, and the eigenvalues in $\Sigma'(H_\lambda)$ are analytic in λ with only algebraic singularities.[1]

2. RAYLEIGH-SCHRÖDINGER PERTURBATION THEORY

Usual perturbation theory starts from the identity

$$R_\lambda(z) = R_0(z)[1 + \lambda V R_0(z)]^{-1}$$

by developing the geometric series

$$[1 + \lambda V R_0(z)]^{-1} = \sum_{n=0}^{\infty} (-\lambda)^n [V R_0(z)]^n.$$

By our assumptions, $VR_0(z)$ is a bounded operator, so this series will converge in norm for $\|\lambda V R_0(z)\| < 1$. For simplicity, we will treat the case of H_0, being a normal operator, in a Hilbert space and V being bounded. The results can be generalized to the general case. Then, $\|\lambda V R_0(z)\| < 1$ can be replaced by $\|\lambda V\| < d(z)$, where $d(z)$ is the distance of z from the spectrum of H_0. Thus the spectrum of H_λ must lie within a distance of $\|\lambda V\|$ from the spectrum of H_0.

Let P_λ be the projection onto the subspace associated with the eigenvalues ϵ_λ branching off from ϵ_0 in $\Sigma'(H_0)$. Then,

$$P_\lambda = \frac{1}{2\pi i} \int_\Gamma R_\lambda(z)\, dz, \tag{2.1}$$

where Γ is a contour encircling ϵ_λ but no other points of $\Sigma(H_\lambda)$.

The perturbation expansion for P_λ results from substituting the above geometric-series expansion in this integral. We obtain

$$P_\lambda = \sum_{n=0}^{\infty} (-\lambda)^n \frac{1}{2\pi i} \int_\Gamma R_0(z)[V R_0(z)]^n\, dz, \tag{2.2}$$

provided that Γ lies in the region of convergence defined by $d(z) > \|\lambda V\|$. The contour integrals on the right-hand side may be evaluated explicitly in terms of P_0, the reduced resolvent, and V; this results in the usual Rayleigh–Schrödinger series.[2]

Now it may be impossible to enclose ϵ_λ but no other points of $\Sigma(H_\lambda)$ by a curve Γ without leaving the region defined by $d(z) > \|\lambda V\|$. The geometric-series development used above represents an approximation of $[1 + \lambda V R_0(z)]^{-1}$ by polynomials in $\lambda V R_0(z)$ uniformly on any compact subset of the region $\{z \mid d(z) > \|\lambda V\|\}$. So one is tempted to look for an approximation by different polynomials, converging in a bigger region, which contains a suitable contour Γ. But there is the following limitation.

Theorem 1: For any approximation of $[1 + \lambda V R_0(z)]^{-1}$ by polynomials in $\lambda V R_0(z)$, uniformly in z on some compact K, the eigenvalues ϵ_0 and ϵ_λ will lie in the same connected component of the complement of K.

Proof: In order that the transition from (2.1) to the analog of (2.2) for some other polynomial approximation is possible, Γ has to be in the region of convergence of this approximation and has to enclose all ϵ_λ. If ϵ_0 and the ϵ_λ group would lay in disconnected components of the complement of K, then Γ could be chosen such that it does not enclose ϵ_0 and thus $P_\lambda = 0$ because the integrands of the analog to (2.2) would have no singularities in the interior of the contour Γ. But $P_\lambda = 0$ is evidently impossible.

3. MITTAG–LEFFLER SUMMABILITY

We will give now an approximation based on the Mittag–Leffler summability method extending the

region of convergence from the set $\{z \mid d(z) > \|\lambda V\|\}$, in the case of the geometric-series development to a region, which in view of Theorem 1 cannot be extended further.

For reference we give a statement of the Mittag–Leffler summability method.[3]

Let $f(z)$ be an analytic function which is regular at the origin, so that it has a power-series development

$$f(z) = \sum_{n=0}^{\infty} c_n z^n, \quad \text{for } |z| < r \neq 0.$$

Then define the entire function

$$\hat{f}(z) = \sum_{n=0}^{\infty} \frac{c_n}{\Gamma(1 + \alpha n)} z^n, \quad 0 < \alpha \leq 1,$$

and let B_α be the set of all points z such that

$$V_z = \{v \mid \operatorname{Re}(z/v)^{1/\alpha} > 1, -\tfrac{1}{2}\alpha\pi \leq \arg(z/v) \leq \tfrac{1}{2}\alpha\pi\},$$

is contained in the principal star of $f(z)$. Then

$$f(z) = \int_0^{\infty} dt e^{-t} \hat{f}(zt^\alpha), \quad \text{for} \quad z \in B_\alpha^i (\text{interior of } B_\alpha).$$

Remark: For each compact $K \subset B_\alpha$ and each $\epsilon > 0$, there is a $T > 0$ and an integer N such that

$$\left| f(z) - \sum_0^N z^n \left(c_n \int_0^T dt e^{-t} t^{\alpha n} \right) \Big/ \Gamma(1 + \alpha n) \right| < \epsilon, \qquad z \in K.$$

Applying the above theorem to the operator-valued function $f(z) = (1 + zA)^{-1}$, where A is bounded operator acting in a Banach space, we get the following corollary.

Corollary: We have that

$$(1 + A)^{-1} = \int_0^{\infty} dt e^{-t} \sum_0^{\infty} \frac{(-At^\alpha)^n}{\Gamma(1 + \alpha n)}, \quad 0 < \alpha \leq 1,$$

if $\Sigma(A)$ is contained in the complement of the set

$$\{z \mid \operatorname{Re}(-z)^{1/\alpha} \geq 1, -\tfrac{1}{2}\alpha\pi \leq \arg(-z) \leq \tfrac{1}{2}\alpha\pi\}.$$

We remark that, for $\alpha = 1$, this takes the form

$$(1 + A)^{-1} = \int_0^{\infty} dt e^{-t(1+A)}, \qquad \text{if} \quad \Sigma(A) \subset \{z \mid \operatorname{Re} z > -1\}.$$

Proof: Because of the Mittag–Leffler theorem, it is true that

$$f(z) = (1 + zA)^{-1} = \int_0^{\infty} dt e^{-t} \sum_{n=0}^{\infty} \frac{(-zAt^\alpha)^n}{\Gamma(1 + \alpha n)}, \qquad \text{for} \quad z \in B_\alpha^i.$$

$z = 1$ is in B_α^i, if there are no singularities of $f(z)$ contained in

$$V_1 = \{v \mid \operatorname{Re}(1/v)^{1/\alpha} \geq 1, -\tfrac{1}{2}\alpha\pi \leq \arg(1/v) \leq \tfrac{1}{2}\alpha\pi\}.$$

But the singularities of $f(z)$ are just $\{v \mid -1/v \in \Sigma(A)\}$, as one sees from

$$f(z) = (1 + zA)^{-1} = (1/z) \cdot [A - (-1/z)]^{-1}.$$

From that the Corollary follows immediately.

4. APPLICATION TO THE RESOLVENT OPERATOR

To apply the result of Sec. 3 to $[1 + \lambda V R_0(z)]^{-1}$ one has to locate the spectrum of $\lambda V R_0(z)$.

Lemma 1: The spectrum of $\lambda V R_0(z)$ for $z \notin \Sigma(H_0)$ is the set of all $\mu \neq 0$ such that $z \in \Sigma(H(-\lambda/\mu))$; $\mu = 0$ may or may not belong to the spectrum.

Proof: $[\lambda V R_0(z) - \mu]^{-1}$ can be rewritten in the following form:

$$[\lambda V R_0(z) - \mu]^{-1} = (-1/\mu)(H_0 - z) \times [H_0 + (-\lambda/\mu)V - z]^{-1}, \quad \mu \neq 0.$$

If $z \in \Sigma[H(-\lambda/\mu)]$ the second factor on the right-hand side is not defined, then $\mu \in \Sigma(\lambda V R_0(z))$. But if $z \notin \Sigma(H(-\lambda/\mu))$ then the second factor is defined and its product with the unbounded operator $H_0 - z$ is a bounded operator on the whole space, because V has relative bound zero.

Lemma 2: We have that

$$[1 + \lambda V R_0(z)]^{-1} = \int_0^{\infty} dt e^{-t} \sum_{n=0}^{\infty} \frac{[-\lambda V R_0(z) t^\alpha]^n}{\Gamma(1 + \alpha n)}$$

for all z which are not in the spectrum of H_ν for

$$\nu \in S_{\lambda,\alpha} = \lambda \cdot \{\kappa \mid \operatorname{Re} \kappa^{-1/\alpha} \geq 1, -\tfrac{1}{2}\alpha\pi \leq \arg \kappa \leq \tfrac{1}{2}\alpha\pi\}.$$

Proof: This follows immediately from the corollary and the Lemma 1.

Theorem 2: If for $\lambda \in \Lambda$ no one of the branches emanating from ϵ_0, since the perturbation parameter ν varies on the straight line segment $[0, \lambda]$, crosses any other branches starting from other eigenvalues in $\Sigma'(H_0)$, then there is an $\alpha > 0$ such that

$$R_\lambda(z) = R_0(z) \int_0^{\infty} dt e^{-t} \sum_{n=0}^{\infty} \frac{[-\lambda t^\alpha V R_0(z)]^n}{\Gamma(1 + \alpha n)},$$

valid for all z on some contour Γ, such that

$$P_\lambda = \int_0^\infty dt e^{-t} \sum_{n=0}^{\infty} \frac{(-\lambda t^\alpha)^n}{\Gamma(1+\alpha n)} R_n,$$

where R_n is the residue of $R_0(z)VR_0(z)\cdots VR_0(z)$ at ϵ_0.

Proof: The only thing to prove is the existence of a contour Γ encircling ϵ_0 and ϵ_λ but no other eigenvalues without leaving the region of convergence. But the set $S_{\gamma,\alpha}$ of Lemma 2 approaches, for $\alpha \to 0$, the straight line segment $[0, \lambda]$ uniformly in α, so that the corresponding set of eigenvalues of H_ν, laying in the interior of Θ, approaches the branches coming from the eigenvalues in $\Sigma'(H_0)$ uniformly in α. This follows for the special case of bounded V from the fact that the spectrum of $H(\nu+\kappa)$ lies within a distance of $|\kappa|\,\|V\|$ from the spectrum of H_ν.

[1] T. Kato, *Perturbation theory for linear operators* (Springer-Verlag, Berlin, 1966), p. 379.
[2] Ref. 1, p. 75.
[3] G. Sansone and J. Gerretsen, *Lectures on the Theory of Functions of a Complex Variable* (P. Noordhoff Ltd., Groningen, The Netherlands, 1960), p. 432.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Analytical Solution of the Boundary-Value Problem for the Nonlinear Helmholtz Equation

H. E. WILHELM
Colorado State University, Fort Collins, Colorado 80521

(Received 17 June 1969; Revised Manuscript Received 1 October 1969)

A rigorous analytical solution in terms of the elliptic integral function of the first kind is derived for the nonlinear inhomogeneous boundary-value problem $\delta(d^2f/dx^2) = \alpha f^2 - \epsilon f - \sigma, f(s) = \tilde{f}$.

I. INTRODUCTION

Various physico-chemical phenomena are described by nonlinear differential equations of second order. Examples of such phenomena are chemical diffusion processes with unimolecular and bimolecular reactions,[1] collective diffusion processes with ionization and recombination reactions in plasmas,[2] the energy release in nuclear systems from power excursions,[3] etc. The boundary-value problem associated with the latter example was treated recently[4] by a method which is known as the expansion in Sturm–Liouville characteristic functions.[5] In this approach, the expansion coefficients are determined by an infinite set of coupled nonlinear algebraic equations,[4] which can only be solved by a finite-mode approximation in any actual evaluation.[5]

In this paper, a rigorous analytical solution in terms of the elliptic integral function of the first kind is established for the boundary-value problem of the nonlinear inhomogeneous Helmholtz equation. Thus the expansion in characteristic functions, the cumbersome determination of the expansion coefficients, and a concession in form of an approximate solution are avoided.

II. FORMULATION OF PROBLEM

Let $f(x)$ be a function defined in the closed interval $-l \le x \le +l$, and let $f(x)$, $df(x)/dx$, and $d^2f(x)/dx^2$ be continuous throughout. The function $f(x)$ is described by the nonlinear differential equation

$$\delta \frac{d^2f}{dx^2} = \alpha f^2 - \epsilon f - \sigma, \tag{1}$$

where

$$f(s) = \tilde{f}, \quad s = \pm l. \tag{2}$$

Equations (1) and (2) represent a nonlinear inhomogeneous boundary-value problem of elliptic type.[6] To relate it to a physical situation, identify $f(x)$ with a particle density, δ, α, and ϵ with the kinetic coefficients of diffusion, two-body and one-body reactions, respectively, and σ with a particle source.

With regard to the physical nature of the density function, it is assumed that the solution of Eqs. (1) and (2) is positive, unique, and uniformly bounded. Since, under physical conditions,

$$\frac{d^2f}{dx^2} = -\frac{1}{\delta}(-\alpha f^2 + \epsilon f + \sigma) < 0, \tag{3}$$

$f(x)$ exhibits no minimum. Accordingly, $f(x)$ has

valid for all z on some contour Γ, such that

$$P_\lambda = \int_0^\infty dt e^{-t} \sum_{n=0}^{\infty} \frac{(-\lambda t^\alpha)^n}{\Gamma(1+\alpha n)} R_n,$$

where R_n is the residue of $R_0(z)VR_0(z)\cdots VR_0(z)$ at ϵ_0.

Proof: The only thing to prove is the existence of a contour Γ encircling ϵ_0 and ϵ_λ but no other eigenvalues without leaving the region of convergence. But the set $S_{\gamma,\alpha}$ of Lemma 2 approaches, for $\alpha \to 0$, the straight line segment $[0, \lambda]$ uniformly in α, so that the corresponding set of eigenvalues of H_ν, laying in the interior of Θ, approaches the branches coming from the eigenvalues in $\Sigma'(H_0)$ uniformly in α. This follows for the special case of bounded V from the fact that the spectrum of $H(\nu+\kappa)$ lies within a distance of $|\kappa|\,\|V\|$ from the spectrum of H_ν.

[1] T. Kato, *Perturbation theory for linear operators* (Springer-Verlag, Berlin, 1966), p. 379.
[2] Ref. 1, p. 75.
[3] G. Sansone and J. Gerretsen, *Lectures on the Theory of Functions of a Complex Variable* (P. Noordhoff Ltd., Groningen, The Netherlands, 1960), p. 432.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Analytical Solution of the Boundary-Value Problem for the Nonlinear Helmholtz Equation

H. E. WILHELM
Colorado State University, Fort Collins, Colorado 80521

(Received 17 June 1969; Revised Manuscript Received 1 October 1969)

A rigorous analytical solution in terms of the elliptic integral function of the first kind is derived for the nonlinear inhomogeneous boundary-value problem $\delta(d^2f/dx^2) = \alpha f^2 - \epsilon f - \sigma, f(s) = \tilde{f}$.

I. INTRODUCTION

Various physico-chemical phenomena are described by nonlinear differential equations of second order. Examples of such phenomena are chemical diffusion processes with unimolecular and bimolecular reactions,[1] collective diffusion processes with ionization and recombination reactions in plasmas,[2] the energy release in nuclear systems from power excursions,[3] etc. The boundary-value problem associated with the latter example was treated recently[4] by a method which is known as the expansion in Sturm–Liouville characteristic functions.[5] In this approach, the expansion coefficients are determined by an infinite set of coupled nonlinear algebraic equations,[4] which can only be solved by a finite-mode approximation in any actual evaluation.[5]

In this paper, a rigorous analytical solution in terms of the elliptic integral function of the first kind is established for the boundary-value problem of the nonlinear inhomogeneous Helmholtz equation. Thus the expansion in characteristic functions, the cumbersome determination of the expansion coefficients, and a concession in form of an approximate solution are avoided.

II. FORMULATION OF PROBLEM

Let $f(x)$ be a function defined in the closed interval $-l \le x \le +l$, and let $f(x)$, $df(x)/dx$, and $d^2f(x)/dx^2$ be continuous throughout. The function $f(x)$ is described by the nonlinear differential equation

$$\delta \frac{d^2f}{dx^2} = \alpha f^2 - \epsilon f - \sigma, \tag{1}$$

where

$$f(s) = \tilde{f}, \quad s = \pm l. \tag{2}$$

Equations (1) and (2) represent a nonlinear inhomogeneous boundary-value problem of elliptic type.[6] To relate it to a physical situation, identify $f(x)$ with a particle density, δ, α, and ϵ with the kinetic coefficients of diffusion, two-body and one-body reactions, respectively, and σ with a particle source.

With regard to the physical nature of the density function, it is assumed that the solution of Eqs. (1) and (2) is positive, unique, and uniformly bounded. Since, under physical conditions,

$$\frac{d^2f}{dx^2} = -\frac{1}{\delta}(-\alpha f^2 + \epsilon f + \sigma) < 0, \tag{3}$$

$f(x)$ exhibits no minimum. Accordingly, $f(x)$ has

only one maximum $f=\hat{f}$ located at $x=\hat{x}$. For symmetry reasons [see Eqs. (1) and (2)], $\hat{x}$ is identical with the center $x=0$ of the interval $-l \leq x \leq +l$, i.e.,

$$\left(\frac{df}{dx}\right)_{x=\hat{x}} = \left(\frac{df}{dx}\right)_{x=0} = 0, \quad f(\hat{x}) = f(0) = \hat{f}. \tag{4}$$

The symmetry property of the boundary-value problem implies that

$$f(+x) = f(-x). \tag{5}$$

III. ANALYTICAL SOLUTION

Equation (1) is solvable by separation of the variables. To this end, Eq. (1) is multiplied by df, and subsequently integrated. Thus, one finds

$$\frac{df}{dx} = \pm\left(\frac{2}{\delta}(\tfrac{1}{3}\alpha f^3 - \tfrac{1}{2}\epsilon f^2 - \sigma f + C_1)\right)^{\frac{1}{2}}. \tag{6}$$

Here, and henceforth, the upper and lower signs apply in the semi-intervals $-l \leq x < 0$ and $0 < x \leq +l$, respectively, since

$$\begin{aligned} df &> 0, \quad \text{for } -l \leq x < 0, \\ dx &< 0, \quad \text{for } 0 < x \leq +l. \end{aligned} \tag{7}$$

According to Eq. (4), the integration constant C_1 and the maximum value $\hat{f}$ are related by

$$C_1 = -(\tfrac{1}{3}\alpha\hat{f}^3 - \tfrac{1}{2}\epsilon\hat{f}^2 - \sigma\hat{f}). \tag{8}$$

Formal integration of Eq. (6) yields, under consideration of Eq. (8),

$$\int_{f(s)}^{f(x)} \frac{df}{[\frac{1}{3}\alpha(f^3-\hat{f}^3) - \frac{1}{2}\epsilon(f^2-\hat{f}^2) - \sigma(f-\hat{f})]^{\frac{1}{2}}} = \pm\left(\frac{2}{\delta}\right)^{\frac{1}{2}}\int_s^x dx. \tag{9}$$

For the following considerations, it is necessary to rewrite Eq. (9) in the form

$$\int_{f(s)}^{f(x)} \frac{df}{[\frac{1}{3}\alpha(f-f_1)(f-f_2)(f-f_3)]^{\frac{1}{2}}} = \pm\left(\frac{2}{\delta}\right)^{\frac{1}{2}}(x-s), \tag{10}$$

where

$$f_1 = f_+, \quad f_2 = \hat{f}, \quad f_3 = f_-, \tag{11}$$

and

$$f_\pm = \frac{1}{2}\left\{\left(\frac{3}{2}\frac{\epsilon}{\alpha} - \hat{f}\right) \pm \left[\left(\frac{3}{2}\frac{\epsilon}{\alpha} - \hat{f}\right)^2 + 4\left(\frac{3}{2}\frac{\epsilon}{\alpha} - \hat{f}\right)\hat{f} + 12\frac{\sigma}{\alpha}\right]^{\frac{1}{2}}\right\}. \tag{12}$$

In Eq. (10), the problem has been reduced to a simple quadrature, and thus essentially solved. The latter can always be carried through to give elliptic integral functions of the first kind.[7]

In kinetic applications, $\epsilon/\alpha \gg 1$ cm^{-3} holds, since two-body reaction coefficients are commonly small compared to one-body reaction coefficients. Furthermore, the (maximum) particle density is generally large, $\hat{f} \gg 1$ cm^{-3}, and the particle source is by definition positive i.e., $\sigma \geq 0$. For these reasons, let the integral in Eq. (10) be evaluated for the case where all roots f_1, f_2, and f_3 are real, and

$$f_1 > f_2 > f_3. \tag{13}$$

The integral in Eq. (10) can then be brought into Legendre's canonical form by means of the substitution

$$\begin{aligned} f &= (f_2 - f_3)\sin^2\phi + f_3, \\ df &= 2(f_2 - f_3)\sin\phi\cos\phi\, d\phi, \end{aligned} \tag{14}$$

which conserves the reality conditions. Thus one finds, for Eq. (10),

$$F(\phi, k) - C_2^{\pm} = \pm[\tfrac{1}{3}\alpha(f_1 - f_3)/2\delta]^{\frac{1}{2}}x, \tag{15}$$

where $F(\phi, k)$ is the elliptic integral function of the first kind[8]

$$F(\phi, k) = \int_0^{\phi} \frac{d\psi}{(1 - k^2\sin^2\psi)^{\frac{1}{2}}} \tag{16}$$

and

$$k^2 = 1 - (f_1 - f_2)/(f_1 - f_3). \tag{17}$$

Equation (15) represents the relation $x = x(\phi)$, i.e., the relation $x = x(f)$, since by Eq. (14)

$$\phi = \arcsin[(f - f_3)/(f_2 - f_3)]^{\frac{1}{2}}. \tag{18}$$

From $x = x(f)$ the desired result $f = f(x)$ is obtained simply by inversion[9]; $C_2^{\pm}$ is determined by the boundary conditions.

(a) Central Boundary Condition:

$$f(x) = \hat{f}, \quad \text{for} \quad x = \hat{x} = 0. \tag{19}$$

According to Eq. (15), the result in this case is

$$\begin{aligned} C_2^{\pm} &= F(\hat{\phi}, k), \\ \hat{\phi} &= \phi(\hat{f}) = \tfrac{1}{2}\pi, \end{aligned} \tag{20}$$

and

$$F(\phi, k) - F(\hat{\phi}, k) = \pm[\tfrac{1}{3}\alpha(f_1 - f_3)/2\delta]^{\frac{1}{2}}x, \tag{21}$$

i.e.,

$$f(x) = (f_2 - f_3)[\mathrm{sn}\,(F(\tfrac{1}{2}\pi, k) \pm [\tfrac{1}{3}\alpha(f_1 - f_3)/2\delta]^{\frac{1}{2}}x)]^2 + f_3. \tag{22}$$

In this solution, the constants f_1, f_2, and f_3 are given by Eqs. (11) and (12), where $\hat{f}$ is the known boundary value of Eq. (19).

(*b*) *Exterior Boundary Condition:*

$$f(x) = \tilde{f}, \quad \text{for} \quad x = \pm l. \tag{23}$$

According to Eq. (15), the result in this case is

$$C_2^{\pm} = F(\tilde{\phi}, k) + [\tfrac{1}{3}\alpha(f_1 - f_3)/2\delta]^{\frac{1}{2}} l, \qquad \tilde{\phi} = \phi(\tilde{f}), \tag{24}$$

and

$$F(\phi, k) - F(\tilde{\phi}, k) = [\tfrac{1}{3}\alpha(f_1 - f_3)/2\delta]^{\frac{1}{2}}(l \pm x), \tag{25}$$

i.e.,

$$f(x) = (f_2 - f_3)\{\operatorname{sn}(F(\tilde{\phi}, k) + [\tfrac{1}{3}\alpha(f_1 - f_3)/2\delta]^{\frac{1}{2}}(l \pm x))\}^2 + f_3. \tag{26}$$

In this solution, f_1, f_2, and f_3 are constants given in terms of $\hat{f}$ by Eqs. (11) and (12), where $\hat{f}$ is defined in terms of the integration constant C_1 by Eq. (8). According to Eqs. (17), (18), and (26), $\hat{f} \equiv f$ $(x = 0)$ is given as the (positive real) root of the transcendental equation

$$\operatorname{sn}^2(F(\tilde{\phi}, k) + (\alpha/6\delta)^{\frac{1}{2}} l \hat{R}^{\frac{1}{2}}) = 1, \tag{27}$$

where

$$\tilde{\phi} \equiv \arcsin\left(\frac{\tilde{f} - \tfrac{1}{2}[(3\epsilon/2\alpha - \hat{f}) - \hat{R}]}{\hat{f} - \tfrac{1}{2}[(3\epsilon/2\alpha - \hat{f}) - \hat{R}]}\right)^{\frac{1}{2}}, \tag{28}$$

$$k^2 \equiv \tfrac{1}{2}[1 - (3\epsilon/2\alpha - 3\hat{f})\hat{R}^{-1}], \tag{29}$$

$$\hat{R} \equiv \left[\left(\frac{3}{2}\frac{\epsilon}{\alpha} - \hat{f}\right)^2 + 4\left(\frac{3}{2}\frac{\epsilon}{\alpha} - \hat{f}\right)\hat{f} + 12\frac{\sigma}{\alpha}\right]^{\frac{1}{2}}. \tag{30}$$

From Eq. (27), $\hat{f}$ is readily obtained by iteration as soon as the physical constants α, δ, ϵ, σ, l, and $\tilde{\phi}$ are specified. A first trial value for $\hat{f}$ is given by $\hat{f}_0 = \frac{1}{2}\{(\epsilon/\alpha) + [(\epsilon/\alpha)^2 + 4(\sigma/\alpha)]^{\frac{1}{2}}\}$, since $\hat{f} \leqslant \hat{f}_0$ for physical reasons.

[Note that the upper sign applies to the semi-interval $-l \leq x \leq 0$, and the lower sign to the semi-interval $0 \leq x \leq +l$; Eqs. (15), (21), (22), (25), and (26).]

If the roots f_1, f_2, and f_3 are real but satisfy inequalities other than those given in Eq. (13), the integral in Eq. (10) can be evaluated by means of a quadratic transformation similar to that of Eq. (14).[7] In the alternative case where one root is real and the other two roots are conjugate complex, the integral in Eq. (10) is again reducible to elliptic integral functions of the first kind by means of a so-called linear transformation.[7]

IV. CONCLUSION

The boundary-value problem for the nonlinear inhomogeneous Helmholtz equation has been solved analytically by the method of separation of variables. In this approach, the independent variable is obtained as a function $x(f)$ of the dependent variable f, the functional relationship involving elliptic integral functions of the first kind. From this result, the dependent variable is obtained as a function $f(x)$ of the independent variable x by inversion, which gives the final results in terms of the sine amplitude function.

ACKNOWLEDGMENT

The investigation was sponsored by the U.S. Office of Naval Research under the auspices of J. A. Satkowski.

[1] E. E. Petersen, *Chemical Reaction Analysis* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1965).

[2] E. W. McDaniel, *Collision Phenomena in Ionized Gases* (John Wiley & Sons, Inc., New York, 1964).

[3] W. K. Ergen, Trans. Am. Nucl. Soc. **8**, 221 (1965).

[4] J. Canosa, J. Math. Phys. **8**, 2180 (1967).

[5] M. R. Siddiqi, Math. Z. **35**, 464 (1932); **40**, 484 (1935).

[6] The special boundary-value problem considered by Canosa (Ref. 4) obtains formally by putting $\delta = 1$, $\alpha = c$, $\epsilon = 1$, $\sigma = 0$, $\tilde{f} = 0$.

[7] J. Edwards, *Treatise on Integral Calculus* (Chelsea Publ. Co., New York, 1922).

[8] There exist comprehensive tables for $F(\varphi, k)$ [for example, *Handbook of Mathematical Functions*, M. Abramowitz and I. E. Stegun, Eds. (Dover Publications, Inc., New York, 1965)], which is given in the series representation by

$$F(\phi, k) = \sum_{m=0}^{\infty} \binom{-\frac{1}{2}}{m} (-k^2)^m \int_0^{\phi} \sin^{2m}\phi \, d\phi,$$

$$\int_0^{\phi} \sin^{2m}\phi \, d\phi = \frac{(1; 2; m)}{2^m m!}\phi - \tfrac{1}{2}\sin\phi\cos\phi \sum_{\nu=1}^{m} \frac{(2m-1; -2; \nu-1)}{2^{\nu-1}(m; -1; \nu)} \sin^{2(m-\nu)}\phi.$$

[9] The elliptic sine function has the series representation

$$\operatorname{sn} u = \frac{2\pi}{kK} \sum_{m=1}^{\infty} \frac{q^{m-\frac{1}{2}}}{1 - q^{2m-1}} \sin(2m-1)\frac{\pi u}{2K},$$

where

$$q = \exp(-\pi K'/K), \quad K = F(\tfrac{1}{2}\pi, k), \quad K' = F(\tfrac{1}{2}\pi, k'), \quad k' = (1 - k^2)^{\frac{1}{2}}$$

[*Tables of Integrals, Series and Products*, I. S. Gradshteyn and I. M. Ryzhik, Eds. (Academic Press Inc., New York, 1966)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Calculation of Lattice Green's Functions*

J. OITMAA
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta, Canada

(Received 22 September 1969)

A proof is given that, in the numerical evaluation of lattice Green's functions, it is possible to restrict the wave-vector summation to an irreducible section of the Brillouin zone. It is shown that, in order to use this simplification, the Green's functions must be appropriately symmetrized.

I. INTRODUCTION

Many quantities of interest in solid state physics can be expressed as sums of the following form:

$$F = \sum_{\mathbf{k}} f(\mathbf{k}), \tag{1}$$

where the allowed values of the wave vector $\mathbf{k}$ form a uniformly distributed set of points throughout the first Brillouin zone of the lattice.

In practice it is not necessary to evaluate the sum (1) over the whole Brillouin zone. By using the symmetry of the crystal, it is possible to restrict the sum to $\mathbf{k}$ values lying within a certain "irreducible section" of the zone, each term being multiplied by an appropriate weighting factor. This procedure is well known for the calculation of phonon frequency distribution functions.[1,2]

Another class of functions which can be expressed in the form of (1) are the lattice Green's functions.[3] These functions are useful in calculating the response of a perfect harmonic crystal to perturbations, such as anharmonic terms or defects. Many authors[4,5] have evaluated the Green's functions numerically by restricting the summation to an irreducible part of the Brillouin zone, but to our knowledge a formal proof that this can be done has never been presented. It is our intention in this paper to present such a proof.

II. THE PROOF

The general lattice dynamical Green's function is given by[3]

$$G_{\alpha\beta}(l\kappa, l'\kappa') = \frac{1}{N(M_\kappa M_{\kappa'})^{\frac{1}{2}}} \sum_{\mathbf{k}j} \frac{w_\alpha(\kappa \mid \mathbf{k}j) w_\beta^*(\kappa' \mid \mathbf{k}j)}{\omega^2 - \omega_j^2(\mathbf{k})} \times e^{i\mathbf{k}\cdot[\mathbf{x}(l\kappa) - \mathbf{x}(l'\kappa')]}, \tag{2}$$

where the symbols have the usual meanings.

The crystal is invariant under operations of the space group. Such a symmetry operation is denoted by $\{\mathbf{S} \mid \mathbf{v}(S) + \mathbf{x}(m)\}$, where $\mathbf{S}$ is a proper or improper rotation, $\mathbf{v}(S)$ is a vector smaller than any primitive translation vector of the crystal, and $\mathbf{x}(m)$ is any translation vector of the crystal. We consider a subgroup G of the space group, which consists of those proper and improper rotations $\mathbf{S}$ for which $\mathbf{v}(S) = \mathbf{0}$. For symmorphic space groups, G is simply the point group of the crystal; for nonsymmorphic space groups G is a subgroup of the point group. The number of elements in G will be denoted by h.

If we apply any symmetry operation $\mathbf{S}$ in G to the crystal, at the lattice site $(l\kappa)$, then the Green's function transforms according to[3]

$$G_{\alpha\beta}(l\kappa, l'\kappa') = \sum_{\gamma\delta} S_{\alpha\gamma} S_{\beta\delta} G_{\gamma\delta}(l\kappa, L'\mathrm{K}'), \tag{3}$$

where $(l'\kappa')$ is the lattice site into which $(L'\mathrm{K}')$ is taken by the operation $\mathbf{S}$, i.e.,

$$\mathbf{x}(l'\kappa') = \mathbf{S}\mathbf{x}(L'\mathrm{K}').$$

Using (3), we then write the Green's function in the symmetrized form

$$\begin{aligned} G_{\alpha\beta}(l\kappa, l'\kappa') &= h^{-1} \sum_{\mathbf{S}} \sum_{\gamma\delta} S_{\alpha\gamma} S_{\beta\delta} G_{\gamma\delta}(l\kappa, L'\mathrm{K}') \\ &= N^{-1} \sum_{\mathbf{k}j} g(\mathbf{k}j), \end{aligned} \tag{4}$$

where

$$\begin{aligned} g(\mathbf{k}j) = (M_\kappa M_{\kappa'})^{-1} &[\omega^2 - \omega_j^2(\mathbf{k})]^{-1} h^{-1} \\ &\times \sum_{\mathbf{S}} e^{i\mathbf{k}\cdot[\mathbf{x}(l\kappa) - \mathbf{S}^{-1}\mathbf{x}(l'\kappa')]} \\ &\times \sum_{\gamma\delta\kappa_1} S_{\alpha\gamma} S_{\beta\delta} \delta(F(\kappa_1; S), \kappa') \\ &\times w_\gamma(\kappa \mid \mathbf{k}j) w_\delta^*(\kappa_1, \mathbf{k}_j), \end{aligned} \tag{5}$$

where $F(\kappa_1; S)$ denotes the position to which an atom of type κ_1 is taken by the operation $\mathbf{S}$.

The sum over $\mathbf{k}$ in (4) is over the whole Brillouin zone. We want to show that we can effectively restrict this to an "irreducible section" equal in volume to h^{-1} of the whole zone. To show this we form the quantity $g(\mathbf{R}\mathbf{k}j)$, where $\mathbf{R}$ is any rotation in G, as follows:

$$\begin{aligned} g(\mathbf{R}\mathbf{k}j) = (M_\kappa M_{\kappa'})^{-1} &[\omega^2 - \omega_j^2(\mathbf{R}\mathbf{k})]^{-1} h^{-1} \\ &\times \sum_{\mathbf{S}} e^{i\mathbf{R}\mathbf{k}\cdot[\mathbf{x}(l\kappa) - \mathbf{S}^{-1}\mathbf{x}(l'\kappa')]} \\ &\times \sum_{\gamma\delta\kappa_1} S_{\alpha\gamma} S_{\beta\delta} \delta(F(\kappa_1; S), \kappa') \\ &\times w_\gamma(\kappa \mid \mathbf{R}\mathbf{k}j) w_\gamma^*(\kappa_1, \mathbf{R}\mathbf{k}j). \end{aligned} \tag{6}$$

The transformation properties of the frequencies and eigenvectors of the normal modes of a crystal

have been worked out by Maradudin and Vosko.[6] For the present case they are

$$\omega_j^2(\mathbf{Rk}) = \omega_j^2(\mathbf{k}), \tag{7a}$$

$$w_\alpha(\kappa \mid \mathbf{Rk}j) = \sum_{\kappa'\beta} R_{\alpha\beta}\delta(\kappa, F(\kappa'; R))w_\beta(\kappa' \mid \mathbf{k}j). \tag{7b}$$

We also simplify the exponent in (6) as follows:

$$\begin{aligned}\mathbf{Rk}\cdot[\mathbf{x}(l\kappa) - \mathbf{S}^{-1}\mathbf{x}(l'\kappa')] &= \mathbf{k}\cdot[\mathbf{R}^{-1}\mathbf{x}(l\kappa) - \mathbf{R}^{-1}\mathbf{S}^{-1}\mathbf{x}(l'\kappa')] \\ &= \mathbf{k}\cdot[\mathbf{x}(l\kappa) - (\mathbf{SR})^{-1}\mathbf{x}(l'\kappa')].\end{aligned}$$

Then (6) becomes

$$\begin{aligned} g(\mathbf{Rk}j) &= (M_\kappa M_{\kappa'})^{-1}[\omega^2 - \omega_j^2(\mathbf{k})]^{-1}h^{-1} \\ &\times \sum_{\mathbf{S}} e^{i\mathbf{k}\cdot[\mathbf{x}(l\kappa)-(\mathbf{SR})^{-1}\mathbf{x}(l'\kappa')]} \\ &\times \sum_{\gamma\delta\kappa_1} S_{\alpha\gamma}S_{\beta\delta}\delta(F(\kappa_1; S), \kappa') \\ &\times \sum_{\lambda\mu\kappa_2\kappa_3} R_{\gamma\lambda}\delta(\kappa, F(\kappa_2; R))w_\lambda(\kappa_2 \mid \mathbf{k}j) \\ &\times R_{\delta\mu}\delta(\kappa_1, F(\kappa_3; \mathbf{R}))w_\mu^*(\kappa_3 \mid \mathbf{k}j) \\ &= (M_\kappa M_{\kappa'})^{-1}[\omega^2 - \omega_j^2(\mathbf{k})]^{-1}h^{-1} \\ &\times \sum_{\mathbf{S}} e^{i\mathbf{k}\cdot[\mathbf{x}(l\kappa)-(\mathbf{SR})^{-1}\mathbf{x}(l'\kappa')]} \\ &\times \sum_{\lambda\mu\kappa_3} (SR)_{\alpha\lambda}(SR)_{\beta\mu}\delta(F(\kappa_3; SR), \kappa') \\ &\times w_\lambda(\kappa \mid \mathbf{k}j)w_\mu^*(\kappa_3 \mid \mathbf{k}j). \end{aligned} \tag{8}$$

We then use the rearrangement theorem from group theory to replace the summation variable $\mathbf{S}$ by $\mathbf{S}' = \mathbf{SR}$. It then follows immediately that

$$g(\mathbf{Rk}j) = g(\mathbf{k}j). \tag{9}$$

Hence to evaluate the sum in Eq. (4) it is necessary to choose a section of the Brillouin zone such that any value of $\mathbf{k}$ outide this "irreducible section" is related to a value of $\mathbf{k}$ inside the section by one of the operations $\mathbf{R}$ of the group G. This irreducible section will have a volume h^{-1} of the volume of the entire zone. The sum in Eq. (4) can then be written as

$$G_{\alpha\beta}(l\kappa, l'\kappa') = n^{-1}\sum_{\mathbf{q}j} w(\mathbf{q})g(\mathbf{q}j), \tag{10}$$

where $\mathbf{q}$ is a $\mathbf{k}$-vector inside the irreducible part of the zone, $w(\mathbf{q})$ is a weight which is equal to h for values of $\mathbf{q}$ inside the irreducible section and smaller for points on the surface of the section, and $n = \sum_{\mathbf{q}} w(\mathbf{q})$.

III. AN EXAMPLE

To illustrate the method we evaluate a Green's function for the diamond structure. For this structure the group G is the point group T_d, with 24 elements. Thus the irreducible section is $\frac{1}{24}$ of the Brillouin zone. A convenient choice for this section is shown in Fig. 1. We will consider the Green's function $G_{xx}(01, 12)$, the labeling scheme for the atoms being shown in Fig. 2. The 24 matrix representations of the rotations of T_d are listed by Kovalev.[7] Using these, we obtain for $g(\mathbf{k}j)$ the result

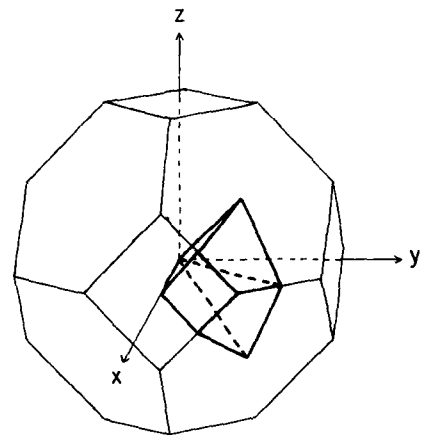


FIG. 1. Irreducible section of Brillouin zone for diamond structure.

$$\begin{aligned} g(\mathbf{k}j) &= M^{-1}[\omega^2 - \omega_j^2(\mathbf{k})]^{-1} \\ &\times \tfrac{1}{8}[w_x(1 \mid \mathbf{k}j)w_x^*(2 \mid \mathbf{k}j) \\ &+ w_y(1 \mid \mathbf{k}j)w_y^*(2 \mid \mathbf{k}j) + w_z(1 \mid \mathbf{k}j)w_z^*(2 \mid \mathbf{k}j)] \\ &\times (\cos\tfrac{1}{4}k_x a_0 \cos\tfrac{1}{4}k_y a_0 \cos\tfrac{1}{4}k_z a_0 \\ &+ i \sin\tfrac{1}{4}k_x a_0 \sin\tfrac{1}{4}k_y a_0 \sin\tfrac{1}{4}k_z a_0). \end{aligned}$$

The Green's function $G_{xx}(01, 12)$ is then given by

$$\begin{aligned} G_{xx}(01, 12) &= (3M)^{-1}n^{-1}\sum_{\mathbf{q}j} w(\mathbf{q})[\omega^2 - \omega_j^2(\mathbf{q})]^{-1} \\ &\times [w_x(1 \mid \mathbf{q}j)w_x^*(2 \mid \mathbf{q}j) \\ &+ w_y(1 \mid \mathbf{q}j)w_y^*(2 \mid \mathbf{q}j) + w_z(1 \mid \mathbf{q}j)w_z^*(2 \mid \mathbf{q}j) \\ &\times (\cos\tfrac{1}{4}q_x a_0 \cos\tfrac{1}{4}q_y a_0 \cos\tfrac{1}{4}q_z a_0 \\ &+ i \sin\tfrac{1}{4}q_x a_0 \sin\tfrac{1}{4}q_y a_0 \sin\tfrac{1}{4}q_z a_0), \end{aligned}$$

where $\mathbf{q}$ is a wave vector in the irreducible section of the Brillouin zone.

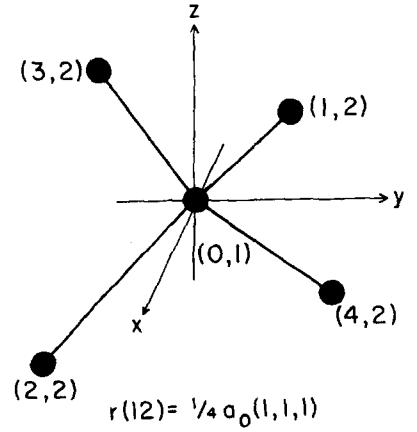


FIG. 2. Labeling of nearest-neighbor atoms for diamond structure.

ACKNOWLEDGMENT

The idea of doing this work arose from conversations with A. Grimm and Dr. A. A. Maradudin.

* Work supported in part by the National Research Council of Canada.

[1] E. W. Kellermann, Phil. Trans. Roy. Soc. (London) **238**, 513 (1940).
[2] A. M. Karo, J. Chem. Phys. **31**, 1489 (1959).
[3] A. A. Maradudin, Rept. Progr. Phys. **28**, 331 (1965).
[4] N. X. Xinh, A. A. Maradudin, and R. A. Coldwell-Horsfall, J. Phys. Radium **26**, 717 (1965).
[5] T. P. Martin, Phys. Rev. **160**, 686 (1967).
[6] A. A. Maradudin and S. H. Vosko, Rev. Mod. Phys. **40**, 1 (1968).
[7] O. V. Kovalev, *Irreducible Representations of the Space Group* (Gordon and Breach, Science Publishers, New York, 1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On the Solution of the Differential Equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0$$

M. KOLSRUD
Institute for Theoretical Physics, University of Oslo, Oslo, Norway

(Received 21 June 1968; Revised Manuscript Received 5 September 1969)

The following solution is obtained:

$$P(x, y, t) = e^{-\gamma xy - \delta} P_0(e^{-\alpha}x, e^{-\beta}y, \epsilon),$$

where $P_0(x, y, t)$ is the solution when the constants a, b, and c are zero, and where the time functions $\alpha, \cdots, \epsilon$ are given explicitly.

In a recent paper,[1] where the above-mentioned equation is solved by a series expansion, the author says: "It appears that the solution of the equation in question is not available in the literature, nor does it seem feasible to solve it by conventional methods." The purpose of the present note is to draw attention to two papers[2] which give the solution of the more general equation ($\partial_i = \partial/\partial x_i$, $\partial_t = \partial/\partial t$):

$$(A_{ij}\partial_i\partial_j + a_i\partial_i + B_{ij}x_ix_j + b_ix_i + C_{ij}x_i\partial_j + \partial_t)\psi = 0, \quad (1)$$

where the coefficients are arbitrary functions of t. Rather than specializing the general solution to the present case, we want to apply our method directly to the equation in question, viz.,

$$(\partial_x\partial_y + ax\partial_x + by\partial_y + cxy + \partial_t)P = 0, \quad (2a)$$

with

$$P(x, y, 0) = \Phi(x, y), \quad (2b)$$

where $\Phi(x, y)$ is given. The coefficients a, b, and c are constants. We introduce the operator T by

$$P(x, y, t) = T(\partial_x, \partial_y, x, y, t)\Phi(x, y), \quad (3)$$

and thus, from (2) and (3)—as Φ is arbitrary[3]—we get

$$(\partial_x\partial_y + ax\partial_x + by\partial_y + cxy)T = -\partial T/\partial t, \quad (4)$$

with

$$T = 1 \quad \text{for} \quad t = 0. \quad (5)$$

The operator T is assumed to be of the form

$$T = e^{-\gamma xy - \delta}e^{-\alpha x\partial_x - \beta y\partial_y}e^{-\epsilon\partial_x\partial_y} = T_1T_2T_3, \quad (6)$$

where the time functions $\alpha, \beta, \cdots, \epsilon$ shall vanish for $t = 0$. By writing $\dot{T} = \partial T/\partial t$, etc., we get from (4) and (6)

$$\begin{aligned}\partial_x\partial_y + ax\partial_x + by\partial_y + cxy &= -\dot{T}T^{-1} \\ &= (\dot{T}_1T_2T_3 + T_1\dot{T}_2T_3 + T_1T_2\dot{T}_3)T_3^{-1}T_2^{-1}T_1^{-1} \\ &= -\dot{T}_1T_1^{-1} - T_1(\dot{T}_2T_2^{-1})T_1^{-1} - T_1(T_2(\dot{T}_3T_3^{-1})T_2^{-1})T_1^{-1}.\end{aligned} \quad (7)$$

Each term is now evaluated:

$$-\dot{T}_1T_1^{-1} = \dot{\gamma}xy + \dot{\delta}, \quad (8)$$

$$\begin{aligned}-T_1(\dot{T}_2T_2^{-1})T_1^{-1} &= e^{-\gamma xy}(\dot{\alpha}x\partial_x + \dot{\beta}y\partial_y)e^{\gamma xy} \\ &= \dot{\alpha}x\partial_x + \dot{\beta}y\partial_y + (\dot{\alpha} + \dot{\beta})\gamma xy,\end{aligned} \quad (9)$$

$$-T_2(\dot{T}_3T_3^{-1})T_2^{-1} = \dot{\epsilon}T_2\partial_xT_2^{-1}T_2\partial_yT_2^{-1},$$

where

$$\begin{aligned}T_2\partial_xT_2^{-1} &= e^{-\alpha x\partial_x}\partial_xe^{\alpha x\partial_x} \\ &= \partial_x - \alpha[x\partial_x, \partial_x] + (\alpha^2/2!)[x\partial_x, [x\partial_x, \partial x]] - \cdots \\ &= \partial_x + \alpha\partial_x + (\alpha^2/2!)\partial_x + \cdots = e^{\alpha}\partial_x.\end{aligned}$$

ACKNOWLEDGMENT

The idea of doing this work arose from conversations with A. Grimm and Dr. A. A. Maradudin.

* Work supported in part by the National Research Council of Canada.

[1] E. W. Kellermann, Phil. Trans. Roy. Soc. (London) **238**, 513 (1940).
[2] A. M. Karo, J. Chem. Phys. **31**, 1489 (1959).
[3] A. A. Maradudin, Rept. Progr. Phys. **28**, 331 (1965).
[4] N. X. Xinh, A. A. Maradudin, and R. A. Coldwell-Horsfall, J. Phys. Radium **26**, 717 (1965).
[5] T. P. Martin, Phys. Rev. **160**, 686 (1967).
[6] A. A. Maradudin and S. H. Vosko, Rev. Mod. Phys. **40**, 1 (1968).
[7] O. V. Kovalev, *Irreducible Representations of the Space Group* (Gordon and Breach, Science Publishers, New York, 1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On the Solution of the Differential Equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0$$

M. KOLSRUD
Institute for Theoretical Physics, University of Oslo, Oslo, Norway

(Received 21 June 1968; Revised Manuscript Received 5 September 1969)

The following solution is obtained:

$$P(x, y, t) = e^{-\gamma xy - \delta} P_0(e^{-\alpha}x, e^{-\beta}y, \epsilon),$$

where $P_0(x, y, t)$ is the solution when the constants a, b, and c are zero, and where the time functions $\alpha, \cdots, \epsilon$ are given explicitly.

In a recent paper,[1] where the above-mentioned equation is solved by a series expansion, the author says: "It appears that the solution of the equation in question is not available in the literature, nor does it seem feasible to solve it by conventional methods." The purpose of the present note is to draw attention to two papers[2] which give the solution of the more general equation ($\partial_i = \partial/\partial x_i$, $\partial_t = \partial/\partial t$):

$$(A_{ij}\partial_i\partial_j + a_i\partial_i + B_{ij}x_ix_j + b_ix_i + C_{ij}x_i\partial_j + \partial_t)\psi = 0, \quad (1)$$

where the coefficients are arbitrary functions of t. Rather than specializing the general solution to the present case, we want to apply our method directly to the equation in question, viz.,

$$(\partial_x\partial_y + ax\partial_x + by\partial_y + cxy + \partial_t)P = 0, \quad (2a)$$

with

$$P(x, y, 0) = \Phi(x, y), \quad (2b)$$

where $\Phi(x, y)$ is given. The coefficients a, b, and c are constants. We introduce the operator T by

$$P(x, y, t) = T(\partial_x, \partial_y, x, y, t)\Phi(x, y), \quad (3)$$

and thus, from (2) and (3)—as Φ is arbitrary[3]—we get

$$(\partial_x\partial_y + ax\partial_x + by\partial_y + cxy)T = -\partial T/\partial t, \quad (4)$$

with

$$T = 1 \quad \text{for} \quad t = 0. \quad (5)$$

The operator T is assumed to be of the form

$$T = e^{-\gamma xy - \delta}e^{-\alpha x\partial_x - \beta y\partial_y}e^{-\epsilon\partial_x\partial_y} = T_1T_2T_3, \quad (6)$$

where the time functions $\alpha, \beta, \cdots, \epsilon$ shall vanish for $t = 0$. By writing $\dot{T} = \partial T/\partial t$, etc., we get from (4) and (6)

$$\begin{aligned}\partial_x\partial_y + ax\partial_x + by\partial_y + cxy &= -\dot{T}T^{-1} \\ &= (\dot{T}_1T_2T_3 + T_1\dot{T}_2T_3 + T_1T_2\dot{T}_3)T_3^{-1}T_2^{-1}T_1^{-1} \\ &= -\dot{T}_1T_1^{-1} - T_1(\dot{T}_2T_2^{-1})T_1^{-1} - T_1(T_2(\dot{T}_3T_3^{-1})T_2^{-1})T_1^{-1}.\end{aligned} \quad (7)$$

Each term is now evaluated:

$$-\dot{T}_1T_1^{-1} = \dot{\gamma}xy + \dot{\delta}, \quad (8)$$

$$\begin{aligned}-T_1(\dot{T}_2T_2^{-1})T_1^{-1} &= e^{-\gamma xy}(\dot{\alpha}x\partial_x + \dot{\beta}y\partial_y)e^{\gamma xy} \\ &= \dot{\alpha}x\partial_x + \dot{\beta}y\partial_y + (\dot{\alpha} + \dot{\beta})\gamma xy,\end{aligned} \quad (9)$$

$$-T_2(\dot{T}_3T_3^{-1})T_2^{-1} = \dot{\epsilon}T_2\partial_xT_2^{-1}T_2\partial_yT_2^{-1},$$

where

$$\begin{aligned}T_2\partial_xT_2^{-1} &= e^{-\alpha x\partial_x}\partial_xe^{\alpha x\partial_x} \\ &= \partial_x - \alpha[x\partial_x, \partial_x] + (\alpha^2/2!)[x\partial_x, [x\partial_x, \partial x]] - \cdots \\ &= \partial_x + \alpha\partial_x + (\alpha^2/2!)\partial_x + \cdots = e^{\alpha}\partial_x.\end{aligned}$$

Correspondingly,

$$T_2\partial_y T_2^{-1} = e^{-\beta y\partial_y}\partial_y e^{\beta y\partial_y} = e^{\beta}\partial_y.$$

The last term in (7) thus becomes

$$\begin{aligned}
&-T_1(T_2(\dot{T}_3T_3^{-1})T_2^{-1})T_1^{-1}\\
&\quad = \dot{\epsilon}e^{\alpha+\beta}(e^{-\gamma xy}\partial_x e^{\gamma xy})(e^{-\gamma xy}\partial_y e^{\gamma xy})\\
&\quad = \dot{\epsilon}e^{\alpha+\beta}(\partial_x + \gamma y)(\partial_y + \gamma x)\\
&\quad = \dot{\epsilon}e^{\alpha+\beta}(\partial_x\partial_y + \gamma x\partial_x + \gamma y\partial_y + \gamma^2 xy + \gamma). \quad (10)
\end{aligned}$$

Inserting (8)–(10) in (7), we get

$$\begin{aligned}
&\partial_x\partial_y + ax\partial_x + by\partial_y + cxy\\
&\qquad = \dot{\epsilon}e^{\alpha+\beta}\partial_x\partial_y + (\dot{\alpha} + \gamma\dot{\epsilon}e^{\alpha+\beta})x\partial_x\\
&\qquad\quad + (\dot{\beta} + \gamma\dot{\epsilon}e^{\alpha+\beta})y\partial_y\\
&\qquad\quad + (\dot{\gamma} + \gamma^2\dot{\epsilon}e^{\alpha+\beta} + \gamma(\dot{\alpha} + \dot{\beta}))xy\\
&\qquad\quad + \gamma\dot{\epsilon}e^{\alpha+\beta} + \dot{\delta}. \quad (11)
\end{aligned}$$

The time functions $\alpha, \cdots, \epsilon$ must therefore satisfy the equations

$$\dot{\epsilon}e^{\alpha+\beta} = 1, \quad (12a)$$

$$\dot{\alpha} + \gamma = a, \quad (12b)$$

$$\dot{\beta} + \gamma = b, \quad (12c)$$

$$\dot{\gamma} + \gamma^2 + \gamma(\dot{\alpha} + \dot{\beta}) = c, \quad (12d)$$

$$\gamma + \dot{\delta} = 0, \quad (12e)$$

and vanish for $t = 0$. [These equations (12) are also valid if a, b, and c are functions of the time t.] With

$$\lambda_{1,2} = -\tfrac{1}{2}(a + b) \pm \{[\tfrac{1}{2}(a + b)]^2 - c\}^{\frac{1}{2}}, \quad (13)$$

we get the solutions of (12) as

$$\alpha = at + \log\frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2 - \lambda_1}, \quad (14a)$$

$$\beta = bt + \log\frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2 - \lambda_1}, \quad (14b)$$

$$\gamma = -\frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{e^{\lambda_1 t}/\lambda_1 - e^{\lambda_2 t}/\lambda_2}, \quad (14c)$$

$$\delta = \log\frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2 - \lambda_1}, \quad (14d)$$

$$\epsilon = -\frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}. \quad (14e)$$

When

$$c = \tfrac{1}{4}(a + b)^2, \quad (15)$$

i.e.,

$$\lambda_1 = \lambda_2 = -\tfrac{1}{2}(a + b), \quad (16)$$

the solutions (14) become

$$\alpha = \tfrac{1}{2}(a - b)t + \log[1 + \tfrac{1}{2}(a + b)t], \quad (17a)$$

$$\beta = \tfrac{1}{2}(b - a)t + \log[1 + \tfrac{1}{2}(a + b)t], \quad (17b)$$

$$\gamma = \frac{\tfrac{1}{4}(a + b)^2 t}{1 + \tfrac{1}{2}(a + b)t}, \quad (17c)$$

$$\delta = -\tfrac{1}{2}(a + b)t + \log[1 + \tfrac{1}{2}(a + b)t], \quad (17d)$$

$$\epsilon = \frac{t}{1 + \tfrac{1}{2}(a + b)t}. \quad (17e)$$

The solution (3) of Eq. (2) shall now be expressed by the function

$$P_0(x, y, \epsilon) = e^{-\epsilon\partial_x\partial_y}\Phi(x, y), \quad (18)$$

which satisfies the equations

$$\left(\frac{\partial^2}{\partial x\partial y} + \frac{\partial}{\partial\epsilon}\right)P_0 = 0, \quad (19)$$

$$P_0(x, y, 0) = \Phi(x, y). \quad (20)$$

This is a special case of Eq. (2), viz., with $a = b = c = 0$, and with t substituted by the symbol ϵ. (P_0 corresponds to P_{II} in Ref. 1.)

From (3), (6), and (18), we now get the solution of (2) as

$$\begin{aligned}
P(x, y, t) &= e^{-\gamma xy-\delta}e^{-\alpha x\partial_x-\beta y\partial_y}e^{-\epsilon\partial_x\partial_y}\Phi(x, y)\\
&= e^{-\gamma xy-\delta}e^{-\alpha x\partial_x}e^{-\beta y\partial_y}P_0(x, y, \epsilon)\\
&= e^{-\gamma xy-\delta}P_0(e^{-\alpha}x, e^{-\beta}y, \epsilon). \quad (21)
\end{aligned}$$

Here we have used

$$\begin{aligned}
e^{-\alpha x\partial_x}f(x) &= \exp[-\alpha\partial/\partial(\log x)]f[\exp(\log x)]\\
&= f[\exp(\log x - \alpha)] = f(e^{-\alpha}x), \quad (22)
\end{aligned}$$

etc. Putting $t = 0$ in the Eqs. (14), and using (20), we obtain (2b).

Neuringer[4] has given the following solution of Eq. (2):

$$P = P'(x, y, t) = e^{\lambda(xy-t)}\iint dk_x\, dk_y G\Psi, \quad (23)$$

where

$$\begin{aligned}
\Psi &= \frac{1}{4\pi^2}\iint dx'\, dy'\\
&\quad \times \exp[i(k_x x' + k_y y') - \lambda x'y']\Phi(x', y'), \quad (24)
\end{aligned}$$

$$G = \exp[-i(k_x\xi + k_y\eta) + k_x k_y\zeta], \quad (25)$$

$$\xi = xe^{-(a+\lambda)t}, \quad \eta = ye^{-(b+\lambda)t}, \quad \zeta = \frac{1 - e^{-(a+b+2\lambda)t}}{a + b + 2\lambda}, \quad (26)$$

and with λ equal to either of the forms (13).

The equivalence of (23) with our solution (21) can be shown in the following way: From (24) we have

$$\Psi = e^{\lambda\partial^2/\partial k_x\partial k_y}\Psi_0, \tag{27}$$

where

$$\Psi_0 = \frac{1}{4\pi^2}\iint dx'\,dy' e^{i(k_x x' + k_y y')}\Phi(x', y'). \tag{28}$$

Hence (23) may be written

$$P'(x, y, t) = e^{\lambda(xy-t)}\iint dk_x\,dk_y G e^{\lambda\partial^2/\partial k_x\partial k_y}\Psi_0 \tag{29}$$

$$= e^{\lambda(xy-t)}\iint dk_x\,dk_y \Psi_0 G', \tag{30}$$

where

$$G' = e^{\lambda\partial^2/\partial k_x\partial k_y}G. \tag{31}$$

Here it is assumed that Ψ_0 has the necessary behavior when $|\mathbf{k}| \to \infty$. Considering (25) and (31), we first observe that

$$\left(\frac{\partial^2}{\partial k_x\partial k_y} + \zeta\frac{\partial}{\partial v}\right) \times \frac{1}{v}\exp\left\{\frac{1}{v}\left[-i(k_x\xi + k_y\eta) + k_x k_y\zeta - \frac{\xi\eta}{\zeta}\right]\right\} = 0. \tag{32}$$

Thus (31) may be evaluated as follows:

$$\begin{aligned} G' &= e^{\lambda\partial^2/\partial k_x\partial k_y}\exp\{-i(k_x\xi + k_y\eta) + k_x k_y\zeta\} \\ &= \lim_{v\to 1} v\exp\left(\frac{\xi\eta}{\zeta v}\right)e^{\lambda\partial^2/\partial k_x\partial k_y} \\ &\quad\times\frac{1}{v}\exp\left\{\frac{1}{v}\left[-i(k_x\xi + k_y\eta) + k_x k_y\zeta - \frac{\xi\eta}{\zeta}\right]\right\} \\ &= \lim_{v\to 1} v\exp\left(\frac{\xi\eta}{\zeta v}\right)e^{-\lambda\zeta\partial/\partial v} \\ &\quad\times\frac{1}{v}\exp\left\{\frac{1}{v}\left[-i(k_x\xi + k_y\eta) + k_x k_y\zeta - \frac{\xi\eta}{\zeta}\right]\right\} \\ &= \lim_{v\to 1} v\exp\left(\frac{\xi\eta}{\zeta v}\right)\frac{1}{v - \lambda\zeta} \\ &\quad\times\exp\left\{\frac{1}{v - \lambda\zeta}\left[-i(k_x\xi + k_y\eta) + k_x k_y\zeta - \frac{\xi\eta}{\zeta}\right]\right\} \\ &= \frac{1}{1 - \lambda\zeta}\exp\left(\frac{\lambda\xi\eta}{\lambda\zeta - 1}\right) \\ &\quad\times\exp\left\{\frac{1}{1 - \lambda\zeta}[-i(k_x\xi + k_y\eta) + k_x k_y\zeta]\right\}. \end{aligned} \tag{33}$$

Inserting (26) in (33), we get, for each λ,

$$e^{\lambda(xy-t)}G' = e^{-\gamma xy-\delta} \times\exp\{-i(k_x x e^{-\alpha} + k_y y e^{-\beta}) + k_x k_y\epsilon\}, \tag{34}$$

where $\alpha, \cdots, \epsilon$ are the expressions (14). Hence the solution (30) becomes

$$P'(x, y, t) = e^{-\gamma xy-\delta}\iint dk_x\,dk_y \times\exp\{-i(k_x x e^{-\alpha} + k_y y e^{-\beta}) + k_x k_y\epsilon\}\Psi_0. \tag{35}$$

To see that this is of our form (21), we consider the case $a = b = c = 0$. Then (23)–(26) give the corresponding solution:

$$P_0'(x, y, t) = \iint dk_x\,dk_y \times\exp\{-i(k_x x + k_y y + k_x k_y t)\}\Psi_0(k_x, k_y), \tag{36}$$

where Ψ_0 is given in (28). Thus (35) may be written as

$$P'(x, y, t) = e^{-\gamma xy-\delta}P_0'(xe^{-\alpha}, ye^{-\beta}, \epsilon), \tag{37}$$

which is the form (21).

We notice that Ψ_0 in (28), compared with Ψ in (24), may be found for a wider class of "suitable" initial functions $\Phi(x, y)$. Therefore the form (37)—with (36)—would seem to be more convenient than (23)–(26).

[1] P. Lambropoulos, J. Math. Phys. **8**, 2167 (1967).
[2] M. Kolsrud, Phys. Rev. **104**, 1186 (1956); Phys. Math. Univ. Oslo, Institute Report No. 28.
[3] Notice that $\partial_x T = \partial T/\partial x + T\partial_x$.
[4] J. L. Neuringer, J. Math. Phys. **10**, 250 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Relativistically Invariant Classical Hamiltonian Mechanics with Extra Variables

PHILIP MARK PEARLE*
Case Western Reserve University, Cleveland, Ohio

(Received 2 May 1969)

It is shown that one can construct a relativistically invariant classical mechanics in Hamiltonian form by utilizing as independent dynamical variables the position of each particle, together with its canonical conjugate variable, and the velocity of each particle, together with its canonical conjugate variable. The ten generators of the Poincaré group, obeying the correct Poisson bracket relations between themselves and the position variables, are constructed in their most general form. By restricting this form, it is possible to construct a Hamiltonian theory where the equations of motion for the particle's positions and velocities depend only upon the positions and velocities (and not upon their conjugate variables): these turn out to be the most general relativistically invariant classical equations of motion of this type. This is not useful as a starting point for constructing a physically interesting quantum theory, since the probability density in the quantized theory obeys the classical Liouville equation. This approach is also applied to the Galilean-invariant classical mechanics.

INTRODUCTION

A completely satisfactory relativistically invariant action-at-a-distance quantum mechanics has not yet been made. Dirac[1] first emphasized the importance of being able to construct the ten generators of the Poincaré group from the dynamical variables. Subsequently, Bakamjian and Thomas,[2] Foldy,[3] and Fong and Sucher[4] showed how the ten generators and also the scattering matrix can be constructed when the dynamical variables are the physical momenta of particles and their conjugate variables (which cannot be interpreted as the physical position variables). However, one may require more of a relativistically invariant quantum mechanics than this. For example, one might ask that the theory also contain physical position variables which transform properly under Lorentz transformations and which have vanishing commutator brackets with each other.

At this point, our discussion switches from the framework of quantum mechanics to the framework of classical mechanics. The classical problem is relevant to the quantum problem because the existence of a Poisson bracket algebra of classical variables is a necessary condition for the existence of the commutator bracket algebra of the analogous operator variables: this follows from the similarity of the two bracket-algebras. The major advantage in working with classical mechanics is that it is often easier to find the solutions to the first-order partial differential equations that arise from Poisson bracket relations than to solve the operator equations which arise from the commutator bracket relations of the quantized theory.

The problem of forming a relativistically invariant classical Hamiltonian mechanics has a certain amount of intrinsic interest of its own. But it should be emphasized that, even if this is accomplished, it does not guarantee the existence of a useful quantum theory. The transition from Poisson brackets to commutator brackets introduces the additional problems of the ordering of operators[5] and the existence of operators. Moreover, even if these difficulties are overcome so that one has achieved operator expressions with the desired commutator bracket relations, there still remain the problem of whether a meaningful interpretation[5] can be given to the quantized theory and the question of whether it is a useful description of nature.

It was shown by Currie,[6] Currie, Jordan and Sudarshan,[7] Cannon and Jordan,[8] and Leutwyler[9] that one cannot construct physical position variables (i.e., functions of the dynamical variables which satisfy the correct Poisson bracket relations with the generators of the Poincaré group in order that they transform as position variables must under Lorentz transformations and which have vanishing Poisson brackets with each other), unless the Hamiltonian function is just the sum of free-particle Hamiltonians so that the theory describes particles which do not interact. Next, Hill[10] and Currie[11] went outside the Hamiltonian framework of classical mechanics and found the most general equations of the type of Newton's second law which are relativistically invariant: their results are nonlinear partial differential equations that the forces (functions of particle positions and velocities) have to satisfy. Following this, Hill[12] investigated a technique of casting these equations of the type of Newton's second law in Hamiltonian form by an appropriate change of variables,[13] and examined the conditions under which the other generators of the Poincaré group can be made to generate canonical

transformations. He emphasized, following Kerner,[14] that it is possible to find position variables which have the proper Poisson bracket relations with the Poincaré generators, without foregoing interaction between the particles, provided one does not require that the Poisson brackets among the position variables must also vanish. Jordan[15] tried to circumvent the Currie–Jordan–Sudarshan no-interaction theorem by removing the requirement that generators of "boosts" be also generators of canonical transformations, but he too encountered a no-interaction theorem.

We would like to suggest a different way in which it is possible to bypass the no-interaction theorem. The idea is to admit more variables into the dynamics than just the usually allowed six variables for each particle.[16] For example, one might try to construct a relativistic theory where the particles interact through the mediation of a finite number of degrees of freedom, instead of interacting through the infinite number of degrees of freedom of a field. Perhaps it may even be possible to choose these extra degrees of freedom in such a way that, in an appropriate limit as their number becomes infinite, the theory approaches electrodynamics.

In the present work, we introduce twelve dynamical variables for each particle: the physical position $\mathbf{x}$ and its canonical conjugate variable $\mathbf{p}$, the physical velocity $\mathbf{v}$ and its canonical conjugate variable which we shall call $\mathbf{r}$. We shall construct the most general form of the generators of the Poincaré group (to within a canonical transformation which does not affect the particle positions or velocities) for the interaction of two particles: the results may be readily extended to any number of particles.

The equations of motion which arise from this theory might be called indeterministic, in the sense that, even when the initial positions and velocities of the particles are specified, the particle trajectories are not determined because the choice of different initial conditions for the $\mathbf{p}$'s and $\mathbf{r}$'s can lead to different particle trajectories. This situation, which always arises when there are "extra variables," is no problem when there is a Lorentz-invariant physical meaning associated with the extra variables (e.g., in electrodynamics the extra variables are the electromagnetic field variables; in Ref. 16, the extra variables are the initial and final masses of the particles). In this work, we are not able to associate such a physical meaning with the extra variables $\mathbf{p}$ and $\mathbf{r}$. Instead, we shall find it possible to eliminate the extra variables from the equations of motion for the particle positions and velocities by suitably restricting the form of the Hamiltonian. The resulting equations of motion are those obtained by Hill[10] and Currie[11]: the interesting feature is that we are able to obtain their results by a Hamiltonian approach. Of course, the choice of initial conditions of the $\mathbf{p}$'s and $\mathbf{r}$'s do not affect the trajectories generated by this restricted Hamiltonian.

In the next section, the Galilean-invariant theory is discussed within this framework because of its simplicity and familiarity. In particular, we concentrate on showing what happens when the theory is quantized. Following that, the Lorentz-invariant theory is presented.

I. THE GALILEAN-INVARIANT THEORY

The time translation generator H, space translation generators $\mathbf{P}$, rotation generators $\mathbf{J}$, and boost generators[17] $\mathbf{K}$ satisfy the following Poisson bracket relations:

$$[P_i, P_j] = 0, \tag{1a}$$
$$[J_i, P_j] = \epsilon_{ijk} P_k, \tag{1b}$$
$$[H, P_j] = 0, \tag{1c}$$
$$[K_i, P_j] = \delta_{ij} M, \tag{1d}$$
$$[J_i, J_j] = \epsilon_{ijk} J_k, \tag{2a}$$
$$[J_i, H] = 0, \tag{2b}$$
$$[J_i, K_j] = \epsilon_{ijk} K_k, \tag{2c}$$
$$[K_i, K_j] = 0, \tag{3a}$$
$$[K_i, H] = P_i \tag{3b}$$

(δ_{ij} is the Kronecker δ function; repeated Latin indices are to be summed over; M is a positive constant).

The physical position variables $\mathbf{x}_\alpha$ and velocity variables $\mathbf{v}_\alpha$ satisfy the usual Poisson bracket relations with their canonical conjugate variables $\mathbf{p}_\alpha$ and $\mathbf{r}_\alpha$, respectively:

$$[x_\alpha^i, p_\beta^j] = \delta_{ij}\delta_{\alpha\beta}, \quad [v_\alpha^i, r_\beta^j] = \delta_{ij}\delta_{\alpha\beta} \tag{4}$$

(all other bracket relations between these variables vanish). Because of the definition of a velocity as the time derivative of a position, we have:

$$\mathbf{v}_\alpha = [\mathbf{x}_\alpha, H]. \tag{5}$$

In order that the positions transform properly under Galilean transformations, it must be true that

$$[x_\alpha^i, P_j] = \delta_{ij}, \tag{6a}$$
$$[J_i, x_\alpha^j] = \epsilon_{ijk} x_\alpha^k, \tag{6b}$$
$$[K_i, x_\alpha^j] = 0. \tag{6c}$$

By taking the Poisson bracket of Eqs. (6) with H and using Eqs. (1)–(5) and the Jacobi identity, we can deduce the following bracket relations involving the

velocities:

$$[v_\alpha^i, P_j] = 0, \tag{7a}$$

$$[J_i, v_\alpha^j] = \epsilon_{ijk} v_\alpha^k, \tag{7b}$$

$$[K_i, v_\alpha^j] = \delta_{ij}. \tag{7c}$$

We now proceed to determine the forms of the ten generators of the Galilean group. It is shown in Appendices A and B that $\mathbf{P}$ and $\mathbf{J}$ may be brought to the forms

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2, \tag{8}$$

$$\mathbf{J} = \mathbf{x}_1 \times \mathbf{p}_1 + \mathbf{x}_2 \times \mathbf{p}_2 + \mathbf{v}_1 \times \mathbf{r}_1 + \mathbf{v}_2 \times \mathbf{r}_2 \tag{9}$$

by a canonical transformation of variables that does not affect the $\mathbf{x}_\alpha$ nor the $\mathbf{v}_\alpha$. The methods used by previous authors[7-9] to obtain "standard forms" for the momentum and angular momentum are perfectly applicable here: our excuses for the demonstrations in the appendices are that we wish this paper to be self contained, that the variables we use are different than those of previous authors, and our methods of proof differ somewhat from those employed by previous authors.

By putting the form for $\mathbf{P}$ in Eq. (8) into Eq. (1c), we find that H depends upon $\mathbf{x}_1$ and $\mathbf{x}_2$ only in the combination $\mathbf{x}_1 - \mathbf{x}_2 \equiv \mathbf{x}$. By putting the form for $\mathbf{J}$ in Eq. (9) into Eq. (2b), we find that H is a rotational scalar, depending only upon scalar products of the vectors $\mathbf{x}$, $\mathbf{p}_\alpha$, $\mathbf{v}_\alpha$, $\mathbf{r}_\alpha$. The dependence of H upon $\mathbf{p}_\alpha$ is determined by the most general solution of Eqs. (5):

$$H = \mathbf{p}_1 \cdot \mathbf{v}_1 + \mathbf{p}_2 \cdot \mathbf{v}_2 + \lambda(\mathbf{x}, \mathbf{v}_\alpha, \mathbf{r}_\beta), \tag{10}$$

where λ is so far an arbitrary scalar function of its arguments.

The boost generator $\mathbf{K}$ does not depend upon $\mathbf{p}_1$ and $\mathbf{p}_2$ at all, according to Eq. (6c). It is a vector under rotations according to Eq. (2c). The partial differential equations (1d) and (7c) further restrict the form of $\mathbf{K}$ to:

$$\mathbf{K} = m_1\mathbf{x}_1 + m_2\mathbf{x}_2 - \mathbf{r}_1 - \mathbf{r}_2 + \boldsymbol{\varkappa}(\mathbf{x}, \mathbf{v}_\alpha), \tag{11}$$

where m_1, m_2 are any numbers such that $m_1 + m_2 = M$ and $\boldsymbol{\varkappa}$ is a vector function of its arguments.

Only Eqs. (3) remain to be satisfied. When our expression (11) for $\mathbf{K}$ is inserted in Eq. (3a), we obtain the partial differential equations

$$\left(\frac{\partial}{\partial v_1^i} + \frac{\partial}{\partial v_2^i}\right)\kappa^j = \left(\frac{\partial}{\partial v_1^j} + \frac{\partial}{\partial v_2^j}\right)\kappa^i. \tag{12}$$

Equation (12) is the familiar statement that the curl of the vector $\boldsymbol{\varkappa}$ vanishes, which is the necessary and sufficient condition for

$$\kappa^i = -\left(\frac{\partial}{\partial v_1^i} + \frac{\partial}{\partial v_2^i}\right)\psi, \tag{13}$$

where ψ is a scalar function of the vectors $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$. According to Eqs. (11) and (13), our form for $\mathbf{K}$ is now

$$\mathbf{K} = m_1\mathbf{x}_1 + m_2\mathbf{x}_2 - (\mathbf{r}_1 + \nabla_{v_1}\psi) - (\mathbf{r}_2 + \nabla_{v_2}\psi). \tag{14}$$

We now make a canonical transformation to new variables $\bar{\mathbf{r}}_1$, $\bar{\mathbf{r}}_2$, $\bar{\mathbf{p}}_1$, $\bar{\mathbf{p}}_2$:

$$\bar{\mathbf{r}}_\alpha \equiv \mathbf{r}_\alpha + \nabla_{v_\alpha}\psi, \quad \bar{\mathbf{p}}_\alpha \equiv \mathbf{p}_\alpha + \nabla_{x_\alpha}\psi. \tag{15}$$

This transformation leaves $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$ unchanged. In the new variables, $\mathbf{P}$ and $\mathbf{J}$ become:

$$\mathbf{P} = \bar{\mathbf{p}}_1 + \bar{\mathbf{p}}_2 - (\nabla_{x_1} + \nabla_{x_2})\psi, \tag{16}$$

$$\mathbf{J} = \mathbf{x}_1 \times \bar{\mathbf{p}}_1 + \mathbf{x}_2 \times \bar{\mathbf{p}}_2 + \mathbf{v}_1 \times \bar{\mathbf{r}}_1 + \mathbf{v}_2 \times \bar{\mathbf{r}}_2 - (\mathbf{x} \times \nabla_\alpha\psi + \mathbf{v}_1 \times \nabla_{v_1}\psi + \mathbf{v}_2 \times \nabla_{v_2}\psi). \tag{17}$$

However, since ψ depends only upon $\mathbf{x}_1 - \mathbf{x}_2$, the last term in Eq. (16) vanishes, and because ψ depends only upon scalar products of the vectors $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$, the last three terms in Eq. (17) also add up to zero. Therefore, our forms for $\mathbf{P}$ and $\mathbf{J}$ are not changed by this transformation. In the new variables, H becomes:

$$H = \bar{\mathbf{p}}_1 \cdot \mathbf{v}_1 + \bar{\mathbf{p}}_2 \cdot \mathbf{v}_2 + (\lambda - \mathbf{v}_1 \cdot \nabla_{x_1}\psi - \mathbf{v}_2 \cdot \nabla_{x_2}\psi). \tag{18}$$

We can define the parenthetical expression in Eq. (18) as a new scalar function $\bar{\lambda}$ of the variables $\mathbf{x}$, $\mathbf{v}_\alpha$, $\mathbf{r}_\alpha$. Thus, we see that the canonical transformation (15) makes no essential change in H either. We henceforth adopt this new set of canonical variables, drop the superscript bars, and write down our general form for $\mathbf{K}$:

$$\mathbf{K} = m_1\mathbf{x}_1 + m_2\mathbf{x}_2 - \mathbf{r}_1 - \mathbf{r}_2. \tag{19}$$

We note in passing that this expression for $\mathbf{K}$ in the Galilean-invariant case is the same for noninteracting particles as for interacting particles, whereas this is not true in the Lorentz-invariant case. As is well known, this accounts for the greater complexity of the latter case when compared with the former.

The only restriction remaining to be satisfied is Eq. (3b). Putting our forms (19), (10), (8) into this equation, we find a restriction on the function λ that appears in the Hamiltonian

$$(\nabla_{v_1} + \nabla_{v_2})\lambda = -m_1\mathbf{v}_1 - m_2\mathbf{v}_2. \tag{20}$$

The most general solution of Eq. (20) is

$$\lambda = -\tfrac{1}{2}m_1v_1^2 - \tfrac{1}{2}m_2v_2^2 + \mu(\mathbf{x}, \mathbf{v}, \mathbf{r}_1, \mathbf{r}_2), \tag{21}$$

where $v_\alpha^2 \equiv \mathbf{v}_\alpha \cdot \mathbf{v}_\alpha$, $\mathbf{v} \equiv \mathbf{v}_1 - \mathbf{v}_2$, and μ is a scalar function of its arguments. Thus, the most general form

for H is

$$H = \mathbf{p}_1 \cdot \mathbf{v}_1 + \mathbf{p}_2 \cdot \mathbf{v}_2 - \tfrac{1}{2}m_1 v_1^2 - \tfrac{1}{2}m_2 v_2^2 + \mu, \quad (22)$$

and our construction of the ten generators is complete.

We first remark that, although $\mathbf{p}_\alpha$ and $\mathbf{r}_\alpha$ transform fairly simply under Galilean transformations (they do not change under translations; they transform as vectors under rotations, and $\mathbf{r}_\alpha$ is unchanged by a boost while $\mathbf{p}_\alpha$ transforms as a momentum usually does under a boost), we cannot associate any simple physical quantities with these variables. This is emphasized by the equations of motion which, according to Eq. (22), are:

$$\dot{\mathbf{x}}_\alpha = \mathbf{v}_\alpha, \quad (23a)$$

$$\dot{\mathbf{v}}_\alpha = \nabla_{r_\alpha}\mu, \quad (23b)$$

$$\dot{\mathbf{p}}_\alpha = -\nabla_{x_\alpha}\mu, \quad (23c)$$

$$\dot{\mathbf{r}}_\alpha = -\mathbf{p}_\alpha + m_\alpha \mathbf{v}_\alpha - \nabla_{v_\alpha}\mu. \quad (23d)$$

Although **P**, **J**, and **H** are constants of the motion, they are not solely functions of the physical variables $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$. This makes them less useful quantities than they are found to be in the usual Galilean-invariant classical mechanics. Indeed, the extra complexity introduced by the additional variables allows us to have trajectories for which the physical quantities which are usually constant are not constant at all. For example, according to Eq. (23b),

$$\frac{d(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)}{dt} = m_1\nabla_{r_1}\mu + m_2\nabla_{r_2}\mu, \quad (24)$$

which does not have to vanish. Thus, our theory, although Galilean-invariant, does not necessarily agree with nature. This has come about because the extra degrees of freedom have permitted us to maintain the role of **P** as a translation generator while eliminating its role as the constant momentum.

II. ELIMINATION OF EXTRA VARIABLES FROM THE DYNAMICS OF PHYSICAL VARIABLES

The interaction is characterized, as in the usual Galilean-invariant Hamiltonian mechanics, by a scalar function. We now observe that it is possible to restrict μ in such a way that the extra variables $\mathbf{p}_\alpha$ and $\mathbf{r}_\alpha$ do not enter into the equations of motion for $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$. The conditions to be satisfied are that the right-hand sides of

$$\dot{\mathbf{x}}_\alpha = [\mathbf{x}_\alpha, H], \quad \dot{\mathbf{v}}_\alpha = [\mathbf{v}_\alpha, H] \quad (25)$$

be independent of $\mathbf{p}_\alpha$ and $\mathbf{r}_\alpha$. As can be seen from Eqs. (23a, b), this simply amounts to the condition that

$$\nabla_{r_\alpha}\mu = \mathbf{a}_\alpha(\mathbf{x}, \mathbf{v}), \quad (26)$$

where the accelerations $\mathbf{a}$ are vector functions of their arguments. The solution of Eqs. (26) is

$$\mu = \mathbf{r}_1 \cdot \mathbf{a}_1 + \mathbf{r}_2 \cdot \mathbf{a}_2 + b, \quad (27)$$

where b is a scalar function of $\mathbf{x}$, $\mathbf{v}$. The Hamiltonian in this restricted case is, therefore,

$$H = \mathbf{p}_1 \cdot \mathbf{v}_1 + \mathbf{p}_2 \cdot \mathbf{v}_2 - \tfrac{1}{2}m_1 v_1^2 - \tfrac{1}{2}m_2 v_2^2 + \mathbf{r}_1 \cdot \mathbf{a}_1 + \mathbf{r}_2 \cdot \mathbf{a}_2 + b. \quad (28)$$

The equations of motion, $\dot{\mathbf{x}}_\alpha = \mathbf{v}_\alpha$ and $\dot{\mathbf{v}}_\alpha = \mathbf{a}_\alpha$, are the most general Galilean-invariant equations of motion for two particles. We do not believe that it has been previously shown that these equations can be obtained through a Hamiltonian formalism.

Again, we remark that $m_1\mathbf{v}_1 + m_2\mathbf{v}_2$ is still not a constant of the motion unless we demand the additional restriction of Newton's third law, $m_1\mathbf{a}_1 + m_2\mathbf{a}_2 = 0$.

III. QUANTIZATION OF THE GALILEAN-INVARIANT THEORY

Since this theory is in Hamiltonian form, we may formally proceed to quantize it, using Dirac's well-known transition from Poisson brackets to commutator brackets. The operators $\mathbf{X}_\alpha$ and $\mathbf{V}_\alpha$ (corresponding to the position and velocity variables) form a complete commuting set. In the representation where these variables are represented by diagonal matrices, the Schrödinger equation for $\psi(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2)$, which follows from Eq. (28), is

$$\frac{\hbar}{i}(\mathbf{v}_1 \cdot \nabla_{x_1} + \mathbf{v}_2 \cdot \nabla_{x_2})\psi + \frac{\hbar}{i}(\mathbf{a}_1 \cdot \nabla_{v_1} + \mathbf{a}_2 \cdot \nabla_{v_2})\psi + (-\tfrac{1}{2}m_1 v_1^2 - \tfrac{1}{2}m_2 v_2^2 + b)\psi = i\hbar\frac{\partial\psi}{\partial t}. \quad (29)$$

If we write $\psi = \rho e^{i\theta/\hbar}$, where ρ and θ are real functions, insert it into Eq. (29), multiply by $e^{-i\theta/\hbar}$, and take the real and imaginary parts of the resulting equation, we find two *uncoupled* equations. ρ is found to satisfy the classical Liouville equation

$$\frac{\partial\rho}{\partial t} + \sum_{\alpha=1}^{2}[\mathbf{v}_\alpha \cdot \nabla_{x_\alpha} + \mathbf{a}_\alpha \cdot \nabla_{v_\alpha}]\rho = 0, \quad (30)$$

while θ satisfies an inhomogeneous Liouville equation

$$\frac{\partial\theta}{\partial t} + \sum_{\alpha=1}^{2}[\mathbf{v}_\alpha \cdot \nabla_{x_\alpha} + \mathbf{a}_\alpha \cdot \nabla_{v_\alpha}]\theta = \frac{1}{h}\left[\tfrac{1}{2}\sum_{\alpha=1}^{2}m_\alpha v_\alpha^2 - b\right].$$

Since $|\psi|^2 = \rho^2$ satisfies the classical Liouville equation, it is difficult to see how a physically interesting quantum theory can arise. The flow of probability in configuration space (the quantum-mechanical configuration space and the classical phase space are

essentially identical: a point is specified by the coordinates $\mathbf{x}_\alpha$, $\mathbf{v}_\alpha$) will not exhibit any of the familiar "wavelike" quantum-mechanical effects such as tunnelling, "interference" of probabilities, etc. We may note in passing, however, that Prigogine[18] and coworkers have found it useful in nonequilibrium classical statistical mechanics to consider the Liouville "operator" and its eigenvalues and eigenvectors, an approach which does not appear too unnatural within the context of the present work.

IV. THE LORENTZ-INVARIANT THEORY

The Poisson bracket relations for the ten Poincaré generators are:

$$[P_i, P_j] = 0, \tag{31a}$$
$$[J_i, P_j] = \epsilon_{ijk} P_k, \tag{31b}$$
$$[H, P_j] = 0, \tag{31c}$$
$$[K_i, P_j] = \delta_{ij} H, \tag{31d}$$
$$[J_i, J_j] = \epsilon_{ijk} J_k, \tag{32a}$$
$$[J_i, H] = 0, \tag{32b}$$
$$[J_i, K_j] = \epsilon_{ijk} K_k, \tag{32c}$$
$$[K_i, K_j] = -\epsilon_{ijk} J_k, \tag{33a}$$
$$[K_i, H] = P_i. \tag{33b}$$

In order that the positions transform properly under Lorentz transformations, it must be true that

$$[x_\alpha^i, P_j] = \delta_{ij}, \tag{34a}$$
$$[J_i, x_\alpha^j] = \epsilon_{ijk} x_\alpha^k, \tag{34b}$$
$$[x_\alpha^i, K_j] = x_\alpha^j v_\alpha^i, \tag{34c}$$

where, as usual, $\mathbf{v}_\alpha \equiv [\mathbf{x}_\alpha, H]$. Pryce[19] first pointed out and Currie, Jordan, and Sudarshan[7] emphasized the importance of Eq. (34c) which, together with Eqs. (31d) and (33a), differ from the corresponding relationships in the Galilean-invariant theory.

P and **J** can be brought to the standard forms (8) and (9) because they are a consequence of Poisson bracket relations (5), (7a, b), (31a, b), (32a), (34a, b), which are the same in the Lorentz-invariant case as in the Galilean-invariant case. Likewise, Eqs. (31c), (32b), and (5) imply the form Eq. (10) for H.

The partial differential equations (34c) for **K** have as their most general solution

$$\mathbf{K} = \mathbf{x}_1 \mathbf{p}_1 \cdot \mathbf{v}_1 + \mathbf{x}_2 \mathbf{p}_2 \cdot \mathbf{v}_2 + \boldsymbol{\varkappa}_1(\mathbf{x}_\alpha, \mathbf{v}_\alpha, \mathbf{r}_\alpha) \tag{35}$$

where, according to Eq. (32c), $\boldsymbol{\varkappa}_1$ is a vector function of its arguments. Putting the forms (35) and (10) into Eqs. (31d), we find that $\boldsymbol{\varkappa}_1$ is restricted by the following condition:

$$\left(\frac{\partial}{\partial x_1^j} + \frac{\partial}{\partial x_2^j}\right) \kappa_1^i = \delta_{ij} \lambda(\mathbf{x}, \mathbf{v}_\alpha, \mathbf{r}_\alpha),$$

whose most general solution, when inserted into Eq. (35), yields

$$\mathbf{K} = \mathbf{x}_1(\mathbf{p}_1 \cdot \mathbf{v}_1 + \tfrac{1}{2}\lambda) + \mathbf{x}_2(\mathbf{p}_2 \cdot \mathbf{v}_2 + \tfrac{1}{2}\lambda) + \boldsymbol{\varkappa}(\mathbf{x}, \mathbf{v}_\alpha, \mathbf{r}_\alpha), \tag{36}$$

where $\boldsymbol{\varkappa}$ is a vector function of its arguments.

Equations (33a, b) remain to be satisfied. When the forms (10), (8), (36) for H, **P**, **K**, respectively, are inserted into Eq. (33b), we obtain the partial differential equation

$$\begin{aligned}&\mathbf{v}_1 \mathbf{p}_1 \cdot \mathbf{v}_1 + \tfrac{1}{2}\mathbf{x}\mathbf{p}_1 \cdot \nabla_{r_1}\lambda - \mathbf{p}_1 \cdot \nabla_{r_1}\boldsymbol{\varkappa} - \tfrac{1}{2}\mathbf{x}\mathbf{v}_1 \cdot \nabla_x \lambda \\ &\quad + \mathbf{v}_2 \mathbf{p}_2 \cdot \mathbf{v}_2 - \tfrac{1}{2}\mathbf{x}\mathbf{p}_2 \cdot \nabla_{r_2}\lambda - \mathbf{p}_2 \cdot \nabla_{r_2}\boldsymbol{\varkappa} - \tfrac{1}{2}\mathbf{x}\mathbf{v}_2 \cdot \nabla_x \lambda \\ &\quad + \tfrac{1}{2}\mathbf{v}_1\lambda + \tfrac{1}{2}\mathbf{v}_2\lambda + \mathbf{v}_1 \cdot \nabla_x \boldsymbol{\varkappa} + - \mathbf{v}_2 \cdot \nabla_x \boldsymbol{\varkappa} \\ &\quad + [\boldsymbol{\varkappa}, \lambda] = \mathbf{p}_1 + \mathbf{p}_2,\end{aligned} \tag{37}$$

which is linear in $\mathbf{p}_1$ and $\mathbf{p}_2$. By taking the partial derivatives of Eq. (37) with respect to these variables, we find the following equations which must be satisfied:

$$\frac{\partial(\frac{1}{2}x^i\lambda - \kappa^i)}{\partial r_1^j} = \delta^{ij} - v_1^i v_1^j, \tag{38a}$$

$$\frac{\partial(-\frac{1}{2}x^i\lambda - \kappa^i)}{\partial r_2^j} = \delta^{ij} - v_2^i v_2^j. \tag{38b}$$

If we define a new vector function $\boldsymbol{\varkappa}'$ by

$$\boldsymbol{\varkappa} \equiv \mathbf{v}_1(\mathbf{v}_1 \cdot \mathbf{r}_1) - \mathbf{r}_1 + \mathbf{v}_2(\mathbf{v}_2 \cdot \mathbf{r}_2) - \mathbf{r}_2 + \boldsymbol{\varkappa}' \tag{39}$$

and insert (39) into Eqs. (38), we find that the latter reduce to

$$\frac{\partial(\frac{1}{2}\mathbf{x}\lambda - \boldsymbol{\varkappa}')}{\partial r_1^j} = 0,$$
$$\frac{\partial(-\frac{1}{2}\mathbf{x}\lambda - \boldsymbol{\varkappa}')}{\partial r_2^j} = 0, \tag{40}$$

whose solutions are

$$\begin{aligned}\tfrac{1}{2}\mathbf{x}\lambda - \boldsymbol{\varkappa}' &= \mathbf{x}\lambda_2 + \mathbf{v}_1 A_2 + \mathbf{v}_2 B_2, \\ -\tfrac{1}{2}\mathbf{x}\lambda - \boldsymbol{\varkappa}' &= -\mathbf{x}\lambda_1 + \mathbf{v}_1 A_1 + \mathbf{v}_2 B_1.\end{aligned} \tag{41}$$

Here λ_1, A_1, B_1 depend upon $\mathbf{r}_1$; λ_2, A_2, B_2 depend upon $\mathbf{r}_2$; and all six functions also depend upon $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$. In writing Eq. (41), we have taken advantage of the fact that any 3-dimensional vector can be written as the sum of three independent vectors multiplied by appropriate scalar coefficients.

Upon subtracting one of Eqs. (41) from the other, we see that

$$\mathbf{x}\lambda = \mathbf{x}(\lambda_1 + \lambda_2) + \mathbf{v}_1(A_2 - A_1) + \mathbf{v}_2(B_2 - B_1), \tag{42}$$

which can be satisfied only if

$$\lambda = \lambda_1 + \lambda_2, \tag{43a}$$
$$A_1 = A_2, \tag{43b}$$
$$B_1 = B_2. \tag{43c}$$

Since A_α, B_α depend upon $\mathbf{r}_\alpha$ only, it must be true that the A_α, B_α are, in fact, independent of $\mathbf{r}_1$, $\mathbf{r}_2$, in order that Eqs. (43b, c) be satisfied. We define $A \equiv -A_1 = -A_2$ and $B \equiv -B_1 = -B_2$, where A and B are scalar functions of $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$. Putting these results into either one of Eqs. (41), we find a form for $\boldsymbol{\varkappa}'$ that may in turn be inserted in Eq. (39) to give us

$$\boldsymbol{\varkappa} = \tfrac{1}{2}\mathbf{x}(\lambda_1 - \lambda_2) + \mathbf{v}_1(\mathbf{v}_1 \cdot \mathbf{r}_1 + A) - \mathbf{r}_1 + \mathbf{v}_2(\mathbf{v}_2 \cdot \mathbf{r}_2 + B) - \mathbf{r}_2. \quad (44)$$

The partial differential equation (37) still has not been completely satisfied. Equations (38a, b), (43a), (44) enable us to write Eq. (37) as follows:

$$\begin{aligned} &-\mathbf{x}(\mathbf{v}_2 \cdot \nabla_x)\lambda_1 + [(\mathbf{r}_1 \cdot \mathbf{v}_1)\nabla_{r_1} + \mathbf{v}_1(\mathbf{r}_1 \cdot \nabla_{r_1})]\lambda_1 + \mathbf{v}_1\lambda_1 \\ &- \mathbf{x}(\mathbf{v}_1 \cdot \nabla_x)\lambda_2 + [(\mathbf{r}_2 \cdot \mathbf{v}_2)\nabla_{r_2} + \mathbf{v}_2(\mathbf{r}_2 \cdot \nabla_{r_2})]\lambda_2 + \mathbf{v}_2\lambda_2 \\ &+ \sum_{\alpha=1}^{2}[-\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) + \nabla_{v_\alpha}](\lambda_1 + \lambda_2) + \mathbf{x}[\lambda_1, \lambda_2] \\ &+ (\mathbf{v}_1 - \mathbf{v}_2) \cdot \nabla_x(\mathbf{v}_1 A + \mathbf{v}_2 B) \\ &+ [\mathbf{v}_1 A + \mathbf{v}_2 B, \lambda_1 + \lambda_2] = 0. \end{aligned} \quad (45)$$

At this point in the argument, our forms for H and $\mathbf{K}$ are [putting (43a) into Eq. (10) and (44) into Eq. (36)]

$$H = \sum_{\alpha=1}^{2} \mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \lambda_\alpha, \quad (46)$$

$$\mathbf{K} = \sum_{\alpha=1}^{2}[\mathbf{x}_\alpha(\mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \lambda_\alpha) + \mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha] + \mathbf{v}_1 A + \mathbf{v}_2 B. \quad (47)$$

It will now be shown that the remaining Poisson bracket equation (33a), which we have yet to discuss, imposes additional restrictions on the functions A and B.

V. SATISFACTION OF THE LAST POISSON BRACKET RELATION

We first observe that Eq. (45) may also be written as

$$\begin{aligned} &-\mathbf{x}[\mathbf{p}_2 \cdot \mathbf{v}_2, \lambda_1] + \left[\sum_{\alpha=1}^{2}\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha, \lambda_1\right] + \mathbf{v}_1\lambda_1 \\ &+ \mathbf{x}[\mathbf{p}_1 \cdot \mathbf{v}_1, \lambda_2] + \left[\sum_{\alpha=1}^{2}\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha, \lambda_2\right] + \mathbf{v}_2\lambda_2 \\ &+ [\mathbf{v}_1 A + \mathbf{v}_2 B, \mathbf{p}_1 \cdot \mathbf{v}_1 + \lambda_1] \\ &+ [\mathbf{v}_1 A + \mathbf{v}_2 B, \mathbf{p}_2 \cdot \mathbf{v}_2 + \lambda_2] + \mathbf{x}[\lambda_1, \lambda_2] = 0. \end{aligned} \quad (48)$$

We see from Eq. (48) that $[\lambda_1, \lambda_2]$ is the sum of two terms, one depending only upon $\mathbf{r}_1$, the other depending only upon $\mathbf{r}_2$: denote these two functions by $[\lambda_1, \lambda_2]_1$ and $[\lambda_1, \lambda_2]_2$, respectively. Then, Eq. (48) may be separated into the following two equations:

$$\begin{aligned} &-\mathbf{x}[\mathbf{p}_2 \cdot \mathbf{v}_2, \lambda_1] \\ &+ \left[\sum_{\alpha=1}^{2}\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha, \lambda_1\right] + \mathbf{v}_1\lambda_1 + \mathbf{x}[\lambda_1, \lambda_2]_1 \\ &+ [\mathbf{v}_1 A + \mathbf{v}_2 B, \mathbf{p}_1 \cdot \mathbf{v}_1 + \lambda_1] \equiv \boldsymbol{\zeta}, \end{aligned} \quad (49a)$$

$$\begin{aligned} &+\mathbf{x}[\mathbf{p}_1 \cdot \mathbf{v}_1, \lambda_2] \\ &+ \left[\sum_{\alpha=1}^{2}\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha, \lambda_2\right] + \mathbf{v}_2\lambda_2 + \mathbf{x}[\lambda_1, \lambda_2]_2 \\ &+ [\mathbf{v}_1 A + \mathbf{v}_2 B, \mathbf{p}_2 \cdot \mathbf{v}_2 + \lambda_2] \equiv -\boldsymbol{\zeta}, \end{aligned} \quad (49b)$$

where $\boldsymbol{\zeta}$ is a vector function of $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$, only. If we calculate $x_1^i\zeta^j - x_1^j\zeta^i$ from Eq. (49a) and calculate $x_2^i\zeta^j - x_2^j\zeta^i$ from Eq. (49b) and subtract, we obtain, after quite a bit of manipulation,

$$\begin{aligned} &[K_i, K_j] + \epsilon_{ijk}J_k \\ &= \left[\sum_{\alpha=1}^{2} v_\alpha^i(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - r_\alpha^i, v_1^j A + v_2^j B\right] \\ &\quad - \left[\sum_{\alpha=1}^{2} v_\alpha^j(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - r_\alpha^j, v_1^i A + v_2^i B\right] \\ &\quad - (x^i\zeta^j - x^j\zeta^i). \end{aligned} \quad (50)$$

The left-hand side of Eq. (50) must vanish according to the Poisson bracket relation (33a). Hence, the right-hand side of Eq. (50) must vanish identically:

$$\begin{aligned} &\sum_{\alpha=1}^{2}\left[v_\alpha^i(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) - \frac{\partial}{\partial v_\alpha^i}\right](v_1^j A + v_2^j B) \\ &\quad - \sum_{\alpha=1}^{2}\left[v_\alpha^j(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) - \frac{\partial}{\partial v_\alpha^j}\right](v_1^i A + v_2^i B) \\ &\quad = x^i\zeta^j - x^j\zeta^i. \end{aligned} \quad (51)$$

By regarding A and B as functions of their six scalar arguments $\frac{1}{2}x^2$, $\frac{1}{2}v_1^2$, $\frac{1}{2}v_2^2$, $\mathbf{x} \cdot \mathbf{v}_1$, $\mathbf{x} \cdot \mathbf{v}_2$, $\mathbf{v}_1 \cdot \mathbf{v}_2$, Eq. (51) may be written as

$$\begin{aligned} &(v_1^i v_2^j - v_1^j v_2^i)\left(\frac{\partial A}{\partial \frac{1}{2}v_2^2} + \frac{\partial A}{\partial(\mathbf{v}_1 \cdot \mathbf{v}_2)} - \mathbf{v}_2 \cdot \nabla_{v_2} A \right. \\ &\qquad \left. - \frac{\partial B}{\partial \frac{1}{2}v_1^2} - \frac{\partial B}{\partial(\mathbf{v}_1 \cdot \mathbf{v}_2)} + \mathbf{v}_1 \cdot \nabla_{v_1} B\right) \\ &= x^i\zeta^j - x^j\zeta^i + (x^i v_1^j - x^j v_1^i)\left(\frac{\partial A}{\partial(\mathbf{x} \cdot \mathbf{v}_1)} + \frac{\partial A}{\partial(\mathbf{x} \cdot \mathbf{v}_2)}\right) \\ &\quad + (x^i v_2^j - x^j v_2^i)\left(\frac{\partial B}{\partial(\mathbf{x} \cdot \mathbf{v}_1)} + \frac{\partial B}{\partial(\mathbf{x} \cdot \mathbf{v}_2)}\right). \end{aligned} \quad (52)$$

Again invoking the independence of the three vectors $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$, we see that both sides of Eq. (52) must independently vanish. The bracketed term on the left-hand side of Eq. (52) is of the form $X_1 A - X_2 B$, where X_1 and X_2 are first-order linear differential

operators which commute. The vanishing of this term thus requires the existence of a function Φ whose arguments are scalar products of $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$, such that

$$A = \left(\mathbf{v}_1 \cdot \nabla_{v_1} - \frac{\partial}{\partial \frac{1}{2}v_1^2} - \frac{\partial}{\partial(\mathbf{v}_1 \cdot \mathbf{v}_2)}\right)\Phi,$$
$$B = \left(\mathbf{v}_2 \cdot \nabla_{v_2} - \frac{\partial}{\partial \frac{1}{2}v_2^2} - \frac{\partial}{\partial(\mathbf{v}_1 \cdot \mathbf{v}_2)}\right)\Phi. \quad (53)$$

With these forms for A and B, our expression (47) for $\mathbf{K}$ becomes

$$\mathbf{K} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha(\mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \lambda_\alpha) + \mathbf{v}_\alpha \mathbf{v}_\alpha \cdot (\mathbf{r}_\alpha + \nabla_{v_\alpha}\Phi) - (\mathbf{r}_\alpha + \nabla_{v_\alpha}\Phi) + \mathbf{x}\frac{\partial \Phi}{\partial(\mathbf{x} \cdot \mathbf{v}_\alpha)}. \quad (54)$$

A canonical transformation to new variables,

$$\bar{\mathbf{r}}_\alpha \equiv \mathbf{r}_\alpha + \nabla_{v_\alpha}\Phi, \quad \bar{\mathbf{p}}_\alpha \equiv \mathbf{p}_\alpha + \nabla_{x_\alpha}\Phi,$$
$$\bar{\lambda}_\alpha \equiv \lambda_\alpha - \mathbf{v}_\alpha \cdot \nabla_{x_\alpha}\Phi - (-1)^\alpha\left[\frac{\partial \Phi}{\partial(\mathbf{x} \cdot \mathbf{v}_1)} + \frac{\partial \Phi}{\partial(\mathbf{x} \cdot \mathbf{v}_2)}\right] \quad (55)$$

as shown in Eqs. (15) et seq., does not affect the forms of $\mathbf{P}$, $\mathbf{J}$, and H. Upon dropping the superscript bars from the new variables, we see that the new form for $\mathbf{K}$ is

$$\mathbf{K} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha(\mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \lambda_\alpha) + \mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha, \quad (56)$$

and there is no loss in generality in choosing

$$A = B = 0. \quad (57)$$

The vanishing of the right-hand side of Eq. (52) requires that

$$\zeta = \mathbf{x}\eta, \quad (58)$$

where η is an arbitrary scalar function. In fact, there is no loss in generality in setting $\eta = 0$, since it can be seen from Eqs. (49) that the part of ζ that is proportional to $\mathbf{x}$ can be altered or made to vanish by an appropriate splitting up of the term $\mathbf{x}[\lambda_1, \lambda_2]$ into $\mathbf{x}[\lambda_1, \lambda_2]_1 + \mathbf{x}[\lambda_1, \lambda_2]_2$. We can uniquely characterize this splitting by requiring it to be such that $\eta = 0$.

We conclude that, when functions λ_1, λ_2 are found such that Eqs. (49) are satisfied (with $A = B = \zeta = 0$), then all ten Poincaré generators can be constructed: in particular, H is given by Eq. (46) and $\mathbf{K}$ is given by Eq. (56).

It is possible to proceed somewhat further with an analysis of Eqs. (49), in order to elucidate further general restrictions on the functions λ_1, λ_2, but since these coupled nonlinear equations do not appear to be solvable in closed form, such an exercise seems to be uncalled for. Accordingly, we will immediately turn to consideration of the interesting situation where the equations of motion for the physical variables $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$ are independent of the "extra" variables $\mathbf{p}_\alpha$ and $\mathbf{r}_\alpha$.

VI. ELIMINATION OF EXTRA VARIABLES FROM THE DYNAMICS OF PHYSICAL VARIABLES

According to Eq. (46), the equations of motion for the physical variables are

$$\frac{d\mathbf{x}_\alpha}{dt} = \mathbf{v}_\alpha, \quad (59a)$$
$$\frac{d\mathbf{v}_\alpha}{dt} = \nabla_{r_\alpha}\lambda_\alpha. \quad (59b)$$

Therefore, Eq. (59b) will be independent of $\mathbf{r}_\alpha$ only if

$$\lambda_\alpha = \mathbf{r}_\alpha \cdot \mathbf{a}_\alpha + b_\alpha, \quad (60)$$

where $\mathbf{a}_\alpha$ is a vector function and b_α is a scalar function of arguments $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$. Upon inserting (60) into Eqs. (49), we find that each equation contains a term independent of $\mathbf{r}_\alpha$ and a term linear in $\mathbf{r}_\alpha$. Since Eqs. (49) must be satisfied identically, we end up with four equations. The two equations arising from terms linear in $\mathbf{r}_1$ and $\mathbf{r}_2$ involve only the accelerations $\mathbf{a}_\alpha$:

$$-x^i(\mathbf{v}_2 \cdot \nabla_x)a_1^j + (a_1^i v_1^j + 2v_1^i a_1^j) + \sum_{\alpha=1}^{2}\left(-v_\alpha^i(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) + \frac{\partial}{\partial v_\alpha^i}\right)a_1^j + x^i(\mathbf{a}_2 \cdot \nabla_{v_2})a_1^j = 0,$$
$$-x^i(\mathbf{v}_1 \cdot \nabla_x)a_2^j + (a_2^i v_2^j + 2v_2^i a_1^j) + \sum_{\alpha=1}^{2}\left(-v_\alpha^i(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) + \frac{\partial}{\partial v_\alpha^i}\right)a_2^j - x^i(\mathbf{a}_1 \cdot \nabla_{v_1})a_2^j = 0. \quad (61)$$

Equations (61) were first obtained by Hill[10] and in 1-dimensional form by Currie.[11] These authors showed that the solutions $\mathbf{a}_\alpha$ are the most general accelerations which lead to Lorentz-invariant equations of motion of the type of Newton's second law.

The two equations arising from terms independent of $\mathbf{r}_1$ and $\mathbf{r}_2$ are

$$-\mathbf{x}(\mathbf{v}_2 \cdot \nabla_x)b_1 + \sum_{\alpha=1}^{2}[-\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) + \nabla_{v_\alpha}]b_1 + \mathbf{v}_1 b_1 + \mathbf{x}\mathbf{a}_2 \cdot \nabla_{v_2}b_1 = 0,$$
$$-\mathbf{x}(\mathbf{v}_1 \cdot \nabla_x)b_2 + \sum_{\alpha=1}^{2}[-\mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \nabla_{v_\alpha}) + \nabla_{v_\alpha}]b_2 + \mathbf{v}_2 b_2 - \mathbf{x}\mathbf{a}_1 \cdot \nabla_{v_1}b_2 = 0. \quad (62)$$

Because the choice of b_1, b_2 does not affect the equations of motion of the physical variables, we

shall select the simplest solution of Eqs. (62); $b_1 = b_2 = 0$.

To summarize, when $\mathbf{a}_1$ and $\mathbf{a}_2$ satisfy Eqs. (61), then

$$H = \sum_{\alpha=1}^{2} \mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \mathbf{r}_\alpha \cdot \mathbf{a}_\alpha, \tag{63}$$

$$\mathbf{K} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha(\mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \mathbf{r}_\alpha \cdot \mathbf{a}_\alpha) + \mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - \mathbf{r}_\alpha \tag{64}$$

are satisfactory forms for the Poincaré generators,[20] leading to self-contained equations of motion for the physical variables.

VII. CONCLUDING REMARKS

It is of some interest to display the forms for H and $\mathbf{K}$ that reduce to the Galilean forms (19) and (28), in the limit of large velocity of light c. To do this, we notice that

$$b_1 = m_1(1 - v_1^2)^{\frac{1}{2}}, \quad b_2 = m_2(1 - v_2^2)^{\frac{1}{2}} \tag{65}$$

satisfy Eqs. (62). With these values for b_1 and b_2 inserted in the expression (60) for λ, which in turn is inserted in Eqs. (46) and (56) for H and $\mathbf{K}$, we have

$$H = \sum_{\alpha=1}^{2} \mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \mathbf{r}_\alpha \cdot \mathbf{a}_\alpha + m_\alpha c^2(1 - v_\alpha^2/c^2)^{\frac{1}{2}},$$

$$\mathbf{K} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha[\mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \mathbf{r}_\alpha \cdot \mathbf{a}_\alpha + m_\alpha c^2(1 - v_\alpha^2/c^2)^{\frac{1}{2}}] + \mathbf{v}_\alpha(\mathbf{v}_\alpha \cdot \mathbf{r}_\alpha) - c^2\mathbf{r}_\alpha, \tag{66}$$

where the proper factors of c have been included. In the limit of large c, Eqs. (66) become

$$H = \sum_{\alpha=1}^{2} m_\alpha c^2 + \mathbf{p}_\alpha \cdot \mathbf{v}_\alpha + \mathbf{r}_\alpha \cdot \mathbf{a}_\alpha - \tfrac{1}{2} m_\alpha v_\alpha^2 + O(v^2/c^2),$$

$$c^{-2}\mathbf{K} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha m_\alpha - \mathbf{r}_\alpha + O(v^2/c^2), \tag{67}$$

which are the proper Galilean-invariant generators. The leading terms in Eqs. (61) for $\mathbf{a}_\beta$ are simply $\sum_\alpha \partial \mathbf{a}_\beta / \partial v_\alpha^i = 0$, which tell us that, to lowest order in c, $\mathbf{a}_\beta$ depends upon $\mathbf{v}_1$ and $\mathbf{v}_2$ only in the combination $\mathbf{v} \equiv \mathbf{v}_1 - \mathbf{v}_2$.

The behavior of the quantized Lorentz-invariant theory differs in no essential way from the behavior of the quantized Galilean-invariant theory: the Schrödinger equation has as its consequence the classical Liouville equation for the modulus of the wave-function, etc. Because of this, the particular theory constructed here does not possess too much physical interest. We believe that its chief interest lies in illustrating the possibility of obtaining a Lorentz-invariant Hamiltonian theory, in which the position variables transform properly, by the introduction of a finite number of "extra" variables.

APPENDIX A

In this appendix we will prove the following: Given that there exist three functions P_i, $i = 1, 2, 3$, of independent dynamical variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ (and canonical conjugates $\mathbf{p}_1, \cdots, \mathbf{p}_n$), $\mathbf{v}_1, \cdots, \mathbf{v}_m$ (and canonical conjugates $\mathbf{r}_1, \cdots, \mathbf{r}_m$), and $\phi_1, \cdots, \phi_s$ (and canonical conjugates $\pi_1, \cdots, \pi_s$) such that the following Poisson bracket relations are satisfied:

$$[P_i, P_j] = 0, \quad [x_\alpha^i, P_j] = \delta_{ij}, \quad [v_\alpha^i, P_j] = 0, \tag{A1}$$

then it is possible to find a new set of momenta $\bar{\mathbf{p}}_1, \cdots, \bar{\mathbf{p}}_n$ canonically conjugate to $\mathbf{x}_1, \cdots, \mathbf{x}_n$ such that

$$\mathbf{P} = \bar{\mathbf{p}}_1 +, \cdots, + \bar{\mathbf{p}}_n. \tag{A2}$$

When this is specialized to the case $n = m = 2$, $s = 0$, we obtain the result that is needed in the body of this paper.

We first observe that if A_α, $\alpha = 1, \cdots, \mu$, are μ known independent functions of N dynamical variables and, furthermore, if

$$[A_\alpha, A_\beta] = c_{\alpha\beta} \tag{A3}$$

(where the $c_{\alpha\beta}$ are constants), then the set of μ homogeneous first-order differential equations,

$$[A_\alpha, f] = 0, \tag{A4}$$

possesses N–μ independent solutions. This is because Eqs. (A4) are a complete set:

$$[A_\alpha, [A_\beta, f]] - [A_\beta, [A_\alpha, f]] = [[A_\alpha, A_\beta], f] = 0$$

[which follows from the Jacobi identity and Eqs. (A3)]; i.e., this introduces no new partial differential equation which the function f must satisfy. If we introduce another independent function B which satisfies

$$[A_\alpha, B] = c_\alpha \tag{A5}$$

(where the c_α are constants) and look for simultaneous solutions of the homogeneous equations (A4) and the additional inhomogeneous equation

$$[B, f] = 1, \tag{A6}$$

we will still find N–μ independent solutions: The conditions (A5) ensure completeness of the enlarged set of equations.

We now proceed to systematically find the new momenta $\bar{\mathbf{p}}_1, \cdots, \bar{\mathbf{p}}_n$. We first determine $\bar{p}_1^1$ by finding any solution of the $(3 + 3m + 3n)$ partial differential equations:

$$[P_i, \bar{p}_1^1] = 0, \quad [v_\alpha^i, \bar{p}_1^1] = 0, \quad [x_\beta^i, \bar{p}_1^1] = 0, \quad [x_1^i, \bar{p}_1^1] = \delta_{i1}, \tag{A7}$$

$i = 1, 2, 3$, $\alpha = 1, \cdots, m$, and $\beta = 2, \cdots, n$. This

set of equations is of the form mentioned in the previous paragraph. We further note that the solution $\bar{p}_1^1$ must be functionally independent of v_α^i, x_α^i, because any function of these variables alone would have a vanishing Poisson bracket with x_1^1.

We next determine $\bar{p}_1^2$ by requiring that it satisfy the first three of Eqs. (A7) (with $\bar{p}_1^2$ replacing $\bar{p}_1^1$) and, in addition, that

$$[x_1^i, \bar{p}_1^2] = \delta_{i2}, \quad [\bar{p}_1^1, \bar{p}_1^2] = 0. \tag{A8}$$

As before, a solution may be found that is functionally independent of v_α^i, x_β^i, and $\bar{p}_1^1$.

Continuing in this way, we can determine $\bar{\mathbf{p}}_1, \cdots, \bar{\mathbf{p}}_{n-1}$ by requiring that each new component $\bar{p}_\gamma^i$ have vanishing Poisson brackets with $\mathbf{P}$, $\mathbf{v}_\alpha$, $\mathbf{x}_\beta$ ($\beta \neq \gamma$), all the previously determined components of the $\bar{\mathbf{p}}$'s, and also satisfy $[x_\gamma^i, p_\gamma^i] = \delta_{ij}$. We next *define*

$$\bar{\mathbf{p}}_n \equiv \mathbf{P} - \bar{\mathbf{p}}_1 - \cdots - \bar{\mathbf{p}}_{n-1}. \tag{A9}$$

It follows from Eqs. (A9) and (A1) and the way in which the $\bar{\mathbf{p}}_1, \cdots, \bar{\mathbf{p}}_{n-1}$ were constructed that

$$[v_\alpha^i, \bar{p}_n^j] = 0, \quad [x_\beta^i, \bar{p}_n^j] = \delta_{\beta n}\delta_{ij}, \quad [\bar{p}_\gamma^i, \bar{p}_n^j] = 0. \tag{A10}$$

Thus, we have constructed a set of $3n$ independent dynamical variables that are canonically conjugate to the $\mathbf{x}_\beta$, and since Eq. (A9) is equivalent to Eq. (A2), our demonstration is virtually complete.

One more consideration is necessary, however. The bracket relations between $\mathbf{P}$ and the $\mathbf{r}$, ϕ, π variables are assumed not important to us, and so far these variables have not been restricted in any way. However, the $\mathbf{p}$ variables we have constructed will not necessarily have vanishing Poisson brackets with the $\mathbf{r}$, ϕ, π variables, and so we do not yet have a canonical set of variables. However, new canonical variables $\bar{\mathbf{r}}$, $\bar{\phi}$, $\bar{\pi}$ can be systematically constructed by an obvious extension of the method we used to construct the $\bar{\mathbf{p}}$ variables. This completes our demonstration: We can remove the bars from the constructed set of canonical variables, and this is done in the body of this paper.

APPENDIX B

In this appendix we shall closely follow Canon and Jordan[8] to show that the angular momentum can be brought to the form

$$\mathbf{J} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha \times \mathbf{p}_\alpha + \mathbf{v}_\alpha \times \mathbf{r}_\alpha \tag{B1}$$

by a canonical transformation on $\mathbf{p}_\alpha$ and $\mathbf{r}_\alpha$ which does not affect $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$ nor the form

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2. \tag{B2}$$

Cannon and Jordan used a theorem by Lomont and Moses[21] which is more general than what we need: It is in the specialization of this theorem that this appendix differs from their work.

There is no loss in generality in writing

$$\mathbf{J} = \sum_{\alpha=1}^{2} \mathbf{x}_\alpha \times \mathbf{p}_\alpha + \mathbf{v}_\alpha \times \mathbf{r}_\alpha + \mathbf{N}, \tag{B3}$$

where $\mathbf{N}$ depends upon all the dynamical variables. However, the physical conditions that $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$ transform under rotations like vectors [the Poisson bracket relations between $\mathbf{J}$ and $\mathbf{x}_\alpha$, $\mathbf{v}_\alpha$, Eqs. (6), (7)] lead to the restrictions

$$[N^i, x_\alpha^j] = 0, \quad [N^i, v_\beta^j] = 0. \tag{B4}$$

Equations (B4) tell us that $\mathbf{N}$ is, in fact, independent of the variables $\mathbf{p}_\alpha$ and $\mathbf{r}_\beta$. As a consequence, we note that

$$[N^i, N^j] = 0. \tag{B5}$$

We further note that, as a consequence of the Poisson bracket relation (1b) between $\mathbf{P}$ and $\mathbf{J}$,

$$[p_1^i + p_2^i, N^j] = 0,$$

and, therefore, each component of $\mathbf{N}$ depends upon $\mathbf{x}_1$ and $\mathbf{x}_2$ only in the combination $\mathbf{x}_1 - \mathbf{x}_2 \equiv \mathbf{x}$.

It follows from Eqs. (B3) and (B5) that the Poisson bracket relation (2a) between components of $\mathbf{J}$ reduces to the following:

$$\left[\sum_{\alpha=1}^{2}(\mathbf{x}_\alpha \times \mathbf{p}_\alpha + \mathbf{v}_\alpha \times \mathbf{r}_\alpha)_i, N_j\right] - \left[\sum_{\alpha=1}^{2}(\mathbf{x}_\alpha \times \mathbf{p}_\alpha + \mathbf{v}_\alpha \times \mathbf{r}_\alpha)_j, N_i\right] = \epsilon_{ijk}N_k,$$

which may be written as

$$\mathbf{L} \times \mathbf{N} = i\mathbf{N}, \tag{B6}$$

where $\mathbf{L}$ is the partial differential operator

$$\mathbf{L} \equiv i^{-1}[\mathbf{x} \times \nabla_x + \mathbf{v}_1 \times \nabla_{v_1} + \mathbf{v}_2 \times \nabla_{v_2}]. \tag{B7}$$

The introduction of the factor i is gratuitous: It is merely so that certain properties of angular-momentum operators, familiar from quantum mechanics, will be called to mind.

We now proceed to show that, as a consequence of Eq. (B6), $\mathbf{N}$ can be written as

$$\mathbf{N} = i\mathbf{L}\phi, \tag{B8}$$

where ϕ is a function of $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$.

By taking the cross product of Eq. (B6) with the operator $\mathbf{L}$, after some elementary manipulations and the use of the operator identity $\mathbf{L} \times \mathbf{L} = i\mathbf{L}$, we obtain

$$L^2\mathbf{N} = \mathbf{L}\mathbf{L} \cdot \mathbf{N}. \tag{B9}$$

Now consider any solution ϕ of the second-order partial differential equation

$$L^2\phi = -i\mathbf{L}\cdot\mathbf{N}. \tag{B10}$$

For example, one can obtain a solution by expanding $-i\mathbf{L}\cdot\mathbf{N}$ in a power series in any suitable set of eigenstates of L^2, using Eq. (B10) to determine the constant coefficients in a similar expansion of ϕ: ϕ will be determined up to an additive function which is an eigenstate of L^2 with eigenvalue zero. By substituting the form for $\mathbf{L}\cdot\mathbf{N}$ in Eq. (B10) into Eq. (B9), realizing that the operators L^2 and $\mathbf{L}$ commute, we have

$$L^2[\mathbf{N} - i\mathbf{L}\phi] = 0, \tag{B11}$$

which tells us that $\mathbf{N} - i\mathbf{L}\phi$ is an eigenstate of L^2 with eigenvalue zero. From elementary properties of angular-momentum operators, we realize that it is also an eigenstate of each component of $\mathbf{L}$ with eigenvalue zero, so that we may write

$$\mathbf{N} = i\mathbf{L}\phi + \mathbf{S}, \quad L_iS_j = 0. \tag{B12}$$

Finally, putting the form (B12) back into Eq. (B6) and again recognizing that $\mathbf{L}\times\mathbf{L}\phi = i\mathbf{L}\phi$ is identically satisfied, we arrive at a restriction on $\mathbf{S}$:

$$\mathbf{L}\times\mathbf{S} = i\mathbf{S}. \tag{B13}$$

The left-hand side of Eq. (B13) vanishes by virtue of Eq. (B12), so we conclude that $\mathbf{S} = 0$ and $\mathbf{N} = i\mathbf{L}\phi$, as was claimed in Eq. (B8).

The remainder of our argument is identical to that of Cannon and Jordan.[8] We have shown, using Eqs. (B3), (B7), (B8), that $\mathbf{J}$ can be written in the form

$$\mathbf{J} = \sum_{\alpha=1}^{2}\mathbf{x}_\alpha\times(\mathbf{p}_\alpha + \nabla_{x_\alpha}\phi) + v_\alpha\times(\mathbf{r}_\alpha + \nabla_{v_\alpha}\phi), \tag{B14}$$

where ϕ depends upon $\mathbf{x}\equiv\mathbf{x}_1 - \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2$. We define a transformation to a new set of variables

$$\bar{\mathbf{p}}_\alpha \equiv \mathbf{p}_\alpha + \nabla_{x_\alpha}\phi, \quad \bar{\mathbf{r}}_\alpha \equiv \mathbf{r}_\alpha + \nabla_{v_\alpha}\phi, \tag{B15}$$

which is readily shown to be a canonical transformation by explicit evaluation of the Poisson brackets of these new variables with themselves and with $\mathbf{x}_\alpha$ and $\mathbf{v}_\alpha$. We observe that

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 = \bar{\mathbf{p}}_1 + \bar{\mathbf{p}}_2 \tag{B16}$$

because $(\nabla_{x_1} + \nabla_{x_2})\phi = 0$; thus, the form of the translation generator $\mathbf{P}$ is not affected by this transformation. Lastly, we see that, from Eqs. (B14) and (B15),

$$\mathbf{J} = \sum_{\alpha=1}^{2}\mathbf{x}_\alpha\times\bar{\mathbf{p}}_\alpha + \mathbf{v}_\alpha\times\bar{\mathbf{r}}_\alpha, \tag{B17}$$

which is the same as the form (B1) which we set out to construct, provided that we drop the superscript bars on the new variables.

* Present address: Hamilton College, Clinton, New York 13323.

[1] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[2] B. Bakamjian and L. H. Thomas, Phys. Rev. **92**, 1300 (1953).
[3] L. L. Foldy, Phys. Rev. **122**, 275 (1961).
[4] R. Fong and J. Sucher, J. Math. Phys. **5**, 956 (1964).
[5] Difficulties involving the position variable are especially anticipated: See Ref. 7, Sec. III.
[6] D. G. Currie, J. Math. Phys. **4**, 1470 (1963).
[7] D. G. Currie, T. F. Jordan, and E. C. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).
[8] J. T. Cannon and T. F. Jordan, J. Math. Phys. **5**, 299 (1964).
[9] H. Leutwyler, Nuovo Cimento **37**, 555 (1965).
[10] R. N. Hill, J. Math. Phys. **8**, 201 (1967).
[11] D. G. Currie, Phys. Rev. **142**, 817 (1966).
[12] R. N. Hill, J. Math. Phys. **8**, 1756 (1967).
[13] A set of n first-order equations of motion in n variables $x^1, \cdots, x^n$ can be converted into any other set of n first-order equations of motion in n variables $z^1, \cdots, z^n$ by any one of many possible transformations of variables. Suppose we solve the first set of equations for $n-1$ constants of the motion $\Phi^1(x), \cdots, \Phi^{n-1}(x)$ which do not involve the independent variable t, and also find a constant of the motion $t - \Phi^n(x)$ which depends linearly on t. Likewise, we solve the second set of equations, obtaining constants of the motion $\Psi^1(z), \cdots, \Psi^{n-1}(z)$, $t - \Psi^n(z)$. Then, the equations of transformation, which are obtained by solving the n equations $\Phi^k(x) = \Psi^k(z)$ for x in terms of z (or vice-versa), will take one set of equations of motion into the other. In particular, if n is even, transformations can be found to convert a set of equations of motion into any Hamiltonian set of equations in n variables.
[14] E. H. Kerner, J. Math. Phys. **6**, 1218 (1965).
[15] T. F. Jordan, Phys. Rev. **166**, 1308 (1968).
[16] P. M. Pearle, Phys. Rev. **168**, 1429 (1968).
[17] In this paper, $\mathbf{K}$ is the time-independent boost generator: it is used to transform dynamical variables from an unprimed reference frame at time $t = 0$ to a relatively moving primed reference frame at time $t' = 0$.
[18] I. Prigogine, *Nonequilibrium Statistical Mechanics* (Interscience, New York, 1962).
[19] M. H. L. Pryce, Proc. Roy. Soc. (London) **A195**, 62 (1948).
[20] Currie (Ref. 6, Appendix A) and Hill (Ref. 12, Footnote 35) have noted that it is possible to find some Lorentz-invariant equations of motion for two particles moving in one spatial dimension, within a Hamiltonian framework without extra variables. Curiously enough, the generators H and $\mathbf{K}$ in their examples are linear in the particle momenta, as are our Eqs. (63), (64). See also E. H. Kerner, J. Math. Phys. **9**, 222 (1968) for a discussion of this 1-dimensional problem.
[21] J. S. Lomont and E. H. Moses, Commun. Pure Appl. Math., **14**, 69 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Exact Finite Method of Lattice Statistics. III. Dimers on the Square Lattice*

L. K. RUNNELS†
Department of Chemistry, Louisiana State University
Baton Rouge, Louisiana

(Received 6 August 1969)

The 2-dimensional square-lattice model of hard dimeric molecules has been studied over the entire range of dimer densities. The effects of an external field have been included by allowing different activities for the two possible molecular orientations. Thermodynamic properties have been obtained by the transfer matrix technique for lattices of increasing finite circumference and infinite length. The convergence with increasing size has been the most rapid of all systems yet studied by this technique, satisfactory convergence usually being realized with a circumference of only ten sites. The results are in agreement with previous zero-field studies employing other methods. There is no phase transition in the presence or absence of an external field, and the equation of state is remarkably insensitive to the field. The limited information obtained regarding the nature of the approach to the close-packed limit is consistent with recent predictions based on the series method.

INTRODUCTION

In earlier papers in this series,[1,2] we have reported studies of 2-dimensional lattice gases of hard "spherical" molecules. The exact finite method (EFM) entails the calculation of thermodynamic properties for lattices of infinite length and a series of increasing finite widths (circumferences). We and others[3] have reported application of the method to "nearest-neighbor" systems, in which the molecules are so small—or the lattice so coarse—that the exclusion shell extends only to the coordinated nearest-neighbor sites of the lattice (square, triangular, honeycomb). The results, confirmed for the first two lattices by the series-expansion method[4] and by other techniques,[5,6] have revealed the existence of second-order phase transitions at densities near 80% of the close-packed density.

The technique also has been applied to 2-dimensional lattice gases with molecules of more extended size.[7,8] For the triangular lattice[8] this modification has yielded a reasonable change in the predicted phase behavior: The transition becomes first order and recalls the first-order phase transition thought to occur for the continuum gas of disk-shaped molecules.[9] For the square lattice,[7] however, increasing molecular size produces somewhat erratic results attributable to abnormal packing situations in the filled lattice.

Here we introduce molecular asymmetry into the lattice-gas model with a return to the problem of dimers on the square lattice, whose initial treatment by Fowler and Rushbrooke[10] anticipated the present procedure by many years. Fowler and Rushbrooke were interested in estimating the entropy—or the "molecular freedom"—of the square lattice *filled* with dimeric molecules twice as long as they are wide. This problem has since, of course, been the subject of several classic publications in which the exact partition function has been obtained for finite and infinite lattices with various boundary conditions, and in the presence of an orienting field.[11] Also of considerable interest is the behavior of this system as a function of the density of the dimeric molecules. Initial computations of Onsager[12] (on a continuum model), reinforced by later work of Zwanzig,[13] suggest strongly that highly asymmetric (needlelike) molecules should show preferential alignment above some critical concentration—at least in three dimensions. Current studies on 3-dimensional lattice[14] and continuum[15] models are directed at a determination of the degree of asymmetry required for the phase separation to occur.

The evidence has been accumulating that a length-to-width ratio of two is not sufficient, at least for the square-lattice model. In a study of dimer correlations in the completely filled lattice, Fisher and Stephenson[16] found a strong tendency toward local alignment but no long-ranged order. Series expansions by Nagle[17] and Gaunt[18] and a variational treatment of the transfer matrix by Baxter[19] have explored the properties of the partially filled lattice with equal horizontal and vertical activities (zero field). No evidence was found for any thermodynamic singularity at finite activity. Our calculations over the entire density range support this position and extend its validity to include an external field favoring one orientation over the other. We find, moreover, that the equation of state is remarkably insensitive to the activity ratio induced by the field (except very near close packing). We obtain some confirmation of Gaunt's[18] prediction of the form of the singularity near the close-packed limit and indirect evidence bearing on the dimer and monomer correlations discussed by Fisher and

Stephenson.[16] Finally, some of the techniques employed represent further development of the computational procedure itself.

METHOD

We follow as closely as possible the formulation previously presented[1,2]; the principal modifications arise from the need to add a second thermodynamic variable (the activity ratio) to the single parameter sufficient to characterize systems of hard, symmetric molecules or asymmetric molecules in zero field. The quadratic lattice under consideration is viewed as in Fig. 1, with a cylindrical boundary condition imposed in the *horizontal* direction (circumference M, assumed even) and with infinite extension in the *vertical* direction. It may be seen that the position and orientation of a dimeric molecule is characterized by the bond it covers.[20] We label the horizontal bonds only, indicated by crosses in the diagram, which represent possible sites for vertically oriented molecules. Vacant locations in the lattice will here be viewed simply as holes, although, equivalently, they may be considered as occupied by monomers.

The transfer-matrix technique introduced by Kramers and Wannier[21] generates thermodynamics of the semi-infinite[22] lattice by means of a matrix $\mathbf{P}(x, y)$, whose Latin subscripts will refer to configurations of adjacent rings of the lattice, partially occupied by vertical molecules (on horizontal bonds); the parameters x and y are the activities of horizontal and vertical dimers, respectively. In this description the matrix element $P_{ij}(x, y)$ is given by

$$P_{ij} = y^{\frac{1}{2}(n_i+n_j)}H_{ij}(x), \tag{1}$$

where n_i is the number of vertical dimers in the ring configuration ψ_i and $H_{ij}(x)$ is a polynomial (degree $\leq \frac{1}{2}M$) representing the partition function of the horizontal dimers occupying the vertical bonds between the two adjacent rings of horizontal bonds. This factor clearly depends on the configurations ψ_i and ψ_j of vertical dimers, and in fact vanishes if these configurations are incompatible (due to overlapping molecules). Otherwise, the projecting vertical molecules from below and above divide the ring of M interstitial squares into disjoint segments k of length l_k, $1 \leq k \leq \frac{1}{2}M$. Since these segments may be paved independently, we can express $H_{ij}(x)$ as a product of factors of the form $h_{l(k)}(x)$, which are polynomials easily seen to be generated by the recursion formula

$$h_l(x) = h_{l-1}(x) + xh_{l-2}(x). \tag{2}$$

The initial conditions are $h_2(x) = 1 + x$ and $h_3(x) = 1 + 2x$. [For closed rings, needed only for $l = M$, the initial conditions are

$$h_2(x) = 1 + 2x, \quad h_3(x) = 1 + 3x.]$$

We are using the grand partition function, for which the dimensionless pressure p/kT for the semi-infinite lattice is given by the dominant eigenvalue λ_1 of $\mathbf{P}$ by[1,2]

$$p/kT = M^{-1} \ln \lambda_1, \tag{3}$$

with

$$\lambda_1(x, y) = \max_{\mathbf{vv}=1} \mathbf{vPv}. \tag{4}$$

Although the eigenvector $\mathbf{v}$ belonging to λ_1 is dependent upon the activities x and y, the densities of horizontal and vertical molecules may be computed as if all activity dependence were contained in the transfer matrix[1]:

$$M\rho_x = \partial \ln \lambda_1/\partial \ln x = (x/\lambda_1)\mathbf{vP}_x\mathbf{v},$$
$$M\rho_y = \partial \ln \lambda_1/\partial \ln y = (y/\lambda_1)\mathbf{vP}_y\mathbf{v}. \tag{5}$$

The elements of the derivative matrices $\mathbf{P}_x$ and $\mathbf{P}_y$ are to be obtained by differentiating Eq. (1) with respect to x and y, respectively. The densities so defined are the mean number of dimers of specified orientation *per lattice site.*

Since the symmetric matrix P is finite and its square has strictly positive elements, we know from Perron's theorem that λ_1 is nondegenerate and an analytic function of x and y. There can be no true singularity in the thermodynamic functions, except possibly in the limit $M \to \infty$, as is usual for lattice systems of this sort.

Since the vertical molecules in any one ring are strictly independent of each other, it follows that there are 2^M ring configurations for a circumference of M sites, or the transfer matrix is $2^M \times 2^M$. It is thus desirable to use the dihedral symmetry of the model to reduce the amount of arithmetic required. Specifically, we group together in an equivalence class K_α

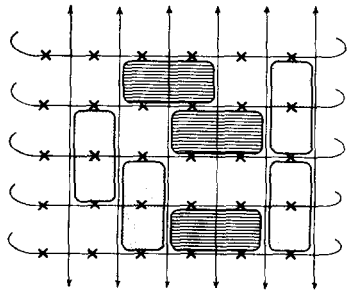


FIG. 1. Dimers on the square lattice, viewed as a "paving" problem. The crosses indicate possible reference sites for molecules oriented vertically (the direction in which the lattice is infinite). The horizontal molecules have a different activity in the presence of a field and in a statistical sense determine the interaction between adjacent rings of vertical dimers.

all of those states ψ_i which are transformed among themselves by the operations of the dihedral group D_M. The actual numerical operations are then performed on the reduced matrix π with elements given by

$$\pi_{\alpha\beta} = \sum_{\psi_i \in K_\beta} P_{ij}, \quad \psi_i \in K_\alpha. \tag{6}$$

The dimension of the matrix π may be determined by an application of Polya's theorem. For details and further explicit formulas for the thermodynamic functions based on the reduced matrix, see Appendix A.

The determinations of the permissible ring configurations and necessary matrix elements, as well as subsequent arithmetic, were accomplished with modifications of the MAP computer program employed in previous investigations. Modifications of three types were important: (1) There are now two independent parameters to vary rather than one; (2) due to the factors $H_{ij}(x)$, previously integral matrix elements are now floating point numbers and can no longer be "packed" to conserve storage; (3) an additional matrix of the same dimension as π is required to determine the horizontal density ρ_x.

These three factors together led us tentatively to discard the use of external storage in favor of the more rapid calculations possible when all storage is in the central processor. This decision implied a significant reduction in accessible lattice size and was adopted firmly only after the rapid convergence discussed in the next section became established.

We investigated 19 different values of the activity ratio $\tau = x/y$, by varying the parameter $\zeta = \tau/(1 + \tau)$ from 0.05 to 0.95 in steps of 0.05. Owing to the nonequivalence of the horizontal and vertical directions imparted by our technique, we felt it desirable to allow τ to assume values greater than one as well as values less than one. At constant field (fixed values of ζ) the logarithm of the vertical activity y was incrementally changed over a wide range of values, in steps of 0.1. At each stage the dominant eigenvector and eigenvalue of π were determined iteratively by repeated multiplication and renormalization of a trial vector.[23] The iterations ceased when the squared length of the vector change between successive trial vectors fell below a predetermined value, set at 10^{-11}. The thermodynamic properties were then determined using the equations found in Appendix A.

RESULTS

Probably the most significant observation was a negative one: the absence of any indication of thermodynamic singularities at finite activities. In this way the dimer system differs from the nearest-neighbor hard-sphere lattice gases. It does, however, call to mind the anomalous behavior of more-extended symmetrical molecules on the square lattice, where the "solid–fluid" transition may be of high order, or absent altogether.[7] The feature held in common by these systems is a lack of a unique close-packed configuration; this is to be contrasted with the specific arrangement required for the nearest-neighbor systems at close packing. Close-packed dimers do possess disorder in *thermodynamic* quantities, while the anomalous quadratic lattice gases do not; the distinction, however, might tend to escape calculations of the present sort on finite strips.

As a general rule, we observed more-rapid convergence with increasing circumference than had been encountered with the symmetric lattice gases. This is undoubtedly due in part to the absence of a phase transition. A very abbreviated sample of calculated total density $\rho_x + \rho_y$ is given in Table I, for three activity ratios. Computations for a circumference greater than 10 would have required the use of external data storage; it was felt, however, that the convergence appeared to be adequate for most purposes without such extension.

As is to be expected, the convergence is best at lower activities. There is also a decided difference between the convergence rate for large values of τ and for small values, although τ and τ^{-1} should be equivalent for the doubly infinite lattice. Of all thermodynamic properties, the density ratio ρ_x/ρ_y is easily the poorest in convergence and will be used to illustrate the nonequivalence of τ and τ^{-1} for the semi-infinite lattice.

At zero activity this ratio should equal τ, regardless of lattice size, and should be approximately τ for the close-packed lattice. In Fig. 2 we show the quantity

TABLE I. Convergence of dimer density with increasing circumference.

ln y[a]	Circumference	Density[b] $\rho_x + \rho_y$ at activity ratio τ[c] $\tau = \frac{1}{3}$	$\tau = 1$	$\tau = 3$
−2	4	0.115314	0.146332	0.211389
	6	0.115314	0.146305	0.210604
	8	0.115314	0.146305	0.210572
	10	0.115314	0.146305	0.210571
0	4	0.290931	0.319797	0.379251
	6	0.290910	0.319105	0.371602
	8	0.290910	0.319064	0.370203
	10	0.290910	0.319062	0.369930
+2	4	0.415453	0.432287	0.469096
	6	0.415274	0.429447	0.458350
	8	0.415263	0.429037	0.453913
	10	0.415263	0.428961	0.452309

[a] Chemical potential of vertical dimer divided by kT.
[b] Estimated uncertainty is five in the last figure.
[c] Ratio of horizontal to vertical activity. The cylindrical lattice is infinite in the vertical direction.

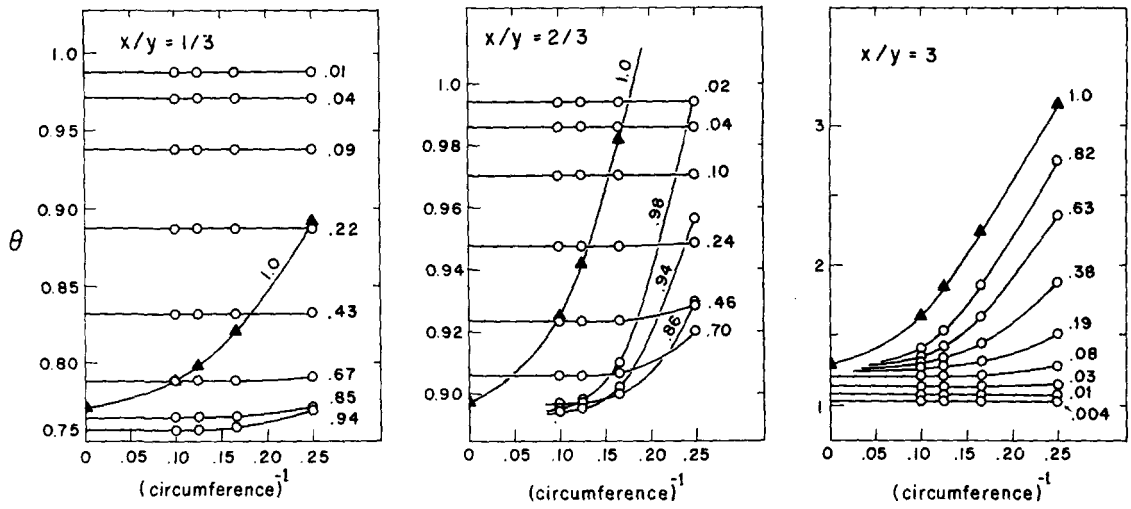


FIG. 2. Convergence of corrected density ratio $\theta = (\rho_x/\rho_y)\tau^{-1} = y\rho_x/x\rho_y$ for three values of the activity ratio x/y. The curves are labeled by rounded values of the activity-related parameter u, which is zero for the empty lattice and one for the close-packed lattice. The curves for $u = 1$ are exact. The difference in vertical scales should be noted. Convergence is clearly enhanced when the field favors alignment along the cylinder axis.

$\theta = (\rho_x/\rho_y)\tau^{-1}$ for each of the four lattice sizes studied (4, 6, 8, 10). The curves are labeled by rounded values of the activity-related parameter $u = z/(1 + z)$, where z is a mean activity,

$$z = (x^x y^y)^{1/(x+y)} = y\tau^{\zeta}. \tag{7}$$

The curves for $u = 1$ are exact, computed from known properties of the close-packed lattice.[11] Calculations were actually performed for many more activities than shown in the figure.

Examination of Fig. 2 reveals interesting detail about the effect of finite circumference and the rate of convergence with increasing circumference. For any external field the corrected density ratio θ is greater for finite circumference than for infinite circumference, showing an enhanced tendency of the dimers to align perpendicular to the cylinder axis. This is due, at least in part, to the added weight which is in effect possessed by configurations containing circumscribing rings of horizontal molecules; each such ring may be rotated by one lattice site without disturbing any other molecules. The extra degeneracy would not apply to circumscribing rings composed entirely of vertical molecules, since the rotated configuration would be indistinguishable from the original.[24] Configurations of either type would occur less frequently for larger lattices, which correlates with the positive slopes shown in Fig. 2.

The slopes, however, are decidedly smaller (corresponding to more rapid convergence) for activity ratios τ less than one than for values of τ greater than one. Two factors contribute to this difference. With $\tau < 1$, the forced predominance of vertical molecules moderates the finite size effect just discussed by rendering rings of horizontal molecules unlikely even for small circumferences. Moreover, the large quantity of vertical molecules required at higher densities can still accommodate occasional horizontal molecules and holes, in any number per horizontal ring; in the corresponding situation for $\tau > 1$ (and even circumference), the predominant horizontal molecules are restricted by the requirement that an even total number of vertical molecules and holes must be encountered in any horizontal ring. This restriction amounts to long-ranged correlations for molecules perpendicular to the cylinder axis and recalls the effect found by Fisher and Stephenson[16]: singlet correlations near an edge (of a close-packed lattice) die out more rapidly for molecules parallel to the edge than they do for molecules perpendicular to the edge.

Regardless of the external field, convergence is always slowest near the close-packed limit. This observation is in accord with the very slow decay of monomer pair correlations at high densities,[16] as well as with the singular nature of the close-packed limit. With due regard for the slower convergence in this region, our results do seem to establish the interesting effect suggested in Fig. 2 for $\tau = \frac{1}{3}$: with sufficiently large field, the corrected density ratio θ is not a monotone function of the total density, even for the doubly infinite lattice. The evidence for this statement is stronger for even smaller values of τ than shown.

Total thermodynamic properties (such as total density or entropy) were found to converge more rapidly than the differential quantity θ (see, for example, Table I). Further discussions will be based

TABLE II. Comparison of EFM with Baxter's variational method[a]

s [b]	κ/s[c]	ρ[d]
0.1	1.38150854	0.48228188
	1.381958	0.482578
0.3	1.488879377	0.4423290570
	1.488912	0.442364
0.5	1.608381870	0.4036382364
	1.608386	0.403646
1.0	1.940215351	0.3190615546
	1.940217	0.319062
2.0	2.700426573	0.2003225902
	2.700426	0.200323
5.0	5.366618412	0.06308145191
	5.366619	0.063082

[a] Reference 19, from which the notation is taken. The first of each pair of figures is from Baxter, the second from this work (with $M = 10$). The activity ratio is one (zero field).
[b] In the present notation, $x = y = s^2$ is the dimer activity. Since the independent variable in our calculations was equally incremented values of 2 ln s, the EFM figures were obtained by Lagrange four-point interpolations. Their estimated uncertainty is 10^{-5}.
[c] In Baxter's notation ln $(\kappa/s^2\rho)$ is the entropy per site, S/kB.
[d] Dimers per site N/B, in agreement with our density definition.

on calculations with $\tau \leq 1$ and, in fact, will use the parameters computed for $M = 10$ with no attempt to extrapolate to infinite circumference. Except for the region very near close packing, where the correlations become long ranged, it is felt that the numbers so produced are valid for the doubly infinite lattice certainly to three significant figures and in most cases more.

This claim is supported by comparison with the results of a recent study by Baxter[19] for $\tau = 1$. Baxter established an exact correspondence between the component of the eigenvector $\mathbf{v}$ and the traces of ordered products of noncommuting infinite matrices. Restricting the matrices to finite arrays and applying the variational principle led to a hierarchy of approximations indexed by matrix size. Convergence with respect to matrix size was found to be very rapid, except at very high activities. Shown in Table II are some of Baxter's results, felt to be valid (for the doubly infinite lattice) to one in the last place shown. Also included for comparison are the values obtained in the present work for $M = 10$; since $\tau = 1$ represents our poorest convergence, the results obtained for $\tau < 1$ should be at least as accurate as suggested by Table II for the zero-field case.

Although similar in appearance, there is no direct connection between the matrix dimensionalities in Baxter's approach and ours. Baxter describes the doubly infinite lattice by one diagonal and two symmetric matrices $(r \times r)$ whose elements are the variational parameters; there are $r^2 + 2r - 2$ parameters after accounting for symmetry and normalization. In the present approach there are $\sigma(M) - 1$ parameters (see Appendix A): the components of the variational eigenvector of our exact matrix π describing an $M \times \infty$ lattice. Baxter's most accurate results $(r = 6)$ corresponded to 46 parameters, while our computations for $M = 10$ in effect used 77 parameters. Judging from Tables I and II, it appears that Baxter's technique is certainly at least as efficient as ours, and probably more efficient. We have not used double-precision arithmetic as Baxter did, but Table I suggests that no great increase in accuracy would result unless the circumference were also increased. Although simplest computationally for $\tau = 1$, there appears to be no fundamental reason why Baxter's approach could not be used for the more general case.

As already indicated, no thermodynamic singularity was discovered, for any finite activity or field; consequently, we shall not attempt to present any complete list of numerical results. We do summarize in Fig. 3 thermodynamic properties calculated for four values of the activity ratio τ, as functions of the convenient parameter u. In addition to the corrected density ratio θ, we show the total density $N/B = \rho = \rho_x + \rho_y$, the dimensionless pressure p/kT, and entropy per site S/kB. To some extent the behavior very near $u = 1$ reflects a finite size effect; for example, θ at $\tau = 1$ should be identically one for all values of u. For small values of τ, however, it has been stated that θ actually is not monotonic for the doubly infinite lattice. On the whole, the somewhat unorthodox approaches to $u = 1$ are manifestations of the singular nature of the functions at the close-packed limit.

Based on series analyses, Gaunt[18] has shown for $\tau = 1$ that the approach of dimers to close packing may be represented by

$$z \sim A(1 - 2\rho)^{-\gamma}, \tag{8}$$

where A and γ are constants and z is the common activity. While the Bethe approximation predicts[17] $\gamma = 2$ and $A = 0.1875$, Gaunt's extrapolations suggested $\gamma = 1.73 \pm 0.04$ and $A = 0.3030 \pm 0.0004$ for the square lattice. We have attempted to explore this asymptotic dependence for several constant values of τ by plotting ln y against ln $(1 - 2\rho)$ and determining the slope of the straight line through about ten points covering the approximate ranges $50 < y < 250$, $0.07 > (1 - 2\rho) > 0.02$. Subsequently, the implied value of the amplitude A was calculated.

The results are found in Table III and show clear substantiation of Gaunt's zero-field values. Several factors should be borne in mind, however, in assessing these parameters. The convergence (with respect to M) is poorest in this density region, and for any circumference finite size effects must appear at extremely high

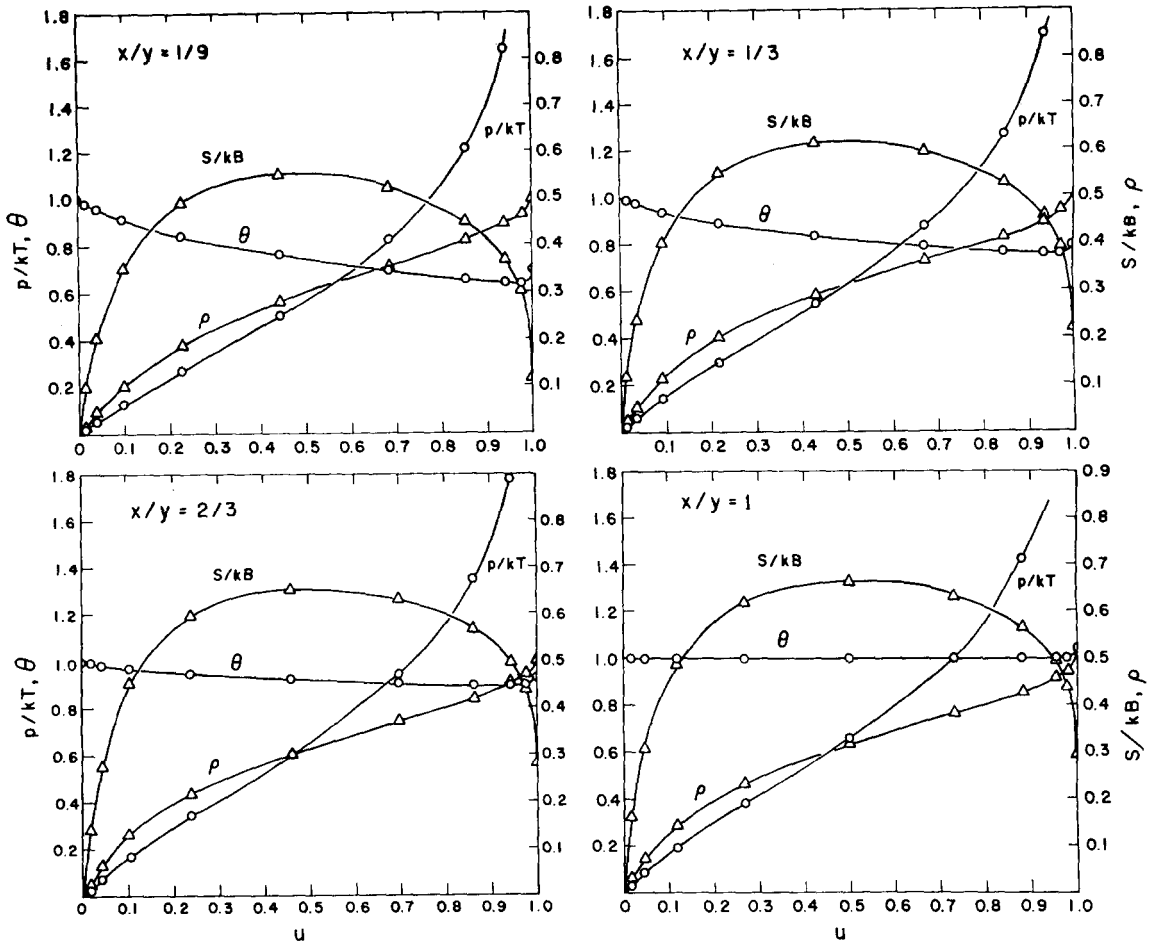


FIG. 3. Thermodynamic properties for circumference $M = 10$. The scale on the left refers to corrected density ratio θ and dimensionless pressure p/kT, while the right-hand scale refers to total density ρ and the entropy per site S/kB. Many more activities (about 100) were included in the calculation of the curves.

densities. Moreover, it is not altogether clear how Eq. (8) should be generalized for the case of unequal activities. Simply by replacing z by the vertical activity y, we have a form which should be exact in the limit $\tau = 0$ as well as reproducing Gaunt's representation at $\tau = 1$. According to the Bethe approximation (Appendix B), such a correlation should be valid with the amplitude approaching 0.25 and the exponent approaching 2 as $\tau \to 0$. Table III bears out the latter but not the former, a fact which may be attributable to convergence difficulties. Another possibility is the inherent inaccuracy of the Bethe approximation at high densities for any nonzero τ, as discussed in Appendix B.[25]

Finally, we show in Fig. 4 the equation of state

TABLE III. Estimates[a] of asymptotic form near close-packing[b] $y \sim A(1 - 2\rho)^{-\gamma}$.

Activity ratio τ	Exponent γ	Amplitude A
1	1.74	0.29
$\frac{2}{3}$	1.79	0.29
$\frac{1}{3}$	1.83	0.32
$\frac{1}{9}$	1.90	0.31
$\frac{1}{19}$	1.93	0.31

[a] Obtained by EFM, with circumference $M = 10$, in the density range 0.465 to 0.49. The uncertainties for the doubly infinite lattice are difficult to estimate, but probably less than 0.02 for both exponent and amplitude.
[b] The Bethe approximation gives $\gamma = 2$ for any τ; it predicts amplitudes of 0.1875 at $\tau = 1$ and 0.25 at $\tau = 0$.

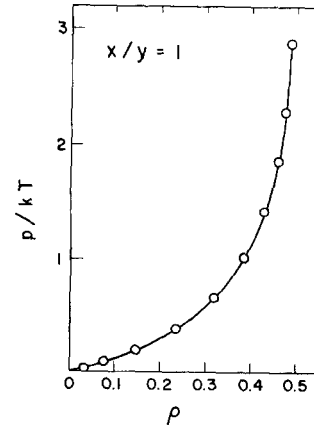


FIG. 4. Zero-field equation of state ($M = 10$). There is no hint of any singularity for any value of the activity ratio τ. Again, many more points were calculated than are shown.

TABLE IV. Effect of activity ratio τ on equation of state.[a]

Density ρ	p/kT at activity ratio $\tau = x/y$ $\tau = 0$[b]	$\tau = \frac{1}{3}$	$\tau = 1$
0.05	0.0541	0.0546	0.0547
0.10	0.1178	0.1198	0.1205
0.15	0.1942	0.1988	0.2007
0.20	0.2877	0.2962	0.2996
0.25	0.4055	0.4189	0.4244
0.30	0.5596	0.5790	0.5869
0.35	0.7732	0.7987	0.8087
0.40	1.0986	1.1275	1.1376
0.45	1.7047	1.7211	1.7235
0.475	2.3512	2.3308	2.3188
0.485	2.8426	2.7815	2.7516

[a] Circumference $M = 10$, for which the pressures are exact to the number of figures shown. For the doubly infinite lattice the last figure is unreliable at the higher densities.
[b] Linear lattice, for which $p/kT = \ln[(1-\rho)/(1-2\rho)]$; see Appendix B.

for $\tau = 1$, obtained by eliminating u and plotting p/kT against ρ. It would in fact be difficult to show on the same graph the equation of state for any other activity ratio, for it is remarkably insensitive to τ. To the extent that this statement is true, the pressure could most simply be predicted by that of the linear lattice,[26] for which Eq. (B5) reduces to

$$p/kT = \ln[(1-\rho)/(1-2\rho)].$$

Table IV shows that over a wide range of densities, this expression is in error by less than 5%.

DISCUSSION

We have concluded that the square-lattice dimer system shows no thermodynamic singularities, regardless of the relative activities of the two possible molecular orientations, except at the close-packing limit. In this sense the asymmetric molecules behave more simply than spherical molecules. The absence of a two-phase region presumably can be laid to the well-known disorder present in the filled lattice. This disorder should decrease with increasing molecular asymmetry, leading to the belief that linear polymers of sufficient length would exhibit an alignment transition; the critical length, however, remains unknown. Alternatively, orientation-specific interactions between dimers could order a dense phase and stabilize it relative to a more dilute, disordered phase.[14,15]

The calculations presented here show that the exact finite method can be applied to model systems containing two independent continuous parameters (x and y). Other related two-parameter systems include lattice gases of symmetric molecules with hard cores and soft interactions of longer range.[27] We have studied such models and will report on them shortly.

ACKNOWLEDGMENTS

Many helpful discussions and correspondences have been appreciated, especially with M. E. Fisher, D. S. Gaunt, and A. G. De Rocco.

APPENDIX A: SYMMETRY AND THERMODYNAMICS

The reduced matrix $\boldsymbol{\pi}$ incorporates into one class α all those states ψ_i which are permuted among themselves by the elements of the dihedral group D_M. The number $\sigma_m(M)$ of such classes containing m (vertical) molecules is counted by Polya's theorem[28]:

$$\sum_m \sigma_m(M) y^m = Z_M(f_1, f_2, \cdots, f_M),$$

where Z_M is the cycle index for the group D_M and $f_k = 1 + y^k$. For even M this cycle index is given by[29]

$$Z_M = f_2^{\frac{1}{2}(M-2)} \tfrac{1}{4}(f_1^2 + f_2) + (2M)^{-1} \sum \varphi(k) f_k^{M/k},$$

the summation being over the divisors of M; $\varphi(k)$ is Euler's totient function and is the number of positive integers less than and prime to k, with the convention that $\varphi(1) = 1$. By setting y equal to one we obtain a simpler formula for the total number of equivalence classes:

$$\sigma(M) = 3 \cdot 2^{\frac{1}{2}(M-4)} + M^{-1} \sum \varphi(k) 2^{(M-k)/k}.$$

Particular values are (6, 13, 30, 78) for $M = (4, 6, 8, 10)$, respectively, in agreement with the findings of the computer program.

The actual computations are then based on the matrix $\boldsymbol{\pi}$. We determine numerically the dominant left eigenvector $\boldsymbol{\xi}$ (belonging to λ_1), with components related to those of normalized eigenvector $\mathbf{v}$ of $\mathbf{P}$ by

$$\xi_\alpha = \left(\sum_\gamma \frac{\xi_\gamma^2}{\omega_\gamma} \right)^{\frac{1}{2}} \omega_\alpha v_i, \quad \psi_i \in K_\alpha, \tag{A1}$$

where ω_α is the number of states in class K_α. Since $\boldsymbol{\pi}$ is not symmetric, the corresponding right eigenvector is not (the transpose of) $\boldsymbol{\xi}$; it may, however, be shown that this right eigenvector is closely related and has components ξ_α/ω_α.

The calculation of the vertical and horizontal densities are based on Eqs. (5). The former is the simplest, since $y(\partial P_{ij}/\partial y) = \frac{1}{2}(n_i + n_j)P_{ij}$; Eqs. (6) and (A1) then lead to

$$\rho_v = M^{-1} \sum_\alpha \frac{n_\alpha \xi_\alpha^2}{\omega_\alpha} \Big/ \sum_\alpha \frac{\xi_\alpha^2}{\omega_\alpha}. \tag{A2}$$

The horizontal density is somewhat more difficult to compute, owing to the more complicated x dependence of $\mathbf{P}$. Following from Eqs. (4), (6), and (A1), we

have

$$\rho_x = (M\lambda)^{-1} \sum_{\alpha,\beta} \xi_\alpha \Lambda_{\alpha\beta} \xi_\beta \Big/ \sum_\alpha \xi_\alpha^2/\omega_\alpha, \qquad \text{(A3)}$$

where

$$\Lambda_{\alpha\beta} = \omega_\beta^{-1} \sum_{\psi_j \in K_\beta} x \frac{\partial P_{ij}}{\partial x}, \quad \psi_i \in K_\alpha. \qquad \text{(A4)}$$

The indicated derivatives $\partial P_{ij}/\partial x$ must be determined from the factors $H_{ij}(x)$, which have the form

$$H_{ij}(x) = \prod_k h_{l(k)}(x).$$

The derivatives dH_{ij}/dx, in turn, are most conveniently expressed as

$$dH_{ij}/dx = H_{ij}(x) \sum_k h'_{l(k)}(x)/h_{l(k)}(x), \qquad \text{(A5)}$$

in terms of the derivatives $h_l'(x)$ of the basic polynomials $h_l(x)$. The computer determination of the matrix elements $\Lambda_{\alpha\beta}$ was based on Eqs. (1), (A4), and (A5), together with a recursion relation for $h_l'(x)$ which follows immediately from Eq. (2).

Once the densities have been determined, all other first-order thermodynamic properties follow immediately. Thus the Gibbs free energy per molecule is given by

$$\mu/kT = (\rho_x \ln x + \rho_y \ln y)/\rho, \qquad \text{(A6)}$$

the entropy per molecule by

$$S/kN = \rho^{-1}(p/kT) - \mu/kT, \qquad \text{(A7)}$$

and the entropy per site by

$$S/kB = p/kT - \rho\mu/kT. \qquad \text{(A8)}$$

APPENDIX B: THE BETHE APPROXIMATION

Following Nagle,[17] we derive the Bethe approximation for the more general case of unequal horizontal and vertical activities by way of a mean field estimate of the entropy $S = k \ln W(B, N_x, N_y)$ of a large quadratic lattice of B sites, with $N_x = \rho_x B$ horizontal and $N_y = \rho_y B$ vertical dimers. Denoting by N the total number of dimers, $N = N_x + N_y$, and by C the combinatorial factor $B!/[(2N_x)!\,(2N_y)!\,(B - 2N)!]$, we have

$$W = [C 2^{2N}][\rho_x \bar\rho_x \bar\rho_y^2]^{N_x}[\rho_y \bar\rho_y \bar\rho_x^2]^{N_y}[\bar\rho_x^2 \bar\rho_y^2]^{\frac{1}{2}(B-2N)}. \qquad \text{(B1)}$$

The abbreviations $\bar\rho_x = 1 - \rho_x$ and $\bar\rho_y = 1 - \rho_y$ have been used. The first factor is the number of choices of the $2N$ lattice sites which are to be covered by an end of a dimer, multiplied by an orientational factor of 2 for each. The next three factors estimate the probabilities that the four neighbors of each x-dimer site, y-dimer site, and vacant site, respectively, will have compatible designations. Each of these latter factors occurs as a power equal to one-half the number of appropriate sites to avoid double counting.

From Eq. (B1) we obtain for the entropy

$$\begin{aligned} S/kB = &-\rho_x \ln \rho_x - \rho_y \ln \rho_y - (1 - 2\rho) \ln (1 - 2\rho) \\ &+ (1 - \rho_x) \ln (1 - \rho_x) \\ &+ (1 - \rho_y) \ln (1 - \rho_y), \end{aligned} \qquad \text{(B2)}$$

in which we can recognize the last two terms as the excess contributions. Equation (B2) now serves as a thermodynamic potential function from which we can by differentiation obtain other properties such as activities and total pressure:

$$x = \rho_x (1 - \rho_x)(1 - 2\rho)^{-2}, \qquad \text{(B3)}$$

$$y = \rho_y (1 - \rho_y)(1 - 2\rho)^{-2}, \qquad \text{(B4)}$$

$$p/kT = \ln (1 - \rho_x)(1 - \rho_y)/(1 - 2\rho). \qquad \text{(B5)}$$

These formulas have as special cases the zero-field equations ($x = y$, $\rho_x = \rho_y = \rho/2$) and the linear lattice description ($x = 0 = \rho_x$, $\rho_y = \rho$).

It should be recognized, however, that although these expressions are exact for the limit $\tau = x/y = 0$, they are seriously in error for the close-packed lattice at *any* finite τ. Thus it may be shown from Eqs. (B3) and (B4) that the corrected density ratio

$$\theta = (\rho_x/\rho_y)\tau^{-1}$$

is predicted to be given by $\theta = 1 - \rho + O(\tau)$ for small τ, or $\theta = \frac{1}{2}$ at close packing and vanishingly small τ. On the other hand, the exact result is[11]

$$\theta = \frac{\tau^{-1} \tan^{-1} \tau}{\frac{1}{2}\pi - \tan^{-1} \tau} \to \frac{2}{\pi}$$

in this limit. It is apparent that for sensitive parameters the Bethe approximation is not reliable at high densities, even for very small values of τ.

* Support for this research was provided by National Science Foundation Grant No. GP-6822. Computer time was also supported in part by National Science Foundation Grant No. GJ-132.

† Alfred P. Sloan Research Fellow.

[1] L. K. Runnels and L. L. Combs, J. Chem. Phys. **45**, 2482 (1966).
[2] L. K. Runnels, L. L. Combs, and J. P. Salvant, J. Chem. Phys. **47**, 4015 (1967).
[3] F. H. Ree and D. A. Chesnut, J. Chem. Phys. **45**, 3983 (1966); N. Karayiania, C. A. Morrison, and D. E. Wortman, J. Math. Phys. **7**, 1458 (1966).
[4] D. S. Gaunt and M. E. Fisher, J. Chem. Phys. **43**, 2840 (1965); D. S. Gaunt, J. Chem. Phys. **46**, 3237 (1967).
[5] D. A. Chesnut, unpublished Monte Carlo studies.
[6] G. W. Woodbury, J. Chem. Phys. **50**, 2247 (1969).
[7] F. H. Ree and D. A. Chesnut, Phys. Rev. Letters **18**, 5 (1967); A. Bellemans and R. K. Nigam, J. Chem. Phys. **46**, 2922 (1967).
[8] J. Orban and A. Bellemans, J. Chem. Phys. **49**, 363 (1968).
[9] B. J. Alder and T. E. Wainwright, Phys. Rev. **127**, 359 (1962).
[10] R. H. Fowler and G. S. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).
[11] P. W. Kasteleyn, Physica **27**, 1209 (1961); H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961); M. E. Fisher, Phys. Rev. **124**, 1664 (1961); A. E. Ferdinand, J. Math. Phys. **8**, 2332 (1967); E. H. Lieb, J. Math. Phys. **8**, 2339 (1967).
[12] L. Onsager, Ann. N. Y. Acad. Sci. **51**, 627 (1949).

[13] R. Zwanzig, J. Chem. Phys. **39,** 1714 (1963).
[14] M. A. Cotter and D. E. Martire (to be published).
[15] A. G. De Rocco, private communication.
[16] M. E. Fisher and J. Stephenson, Phys. Rev. **132,** 1411 (1963).
[17] J. F. Nagle, Phys. Rev. **152,** 190 (1966).
[18] D. S. Gaunt, Phys. Rev. **179,** 174 (1969).
[19] R. J. Baxter, J. Math. Phys. **9,** 650 (1968).
[20] For clarity slight notational differences should be mentioned. Many authors picture a dimer as extending along a bond from lattice site to lattice site; thus an allowed configuration is a "matching" in the language of graph theory—a perfect matching for the filled lattice. We prefer instead to view a dimer configuration as a "paving"; of course, the distinction is devoid of mathematical content for the self-dual square lattice. Additionally, it should be remarked that the cylindrical lattice described by Kasteleyn (Ref. 11) is infinite in the horizontal direction and periodic in the vertical direction. Different definitions have been devised for dimer densities (all denoted ρ); ours agrees with that of Fisher (Ref. 11) and Baxter (Ref. 19).
[21] H. A. Kramers and G. H. Wannier, Phys. Rev. **60,** 252 (1941).
[22] To avoid confusion, the somewhat awkward terminology "semi-infinite" ($M \times \infty$) and "doubly infinite" ($\infty \times \infty$) seems unavoidable.
[23] This trial vector was the final vector for the preceding value of y except for the initial calculation, for which the vector corresponding to the vacuum state served as the trial vector.
[24] Molecules are automatically treated as indistinguishable, since the configurations ψ_i are determined solely by which sites are occupied.
[25] Another approach would be to interpret the activity in Eq. (8) as the mean activity defined by Eq. (7) or, alternatively, by $\exp(2\rho_x \ln x + 2\rho_y \ln y)$, where the densities are the known values at close packing for specified τ. Both of these necessarily lead to the same exponents as given in Table III, but imply smaller amplitudes. Neither, however, appears to approach the Bethe amplitude; rather, these amplitudes show a minimum (about 0.25) at $\tau \approx \frac{1}{3}$ and then increase with further decrease in τ.
[26] M. E. Fisher and H. N. V. Temperley, Rev. Mod. Phys. **32,** 1029 (1960).
[27] J. Orban, J. Van Craen, and A. Bellemans, J. Chem. Phys. **49,** 1778 (1968).
[28] See for example G. E. Uhlenbeck and G. W. Ford in J. de Boer and G. E. Uhlenbeck, Eds., *Studies in Statistical Mechanics* (North-Holland Publishing Co., Amsterdam, 1962), Vol. I, p. 167.
[29] J. Riordan, *An Introduction to Combinatorial Analysis* (John Wiley & Sons, Inc., New York, 1958), p. 150.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Toward a General Theory of Measurement. I

DAVID R. HORST
Department of Physics and Astronomy, State University of New York at Buffalo, Buffalo, New York 14214

(Received 5 September 1969)

A self-consistent measurement theory is set forth in this paper, which can be applied to the problem of measurement in field theory. Following the formulation of Birkhoff and von Neumann, we construct a system of axioms sufficient to ensure that it is possible to find an implicit mechanism within each theory for the experimental procedure corresponding to each observable proposition of that theory. It is shown that any theory of a type similar to quantum mechanics cannot consistently describe a closed system from the point of view of a single observer.

I. INTRODUCTION

For some time it has been generally recognized that the quantum theory of measurement, as applied to field theory, is self-contradictory, and Schwinger[1] suggested that this fact may be the cause of the persistent problems of field theory. In the original article on the measurement theory of quantum electrodynamics, Bohr and Rosenfeld[2] assumed that the experimentalist has at his disposal test bodies of arbitrarily large mass and charge density, which could be coupled directly to the electromagnetic field. This assumption is questionable in several respects. In the first place, interactions between the field and macroscopic objects can never be precisely localized, so that the fundamental dynamical quantities cannot have direct observational significance. Secondly, although such interactions involve collective effects, ultimately they are mediated by the primary interactions, which are known to be strictly limited in number, strength, and character. Indeed, no entirely consistent measurement scheme for field theory is yet known, i.e., one which does not involve concepts alien to field theory, such as rigid charge distributions with a finite number of degrees of freedom. In the foregoing example, it is problematical whether all the quantities which we asserted at the outset to be observable can actually be measured by any physically realizable mechanism. Thus, there might arise doubt concerning the internal consistency of such theories.

However, in the meantime not much progress has been achieved in eliminating these contradictions. Some of them have been tacitly conceded in the axiomatic formulation of field theory, but as yet there has been no systematic attempt to resolve them in a fundamental way. Certainly, this problem is no simple one, for it is complicated by the reciprocal relation between a measurement theory and the physical properties which it describes; at the same time that it underlies all of physical theory, it is also a consequence of that theory. Thus, it is extremely difficult to find a suitable language in which the ideas of a general theory of measurement can be expressed.

Recently, however, there has been a resurgence of interest in the conceptual foundations and physical uniqueness of quantum mechanics along the line of investigation initiated by Birkhoff and von Neumann[3] over thirty years ago, the so-called axiomatic quantum mechanics; several books[4] and scores of articles have been published lately on this topic. The same method of analysis can be extended to the case of field theory in a straightforward manner. Its great generality makes the axiomatic method a convenient medium for the expression of ideas in physical epistemology.

The object of the present program is not to reproduce the features of quantum field theory, but rather to seek a self-consistent measurement theory for systems of this kind. This means that in every admissible theory we must require that an implicit mechanism exists for the experimental procedure corresponding to each observable proposition. The purpose of this paper is to indicate how one might construct a system of axioms, similar to that of axiomatic quantum mechanics, which ensures that the foregoing requirement holds, even without specific information about the nature of the fundamental interactions. Of course, it may be desirable to impose additional conditions upon any proposed description of physical reality. It should be noted that, while this metatheory was conceived in connection with field theory, it is quite general and therefore applicable to any aspect of theoretical physics, for example, ordinary quantum mechanics, although that was not the original intention. The application of this method to problems of field theory is reserved for subsequent publications.

II. FUNDAMENTAL PRINCIPLES

Empirical evidence supporting the quantum theory suggests that a complete description of physical reality is necessarily a probabilistic picture. At any

rate, a deterministic theory can always be represented as a special case of a statistical one. If we abstract a portion from the physical universe and restrict our attention to it, we can prepare an ensemble of representatives of this system and employ the relative-frequency interpretation of probability. Therefore, to a physical system in any given state α, we can attach a probability that any empirical proposition a about the system is true; this probability we denote $p_\alpha(a)$. It is easy to define a partial ordering on the propositions: $a \leq b$ if and only if $p_\alpha(a) \leq p_\alpha(b)$ for all states α. This relation corresponds closely to the implication relation of ordinary logic. Given any pair of propositions a, b, it is reasonable to assume that a supremum and infimum with respect to this ordering can be found, with the obvious interpretation. This is precisely the definition of a lattice. Therefore, we assume that the set of propositions characterizing the given system forms an orthocomplemented complete lattice L, called the *logic* of the system, which defines the latter physically. With each system we also associate an abstract set of states S_L, although strictly these are analogous to classes of partial states. Then we can write ${}^Lp\colon S_L \times L \to [0, 1] \subset \mathbf{R}$, making the dependence on L explicit, to indicate the fact that ${}^Lp$ is a map which assigns to each physical state and proposition about the system a definite probability in the closed interval [0, 1]. In accordance with the Kolmogorov axioms of probability theory, we postulate the first axiom.

Axiom 1: $\mathfrak{L}$ is a family of orthocomplemented complete lattices. For each $L \in \mathfrak{L}$, there is an abstract set S_L of (relative) states and a function ${}^Lp\colon S_L \times L \to [0, 1] \subset \mathbf{R}$ such that if ${}^Lp_\alpha(a) \leq {}^Lp_\alpha(b)$, for all $\alpha \in S_L$, then $a \leq b$. There are idealized propositions $0, 1 \in \bigcap_{L\in\mathfrak{L}} L$ such that ${}^Lp_\alpha(0) = 0$ and ${}^Lp_\alpha(1) = 1$ for all L and $\alpha \in S_L$. Each map ${}^Lp_\alpha$ is countably additive, that is

$$ {}^Lp_\alpha\left(\bigcup_{i=1}^{\infty} a_i\right) = \sum_{i=1}^{\infty} {}^Lp_\alpha(a_i), \quad \text{if} \quad i \neq j \Rightarrow a_i \leq a_j' $$

for all $\alpha \in S_L$ and countable families of propositions $a_i \in L$.

In accordance with our assumption that the probability measure ${}^Lp$ contains all possible physical information about the system, we must insist that distinct states and propositions yield different probability distributions, since this is the only method of distinguishing between them. To assure this uniqueness, we require the axiom:

Axiom 2: If ${}^Lp_\alpha(a) = {}^Lp_\alpha(b)$ for all $\alpha \in S_L$, then $a = b$. Likewise if ${}^Lp_\alpha(a) = {}^Lp_\beta(a)$ for all $a \in L$, then $\alpha = \beta$.

Part of the dispersion in these probabilities may arise from the fact that the ensembles of systems on which they are based are statistical combinations originating from several different sources or modes of preparation. To provide for the existence of such mixed states, we need the following:

Axiom 3: For every countable family of states $\alpha_j \in S_L$ and every convex partition of unity t_j, there is a state $\alpha \in S_L$ such that

$$ {}^Lp_\alpha(a) = \sum_{j=1}^{\infty} t_j \, {}^Lp_{\alpha_j}(a), \quad \text{for all} \quad a \in L. $$

It is natural to write the mixture α as a formal sum of the states α_j and to denote it as $\Sigma t_j\alpha_j$. States which cannot be formed from such linear combinations of other states are called pure; presumably the preparation of all representatives of such states is uniform, and any dispersion in the probabilities arising from pure states may be ascribed solely to the essentially statistical character of the theory. In classical mechanics, all pure states are dispersion free; that is, all probabilities computed for pure states have either of the values 0 or 1.

III. MEASUREMENT THEORY

Up to this point, we have refrained from innovation, and the preceding axiomatic formulation merely duplicates that of Birkhoff and von Neumann, and also Mackey and Jauch and others, suitably generalized. However, our course now diverges from theirs, since our objective differs. Since we intend to construct a self-consistent measurement theory, we must first inquire into the nature of physical observation. For the purpose of conceptual clarity, we abstract a portion of the physical universe as the system under consideration, which we observe by means of experimental apparatus which ordinarily do not enter into the description of the system. On the other hand it is possible, by enlarging the system under consideration, to include part of the apparatus in our description of physical reality, and thereby construct a model of the measuring process. Interactions between the observed system and the measuring apparatus produce correlations which make physical measurement possible. Of course, this procedure can be repeated arbitrarily many times, and the system described can thus be made as large as desired. Since the division of the

universe into parts is entirely factitious, it must make no difference how this partition is made, consistent with the requirement of observability. In particular, if we are given two different domains on which physical systems can be described, there must also be a third domain which contains and encompasses each of the other two. Also, it may be possible to describe a certain subsystem on a restricted domain. In either case, the description of the physical system must coincide in the regions common to the various pairs of domains. This is expressed formally as the axiom:

Axiom 4: For any pair of lattices $L_\alpha, L_\beta \in \mathfrak{L}$, there is $L_\gamma \in \mathfrak{L}$ with $L_\alpha \subset_L L_\gamma$, $L_\beta \subset_L L_\gamma$ such that, given any $\gamma \in S_{L_\gamma}$ there is $\alpha \in S_{L_\alpha}$ and $\beta \in S_{L_\beta}$ such that

$$^{L_\alpha}p_\alpha(a) = {}^{L_\gamma}p_\gamma(a), \quad \text{for all} \quad a \in L_\alpha$$

and

$$^{L_\beta}p_\beta(b) = {}^{L_\gamma}p_\gamma(b), \quad \text{for all} \quad b \in L_\beta.$$

Likewise for any state $\alpha \in S_L$ and lattice $L' \subset_L L$, there is $\delta \in S_{L'}$ such that $^{L'}p_\delta(d) = {}^{L}p_\alpha(d)$ for all $d \in L'$.

Thus, we see that the states of $S_{L'}$, where $L' \subset_L L$, are actually equivalence classes of states of S_L, by Axiom 2.

Now as a rule, in microscopic physics, the experimental apparatus consists of some sort of detector coupled to an amplifier which is capable of displaying the results of measurement on a macroscopic scale, for example the pointer of an instrument. We assume that the indications of these instruments are essentially independent and simultaneously observable in themselves; that is, that propositions concerning the indications of these instruments, e.g., positions of pointers, conform to the rules of ordinary logic. Furthermore, since these indications are directly perceptible, we assume that ideally they are uniquely correlated with the corresponding states of consciousness, so that they can be conceived as identical with the act of cognition. This last assumption has the double advantage of eliminating the need for any subsequent observers, and also of replacing conscious phenomena with relatively simple mechanical analogs of them; however, it can be modified somewhat without serious consequences. To phrase these assumptions more precisely, the propositions about the readings of indicators form a Boolean algebra which is a sublattice of the composite system of object and apparatus. Probabilities of microscopic events are operationally defined by equating them to the probabilities of the corresponding macroscopic events. When we have been able to correlate propositions referring to the experimental system uniquely with the propositions of this Boolean algebra, we shall consider that the description of the measurement procedure is complete. The content of the foregoing discussion is summarized in the following axiom.

Axiom 5: For every $L \in \mathfrak{L}$, there is a lattice $L' \in \mathfrak{L}$ such that $L \subset_L L'$, a Boolean algebra $L_B \subset_L L'$, and a map h from a subset $M \subset_S L_B$ *onto* L, such that for every $\alpha \in S_L$ there is $\gamma \in S_L$, such that,[5] for all $b \in M$,

$$^{L}p_\alpha[h(b)] = {}^{L'}p_\gamma[h(b)] = {}^{L'}p_\gamma[h(b) \cap b] = {}^{L'}p_\gamma(b).$$

Now let us prove a theorem.

Theorem 1: If $\mathfrak{L}$ has a maximal element L_M, then L_M is Boolean.

By Axiom 5, there is $L_B \subset_L L_M$ and a map h_M from a subset $M \subset_S L_B$ *onto* L_M, such that for every $\alpha \in S_{L_M}$ there is $\gamma \in S_{L_M}$, such that for all $b \in M$,

$$^{L_M}p_\alpha[h_M(b)] = {}^{L_M}p_\gamma[h_M(b)] = {}^{L_M}p_\gamma(b).$$

But, since h_M is onto L_M, we must have $\alpha = \gamma$ and $h_M(b) = b$ by Axiom 2. Therefore, h_M is just the identity map on L_M, implying that $M = L_B = L_M$ and L_M is Boolean. Q.E.D.

The preceding theorem has rather far-reaching consequences. Since every orthocomplemented sublattice of a Boolean algebra is Boolean, it is clear that if any of the lattices in $\mathfrak{L}$ are not Boolean, then these lattices cannot be contained in a maximal element. Hence, by Zorn's lemma, $\mathfrak{L}$ has no upper bound in this case. This means that for such systems there is no totality of physical knowledge.

We shall not consider systems in which all of the lattices in $\mathfrak{L}$ are Boolean, inasmuch as clearly the theory is not intended to treat this case, which always leads to a trivial measurement theory. This case is precluded by the empirical evidence supporting the quantum theory.

Next we need to organize these logics L somewhat. Characteristically, physical measurements can be quantified; that is, the results of measurement can be disposed into suitably topologized scales. A probability measure associated with one of these scales assigns a probability that a physical observation yields a result that lies within each measurable subset of this scale. Thus, by Axiom 1, each probability measure corresponds to a σ-homomorphism into L. The set of measurable subsets of each scale forms a sublattice of

the subset lattice of that scale. Since the subset lattice of every set is distributive, so is every sublattice, and in particular the lattice of measurable sets. Furthermore, the homomorphic image of this lattice is also distributive. The image of this homomorphism is a sublattice of L and has the orthocomplementation of L induced on it. Thus, it is a Boolean algebra. But generally speaking, and indeed in the case of interest here, L is not Boolean. In this case, none of these homomorphisms is onto. The image of each σ-homomorphism will be called a physical observable. The condition that every proposition in L lie in the range of at least one observable, has been obtained by Varadarajan,[6] who found that it is just the separability of L. In accordance with his results, it is reasonable to define an observable in L as a maximal Boolean sublattice of L. Consequently, for the present purposes, we shall assume that each logic L is separable. Let us denote the observables of L as A_i^* and the corresponding σ-homomorphisms A_i. Then we have $A_i: M(X_i) \to A_i^*$ for all observables, where X_i is the scale of the observable A_i^* and $M(X_i)$ is a σ-algebra of subsets of X_i, the measurable subsets. As a rule, we will have the lattice-theoretic relation

$$T(X_i) \subset_L M(X_i) \subset_L P(X_i),$$

where $P(X_i)$ is the power set of X_i, made into a lattice by the obvious construction, and $T(X_i)$ is the topology of X_i, for all observables A_i^*. (Note that the topology of every space is a join-complete, distributive sublattice of its subset lattice.) In like manner, we denote the observables of L_B, given by Axiom 5, as C_k^*.

A common feature of all known physical theories is the existence of algebraic relations between observable quantities. We may reasonably expect that, as in the semiclassical analysis of the quantum theory of measurement in field theory,[7] simple formulas can be found which express the relation between the probability distribution of each observable A_i^*, over its domain X_i, and the probability distributions of various observables of L_B, representing indications of the experimental apparatus. In particular, suppose that there is a pointwise correspondence; that is, there is a function, to be specified further below, for each observable A_i^*, say $g_i: \prod_k Y_k \to X_i$, relating measurements corresponding to points on the scale X_i to points on the scales of the experimental apparatus Y_k. Then we can formalize this general conception of a correspondence in the following axiom:

Axiom 6: Each lattice $L \in \mathfrak{L}$ is separable. For every observable A_i^* in L, there is a measurable function $g_i: \prod_k Y_k \to X_i$, such that

$$A_i(E) = h\left[\bigcup_{\vec{g}_i(\Pi_k F_k) \subset E} \bigcap_k C_k(F_k)\right]$$

for all $E \in M(X_i)$, where h is given by Axiom 5.

The completeness of L ensures the existence of the argument in the right member of the preceding equation. This is the most general condition which ensures the existence of recognizable relations between the observables of the experimental system and those of the experimental apparatus. If, for a moment, we restrict our attention to the case where g_i is a function of one variable only, we see then by Axiom 5 that such correspondences induce homomorphisms from sublattices of L_B to the observables A_i^*. It is natural, therefore, to speak of the general correspondences defined above as generalized homomorphisms.

IV. CONCLUSIONS

We have analyzed the experimental procedure and have attempted to formulate intuitively plausible axioms which appear to be absolutely necessary conditions for a physical theory which yields a consistent measurement scheme. It is easy to verify the consistency of the axiom system by considering classical mechanics. In this case, $\mathfrak{L}$ contains a single element, the lattice of measurable subsets of phase space. Although this example is not included in the case of interest here, nonetheless, it suffices to illustrate the consistency of the axioms. Jauch[8] has shown that the superposition principle precludes Boolean logics. However, it would be desirable to discover a more intuitive principle which explains the non-Boolean nature of quantum mechanics. For the present, it seems impossible to find such a principle.

In this connection, the result of Theorem 1 is significant. If any of the logics is not Boolean, then there exists no totality of physical knowledge and no total system. In particular, if the universe is assumed to be a closed physical system comprised entirely of elementary particles which are individually indistinguishable, then it seems possible to describe it by means of a maximal logic, which by Theorem 1 would have to be Boolean. This apparently contradicts the possibility of quantizing such closed systems as the universe from the point of view of a single observer. However, such cosmological questions are purely incidental to the purpose of the present study, and we merely note one implication of this result, without further comment.

Implications of the present formulation are the subject for further study. It remains to be determined

whether conventional quantum field theory, or any similar theory, obeys the system of axioms which we have constructed. These questions will be investigated in subsequent articles in this series.

[1] See the editor's preface in *Quantum Electrodynamics*, J. Schwinger, Ed. (Dover Publications, Inc., New York, 1958), p. xv ff., which contains, I think the most succinct summary of the situation, which is essentially unchanged despite the progress of over a decade.

[2] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **12**, 8 (1933).

[3] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).

[4] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, New York, 1963); J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1968). The latter reference contains a comprehensive bibliography, which the reader should consult for a survey of recent research.

[5] In many presentations of axiomatic quantum mechanics, the lattice operations $\cap$ and $\cup$ are identified with the operations of conjunction and disjunction, respectively, of ordinary logic. Although for reasonable physical systems this interpretation is satisfactory, simple examples suffice to show that in general, this correspondence is no more than formal. Indeed, if c is the proposition which is defined to mean "a and b are true," then one can only prove that $c \leq a \cap b$. It is worth remembering this fact when reading Jauch's book. Nonetheless, one can still see the necessity of the condition in Axiom 5.

[6] V. S. Varadarajan, Commun. Pure Appl. Math. **15**, *2*, 189 (1962).

[7] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach Publishers, Inc., New York, 1965).

[8] J. M. Jauch, ref. 4, p. 105ff.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Theory of Neutron-Wave Propagation through an Interface

O. C. BALDONADO* AND R. C. ERDMANN
University of California, Los Angeles, California 90024

(Received 25 November 1968)

Neutron-wave propagation in a system consisting of two adjacent half-spaces having different properties is investigated using four models: (A) one-speed diffusion theory; (B) one-speed transport theory; (C) energy-dependent diffusion theory; (D) energy-dependent transport theory. An oscillating exponential source of neutrons is located in the left medium. Isotropic scattering is assumed for the transport models B and D, and a one-term degenerate kernel is used for the energy-dependent models C and D. Elementary solutions for each model are obtained and are used to construct the general solution. The resulting solutions are identified in terms of incident, reflected, and transmitted components of the wave. Both discrete and continuum solutions are seen to exist for models B, C, and D. The analyses given are exact except for model C, where the continuum solutions are neglected in the evaluation of the coefficients. The expansion coefficients in model B are evaluated by solving a Riemann–Hilbert problem with discontinuous coefficients. A similar, but generalized Riemann–Hilbert problem was solved to obtain the expansion coefficients in model D.

I. INTRODUCTION

The problem investigated in this paper is how a discontinuity in material properties affects the propagation of a neutron wave. The system is made up of two adjacent half-spaces which have different neutronic properties (see Fig. 1). The problem of neutron-wave propagating in an infinite medium, a special case of this work, has been studied in great detail, even to the extent of including energy-dependent scattering kernels to represent crystalline media. The emphasis in the single-medium problem is centered on the propagation constant. Two definite works on this single-medium problem are those of Duderstadt[1] and Warner.[2] For this two-region problem, the wave properties which depend upon the properties of the two half-spaces will be emphasized.

Particular attention is devoted to the structure of the incident, reflected, and transmitted components of the wave. Decay constants, phase shifts, and amplitudes of the waves will be analyzed. It is expected that some of these quantities will depend only upon the local or single-medium properties, while others will depend upon global properties. These global characteristics depend upon the properties of both media. Discrete and continuum solutions will be generated, and the magnitudes of the continuum effects will be obtained by examining the discontinuity in the discrete components of the wave.

The models used in the investigation are:

model A—one-speed diffusion theory;
model B—one-speed transport theory;
model C—energy-dependent diffusion theory;
model D—energy-dependent transport theory.

It is assumed throughout that scattering is isotropic in the CM system. The equations appropriate to these

whether conventional quantum field theory, or any similar theory, obeys the system of axioms which we have constructed. These questions will be investigated in subsequent articles in this series.

[1] See the editor's preface in *Quantum Electrodynamics*, J. Schwinger, Ed. (Dover Publications, Inc., New York, 1958), p. xv ff., which contains, I think the most succinct summary of the situation, which is essentially unchanged despite the progress of over a decade.

[2] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **12**, 8 (1933).

[3] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).

[4] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, New York, 1963); J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1968). The latter reference contains a comprehensive bibliography, which the reader should consult for a survey of recent research.

[5] In many presentations of axiomatic quantum mechanics, the lattice operations $\cap$ and $\cup$ are identified with the operations of conjunction and disjunction, respectively, of ordinary logic. Although for reasonable physical systems this interpretation is satisfactory, simple examples suffice to show that in general, this correspondence is no more than formal. Indeed, if c is the proposition which is defined to mean "a and b are true," then one can only prove that $c \leq a \cap b$. It is worth remembering this fact when reading Jauch's book. Nonetheless, one can still see the necessity of the condition in Axiom 5.

[6] V. S. Varadarajan, Commun. Pure Appl. Math. **15**, *2*, 189 (1962).

[7] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach Publishers, Inc., New York, 1965).

[8] J. M. Jauch, ref. 4, p. 105ff.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Theory of Neutron-Wave Propagation through an Interface

O. C. BALDONADO* AND R. C. ERDMANN
University of California, Los Angeles, California 90024

(Received 25 November 1968)

Neutron-wave propagation in a system consisting of two adjacent half-spaces having different properties is investigated using four models: (A) one-speed diffusion theory; (B) one-speed transport theory; (C) energy-dependent diffusion theory; (D) energy-dependent transport theory. An oscillating exponential source of neutrons is located in the left medium. Isotropic scattering is assumed for the transport models B and D, and a one-term degenerate kernel is used for the energy-dependent models C and D. Elementary solutions for each model are obtained and are used to construct the general solution. The resulting solutions are identified in terms of incident, reflected, and transmitted components of the wave. Both discrete and continuum solutions are seen to exist for models B, C, and D. The analyses given are exact except for model C, where the continuum solutions are neglected in the evaluation of the coefficients. The expansion coefficients in model B are evaluated by solving a Riemann–Hilbert problem with discontinuous coefficients. A similar, but generalized Riemann–Hilbert problem was solved to obtain the expansion coefficients in model D.

I. INTRODUCTION

The problem investigated in this paper is how a discontinuity in material properties affects the propagation of a neutron wave. The system is made up of two adjacent half-spaces which have different neutronic properties (see Fig. 1). The problem of neutron-wave propagating in an infinite medium, a special case of this work, has been studied in great detail, even to the extent of including energy-dependent scattering kernels to represent crystalline media. The emphasis in the single-medium problem is centered on the propagation constant. Two definite works on this single-medium problem are those of Duderstadt[1] and Warner.[2] For this two-region problem, the wave properties which depend upon the properties of the two half-spaces will be emphasized.

Particular attention is devoted to the structure of the incident, reflected, and transmitted components of the wave. Decay constants, phase shifts, and amplitudes of the waves will be analyzed. It is expected that some of these quantities will depend only upon the local or single-medium properties, while others will depend upon global properties. These global characteristics depend upon the properties of both media. Discrete and continuum solutions will be generated, and the magnitudes of the continuum effects will be obtained by examining the discontinuity in the discrete components of the wave.

The models used in the investigation are:

model A—one-speed diffusion theory;
model B—one-speed transport theory;
model C—energy-dependent diffusion theory;
model D—energy-dependent transport theory.

It is assumed throughout that scattering is isotropic in the CM system. The equations appropriate to these

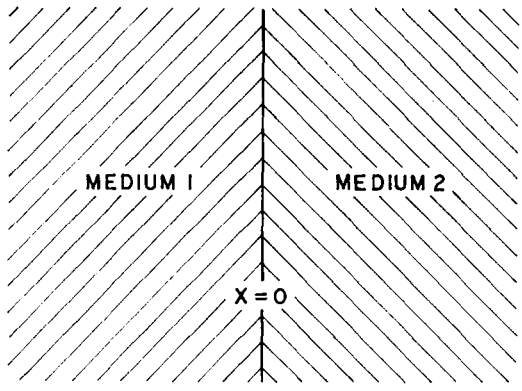


FIG. 1. Problem geometry.

four models are:

(A)

$$\frac{1}{v}\frac{\partial\psi(x,t)}{\partial t} - D\frac{\partial^2\psi(x,t)}{\partial x^2} + \sigma_a\psi(x,t) = S(x,t); \quad (1)$$

(B)

$$\frac{1}{v}\frac{\partial N(x,\mu,t)}{\partial t} + \mu\frac{\partial N(x,\mu,t)}{\partial x} + \sigma_t N(x,\mu,t) = \tfrac{1}{2}\sigma_s\int_{-1}^{1} N(x,\mu',t)\,d\mu' + \tfrac{1}{2}S(x,\mu,t); \quad (2)$$

(C)

$$\frac{1}{v}\frac{\partial\psi(x,E,t)}{\partial t} - D\frac{\partial^2\psi(x,E,t)}{\partial x^2} + \Sigma_t(E)\psi(x,E,t) = \int_0^\infty dE'K(E',E)\psi(x,E',t) + S(x,E,t); \quad (3)$$

(D)

$$\frac{1}{v}\frac{\partial N}{\partial t}(x,\mu,E,t) + \mu\frac{\partial N}{\partial x}(x,\mu,E,t) + \Sigma_t(E)N(x,\mu,E,t) = \tfrac{1}{2}\int_0^\infty dE'\int_{-1}^{1} d\mu' K(E',E)N(x,\mu',E',t) + \tfrac{1}{2}S(x,\mu,E,t). \quad (4)$$

In the above equations, ψ and N denote the diffusion and transport distributions functions, respectively, and the rest of the notation is standard.[3] As yet, no assumption has been made of the structure of the scattering kernel in models C and D, except that it does not depend upon μ (i.e., it is isotropic).

Equations (1)–(4) shall be solved for the oscillating component of the distribution function only. For this wave problem, $e^{i\omega t}$ is assumed for the time dependence. In the discussion to follow, a neutron-wave function is written as $Ae^{kx+i\omega t}$ with the complex amplitude A independent of x and t. ω denotes the oscillation frequency, and k is the propagation constant.

Also in subsequent equations, the subscript is dropped if the discussion applies to either medium. A subscript 1 (2) indicates the medium on the left (right).

II. DEVELOPMENT OF THE GENERAL SOLUTIONS

Model A: For this model, the elementary solutions are $\exp(\pm x/L_j)$, where

$$\frac{1}{L_j^2(\omega)} = \frac{\sigma_{a_j} + i\omega/v}{D_j}, \quad j = 1, 2. \quad (5)$$

The general solution is written as

$$\psi(x,t) = e^{i\omega t}[A_1e^{-x/L_1} + B_1e^{x/L_1}], \quad x < 0,$$
$$= e^{i\omega t}[A_2e^{-x/L_2} + B_2e^{x/L_2}], \quad x > 0. \quad (6)$$

A_1, A_2, B_1, and B_2 are constants to be evaluated later.

Model B: After assuming for either medium

$$N(x,\mu,t) = e^{i\omega t}\exp\left(-\frac{(\sigma_t + i\omega/v)x}{\nu}\right)\phi_\nu(\mu), \quad (7)$$

$$\int_{-1}^{+1}\phi_\nu(\mu)\,d\mu = \sigma_t + i\omega/v, \quad (8)$$

and using the method described by Case,[4] two types of solutions result:

B1—Discrete:

$$\phi_\pm(\mu) = \tfrac{1}{2}c\sigma_t\frac{\nu_0}{\nu_0 \mp \mu}, \quad \nu_0 \notin (-1, 1); \quad (9)$$

B2—Continuous:

$$\phi_\nu(\mu) = \tfrac{1}{2}c\sigma_t P\frac{\nu}{\nu - \mu} + \lambda(\nu)\delta(\nu - \mu), \quad \nu \in (-1, 1). \quad (10)$$

ν_0 and $\lambda(\nu)$ are obtained by application of the normalization condition (8):

$$\sigma_t + \frac{i\omega}{v} = \tfrac{1}{2}c\sigma_t\nu_0\ln\left(\frac{\nu_0 + 1}{\nu_0 - 1}\right), \quad (11)$$

$$\lambda(\nu) = \sigma_t + \frac{i\omega}{v} - \tfrac{1}{2}c\sigma_t\nu\ln\left(\frac{1+\nu}{1-\nu}\right). \quad (12)$$

After defining

$$\xi_{0j} = (i\omega/v + \sigma_{tj})/\nu_{0j}, \quad j = 1, 2, \quad (13)$$

$$\xi_j = (i\omega/v + \sigma_{tj})/\nu_j, \quad j = 1, 2, \quad (14)$$

one can write the general solution for this model as

$$N(x,\mu,t) = e^{i\omega t}\Big[-a_+e^{-\xi_{01}x}\phi_{1+}(\mu) - a_-e^{\xi_{01}x}\phi_{1-}(\mu) - \int_{-1}^{1} A_1(\nu)e^{-\xi_1 x}\phi_{1\nu}(\mu)\,d\nu\Big], \quad x < 0,$$
$$= e^{i\omega t}\Big[b_+e^{-\xi_{02}x}\phi_{2+}(\mu) + b_-e^{\xi_{02}x}\phi_{2-}(\mu) + \int_{-1}^{1} A_2(\nu)e^{-\xi_2 x}\phi_{2\nu}(\mu)\,d\nu\Big], \quad x > 0, \quad (15)$$

where $a_\pm$, $b_\pm$, $A_1(\nu)$, and $A_2(\nu)$ are unknown expansion constants.

Model C: Setting

$$\psi(x, E, t) = \psi(x, E)e^{i\omega t} = X(x)F(E)e^{i\omega t}, \quad (16)$$

one gets

$$\frac{d^2x}{dx^2} - \xi^2 X(x) = 0, \quad (17)$$

$$\left(\frac{1}{L^2(E)} - \xi^2\right)F(E) = \int_0^\infty K(E', E)F(E')\,dE', \quad (18)$$

where

$$\frac{1}{L^2(E)} = \frac{\Sigma_t(E) + i\omega/v}{D(E)}. \quad (19)$$

Again, two cases arise resulting in two types of solution. Assuming

$$K(E', E) = f(E)g(E'), \quad (20)$$

with $F(E)$ normalized such that

$$\int_0^\infty g(E')F(E')\,dE' = 1, \quad (21)$$

the two types of solutions are as follows:

C1—Discrete:

$$\xi^2 - L^{-2}(E) \neq 0, \quad \forall\, E \text{ and } \omega,$$

$$F(E) = F(\xi_0, E) = \frac{f(E)}{D(E)[L^{-2}(E) - \xi_0^2]}, \quad (22)$$

where ξ_0 comes from condition (21) applied to Eq. (22);

C2—Continuous:

$$\xi^2 - L^{-2}(E) = 0, \quad \text{for a range of } E \text{ and } \omega, \quad (23)$$

$$F(E) = F(\xi, E) = F(\xi^+, E), \quad \xi \in C_+(\xi),$$
$$= F(\xi^-, E), \quad \xi \in C_-(\xi), \quad (24)$$

where

$$\left\{C_\pm(\xi) \mid \xi = \pm\left(\frac{v\Sigma_t(E) + i\omega}{vD(E)}\right)^{\frac{1}{2}}, 0 < E < \infty\right\}, \quad (25)$$

$$F(\xi^\pm, E) = P\frac{f(E)}{D(E)[L^{-2}(E) - \xi^2]} + \lambda_\pm(\xi)\delta[\xi \mp L^{-1}(E)], \quad (26)$$

$$\lambda_\pm(\xi) = 1 - P\int_0^\infty \frac{f(E)\,dE}{D(E)[L^{-1}(E) - \xi][L^{-1}(E) + \xi]}. \quad (27)$$

Two discrete solutions exist in each medium, so that the general solution is written as

$$\psi(x, E, t) = e^{i\omega t}\left\{F(\xi_{01}, E)[-a_+e^{-\xi_{01}x} - a_-e^{\xi_{01}x}] - \int_{C_1} A(\xi_1)e^{-\xi_1 x}F(\xi_1, x)\,d\xi_1\right\}, \quad x < 0,$$
$$= e^{i\omega t}\left\{F(\xi_{02}, E)[b_+e^{-\xi_{02}x} + b_-e^{\xi_{02}x}] + \int_{C_2} B(\xi_2)e^{-\xi_2 x}F(\xi_2, x)\,d\xi_1\right\}, \quad x > 0. \quad (28)$$

$a_\pm$, $b_\pm$, $A(\xi_1)$, and $B(\xi_2)$ are expansion coefficients which do not depend upon either E or x. $F(\xi_{0j}, E)$, $j = 1, 2$, are even functions in ξ_0, allowing one to factor them from the discrete solution.

Model D: Setting

$$N(x, \mu, E, t) = N(x, \mu, E)e^{i\omega t} \quad (29)$$

and defining new variables[5,6] Z and u, where

$$u = \frac{v\mu}{v\Sigma_t(E) + i\omega}, \quad (30)$$

$$Z(x, u) = \frac{v\Sigma_t(E) + i\omega}{vM(E)}N(x, \mu, E), \quad (31)$$

yields

$$u\frac{\partial Z(x, u)}{\partial x} + Z(x, u) = \iint_{G(\omega)} \phi(u')Z(x, u')\,du', \quad (32)$$

where

$$\phi(u) = \frac{1}{2}\frac{v\Sigma_s M(E)}{v\Sigma_t(E) + i\omega}. \quad (33)$$

$G(\omega)$ is the area in the u plane which is generated when E and μ take on their range of values. To solve Eq. (32), we follow a separation of variables technique similar to that in the one-speed problem.

Letting

$$Z(x, u) = e^{-x/\nu}f_\nu(u) \quad (34)$$

and requiring that

$$\iint_{G(\omega)} \phi(u')f_\nu(u')\,du' = 1 \quad (35)$$

gives

$$(1 - u/\nu)f_\nu(u) = 1. \quad (36)$$

Again, two types of solutions arise.

D1—Discrete:

$$\nu \neq u \quad \text{or} \quad \nu \notin G(\omega), \quad (37)$$

$$f_{\nu_0}(u) = f_\pm(u) = \frac{\nu_0}{\nu_0 \mp u}, \quad \nu_0 \notin G(\omega). \quad (38)$$

D2—Continuous:

$$\nu = u \quad \text{or} \quad \nu \in G(\omega), \tag{39}$$

$$f_\nu(u) = \frac{\nu}{\nu - u} + p(\nu)\delta(\nu - u), \quad \nu \in G(\omega), \tag{40}$$

$$p(\nu) = \frac{1}{\phi(\nu)}\left(1 - \iint_{G(\omega)} \phi(u)\frac{\nu}{\nu - u}\,du\right). \tag{41}$$

There exist[5] two values $\pm\nu_0$ for Case D1. Hence, the general solution is written as

$$\begin{aligned} e^{i\omega t}Z(x, u) &= -a_+Z_{1+}(x, u) - a_-Z_{1-}(x, u) \\ &\quad - \iint_{G_1(\omega)} A(\nu)Z_{1\nu}(x, u)\,d\nu, \quad x < 0, \\ &= b_+Z_{2+}(x, u) + b_-Z_{1-}(x, u) \\ &\quad + \iint_{G_2(\omega)} B(\nu)Z_{2\nu}(x, u)\,d\nu, \quad x > 0, \end{aligned} \tag{42}$$

where $a_\pm$, $b_\pm$, $A(\nu)$, and $B(\nu)$ are the expansion coefficients.

The conversion to $N(x, \mu, E, t)$ is easily made. However, in evaluating the coefficients, $Z(x, u)$ is easier to deal with.

III. DISCRETE PROPAGATION CONSTANTS

The discrete wave was written as $e^{\pm kx+i\omega t}$, where k is given as follows for the different models:

Model A:

$$k = \left(\frac{\sigma_a + i\omega/v}{D}\right)^{\frac{1}{2}}; \tag{43}$$

Model B:

$$k = \frac{(\sigma_t + i\omega/v)}{\nu_0}, \tag{44}$$

where

$$\sigma_t + \frac{i\omega}{v} = \tfrac{1}{2}c\sigma_t\nu_0 \ln\left(\frac{\nu_0 + 1}{\nu_0 - 1}\right);$$

Model C:

$$k = \xi_0, \tag{45}$$

where

$$\int_0^\infty \frac{g(E)f(E)\,dE}{D(E)[L^{-2}(E) - \xi_0^2]} = 1; \tag{46}$$

Model D:

$$k = \nu_0^{-1}, \tag{47}$$

where

$$\iint_{G(\omega)} \frac{\nu_0}{\nu_0 - u}\phi(u)\,du = 1. \tag{48}$$

Since models A and B are one-speed models, let us examine under what assumptions they are compatible. Expansion of the logarithmic term in model B reveals that

$$\left(\frac{i\omega}{v} + \sigma_t\right)\nu_0^{-1} \simeq \left[3\sigma_t\left(\frac{\sigma_t}{\sigma_s}\right)\left(\sigma_a + \frac{i\omega}{v}\right)\left(1 + \frac{i\omega}{v\sigma_t}\right)^2\right]^{\frac{1}{2}}. \tag{49}$$

If one writes $D = \frac{1}{3}\sigma_t^{-1}$ in Eq. (43), the result is

$$k = \frac{1}{L} = \left[3\sigma_t\left(\sigma_a + \frac{i\omega}{v}\right)\right]^{\frac{1}{2}}. \tag{50}$$

Hence, k of model B reduces to k of model A if $\sigma_s/\sigma_t = c \simeq 1$ and $\omega/v\sigma_t \ll 1$. This means that the one-speed diffusion-theory value for k is obtained from the one-speed transport-theory result, provided that absorption is small, and the collision frequency $v\sigma_t$ is much greater than the source oscillation frequency ω.

Next, the equivalence of the two energy-dependent models is shown. If the Corngold kernel[7] is used, Eq. (46) becomes

$$1 = \int_0^\infty \frac{\Sigma_s M(E)\,dE}{D(E)[L^{-2}(E) - \xi_0^2]}. \tag{51}$$

Equation (48) is given explicitly by

$$\begin{aligned} 1 &= \tfrac{1}{2}\int_0^\infty dE \int_{-1}^1 d\mu \frac{v\Sigma_t M(E)}{v\Sigma_t(E) + i\omega} \\ &\quad \times \nu_0\left(\nu_0 - \frac{v\mu}{v\Sigma_t(E) + i\omega}\right)^{-1}. \end{aligned} \tag{52}$$

The μ integrates at once:

$$\tfrac{1}{2}\nu_0\int_0^\infty dE\Sigma_s M(E) \ln\left[\frac{v\Sigma_t(E) + i\omega + v/\nu_0}{v\Sigma_t(E) + i\omega - v/\nu_0}\right] = 1. \tag{53}$$

The logarithm may be expanded to give

$$\tfrac{1}{2}\nu_0 \ln\left[\ \right] \simeq \frac{v}{v\Sigma_t(E) + i\omega}\left[1 + \frac{1}{3\nu_0^2}\left(\frac{v}{v\Sigma_t(E) + i\omega}\right)^2\right]. \tag{54}$$

For the diffusion theory (Eq. 51),

$$\begin{aligned} &\left(\frac{v}{v\Sigma_t(E) + i\omega}\right)\left(1 - \frac{\xi_0^2 v}{3\Sigma_t(E)[v\Sigma_t(E) + i\omega]}\right)^{-1} \\ &\simeq \left(\frac{v}{v\Sigma_t(E) + i\omega}\right)\left(1 + \frac{\xi_0^2 v^2}{3v\Sigma_t(E)[v\Sigma_t(E) + i\omega]}\right). \end{aligned} \tag{55}$$

Therefore, the integrands reduce to one another provided that $\omega/v\Sigma_t(E)$ is negligible; that is, for the two energy-dependent models to have equivalent separation parameters (or equivalent discrete propagation constants), one must have $(v\Sigma_t)_{\min} \gg \omega$. For the case of $\Sigma_t(E) = \sigma_s + \sigma_a v_0/v$, it is only necessary that $\sigma_a v_0 \gg \omega$. A numerical comparison of the two one-speed theories (A and B) with the two energy-dependent theories (C and D) is simple, provided that

the cross sections are properly normalized. One sees that the spatial dependence of these waves is not energy dependent for these discrete terms, since, in the $e^{kx+i\omega t}$ term, k is independent of energy. The case of the continuum terms is, however, different. The results of some numerical calculations quantifying these effects are shown elsewhere.[8]

IV. CONTINUUM PROPAGATION CONSTANT

The propagation constant k assumes a continuum of values in all models except model A. The set of k is a line set for models B and C, and is an area set in model D. The expressions for these continuum propagation constants are given below:

Model B:

$$k = \left(\frac{i\omega}{v} + \sigma_t\right)\nu^{-1}, \quad \nu \in (-1, 1); \tag{56}$$

Model C:

$$k = \xi, \quad \xi \in C_{\pm}(\xi), \tag{57}$$

$$\left\{C_{\pm}(\xi) \mid \xi = \pm\left(\frac{v\Sigma_t(E) + i\omega}{vD(E)}\right)^{\frac{1}{2}}, 0 < E < \infty\right\}; \tag{58}$$

Model D:

$$k = \nu^{-1}, \quad \nu \in G(\omega), \tag{59}$$

$$\left\{G(\omega) \mid u = \frac{v\mu}{v\Sigma_t(E) + i\omega}, 0 < E < \infty, -1 \leq \mu \leq 1\right\}. \tag{60}$$

In model B, the μ (in model C, the E) variable, gives rise to the continuum solution and, in model D, both E and μ are present, resulting in an area continuum. The regions where these continua exist are shown in Figs. 2–4.

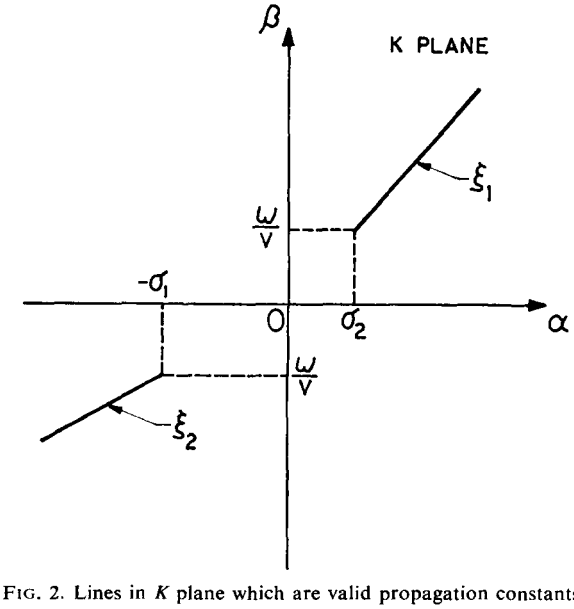


FIG. 2. Lines in K plane which are valid propagation constants.

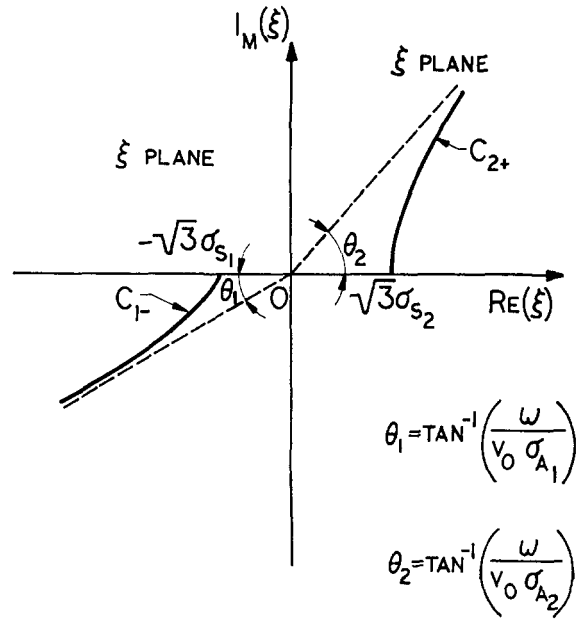


FIG. 3. The open curve $C = C_{1-} \cup C_{2+}$.

V. APPLICATION OF THE BOUNDARY AND INTERFACE CONDITIONS

The source is assumed on the left, and is one of the elementary solutions in each model. The neutron distribution function vanishes at $x = \infty$ and is continuous at the interface. In addition, for model A, continuity of current is required. Thus, we have the following:

Model A:

$$\lim_{x \to \infty} \psi(x, t) = 0, \tag{61}$$

$$\lim_{x \to -\infty} \psi(x, t) = S_0 e^{i\omega t} e^{-x/L_1}, \tag{62}$$

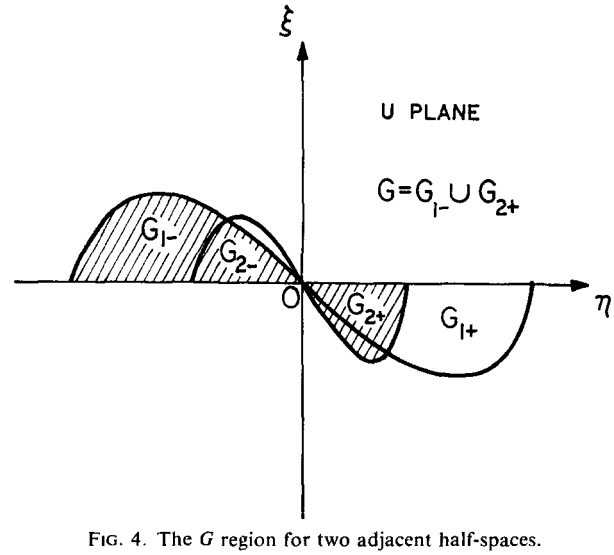


FIG. 4. The G region for two adjacent half-spaces.

$$\psi(0+, t) = \psi(0-, t), \tag{63}$$

$$D_1 \frac{\partial \psi(0-, t)}{\partial x} = D_2 \frac{\partial \psi(0+, t)}{\partial x}; \tag{64}$$

Model B:

$$\lim_{x \to \infty} N(x, \mu, t) = 0, \tag{65}$$

$$\lim_{x \to -\infty} N(x, \mu, t) = S_0 e^{i\omega t} e^{-\xi_{01} x} \phi_{1+}(\mu), \tag{66}$$

$$N(0^+, \mu, t) = N(0^-, \mu, t); \tag{67}$$

Model C:

$$\lim_{x \to \infty} \psi(x, E, t) = 0, \tag{68}$$

$$\lim_{x \to -\infty} \psi(x, E, t) = S_0 e^{i\omega t} F(\xi_{01}, E) e^{-\xi_{01} x}, \tag{69}$$

$$\psi(0^+, E, t) = \psi(0^-, E, t); \tag{70}$$

Model D:

$$\lim_{x \to +\infty} N(x, \mu, E, t) = 0, \tag{71}$$

$$\lim_{x \to -\infty} N(x, \mu, E, t) = S_0 e^{i\omega t} \frac{vM(E)}{v\Sigma_t(e) + i\omega} e^{-x/v_{01}} f_{v_{01}}(u), \tag{72}$$

$$N(0^+, \mu, E, t) = N(0^-, \mu, E, t). \tag{73}$$

Upon application of these boundary and interface conditions, the expansion coefficients in each case can be written down.

For model A, the solution is

$$\begin{aligned} \psi(x, t) &= S_0 e^{i\omega t}[e^{-x/L_1} + (r - 1)e^{x/L_1}], \quad x < 0, \\ &= S_0 e^{i\omega t} r e^{-x/L_2}, \qquad\qquad x > 0, \end{aligned} \tag{74}$$

where

$$r = 2(1 + L_1 D_2 / L_2 D_1)^{-1}. \tag{75}$$

For model B, the vanishing condition at $+\infty$ and the source condition imply

$$\begin{aligned} a_+ &= -S_0, \quad A_1(\nu) = 0, \quad \nu > 0, \\ b_- &= 0, \qquad A_2(\nu) = 0, \quad \nu < 0. \end{aligned} \tag{76}$$

The interface condition leads to

$$\int_0^1 A_2(\nu)\phi_{2\nu}(\mu)d\nu + \int_{-1}^0 A_1(\nu)\phi_{1\nu}(\mu)\, d\nu = h(\mu), \tag{77}$$

where

$$h(\mu) = S_0\phi_{1+}(\mu) - a_-\phi_{1-}(\mu) + b_+\phi_{2+}(\mu). \tag{78}$$

Equation (77) is a singular integral equation for the coefficients $A_1(\nu)$ and $A_2(\nu)$, which can be solved; see, for example, Ref. 9. The results are

$$a_- = -S_0 \frac{\nu_{02} - \nu_{01}}{\nu_{02} + \nu_{01}} \frac{X(\nu_{01})}{X(-\nu_{01})}, \tag{79}$$

$$b_+ = S_0 \frac{2\nu_{01}^2}{\nu_{02}\nu_{01} + \nu_{02}^2} \frac{X(\nu_{01})}{X(\nu_{02})}, \tag{80}$$

$$A_1(\mu) = -\frac{\nu_{01} S_0 X(\nu_{01})}{X^-(\mu)L^+(\mu)}\left(\frac{1}{\mu - \nu_{01}} - \frac{\nu_{02} - \nu_{01}}{\nu_{02} + \nu_{01}} \frac{1}{\mu + \nu_{01}} - \frac{2\nu_{01}}{\nu_{01} + \nu_{02}} \frac{1}{\mu - \nu_{02}}\right), \tag{81}$$

$$A_2(\mu) = -\frac{\nu_{01} S_0 X(\nu_{01})}{X^-(\nu)L^+(\nu)} \frac{c_1\sigma_{t1}}{c_2\sigma_{t2}}\left(\frac{1}{\mu - \nu_{01}} - \frac{\nu_{02} - \nu_{01}}{\nu_{02} + \nu_{01}} \frac{1}{\mu + \nu_{01}} - \frac{2\nu_{01}}{\nu_{01} + \nu_{02}} \frac{1}{\mu - \nu_{02}}\right), \tag{82}$$

where

$$X_0(z) = \exp\left(\frac{1}{2\pi i} \int_{-1}^{1} \frac{\ln G(\mu)\, d\mu}{\mu - z}\right) \tag{83}$$

and the remaining symbols are defined in Ref. 9.

In model C, conditions (1) and (2) lead to

$$\begin{aligned} a_+ &= -S_0, \quad A(\xi_1) = 0, \quad \text{for} \quad \operatorname{Re} \xi_1 > 0, \\ b_- &= 0, \qquad B(\xi_2) = 0, \quad \text{for} \quad \operatorname{Re} \xi_2 < 0. \end{aligned} \tag{84}$$

The interface condition leads to

$$\int_{c_{1-}} A(\xi_1)F(\xi_{1-}, E)\, d\xi_1 + \int_{c_{2+}} A(\xi_2)F(\xi_{2+}, E) = h(\xi_0, E), \tag{85}$$

where

$$h(\xi_0, E) = -F(\xi_{02}, E)b_+ + F(\xi_{01}, E)(S_0 - a_-). \tag{86}$$

This is also a Rieman–Hilbert problem; but, since this is a diffusion model, this line of investigation was discontinued and the continuum solutions were neglected. Assuming that only the discrete part of the spectrum exists and that diffusion-type boundary conditions are employed, another condition is to be added—that the integrated net current at the interface is zero:

$$\int_0^\infty D_1(E) \frac{\partial \psi(0^-, E, t)}{\partial x} dE = \int_0^\infty D_2(E) \frac{\partial \psi(0^+, E, t)}{\partial x} dE. \tag{87}$$

This extra condition allows us to evaluate the discrete coefficients:

$$a_- = S_0\left[\frac{\xi_{01} - \xi_{01}d}{\xi_{01} + \xi_{02}d}\right], \tag{88}$$

$$b_+ = \frac{S_0 2\xi_{01}}{\xi_{01} + \xi_{02}d}, \tag{89}$$

where

$$d = \int_0^\infty D_2(E)F(\xi_{02}, E)\,dE \Big/ \int_0^\infty D_1(E)F(\xi_{01}, E)\,dE. \tag{90}$$

For model D, the vanishing condition at ∞, the source condition at $-\infty$, and the continuity condition at the interface lead to

$$\iint_{G_{2+}} B(\nu)f_{2\nu}(u)\,d\nu + \iint_{G_{1-}} A(\nu)f_{1\nu}(u)\,d\nu = \xi(\mu), \tag{91}$$

where

$$\xi(\mu) = S_0 Z_{1+}(0, u) - b_+ Z_{2+}(0, u) - a_- Z_{1-}(0, u).$$

The details of the solution to Eq. (91) are in the Appendix. The results are

$$a_- = \frac{I_{S_1}I_{20} - I_{S_0}I_{21}}{I_{10}I_{21} - I_{20}I_{11}}, \quad b_+ = \frac{I_{S_1}I_{10} - I_{S_0}I_{11}}{I_{10}I_{21} - I_{20}I_{11}}, \tag{92}$$

$$I_{Sl} = \iint_G \frac{\nu^l X(\nu) S_0 \nu_{01}\,d\nu}{(\nu_{01} - \nu)p(\nu)},$$

$$I_{1l} = \iint_G \frac{\nu^l X(\nu)\nu_{01}\,d\nu}{(\nu_{01} + \nu)p(\nu)}, \tag{93}$$

$$I_{2l} = \iint_G \frac{\nu^l X(\nu)\nu_{02}\,d\nu}{(\nu_{02} - \nu)p(\nu)}, \quad l = 1, 2.$$

The solutions can be written down in all four cases in terms of incident, reflected, and transmitted components:

$$\begin{aligned} \psi &= I + R, \quad x < 0, \\ &= T, \qquad x > 0. \end{aligned} \tag{94}$$

Here I is the incident component, R is the reflected component, and T is the transmitted component. We obtain the following results:

Model A:

$$\psi = \psi(x, t), \quad I = I(x, t), \quad R = R(x, t), \quad T = T(x, t), \tag{95}$$

$$I(x, t) = S_0 e^{i\omega t - (\alpha_1 + i\beta_1)x}, \tag{96}$$

$$R(x, t) = S_0[2(1 + L_1D_2/L_2D_1)^{-1} - 1]e^{i\omega t + (\alpha_1 + i\beta_1)x}, \tag{97}$$

$$T(x, t) = S_0[2(1 + L_1D_2/L_2D_1)^{-1}]e^{i\omega t - (\alpha_2 + i\beta_2)x}; \tag{98}$$

Model B:

$$\begin{aligned} \psi &= N(x, \mu, t), \quad I = I(x, \mu, t), \\ R &= R(x, \mu, t), \quad T = T(x, \mu, t), \end{aligned} \tag{99}$$

$$I = S_0 e^{i\omega t} \exp\left(-\frac{(i\omega/v + \sigma_{t1})x}{\nu_{01}}\right)\phi_{1+}(\mu), \tag{100}$$

$$R = a_- e^{i\omega t} e^{(i\omega/v + \sigma_{t1})x\nu_{01}^{-1}}\phi_{1-}(\mu) + e^{i\omega t}\int_{-1}^{0} A(\nu)\exp\left(-\frac{(i\omega/v + \sigma_{t1})x}{\nu}\right)\phi_{1\nu}(\mu)\,d\nu, \tag{101}$$

$$T = b_+ e^{i\omega t}\exp\left(-\frac{(i\omega/v + \sigma_{t2})x}{\nu_{02}}\right) + e^{i\omega t}\int_0^1 B(\nu)\exp\left(-\frac{(i\omega/v + \sigma_{t2})x}{\nu}\right)\phi_{2\nu}(\mu)\,d\nu; \tag{102}$$

Model C:

$$\begin{aligned} \psi &= \psi(x, E, t), \quad I = I(x, E, t), \\ R &= R(x, E, t), \quad T = T(x, E, t), \end{aligned} \tag{103}$$

$$I = S_0 e^{i\omega t} e^{-\xi_{01}x} F(\xi_{01}, E), \tag{104}$$

$$R = -A_- e^{i\omega t} e^{\xi_{01}x} F(\xi_{01}, E) - e^{i\omega t}\int_{C_{1-}} A(\xi_1)e^{-\xi_1 x}F(\xi_1, E)\,d\xi, \tag{105}$$

$$T = B_+ e^{i\omega t} e^{-\xi_{02}x} F(\xi_{02}, E) + e^{i\omega t}\int_{C_{2+}} A(\xi_2)e^{-\xi_2 x}F(\xi_2, E)\,d\xi_2; \tag{106}$$

Model D:

$$\begin{aligned} \psi &= N(x, \mu, E, t), \quad I = I(x, \mu, E, t), \\ R &= R(x, \mu, E, t), \quad T = T(x, \mu, E, t), \end{aligned} \tag{107}$$

$$I = S_0 e^{i\omega t}\frac{vM(E)}{v\Sigma_t(E) + i\omega} \times \nu_{01}e^{-x/\nu_{01}}\left(\nu_{01} - \frac{v\mu}{v\Sigma_t(E) + i\omega}\right)^{-1}, \tag{108}$$

$$R = e^{i\omega t}\frac{vM(E)}{v\Sigma_t(E) + i\omega} \times \left\{A_-\nu_{01}e^{x/\nu_{01}}\left(\nu_{01} + \frac{v\mu}{v\Sigma_t(E) + i\omega}\right)^{-1} - \iint_{G_{1-}} A(\nu)e^{-x/\nu}f_{1\nu}(u)\,d\nu\right\}, \tag{109}$$

$$T = e^{i\omega t}\frac{vM(E)}{v\Sigma_t(E) + i\omega} \times \left\{B_+\nu_{02}e^{-x/\nu_{02}}\left(\nu_{02} - \frac{v\mu}{v\Sigma_t(E) + i\omega}\right)^{-1} + \iint_{G_{2+}} B(\nu)e^{-x/\nu}f_{2\nu}(u)\,d\nu\right\}. \tag{110}$$

I, which is the source term, was chosen to be discrete only. In models B and D, there is a limiting frequency above which only continuum waves can propagate in a given medium. It is assumed that $\omega < \min[\omega_{L_1}, \omega_{L_2}]$ in both models to be consistent with the assumption that a discrete incident wave coming from $x = -\infty$ can exist. I is a wave which propagates to the right, up to the interface. A component R is identified as a wave propagating to the left; R is purely discrete for model A and is a mixture of both discrete and continuum terms for models B, C, and D. To the right of the interface is the transmitted component T. T is purely discrete for model A but, for $\omega < \omega_{L_2}$, T is a mixture of discrete and continuum terms in models B, C, and D.

VI. CONCLUSIONS

The purpose of this paper has been to give the theory of neutron-wave propagation through an interface. The expressions for the incident, reflected, and transmitted waves are given. Numerical comparisons are shown elsewhere.[10]

From this study one can see that the analysis quickly becomes very complicated as the theory applied becomes more sophisticated. However, the basic concepts of incidence, reflection, and transmission of waves are found to hold in all theories. Extensions to include more general scattering kernels, though not possible in model D, are not expected to change the conceptual results obtained here.

APPENDIX: SOLUTION OF THE GENERALIZED RIEMANN-HILBERT PROBLEM

The problem[11] is to solve

$$\xi(u) = \iint_{G_{2+}} A_2(v) f_{2v}(u)\, dv + \iint_{G_{1-}} A_1(v) f_{1v}(u)\, dv \quad \text{(A1)}$$

for $A_1(v)$ and $A_2(v)$, where u and v are complex variables:

$$u = \alpha + i\beta,$$
$$v = \gamma + i\delta. \quad \text{(A2)}$$

The integrals are integrals in the complex plane and are interpreted to mean

$$\iint_{G_{j\pm}} h(v, u)\, dv = \iint_{G_{j\pm}} h(\gamma + i\delta, u)\, d\gamma\, d\sigma. \quad \text{(A3)}$$

G_j is a region which is antisymmetric about the origin (see Fig. 5). For definition, let G_{j+} equal that part of region G_j where Re $(v) > 0$, and G_{j-} that part where Re $(v) < 0$; that is,

$$G_j = G_{j+} \cup G_{j-}, \quad j = 1, 2. \quad \text{(A4)}$$

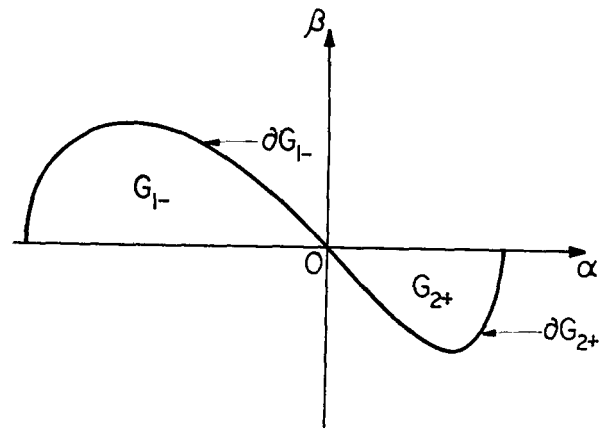


FIG. 5. The G region in the u plane.

If D is a domain, then let $\bar{D}$ denote the closure of D and ∂D denote the boundary of D. The following definitions of the functions are taken from

$$f_{jv}(u) = \frac{v}{v - u} + p_j(u)\delta(v - u), \quad \text{(A5)}$$

$$p_j(v) = \frac{1}{\varphi_j(v)}\left(1 - \iint_{G_j} \frac{\varphi_j(u)u}{u - v}\, du\right). \quad \text{(A6)}$$

For purposes of discussion, $\xi(u)$ and $\varphi(u)$ are considered to be known functions, and we shall impose conditions on them necessary to solve the problem.

The regions G_1 and G_2 are, in general, different from each other; thus, the region of integration specified in Eq. (A1) is

$$G = G_{1-} \cup G_{2+},$$

which is given schematically in Fig. 5.

Two theorems which are needed in the development are next stated. They are proved by Vekna.[11]

Theorem 1: We take $f(u) \in L_1(\bar{G})$, or

$$\iint_G f(u)\, du < \infty,$$

and define the T_G operator as

$$T_G f = -\frac{1}{\pi} \iint_G \frac{f(v)}{v - u}\, dv. \quad \text{(A7)}$$

Then,

(1) $T_G f$ exists $\forall u \notin \bar{G}$;
(2) $T_G f$ is holomorphic $\forall u \notin \bar{G}$;
(3) $T_G f$ behaves like u^{-1} as $u \to \infty$;
(4) $T_G f$ is continuous on ∂G, except perhaps at a finite number of points;

(5) $T_G f$ possesses a generalized derivative with respect to $\bar{u}$ in the Sobolev sense[12]; that is,

$$\frac{\partial}{\partial \bar{u}} T_G f = f(u), \quad u \in G,$$
$$= 0, \qquad u \notin \bar{G}. \tag{A8}$$

Theorem 2: Suppose $f(u) \in L_1(G)$ and

$$T_G f = -\frac{1}{\pi} \iint_G \frac{f(v)}{v-u}\, dv.$$

Suppose further that:

(1) $g(u)$ is holomorphic $\forall u \notin \bar{G}$;
(2) $g(u) \to u^{-1}$ as $u \to \infty$;
(3) $g(u)$ is continuous almost everywhere at ∂G;
(4) $g(u)$ has the generalized derivative

$$\frac{\partial T_G f}{\partial \bar{u}} = f(u), \quad u \in G,$$
$$= 0, \qquad u \notin G.$$

Then,

$$g(u) = T_G f. \tag{A9}$$

In the above, $\bar{u}$ denotes the complex conjugate of u. We now proceed with the solution. From Eq. (A5) we substitute $f_{jv}(u)$, $j = 1, 2$, in Eq. (A1), giving

$$\xi(u) = \iint_{G_{2+}} \frac{A_2(v) v\, dv}{v-u} + \iint_{G_{1-}} \frac{A_1(v) v\, dv}{v-u} + \begin{cases} A_2(u)p_2(u), & u \in G_{2+}, \\ A_1(u)p_1(u), & u \in G_{1-}. \end{cases} \tag{A10}$$

Using the definition of $T_G f$, we write Eq. (A10) as follows:

$$\xi(u) = \pi T_{G_{2+}}(A_2 u) - \pi T_{G_{1-}}(A_1 u) + \begin{cases} A_2(u)p_2(u), & u \in G_{2+}, \\ A_1(u)p_1(u), & u \in G_{1-}. \end{cases} \tag{A11}$$

Also by the same definition, from Eq. (A6), we have

$$p_j(u) = \frac{1 - \pi u T_{Gj}(\varphi_j)}{\varphi_j(u)}, \quad j = 1, 2. \tag{A12}$$

Substituting Eq. (A12) into Eq. (A11) gives

$$\xi(u) = -\pi[T_{G_{2+}}(A_2 u) + T_{G_{1-}}(A_1 u)] + \begin{cases} \dfrac{1 - \pi u T_{G_2}(\varphi_2)}{\varphi_2} A_2(u), & u \in G_{2+}, \\ \dfrac{1 - \pi u T_{G_1}(\varphi_1)}{\varphi_1} A_1(u), & u \in G_{1+}. \end{cases} \tag{A13}$$

The equations simplify considerably with the following notation:

$$G = G_{2-}, \quad u \in G_{2+},$$
$$= G_{1-}, \quad u \in G_{1-}, \tag{A14}$$

$$A(v) = A_2(v), \quad v \in G_{2+},$$
$$= A_1(v), \quad v \in G_{1-}, \tag{A15}$$

$$p(u) = p_2(u), \quad u \in G_{2+},$$
$$= p_1(u), \quad u \in G_{1-}, \tag{A16}$$

$$T_G f = -\frac{1}{\pi} \iint_G \frac{f(v)\, dv}{v-u} = T_{G_{2+}} f + T_{G_{1-}} f. \tag{A17}$$

Accordingly, Eq. (A13) becomes

$$\xi(u) = -\pi T_G(uA) + p(u)A(u), \quad u \in G. \tag{A18}$$

Suppose that

$$\iint_G uA(u)\, du < \infty$$

or

$$\iint_{G_{2+}} uA_2(u)\, du + \iint_{G_{1-}} uA_1(u)\, du < \infty.$$

This is satisfied, if

$$\iint_{G_{2+}} uA_2(u)\, du < \infty,$$
$$\iint_{G_{1-}} uA_1(u)\, du < \infty. \tag{A19}$$

By definition of $T_G(uA)$ and Eq. (A8), there results

$$\frac{\partial}{\partial \bar{u}} T_G(uA) = uA(u), \quad u \in G,$$
$$= 0, \qquad u \notin \bar{G}. \tag{A20}$$

More explicitly,

$$\frac{\partial}{\partial \bar{u}} T_G(uA) = uA_1(u), \quad u \in G_{1-},$$
$$= uA_2(u), \quad u \in G_{2+},$$
$$= 0, \qquad u \notin G_{2+} \cup G_{1-} + \text{boundaries}.$$

To solve Eq. (A18), we define a function $X(u)$ such that

$$X(u) \in L_1(\bar{G}) \tag{A21}$$

and form the function

$$X(u) T_G(uA).$$

By the chain rule for differentiation,

$$\frac{\partial}{\partial \bar{u}} [X(u) T_G(uA)] = \frac{\partial X}{\partial \bar{u}} T_G(uA) + X(u) \frac{\partial}{\partial \bar{u}} T_G(uA). \tag{A22}$$

By combining Eq. (A20) and Eq. (A22), one gets

$$\frac{\partial}{\partial \bar{u}}[X(u)T_G(uA)] = \frac{\partial X}{\partial \bar{u}} T_G(uA) + X(u)uA(u), \quad u \in G,$$
$$= \frac{\partial X}{\partial \bar{u}} T_G(uA), \quad u \notin \bar{G}. \quad \text{(A23)}$$

From Eq. (A18) we solve for $uA(u)$:

$$uA(u) = \frac{u\xi(u) + \pi u T_G(uA)}{p(u)} \quad \text{(A24)}$$

for $p(u) \neq 0$ [this is also necessary since it was required that $uA(u) \in L_1(\bar{G})$]. This restriction imposes certain conditions on $\xi(u)$ and $uT_G(uA)$.

By substituting Eq. (A24) into Eq. (A23), there results

$$\frac{\partial}{\partial \bar{u}}[X(u)T_G(uA)] = \frac{\partial X}{\partial \bar{u}} T_G(uA) + \frac{X(u)[u\xi(u) + \pi u T_G(uA)]}{p(u)}, \quad u \in G,$$
$$= \frac{\partial X}{\partial \bar{u}} T_G(uA), \quad u \notin \bar{G}. \quad \text{(A25)}$$

Equation (A25) is already in a form which allows the use of some of the techniques of generalized analytic functions. In particular, $X(u)$ is arbitrary, except for the requirement that it is analytic everywhere so that $\partial X/\partial \bar{u}$ is well defined. The procedure is to construct $X(u)$ so that Eq. (A25) can be solved. Note that Eq. (A25) already contains the essence of the original Eq. (A18). Now we suppose that $X(u)$ is defined such that

$$\frac{\partial X}{\partial \bar{u}} = \frac{-X(u)\pi u}{p(u)}, \quad u \in G,$$
$$= 0, \quad u \notin \bar{G}. \quad \text{(A26)}$$

Equation (A26), which $X(u)$ has to obey, is one requirement of Theorem 2. The actual construction of $X(u)$ is delayed until later. Using Eq. (A26) into Eq. (A25), one gets

$$\frac{\partial}{\partial \bar{u}}[X(u)T_G(uA)] = \frac{X(u)u\xi(u)}{p(u)}, \quad u \in G,$$
$$= 0, \quad u \notin \bar{G}. \quad \text{(A27)}$$

Theorem 2 shall now be used. All the requirements of the theorem have to be satisfied. It is, therefore, necessary to suppose that:

(1) $X(u)u\xi(u)/p(u) \in L_1(\bar{G})$;
(2) $X(u)T_G(uA)$ is holomorphic $\forall u \notin G$;
(3) $X(u)T_G(uA) \to 0$ as $u \to \infty$;
(4) $X(u)T_G(uA)$ is continuous almost everywhere on ∂G.

Because of Eq. (A27), the above are just the assumptions of Theorem 2; and, therefore, using the conclusion of the theorem, one gets

$$X(u)T_G(uA) = T_G\left[\frac{X(u)u\xi(u)}{p(u)}\right], \quad \text{(A28)}$$

which can be solved for $T_G(uA)$ provided $X(u)$ is a nonvanishing function

$$T_G(uA) = \frac{1}{X(u)} T_G\left[\frac{X(u)u\xi(u)}{p(u)}\right]. \quad \text{(A29)}$$

Equation (A29) is in a form which allows one to solve for $uA(u)$. We had already seen that $uA(u)$ obeys Eq. (A20), provided that $uA(u) \in L_1(G)$ by Theorem 1. Hence, by using the aerolar derivative in the Sobolev sense,

$$uA(u) = \frac{\partial}{\partial \bar{u}}\left[\frac{1}{X(u)} T_G\left(\frac{X(u)u\xi(u)}{p(u)}\right)\right], \quad \text{(A30)}$$

which gives the desired coefficients. In particular,

$$A_1(u) = \frac{1}{u}\frac{\partial}{\partial \bar{u}}\left[\frac{-1}{\pi X(u)} \iint_G \frac{X(u')u'\xi(u')\,du'}{p(u')(u'-u)}\right],$$
$$A_2(u) = \frac{1}{u}\frac{\partial}{\partial \bar{u}}\left[\frac{-1}{\pi X(u)} \iint_G \frac{X(u')u'\xi(u')\,du'}{p(u')(u'-u)}\right], \quad \text{(A31)}$$

where $X(u)$ is different for each u and so with $p(u)$.

The X function is considered next, since it was seen that this function is crucial in order to solve for $A_1(u)$ and $A_2(u)$. The following conditions were imposed on $X(u)$:

(1) By Eq. (A26),

$$\frac{\partial X}{\partial \bar{u}} = \frac{-X(u)\pi u}{p(u)}, \quad u \in G,$$
$$= 0, \quad u \notin \bar{G}; \quad \text{(A32)}$$

(2) $X(u)$ is a nonvanishing function in G; this is necessary to get Eq. (A29);

(3) $X(u)$ must be holomorphic, $\forall u \in \bar{G}$, since (a) $X(u)T_G(uA)$ is holomorphic, $\forall u \notin \bar{G}$, and (b) $uA(u) \in L_1(\bar{G})$ so that, by Theorem 1, $T_G(uA)$ is holomorphic in $\bar{G}$; hence, this is required of $X(u)$ as well;

(4) $X(u)$ is continuous everywhere on ∂G; this follows from the requirement that $X(u)T_G(uA)$ is

everywhere continuous on ∂G and $T_G(uA)$ is everywhere continuous, again by Theorem 1.

The construction of the X function is done by the use of Theodorescus's formula in Vekua.[13] Theodorescus's formula gives a solution to the following equation:

$$\frac{\partial W}{\partial \bar{u}} + A(u)W = 0, \tag{A33}$$

where W is a function of the complex variable u. The solution to Eq. (A33) is

$$W = \psi(u)e^{-T_G(A)}, \tag{A34}$$

where $\psi(z)$ is an arbitrary analytic function. To use this result, we consider a function $X_0(u)$ defined by

$$X_0(u) = \exp T_G\left(\frac{-\pi u}{p(u)}\right) = \exp T_G\left(\frac{-\pi u\varphi(u)}{1 - \pi u T_H(\varphi)}\right), \tag{A35}$$

where H implies the whole region G_1 or G_2, depending upon whether Re $(u) < 0$ or Re $(u) > 0$, respectively.

If the derivative of Eq. (A35) is taken,

$$\frac{\partial X_0}{\partial \bar{u}} = X_0(u)\frac{\partial}{\partial \bar{u}} T_G\left(\frac{-\pi u}{p(u)}\right). \tag{A36}$$

To solve Eq. (A36), we impose

$$\frac{u}{p(u)} \in L_1(\tilde{G}). \tag{A37}$$

This will have the effect of requiring $p(u)$ not to vanish in $\tilde{G}$, except perhaps at $u = 0$. By Theorem 1,

$$\frac{\partial}{\partial \bar{u}} T_G\left(\frac{-\pi u}{p(u)}\right) = \frac{-\pi u}{p(u)}, \quad u \in G,$$
$$= 0, \qquad u \notin \tilde{G}. \tag{A38}$$

Therefore,

$$\frac{\partial X_0}{\partial \bar{u}} = \frac{X_0 \pi u}{p(u)}, \quad u \in G,$$
$$= 0, \qquad u \notin \tilde{G}. \tag{A39}$$

Hence, $X_0(z)$ satisfies the derivative condition (1). To construct an X function which satisfies also the rest of the conditions, we assume that

$$X(u) = \psi(u)X_0(u), \tag{A40}$$

where $\psi(u)$ is an analytic function. Since ψ is analytic,

$$\frac{\partial \psi}{\partial \bar{u}} = \frac{\partial u}{\partial \bar{u}}\frac{\partial \psi}{\partial u} = 0; \tag{A41}$$

hence,

$$\frac{\partial X}{\partial u} = \psi(u)\frac{\partial X_0}{\partial \bar{u}} = \frac{-\psi(u)X_0(u)\pi u}{p(u)}, \quad u \in G,$$
$$= 0, \qquad u \notin \tilde{G}. \tag{A42}$$

Hence, the $X(u)$ of Eq. (A40) still satisfies the differentiation condition (1).

Now $\psi(u)$ shall next be determined by the remaining conditions that were imposed on $X(u)$. To do this, we observe that

$$T_G\left(\frac{-\pi u}{p(u)}\right) = -\frac{1}{\pi}\iint_G \frac{-\pi u'\,du'}{p(u')(u' - u)}; \tag{A43}$$

thus,

$$T_H\left(\frac{-\pi u}{p(u)}\right) = \iint_{G_{2+}} \frac{u'\varphi_2(u')\,du'}{1 - \pi u' T_{G_2}(\varphi_2)(u' - u)} + \iint_{G_{1-}} \frac{u'\varphi_2(u')\,du'}{1 - \pi u' T_{G_1}(\varphi_1)(u' - u)}. \tag{A44}$$

By defining

$$\frac{\varphi(u)}{1 - \pi u T_H(\varphi)} = \frac{\varphi_1(u)}{1 - \pi u T_{G_1}(\varphi_1)}, \quad u \in G_{1-},$$
$$= \frac{\varphi_2(u)}{1 - \pi u T_{G_2}(\varphi_2)}, \quad u \in G_{2+}, \tag{A45}$$

we obtain

$$T_G\left(\frac{-\pi u}{p(u)}\right) = T_G\left[\frac{-\pi u\varphi}{1 - \pi u T_H\varphi}\right]. \tag{A46}$$

Now, since

$$\frac{\partial}{\partial \bar{u}} \ln f(u) = \frac{\partial f}{\partial \bar{u}}\frac{1}{f(u)},$$

there results

$$T_G\left[\frac{-\pi u\varphi}{1 - \pi u T_H(\varphi)}\right] = T_G\left\{\frac{\partial}{\partial \bar{u}} \ln\left[1 - \pi u T_H(\varphi)\right]\right\}. \tag{A47}$$

The last inequality follows since

$$\frac{\partial}{\partial \bar{u}}\left[u T_H(\varphi)\right] = u\frac{\partial}{\partial \bar{u}}\left(T_H(\varphi)\right) = u\varphi, \quad u \in H,$$
$$= 0, \qquad u \notin \tilde{H}. \tag{A48}$$

By definition,

$$T_G\left\{\frac{\partial}{\partial \bar{u}} \ln\left[1 - \pi u T_H(\varphi)\right]\right\} = -\frac{1}{\pi}\iint \frac{1}{(u' - u)}\frac{\partial}{\partial \bar{u}'} \ln\left[1 - \pi u' T_H(\varphi)\right] du'. \tag{A49}$$

Also, by a generalized Green's formula,[14]

$$f(\xi) = \frac{1}{2\pi}\int \frac{f(z)\,dz}{z - \xi} - \frac{1}{\pi}\iint \frac{\partial f}{\partial \bar{z}}\frac{dz}{z - \xi} \tag{A50}$$

and, therefore,

$$T_G\left[\frac{-\pi u\varphi(u)}{1-\pi uT_H(\varphi)}\right] = \ln\left[1-\pi uT_H(\varphi)\right] - \frac{1}{2\pi i}\int_{\partial G}\frac{\ln\left[1-\pi zT_H(\varphi)\right]}{z-u}dz, \quad u\in G,$$

$$= -\frac{1}{2\pi i}\int_{\partial G}\frac{\ln\left[1-\pi zT_H(\varphi)\right]}{z-u}dz, \quad u\notin \bar{G}. \qquad \text{(A51)}$$

The purpose of introducing the logarithm is to convert the expression in the exponential of the $X_0(u)$ function to a form where one can use the generalized Green's formulas.

Because of the presence of the logarithm, there is a change in the value across the branch when there exists a discrete spectrum. If we are talking about the region H, with ∂H as its boundary, we define

$$L(u) = \ln\left[1-\pi uT_H(\varphi)\right]. \qquad \text{(A52)}$$

Then the change in the imaginary part of $[1-\pi uT_H(\varphi)]$ is equal to $2\pi i$ (number of zeros–number of poles) outside H. But $\ln[1-\pi uT_H(\varphi)]$ has no zeros, only poles; hence,

$$\Delta I_m[L(u)] = 0, \quad 1-\pi uT_H(\varphi)\neq 0,$$
$$= -2n\pi i, \quad 1-\pi uT_H(\varphi) = 0. \qquad \text{(A53)}$$

The value $2n\pi i$ is used when there exist $2n$ possible solutions.

From the structure of the regions, there are four possibilities concerning the position of the zeros. Let u_0 be a zero of $1-\pi uT_H(\varphi)$. Then, we have the following:

Case 1: $u_0\notin G$;
Case 2: $u_0\in G_{2+}$, but $u_0\notin G_{1-}$;
Case 3: $u_0\in G_{1-}$, but $u_0\notin G_{2+}$;
Case 4: $u_0\in G_{1-}, G_{2+}$.

Since

$$H = G_1, \quad \text{Re}\,(u) < 0,$$
$$= G_2, \quad \text{Re}\,(u) > 0,$$

then

$$\Delta I_m[L(u)] = -4\pi i, \quad \text{for Case 1},$$
$$= -2\pi i, \quad \text{for Case 2},$$
$$= -2\pi i, \quad \text{for Case 3},$$
$$= 0, \quad \text{for Case 4};$$

so, for $X(u)$ to satisfy property (4), it is necessary that

$$\psi(u) = u^{-2}, \quad \text{for Case 1},$$
$$= u^{-1}, \quad \text{for Case 2},$$
$$= u^{-1}, \quad \text{for Case 3},$$
$$= u^{0}, \quad \text{for Case 4}.$$

Thus, the $X(u)$ function we need is

$$X(u) = u^{-h}\exp\left[T_H\left(\frac{-\pi u}{p(u)}\right)\right], \qquad \text{(A54)}$$

where h assumes values 2, 1, 1, and 0 for Cases 1–4, respectively. Knowing $X(u)$, we can evaluate the coefficients, and the solution is complete.

Where the zero of $[1-\pi uT_H(\varphi)]$ exists is determined by the function $\varphi(u)$. $\varphi(u)$, from Eq. (33), is a function of the frequency ω and the cross sections.

* Present address: Holmes & Narver, Inc., 828 South Figueroa Street, Los Angeles, California 90017

[1] J. J. Duderstadt, J. Math. Phys. **10**, 266 (1969).
[2] J. Warner, Ph.D. thesis, University of California, Los Angeles, 1968.
[3] K. H. Beckurts and K. Wirts, *Neutron Physics* (Springer-Verlag, New York, 1964).
[4] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[5] C. Cercignani, Ann. Phys. (N.Y.) **40**, 454 (1966).
[6] C. Cercignani, Ann. Phys. (N.Y.) **40**, 469 (1966).
[7] N. Corngold, Nucl. Sci. Eng. **19**, 80 (1964).
[8] O. C. Baldonado and R. C. Erdmann, Nukleonik **12**, 57 (1969).
[9] O. C. Baldonado, thesis, University of California, Los Angeles, California, 1968.
[10] O. C. Baldonado and R. C. Erdmann, Nucl. Sci. Eng. **37**, 59 (1969).
[11] This Appendix is based on the work of Cercignani, Ref. 5 and I. Vekua, *Generalized Analytic Functions* (Pergamon Press, New York, 1962).
[12] I. Vekua, Ref. 11, Sec. 5, Chap. 1.
[13] I. Vekua, Ref. 11, Sec. 4, Chap. 3.
[14] I. Vekua, Ref. 11, Sec. 7, Chap. 1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Birkhoff Gravitational Field of Milne's Universe

J. B. BOYLING
Department of Applied Mathematics, University of Leeds, Leeds 2, England

(Received 18 June 1969)

It is shown that Milne's special relativistic cosmology is consistent with Birkhoff's flat-space theory of gravitation and maintains its stability in the presence of gravity. The gravitational correction to the redshift in the light from distant galaxies is calculated.

1. INTRODUCTION

It has been pointed out by Milne[1] that his special relativistic uniformly expanding model of the universe must be consistent with any conceivable flat-space theory of gravitation, in the sense that gravity will not exert any net force on the fundamental particles (galaxies) to disturb their uniform motion. This follows at once from symmetry considerations, since an observer attached to a fundamental particle will always see the universe as being spherically symmetric about himself. In the case of Whitehead's theory of gravitation,[2] this consistency property has been demonstrated explicitly by Rayner,[3] who also proves that the model is stable against disturbances of the motion of its fundamental particles. In fact, a test particle projected with arbitrary velocity ultimately comes to rest relative to its local surroundings in the universe. The behavior of photons in this universe is also investigated. The presence of gravity results in a small correction to the formula for the redshift in light coming from distant galaxies.

In this paper we consider the effect on Milne's cosmology of Birkhoff's flat-space theory of gravitation.[4] We find that the consistency and stability properties again hold, but there is no effect corresponding to the "visual barrier" of the Whitehead theory.[3]

2. THE GRAVITATIONAL POTENTIAL

We shall assume that the special theory of relativity holds and that the universe is populated by a dust of fundamental particles which originated from a primeval fireball. Only those inertial frames will be considered in which the explosion of this fireball is taken as the origin of space–time. It follows that the fundamental particles are confined within the forward light-cone of this initial catastrophe, i.e., they can only occupy world points whose space–time coordinates $x^0 = ct$, x^1, x^2, and x^3 satisfy the conditions

$$x^0 = ct > 0$$

and

$$x^2 = x^\mu x_\mu = g_{\mu\nu}x^\mu x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = c^2t^2 - r^2 > 0.$$

We assume[1] that the 4-velocity of the fundamental particle at the world point x^μ is

$$u^\mu = c(x^2)^{-\frac{1}{2}}x^\mu, \tag{1}$$

and that the density ρ of the fundamental particles (i.e., rest-mass per unit volume) is given by

$$\rho(x) = Bc^{-1}t(c^2t^2 - r^2)^{-2} = Bc^{-2}x^0(x^2)^{-2},$$

where B is a constant with the dimensions of energy. The energy tensor must therefore be

$$T^{\mu\nu} = B(x^2)^{-\frac{5}{2}}x^\mu x^\nu. \tag{2}$$

We shall further suppose that gravity may be accounted for by Birkhoff's theory of gravitation[4] in the modified form proposed by Weyl.[5] Here the gravitational field is derived from a symmetric second-order tensor potential $h^{\mu\nu}$ which satisfies the field equations

$$\Box h^{\mu\nu} = \kappa\tilde{T}^{\mu\nu}, \tag{3}$$

$$\partial_\nu h^{\mu\nu} = \tfrac{1}{2}\partial^\mu h, \tag{4}$$

where

$$\Box = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2,$$

$$\partial_\nu = \frac{\partial}{\partial x^\nu},$$

and we have used the notations

$$h^\mu_\mu = h,$$
$$T^\mu = T,$$
$$\tilde{T}^{\mu\nu} = T^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}T.$$

The constant κ is related to Newton's universal constant of gravitation G by

$$\kappa = 8\pi Gc^{-4}.$$

The gravitational 4-force F_μ acting on a test particle of (rest) mass m and 4-velocity u^μ is given by

$$F_\mu = m(\partial_\lambda h_{\mu\nu} - \partial_\mu h_{\lambda\nu})u^\nu u^\lambda. \tag{5}$$

We note that this 4-force satisfies the orthogonality condition

$$F_\mu u^\mu = 0.$$

Now the world model under consideration has the property of looking exactly the same in every inertial frame in which time moves forwards and the initial catastrophe is taken as space-time origin. It follows that the symmetric tensor field $h^{\mu\nu}$ must take the same form in every such inertial frame. This can only happen if

$$h^{\mu\nu} = \phi(x^2)x^\mu x^\nu + \psi(x^2)g^{\mu\nu} \tag{6}$$

for suitable functions ϕ and ψ of the single variable

$$x^2 = X.$$

We shall now investigate the conditions imposed on the functions ϕ and ψ by the field equations (3) and (4) with the energy tensor (2).

Substituting (6) and (2) into (3) gives

$$4(4\phi' + x^2\phi'')x^\mu x^\nu + 2(\phi + 4\psi' + 2x^2\psi'')g^{\mu\nu} = \kappa B(x^2)^{-\frac{5}{2}}(x^\mu x^\nu - \tfrac{1}{2}x^2 g^{\mu\nu}),$$

where the primes denote differentiation with respect to the variable X. Consequently,

$$4\phi' + x^2\phi'' = \tfrac{1}{4}\kappa B(x^2)^{-\frac{5}{2}} \tag{7}$$

and

$$\phi + 4\psi' + 2x^2\psi'' = -\tfrac{1}{4}\kappa B(x^2)^{-\frac{3}{2}}. \tag{8}$$

Similarly, substitution of (6) into (4) yields

$$x^2\phi' + 4\phi - 2\psi' = 0. \tag{9}$$

Eliminating the constant B from Eqs. (7) and (8), we obtain

$$X^2\phi'' + 4X\phi' + \phi + 4\psi' + 2X\psi'' = 0.$$

Substituting for ψ' and ψ'' from (9), we find that ϕ satisfies the homogeneous differential equation

$$2X^2\phi'' + 11X\phi' + 9\phi = 0$$

whose general solution is

$$\phi = PX^{-\frac{3}{2}} + QX^{-3}, \tag{10}$$

where P and Q are constants. Substituting (10) into (7) we find that

$$P = -\tfrac{1}{9}\kappa B. \tag{11}$$

Substituting (10) and (11) into (9) gives

$$\psi' = -\tfrac{5}{36}\kappa B X^{-\frac{3}{2}} + \tfrac{1}{2}QX^{-3},$$

whence

$$\psi = \tfrac{5}{18}\kappa B X^{-\frac{1}{2}} - \tfrac{1}{4}QX^{-2} + R \tag{12}$$

for some constant R. Substituting (10), (11), and (12) into (6), we see that the gravitational potential $h_{\mu\nu}$ must be of the form

$$h_{\mu\nu} = \tfrac{1}{18}\kappa B(x^2)^{-\frac{3}{2}}(5x^2 g_{\mu\nu} - 2x_\mu x_\nu) + \partial_\mu\partial_\nu\Lambda, \tag{13}$$

where

$$\Lambda = Q/8x^2 + \tfrac{1}{2}Rx^2$$

is a solution of the wave equation

$$\Box\,\Lambda = 0.$$

Since the last term on the right-hand side of (13) makes no contribution to the force (5), we may assume $\Lambda = 0$ and

$$h_{\mu\nu} = \tfrac{1}{18}\kappa B(x^2)^{-\frac{3}{2}}(5x^2 g_{\mu\nu} - 2x_\mu x_\nu). \tag{14}$$

In this last step we have simply used the gauge invariance of the Birkhoff theory.

3. THE MOTION OF TEST PARTICLES

The equation of motion for a test particle of mass m is

$$m\frac{d^2x^\mu}{d\tau^2} = F^\mu,$$

where $d\tau$ denotes the interval of proper time along the world path of the particle, and F^μ is the 4-force given by (5). Substituting (14) into (5), we find that the equation of motion for a massive test particle in Milne's universe is

$$\frac{d^2x^\mu}{d\tau^2} = -\tfrac{1}{6}\kappa B(x^2)^{-\frac{3}{2}}[(x_\lambda u^\lambda)u^\mu - c^2x^\mu]. \tag{15}$$

Since the 4-velocity of a fundamental particle is given by (1), we see at once that such a particle is not accelerated by gravity, so that the model is indeed consistent. To prove that the model is also stable, we must show more generally that the motion of a test particle always becomes asymptotically that of a fundamental particle. Let v be the instantaneous speed of the test particle at a general world point x^μ on its world path as seen by an observer attached to a fundamental particle at x^μ. We must show that v tends to zero as the proper time τ of the test particle (measured from some arbitrary "initial event" on its world path) tends to (plus) infinity. This is equivalent to showing that $\gamma \to 1$, where

$$\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}.$$

Now the observer attached to the fundamental particle at x^μ may use an inertial frame in which he is

permanently at rest at the origin of space. In such an inertial frame

$$x^0 = (x^2)^{\frac{1}{2}} = X^{\frac{1}{2}}, \quad x^1 = x^2 = x^3 = 0,$$

and the time component of the instantaneous 4-velocity of the test particle is

$$u^0 = \gamma c.$$

Therefore

$$x_\mu u^\mu = \gamma c (x^2)^{\frac{1}{2}},$$

which gives a Lorentz-invariant expression for the quantity γ. Now

$$u^\mu = \frac{dx^\mu}{d\tau}$$

and

$$X = x_\mu x^\mu,$$

so that

$$\frac{dX}{d\tau} = 2x_\mu u^\mu, \tag{16}$$

whence

$$\gamma = c^{-1}(x^2)^{-\frac{1}{2}} x_\mu u^\mu = (2c)^{-1} X^{-\frac{1}{2}} \frac{dX}{d\tau}. \tag{17}$$

Differentiation of (16) yields

$$\begin{aligned} \frac{d^2X}{d\tau^2} &= 2\left(x_\mu \frac{du^\mu}{d\tau} + u_\mu u^\mu\right) \\ &= 2\left(x_\mu \frac{d^2x^\mu}{d\tau^2} + c^2\right). \end{aligned} \tag{18}$$

Contracting both sides of the equation of motion (15) with $2x_\mu$ and substituting from (16) and (18), we obtain

$$\frac{d^2X}{d\tau^2} + \tfrac{1}{12}\kappa B X^{-\frac{3}{2}} \left(\frac{dX}{d\tau}\right)^2 = 2c^2\{1 + \tfrac{1}{6}\kappa B X^{-\frac{1}{2}}\}. \tag{19}$$

Putting

$$\frac{dX}{d\tau} = w,$$

we may rewrite (19) in the form

$$w\frac{dw}{dX} + \tfrac{1}{12}\kappa B X^{-\frac{3}{2}} w^2 = 2c^2\{1 + \tfrac{1}{6}\kappa B X^{-\frac{1}{2}}\}.$$

Multiplying both sides by the integrating factor

$$\exp(-\tfrac{1}{3}\kappa B X^{-\frac{1}{2}}),$$

we find that

$$\left(\frac{dX}{d\tau}\right)^2 = w^2 = 4c^2 X + A \exp(\tfrac{1}{3}\kappa B X^{-\frac{1}{2}}),$$

where A is a constant. It follows from (17) that

$$\gamma = [1 + (A/4c^2X)\exp(\tfrac{1}{3}\kappa B X^{-\frac{1}{2}})]^{\frac{1}{2}}. \tag{20}$$

As $\tau \to +\infty$, it follows from (17) since $\gamma \geq 1$, that $X \to +\infty$ and hence from (20), that $\gamma \to 1$. The stability of the model is therefore proved.

4. REDSHIFTS IN RADIATION FROM FUNDAMENTAL PARTICLES

The preceding argument used for a massive test particle cannot be applied directly to the motion of a photon, because 4-force is not a meaningful concept for a particle of zero mass. However, we can rewrite the equation of motion (15) for a massive particle in the differential form

$$dp^\mu = -\tfrac{1}{6}\kappa B (x^2)^{-\frac{3}{2}}\{p^\mu\, d(\tfrac{1}{2}x^2) - x^\mu p_\lambda\, dx^\lambda\}, \tag{21}$$

where

$$p^\mu = mu^\mu = m\frac{dx^\mu}{d\tau}$$

is the 4-momentum of the particle. The differential equation of motion (21) still makes sense when $m = 0$ and may be used to determine the motion of a photon in the Birkhoff gravitational potential (14). Now, for a photon of frequency ν, we have

$$dx^\lambda = c^2(h\nu)^{-1} p^\lambda\, dt$$

and

$$p^\lambda p_\lambda = 0.$$

Consequently,

$$p_\lambda\, dx^\lambda = 0$$

and

$$dp^\mu = -\tfrac{1}{12}\kappa B X^{-\frac{3}{2}} p^\mu\, dX,$$

whence

$$p^\mu \exp(-\tfrac{1}{6}\kappa B X^{-\frac{1}{2}}) = \text{const.}$$

In particular, the frequency of the photon in any given inertial frame will decrease as time proceeds since

$$\nu \exp(-\tfrac{1}{6}\kappa B X^{-\frac{1}{2}}) = \text{const.} \tag{22}$$

We shall now consider the light seen by an observer attached to a fundamental particle, and shall use an inertial frame in which this observer is permanently at rest at the origin of space. Suppose that at time t_0 he observes light which originated at time $t_0 - r/c$ in a galaxy whose distance from him at the instant of emission of the light was r. At that instant the galaxy was receding from him at speed $r(t_0 - r/c)^{-1}$. Suppose that the light emitted by the galaxy has frequency ν in an inertial frame in which the galaxy is at rest. The frequency $\bar{\nu}$ of this same light in the rest frame of the observer is given by the Doppler formula

$$\bar{\nu} = \nu(1 - 2r/ct_0)^{\frac{1}{2}}. \tag{23}$$

It follows from (22) that the frequency ν' of the light actually seen by the observer is related to $\bar{\nu}$ by

$$\nu' \exp(-\kappa B/6ct_0) = \bar{\nu} \exp\{-\tfrac{1}{6}\kappa B[(ct_0 - r)^2 - r^2]^{-\frac{1}{2}}\}.$$

Substituting for $\bar{\nu}$ from (23), we finally obtain

$$\nu' = \nu\left(1 - \frac{2r}{ct_0}\right)^{\frac{1}{2}} \exp\left\{-\frac{\kappa B}{6ct_0}\left[\left(1 - \frac{2r}{ct_0}\right)^{-\frac{1}{2}} - 1\right]\right\},$$

in which the exponential factor gives the gravitational correction to the usual redshift formula of the Milne cosmology. This factor differs significantly from unity only when r is close to $\frac{1}{2}ct_0$, i.e., only for light from a galaxy close to the edge of the visible universe. Normally the gravitational correction is quite negligible.

[1] E. A. Milne, *Relativity, Gravitation and World-Structure* (Oxford University Press, London, 1935).
[2] A. N. Whitehead, *The Principles of Relativity* (Cambridge University Press, Cambridge, 1922).
[3] C. B. Rayner, Proc. Roy. Soc. (London) **A222**, 509 (1954).
[4] G. D. Birkhoff, Proc. Natl. Acad. Sci. U.S. **29**, 231 (1943).
[5] H. Weyl, Am. J. Math. **66**, 591 (1944).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Search for a Universal Symmetry Group in Two Dimensions

ARTURO CISNEROS* AND HAROLD V. MCINTOSH
Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional
México 14, D.F., México

(Received 20 June 1968)

Recently several authors have proposed a universal symmetry group and demonstrated the classical validity of the concept. Supposedly, an $SU(n)$ symmetry group could be constructed for whatsoever system of n degrees of freedom. The claim is assuredly valid for all classically degenerate systems, but is in contradiction with most of the well-known and widely accepted solutions of Schrödinger's equation. We examine the reasons for this discrepancy on the quantum-mechanical level. The construction of the universal symmetry group requires ladder operators, which in most cases are the ladder operators of Infeld and Hill. Complications which owe their origin entirely to numerical relationships imposed by quantization prevent these operators from forming a von Neumann algebra and, in turn, an $SU(n)$ group of constants of the motion. Two important effects are those imposed by anisotropy, wherein not all quanta have the same size, and by non-Cartesian coordinates, wherein quanta in some dimensions are restricted in size by those in other dimensions. These effects can be seen quite clearly in two-dimensional systems: anisotropy in the Cartesian anisotropic harmonic oscillator, and conflict between dimensions in the polar form of the isotropic oscillator. Further complications arise when the two effects are combined, as in the harmonic oscillator or hydrogen atom with "excess" angular momentum. Enough residue of the universal symmetry concept remains that many similarities between the hydrogen atom and harmonic oscillator may be understood, including the fact that some levels of the hydrogen atom, which ordinarily transform according to an orthogonal group, may form irreducible representations of the unitary group.

I. INTRODUCTION

Shortly after Fock's[1] remarkable analysis of the symmetries and degeneracies of the quantum-mechanical hydrogen atom, similar methods were applied to the study of the isotropic harmonic oscillator, a second quantum-mechanical system of major importance. In both cases, the apparent symmetry of the problem was spherical, due to the central forces present, for which group-theoretical reasoning predicted a degeneracy in the z component of the angular momentum. While it is true that such a degeneracy is always present in the solutions of central force problems, the degeneracy occurring in the hydrogen atom and the harmonic oscillator was far greater. Such additional degeneracies, which are observed to occur in a few other situations as well, had been widely considered to have been "accidental," since their occurrence was not foreseen from group-theoretical reasoning based on the ostensible spherical symmetry. They had to be accepted on the grounds that amongst all possible potentials there were certainly some for which the energies of wavefunctions belonging to different symmetry types might just happen to coincide.

Thus there was a certain satisfaction when Fock showed that certain potentials might possess "hidden" symmetries, which is to say, symmetries arising from regularities in the phase space of Hamiltonian dynamics which were not immediately apparent from an inspection of the geometric symmetries of the

It follows from (22) that the frequency ν' of the light actually seen by the observer is related to $\bar{\nu}$ by

$$\nu' \exp(-\kappa B/6ct_0) = \bar{\nu} \exp\{-\tfrac{1}{6}\kappa B[(ct_0 - r)^2 - r^2]^{-\frac{1}{2}}\}.$$

Substituting for $\bar{\nu}$ from (23), we finally obtain

$$\nu' = \nu\left(1 - \frac{2r}{ct_0}\right)^{\frac{1}{2}} \exp\left\{-\frac{\kappa B}{6ct_0}\left[\left(1 - \frac{2r}{ct_0}\right)^{-\frac{1}{2}} - 1\right]\right\},$$

in which the exponential factor gives the gravitational correction to the usual redshift formula of the Milne cosmology. This factor differs significantly from unity only when r is close to $\frac{1}{2}ct_0$, i.e., only for light from a galaxy close to the edge of the visible universe. Normally the gravitational correction is quite negligible.

[1] E. A. Milne, *Relativity, Gravitation and World-Structure* (Oxford University Press, London, 1935).
[2] A. N. Whitehead, *The Principles of Relativity* (Cambridge University Press, Cambridge, 1922).
[3] C. B. Rayner, Proc. Roy. Soc. (London) **A222**, 509 (1954).
[4] G. D. Birkhoff, Proc. Natl. Acad. Sci. U.S. **29**, 231 (1943).
[5] H. Weyl, Am. J. Math. **66**, 591 (1944).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Search for a Universal Symmetry Group in Two Dimensions

ARTURO CISNEROS* AND HAROLD V. MCINTOSH
Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional
México 14, D.F., México

(Received 20 June 1968)

Recently several authors have proposed a universal symmetry group and demonstrated the classical validity of the concept. Supposedly, an $SU(n)$ symmetry group could be constructed for whatsoever system of n degrees of freedom. The claim is assuredly valid for all classically degenerate systems, but is in contradiction with most of the well-known and widely accepted solutions of Schrödinger's equation. We examine the reasons for this discrepancy on the quantum-mechanical level. The construction of the universal symmetry group requires ladder operators, which in most cases are the ladder operators of Infeld and Hill. Complications which owe their origin entirely to numerical relationships imposed by quantization prevent these operators from forming a von Neumann algebra and, in turn, an $SU(n)$ group of constants of the motion. Two important effects are those imposed by anisotropy, wherein not all quanta have the same size, and by non-Cartesian coordinates, wherein quanta in some dimensions are restricted in size by those in other dimensions. These effects can be seen quite clearly in two-dimensional systems: anisotropy in the Cartesian anisotropic harmonic oscillator, and conflict between dimensions in the polar form of the isotropic oscillator. Further complications arise when the two effects are combined, as in the harmonic oscillator or hydrogen atom with "excess" angular momentum. Enough residue of the universal symmetry concept remains that many similarities between the hydrogen atom and harmonic oscillator may be understood, including the fact that some levels of the hydrogen atom, which ordinarily transform according to an orthogonal group, may form irreducible representations of the unitary group.

I. INTRODUCTION

Shortly after Fock's[1] remarkable analysis of the symmetries and degeneracies of the quantum-mechanical hydrogen atom, similar methods were applied to the study of the isotropic harmonic oscillator, a second quantum-mechanical system of major importance. In both cases, the apparent symmetry of the problem was spherical, due to the central forces present, for which group-theoretical reasoning predicted a degeneracy in the z component of the angular momentum. While it is true that such a degeneracy is always present in the solutions of central force problems, the degeneracy occurring in the hydrogen atom and the harmonic oscillator was far greater. Such additional degeneracies, which are observed to occur in a few other situations as well, had been widely considered to have been "accidental," since their occurrence was not foreseen from group-theoretical reasoning based on the ostensible spherical symmetry. They had to be accepted on the grounds that amongst all possible potentials there were certainly some for which the energies of wavefunctions belonging to different symmetry types might just happen to coincide.

Thus there was a certain satisfaction when Fock showed that certain potentials might possess "hidden" symmetries, which is to say, symmetries arising from regularities in the phase space of Hamiltonian dynamics which were not immediately apparent from an inspection of the geometric symmetries of the

force field itself. That such an accounting could be given for the degeneracies of the harmonic oscillator was demonstrated by Jauch and Hill,[2] and in the course of time has been elaborated by other authors. The isotropic harmonic oscillator with n degrees of freedom was found to have the symmetry group $SU(n)$, whose symmetric tensor representations gave an accounting for all the degeneracy observed in the oscillator. Such groups are generated by a collection of tensor constants of the motion, much as the $O(4)$ group of Fock was generated by the components of two vector constants of the motion for the hydrogen atom.

Interest has centered on the isotropic harmonic oscillator, in great part because it has lately found favor in modelling collective motions in nuclear theory, but no doubt also because results for the anisotropic harmonic oscillator have been somewhat inconclusive. It was already noticed by Jauch and Hill that the anisotropic oscillator possessed an accidental degeneracy when its decomposition into one-dimensional eigenoscillators produced components with commensurable frequencies. Only very recently did Demkov[3] begin a systematic study of the quantum-mechanical symmetry operators for such oscillators.

Again the symmetry group was found to be $SU(n)$, but its generators could not be given so simply as the components of a tensor, and there did not appear to be any way to treat the incommensurable case. Most of the analyses were made for the classical version of the system. Once classical constants of the motion were found, they were transcribed into quantum-mechanical terms by the usual device of replacing the canonical coordinates by coordinate multipliers, and their conjugate momenta by derivatives with respect to the coordinates. There was little difficulty in effecting such a procedure if the operators were linear functions of the canonical coordinates and momenta, quadratic functions which could be symmetrized, functions of constants of the motion, or algebraic combinations of the foregoing.

More recently, Dulock[4] succeeded in constructing classical constants of the motion for the anisotropic oscillator with arbitrary, and in particular, possibly incommensurable eigenfrequencies. The constants so constructed were transcendental functions in the most general case, and anyway there was a multiple valuedness present in the definitions of the operators which seemed to preclude the formation of any quantum-mechanical analogs of his constants of the motion. However, the fact that his operators also generated a symmetry group of the type $SU(n)$ was quite disturbing, since one can so easily determine the spectrum of the oscillator with general frequencies and see that when they were incommensurable there would be no degeneracy whatsoever, a result apparently incompatible with the occurrence of an $SU(n)$ group. There was some consolation in thinking that the multiple valuedness of the functions determining the constants would somehow prevent the formation of quantum-mechanical operators.

He applied his methods to the analysis of the problem of the cyclotron motion of a charged particle in a uniform magnetic field.[5] This problem involves an essentially quadratic Hamiltonian and is mathematically identical to the problem of the harmonic oscillator. There was therefore no loss in generality to consider a superimposed isotropic harmonic oscillator potential, altogether, the Zeeman effect for a harmonic oscillator. Again it was found that there would be an $SU(3)$ symmetry group, whatever combinations of magnetic field strength and oscillator constants were used, save in the limiting case where the oscillator potential was zero. In this latter case, one obtained a limiting case of the $SU(3)$ group, whose generators obeyed the commutation rules of ladder operators.[6] However, when the problem is solved quantum mechanically, it is found that degeneracy actually exists only for certain combinations of field strength and oscillator frequencies. These are the same combinations which produce closed orbits in the classical version.

From these historical precedents, it appears that there are actually two effects to be dealt with. One is the effect of anisotropy, which is to say that after the reduction of the problem to action-angle variables, it is found that the Hamiltonian is a linear combination of the action variables in which the coefficients are not necessarily equal, or even commensurable. The second effect arises from the transformation to action-angle variables, in which classical functions may arise which cannot be realized in terms of the noncommuting quantum-mechanical operators. Or, what will be more relevant to our present interests, the particular commutation rules of the quantum-mechanical operators may produce some slight discrepancies, negligible in the limit of large quantum numbers, but adequate to alter substantially the apparent symmetry.

The second effect derives its importance from the fact that there has been a general realization that a very large class of problems is somehow equivalent to an appropriate harmonic oscillator. In celestial mechanics there has long existed a transformation

connecting the Kepler problem and harmonic oscillator, whose quantum-mechanical residue may be perceived in the similarities in the wavefunctions and degeneracy patterns of the two systems. In Hamiltonian mechanics such relationships are readily perceived in terms of action-angle variables, to the extent that several authors[7,8,9] have recently shown how one may construct a classical $SU(n)$ symmetry group whenever the requisite action-angle variables may be formed. Such a viewpoint is considerably more general than the specific geometric constructions in phase space by which Fock, Laporte and Rainich,[10] or Jauch and Hill exposed the hidden symmetry of the systems which they treated.

We now undertake to clarify the role of $SU(n)$ as a symmetry group of the anisotropic quantum oscillator by considering in detail the wavefunctions of the system and how they combine to form degenerate energy levels. The general characteristics of the situation may be amply described in two dimensions, where one would expect to obtain an $SU(2)$ group as the symmetry group for all two-dimensional oscillators.

Among those problems for which a complete set of action-angle variables exist, there does not seem to be any inconsistency to the claim that $SU(n)$ is a classical universal symmetry group, applicable to all systems with n degrees of freedom. Actually, there are several degrees to which one might wish to carry such a claim: for all spherical potentials, for all systems with separable Hamilton–Jacobi equation, or indeed for each and every classical system. It is certainly applicable to classically degenerate systems, since all the authors whom we have mentioned have exhibited the necessary operators. For the most part, this paper will be restricted to this first level. Most of our authors have speculated, and Mukunda[8] has explicitly shown, how the claim could reach its utmost generality.

The quantum-mechanical situation is anomalous in the respect that an $SU(n)$ group is likewise predicted for all the classically degenerate problems, and the occurrence of a degeneracy in the energy levels induced by irreducible representations of an $SU(n)$ is contrary both to the experimental evidence, and to the traditional and familiar solutions of Schrödinger's equations for these problems.

In order to examine this anomaly in careful detail, we shall discuss the solution of two familiar problems in polar coordinates, again in two dimensions for simplicity. These problems are the hydrogen atom and the harmonic oscillator, the two standard examples of accidental degeneracy. In the first case the discrepancy between the predicted $SU(2)$ and observed $O(3)$ is slight, but is nevertheless fundamental. The mechanism for the discrepancies which we shall encounter is quite general, applying to the analogous problems in any number of dimensions. In the second case the predicted groups are the same, but when the ladder operators are carefully examined the same discrepancy is found as for the hydrogen atom, and the operators do not directly generate an $SU(2)$ group, or for that matter a group of finite rank at all.

The discrepancy to which we refer originates from the fact that the ladder operators which Dulock prescribes, although they satisfy the proper Poisson-bracket relations in classical mechanics to generate an $SU(n)$ group, do not do so quantum mechanically. It is still true that one can form a pair of ladder operators, each of which is an eigenfunction of the Hamiltonian, and such that the members of the pair have eigenvalues which are negatives of one another. These pairs of ladder operators belong to supposedly independent canonical coordinates, and so members of different pairs should commute with one another. However, it is found that when applied to wavefunctions with extreme eigenvalues, the commutation is no longer valid, and that bilinear products of ladder operators no longer generate the Lie algebra of an $SU(n)$ group. In particular, they may allow more levels to be degenerate than called for by $SU(n)$ symmetry, and in the case of the hydrogen atom the resultant degeneracy is actually appropriate to the orthogonal group $O(n+1)$.

One might think that the failure to form a canonical set of ladder operators is a fault of Dulock's prescription, and that there might conceivably exist a set of operators chosen differently which would correspond to the $SU(n)$ group. In part the supposition must be correct, since another transformation given by Dulock will actually map the misbehaving operators into generators of a symmetry group. In the case of the harmonic oscillator written in polar coordinates, they yield the legitimate unitary group, and in the case of the hydrogen atom in polar coordinates, they yield the Runge vector and the angular-momentum operators generating the orthogonal group responsible for the accidental degeneracy. However, as Susskind and Glogower[11] have shown, there may already occur difficulties in trying to define quantum-mechanical action-angle variables. Fortunately, even though the cyclic angle variables and their conjugate action variables may not be able to correspond to a pair of operators satisfying the commutation rules of a coordinate–momentum pair, it seems that it may still be possible to form usable ladder operators from such quantities. Such was certainly true for the specific

case they studied, because even though one cannot form a phase operator for a unidimensional harmonic oscillator, its ladder operators are in no way deficient. For such a ladder always to be infinite in extent is another matter.

In the restricted class of problems which we treat in this paper, namely those whose Hamilton–Jacobi equation is separable, and which moreover are classically degenerate, the commutation difficulties with the ladder operators in different coordinates have a much more prosaic origin. The ladder operators are defined in terms of the action-angle variables, which we will actually express in terms of polar coordinates and their associated momenta. Even in higher dimensions, the separation process is such that one momentum is first discovered to be a constant of the motion, say the azimuthal angular momentum. With this knowledge, another constant may be separated, say the total angular momentum. Thus there will arise a chain of constants, each depending upon the previous one, as well as the coordinates and momenta.

We might refer to this process as a "cumulative" separation. It results in a cumulative dependence of the angle variables in the opposite direction. For instance, in three dimensions with a symmetrical potential, the radial angle variable depends only on r, p_r, and constants of the motion. However, the theta angular variable depends on r, θ, p_r, p_θ, and constants of the motion, while the ϕ angle variable is by no means ϕ itself, but depends upon all the canonical variables, combined as constants of the motion and otherwise.

The result is that when the Schrödinger equation is separated in polar coordinates, the r-ladder operators will modify only the radial quantum number. However, the θ-ladder operators will modify simultaneously the principal quantum number n arising from the radial equation, and the angular momentum l, which is the quantum number depending on the theta coordinate. Finally, the ϕ ladder will modify both these quantum numbers, and the azimuthal quantum number m as well. There are also restrictions on the ranges of these quantum numbers which make them interdependent, $|m| \leq l \leq n$. The commutation failures which we encounter all arise from conflicts among these restrictions.

The two effects which we have mentioned (the pattern of degeneracy appropriate to anisotropy, and the complications due to cumulative separation in non-Cartesian coordinates) occur simultaneously for Hamiltonians of the form

$$H = \tfrac{1}{2}p^2 + V(r) + \delta L^2/2r^2,$$

because their separation in polar coordinates produces

$$H = H(\beta J_r + \alpha J_\theta)$$

when $V(r)$ is either $-1/r$ or $r^2/2$. This is the same functional dependence as for the anisotropic oscillator in Cartesian coordinates. Such Hamiltonians came to our attention in a four-dimensional form because they arise from trying to treat the magnetic monopole in quaternionic coordinates; nevertheless, the two-dimensional version is quite instructive and illustrates the principles involved.

Newton's theorem of revolving orbits[12] explains how to interpret such Hamiltonians. The effect of a centrifugal potential proportional to the square of the angular momentum is to make the motion a function of θ/γ rather than θ, where γ is a factor depending upon δ, while the angular momentum has to be multiplied by the same factor. This is a canonical transformation which causes precession of the orbits. It is not the same as a transformation to rotating coordinates, because in that case the difference between the real and apparent angle is a quantity which increases linearly with time, while in this case, independently of the time, the angle is multiplied by a factor.

The precession is particularly interesting in two cases. First, for the harmonic oscillator when the factor δ is 3, in which case an elliptical orbit with center at the origin is transformed into an oval with focus at the origin. Second, for the Kepler problem when the constant is $-\frac{3}{4}$, in which case the opposite transformation takes place. It is interesting that, in these two examples, the symmetry group which is appropriate for the one problem shifts into that appropriate for the other, just as does the form of the orbits. Again the result is applicable to higher-dimensional problems, and gives an interesting relationship between unitary and orthogonal groups. Actually the transformation between the harmonic oscillator and the Kepler problem is slightly more complicated than this, because a change in the functional form of the Hamiltonian is also necessary. Nevertheless, it shows that there is a very close relation between the two systems, to which we have already alluded.

The objective of this paper is to determine, at least in two dimensions, to what extent the expectation that $SU(n)$ might be a universal symmetry group for all degenerate systems might be realized. Even without going beyond the classically degenerate systems, we see that quantization imposes numerical relationships governing the size of quanta that cannot be met by ladder operators which in effect require arbitrary

subdivisions of the quanta. To the extent that the generators of the unitary groups are based on ladder operators,[13] their quantum-mechanical occurrence is precluded. At the same time, the results are not entirely negative, since there remains an operator calculus of ladder operators which is very versatile and effective in explaining the degeneracies, energy levels, wavefunctions, and symmetries which are actually observed to exist. One might hope that such results might be extended to systems which are not classically degenerate, and that the whole theory might be applied to relativistic quantum mechanics and field theory.

II. DEGENERACY OF THE ANISOTROPIC HARMONIC OSCILLATOR

The Hamiltonian for the two-dimensional harmonic oscillator in its most general form, wherein the potential is not assumed to be isotropic, is

$$H = (p_x^2 + p_y^2 + \alpha^{-2}x^2 + \beta^{-2}y^2)/2. \qquad (1)$$

Units have been chosen which will avoid inconvenient scale factors; the anisotropy is reflected by the two constants α and β. The corresponding Schrödinger equation is

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2}\right)\Psi = -2E\Psi, \qquad (2)$$

whose solutions[14] are expressible in terms of the Hermite polynomials H_n,

$$\Psi_{n_1,n_2} = \frac{e^{-(\alpha^{-1}x^2+\beta^{-1}y^2)/2}H_{n_1}(\alpha^{-\frac{1}{2}}x)H_{n_2}(\beta^{-\frac{1}{2}}y)}{(2^{n_1+n_2}n_1!\,n_2!\,\pi(\alpha\beta)^{\frac{1}{2}})^{\frac{1}{2}}}, \qquad (3)$$

and whose energy eigenvalues are represented in terms of the quantum numbers n_1 and n_2 by the formula

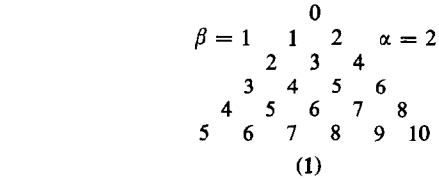
$$\alpha\beta(E - E_0) = \beta n_1 + \alpha n_2,$$
$$E_0 = (\alpha + \beta)/2\alpha\beta, \quad n_1, n_2 = 0, 1, 2, \cdots. \qquad (4)$$

Determination of the degeneracies is then a matter of ascertaining the number of ways that the same value of $N = (E - E_0)\alpha\beta$ can be formed from the two individual quantum numbers. In the case that $\alpha = \beta = 1$, one finds the isotropic oscillator, in which the level N is $(N + 1)$-fold degenerate, since n_1 may range from 0 to N, thereby completely fixing n_2. Therefore, every positive integer represents a possible multiplicity of a degenerate state, and each such multiplicity occurs only once in the spectrum.

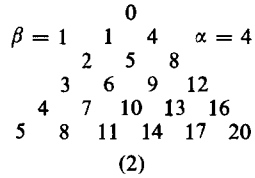
Matters are slightly different when $\alpha = 2$, $\beta = 1$, for which the frequency ratio is 2:1, and the energy formula now takes the form $N = n_1 + 2n_2$. Again a degeneracy prevails, which is now according to the number of ways that a given integer may be written as the sum of two nonnegative integers, one of which is *even*. Considering the levels one by one, we see that the level for $N = 0$ can only be realized by the wavefunction with quantum numbers (0, 0), while $N = 1$ can only be realized in the state (1, 0). Thus, unlike the isotropic case, we find that the one-dimensional representation occurs *twice*. Moreover, the value $N = 2$ can be realized by either of the states (2, 0) or (0, 1), while $N = 3$ can be realized by either (1, 1) or (3, 0). In general, every n-fold degenerate set of states will occur twice, once for an even value of N and once for an odd value. These results were already obtained by Demkov, but allow us to fix a particular example in our mind.

One can express the general situation with the help of a triangle diagram.[15] Starting with the single integer 0, one writes lines of integers. The kth line commences with $k\beta$, while $(\alpha - \beta)$ is to be added to obtain successive integers in the line, until $k\alpha$ is reached. Each line is offset by a half space to center it with the line above. One may read the possible values of N from the body of the table, while to determine the values of n_1 and n_2 to which they correspond one may number the diagonals. Taking as examples $\alpha = 2$ and $\beta = 1$, $\alpha = 4$ and $\beta = 1$, $\alpha = 5$ and $\beta = 2$, we obtain the diagrams of Fig. 1.

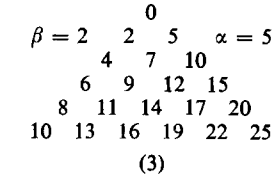
The general aspect of the pattern of multiplicities is that the least common multiple of the two frequencies will determine the first degenerate energy level. Hence, the less commensurable two frequencies, as measured by their least common multiple, the more groups of states with a given multiplicity there will be, and the higher the energy of the first degenerate state. The lower-energy states will generally have lesser degeneracy, although the energy is not strictly ordered according to the degeneracy.



(1)



(2)



(3)

FIG. 1. Triangle diagram of energies for anisotropic harmonic oscillator: (1) $E = n_1 + 2n_2$, (2) $E = n_1 + 4n_2$, (3) $E = 2n_1 + 5n_2$.

In the "limit" as the frequencies become more and more incompatible while their ratio remains nearly fixed, the number of nondegenerate states increases and the energy gap to the first degenerate state becomes greater and greater. Consequently, if we approximate an irrational frequency ratio by rational fractions with greater and greater numerator and denominator, we will find that we approach a situation in which all the finite energy levels are nondegenerate and that there is an infinite gap to what would have been the more highly degenerate states.

In order to explain such patterns of degeneracy by means of a symmetry group, we apply the techniques of Dulock, which are similar to the method of Fradkin, Mukunda, and others. We have described the process in considerable detail in an earlier paper,[16] so that we will here only exhibit the principal quantities necessary to construct the ladder operators from which the constants of the motion and generators of the prospective symmetry group will eventually be formed.

Supposing that we have a two-dimensional problem whose Hamiltonian–Jacobi is separable and for which there exist action-angle variables, we suppose that the Hamiltonian takes the form

$$H = H(\beta J_1 + \alpha J_2), \tag{5}$$

where α and β are numerical constants. If one then constructs the functions

$$a_1^{\pm} = (J_1/2\pi\alpha)^{\frac{1}{2}} e^{\pm 2\pi i \alpha w_1}, \tag{6a}$$

$$a_2^{\pm} = (J_2/2\pi\beta)^{\frac{1}{2}} e^{\pm 2\pi i \beta w_2}, \tag{6b}$$

where w_1 and w_2 are the angle variables, he may show that they satisfy the Poisson-bracket relations

$$\{a_1^+, a_1^-\} = \{a_2^+, a_2^-\} = i; \tag{7}$$

the remaining pairwise brackets vanish.

From these functions one forms the constants of the motion

$$a_1^+ a_1^-, \quad a_1^+ a_2^-, \quad a_2^+ a_1^-, \quad a_2^+ a_2^- \tag{8}$$

from which an $SU(2)$ Poisson-bracket Lie algebra can be constructed. This construction is possible whether the system is classically degenerate or not, that is, whether the frequency ratio α/β is a rational number or not.

Writing the Hamilton–Jacobi equation of the two-dimensional anisotropic harmonic oscillator (1) in Cartesian coordinates,

$$2E = \left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2}. \tag{9}$$

The principal function should be written in the form $S(x, y) = S_x(x) + S_y(y)$ to obtain separated equations

$$\left(\frac{\partial S_x}{\partial x}\right)^2 + \frac{x^2}{\alpha^2} = 2E_x, \tag{10a}$$

$$\left(\frac{\partial S_y}{\partial y}\right)^2 + \frac{y^2}{\beta^2} = 2E_y, \tag{10b}$$

where $E_x + E_y = E$.

The action variables defined by $J_j = \oint p_j\, dq_j$ are obtained through the use of Eqs. (10) to yield

$$J_x = \oint (2E_x - x^2/\alpha^2)^{\frac{1}{2}}\, dx = 2\pi\alpha E_x, \tag{11a}$$

$$J_y = \oint (2E_y - y^2/\beta^2)^{\frac{1}{2}}\, dy = 2\pi\beta E_y. \tag{11b}$$

From the above result, the energy of the system can be written in terms of the action variables in the required form

$$E = (\beta J_x + \alpha J_y)/2\pi\alpha\beta. \tag{12}$$

Therefore, we proceed to the construction of the angle variables, which are

$$w_x = \frac{\partial S}{\partial J_x} = \int \left(\frac{J_x}{\alpha\pi} - \frac{x^2}{\alpha^2}\right)^{-\frac{1}{2}} \frac{dx}{2\pi\alpha}$$

$$= \frac{1}{2\pi} \arcsin \frac{x}{(\alpha J_x/\pi)^{\frac{1}{2}}}, \tag{13a}$$

$$w_y = \frac{\partial S}{\partial J_y} = \int \left(\frac{J_y}{\beta\pi} - \frac{y^2}{\beta^2}\right)^{-\frac{1}{2}} \frac{dy}{2\pi\beta}$$

$$= \frac{1}{2\pi} \arcsin \frac{y}{(\beta J_y/\pi)^{\frac{1}{2}}}, \tag{13b}$$

from which we finally find

$$e^{\pm 2\pi i w_x} = (2\alpha E_x)^{-\frac{1}{2}} (\alpha^{\frac{1}{2}} p_x \pm i\alpha^{-\frac{1}{2}} x), \tag{14a}$$

$$e^{\pm 2\pi i w_y} = (2\beta E_y)^{-\frac{1}{2}} (\beta^{\frac{1}{2}} p_y \pm i\beta^{-\frac{1}{2}} y), \tag{14b}$$

and ultimately, the ladder operators of Eq. (6),

$$a_x^{\pm} = (J_x/2\pi\alpha)^{\frac{1}{2}} (\mp i e^{\pm 2\pi i w_x})^{\alpha}$$

$$= E_x^{\frac{1}{2}} [(2\alpha E_x)^{-\frac{1}{2}} (\alpha^{-\frac{1}{2}} x \mp i\alpha^{\frac{1}{2}} P_x)]^{\alpha}, \tag{15a}$$

$$a_y^{\pm} = (J_y/2\pi\beta)^{\frac{1}{2}} (\mp i e^{\pm 2\pi i w_y})^{\beta}$$

$$= E_y^{\frac{1}{2}} [(2\beta E_y)^{-\frac{1}{2}} (\beta^{-\frac{1}{2}} y \mp i\beta^{\frac{1}{2}} p_y)]^{\beta}. \tag{15b}$$

An appropriate phase factor, which does not change the commutation relations among the a's, has been included for convenience.

To find the corresponding quantum-mechanical

operators, we note that the operators

$$\mathcal{K}_x^+ = (2n_1 + 2)^{-\frac{1}{2}}\left(\alpha^{-\frac{1}{2}}x - \alpha^{\frac{1}{2}}\frac{\partial}{\partial x}\right), \tag{16a}$$

$$\mathcal{K}_x^- = (2n_1)^{-\frac{1}{2}}\left(\alpha^{-\frac{1}{2}}x + \alpha^{\frac{1}{2}}\frac{\partial}{\partial x}\right), \tag{16b}$$

$$\mathcal{K}_y^+ = (2n_2 + 2)^{-\frac{1}{2}}\left(\beta^{-\frac{1}{2}}y - \beta^{\frac{1}{2}}\frac{\partial}{\partial y}\right), \tag{16c}$$

$$\mathcal{K}_y^- = (2n_2)^{-\frac{1}{2}}\left(\beta^{-\frac{1}{2}}y + \beta^{\frac{1}{2}}\frac{\partial}{\partial y}\right) \tag{16d}$$

are known to be normalized ladder operators for the wavefunctions[17] of (3). In other words,

$$\mathcal{K}_x^{\pm}\Psi_{n_1,n_2} = \Psi_{n_1\pm1,n_2},$$
$$\mathcal{K}_y^{\pm}\Psi_{n_1,n_2} = \Psi_{n_1,n_2\pm1}.$$

In the classical limit, we have

$$\mathcal{K}_x^{\pm} \to \mp ie^{\pm 2\pi i w_x},$$
$$\mathcal{K}_y^{\pm} \to \mp ie^{\pm 2\pi i w_y}.$$

By observing the functions (15), we find that the quantum-mechanical construction which we seek will be possible only for integral α and β.

A slightly different normalization is necessary to make the ladder operators for more than a single step compatible, and to ensure that they will satisfy the commutation rules of the generators of $SU(2)$:

$$\mathcal{A}_x^+ = [(n_1 - k + \alpha)/\alpha]^{\frac{1}{2}}(\mathcal{K}_x^+)^\alpha, \tag{17a}$$
$$\mathcal{A}_x^- = [(n_1 - k)/\alpha]^{\frac{1}{2}}(\mathcal{K}_x^-)^\alpha, \tag{17b}$$
$$\mathcal{A}_y^+ = [(n_2 - j + \beta)/\beta]^{\frac{1}{2}}(\mathcal{K}_y^+)^\beta, \tag{17c}$$
$$\mathcal{A}_y^- = [(n_2 - j)/\beta]^{\frac{1}{2}}(\mathcal{K}_y^-)^\beta. \tag{17d}$$

In these definitions, $k = n_1 \pmod{\alpha}$, and $j = n_2 \pmod{\beta}$.

Routine verification establishes that the operators (17) satisfy the commutation relations,

$$[\mathcal{A}_x^-, \mathcal{A}_x^+] = 1, \tag{18a}$$
$$[\mathcal{A}_y^-, \mathcal{A}_y^+] = 1, \tag{18b}$$

all other commutator pairs being zero.

The numbers k and j have been included in the definition of the operators (17) so that the commutation relations (18) would be satisfied for all eigenfunctions of our Hamiltonian. This may be verified explicitly by applying the commutator (18a) to an eigenfunction:

$$[\mathcal{A}_x^-, \mathcal{A}_x^+]\Psi_{n_1,n_2} = \left(\frac{n_1 - k + \alpha}{\alpha} - \frac{n_1 - k}{\alpha}\right)\Psi_{n_1,n_2}. \tag{19}$$

For $n_1 \geq \alpha$ both terms in parentheses in Eq. (19) appear, so that we have the identity operator, and the value of k makes no difference in this case. Nevertheless, for states such that $n_1 < \alpha$, the second term in the parentheses does not appear because $\mathcal{A}_x^-$ applied to such a state vanishes; the result is $[(n_1 - k + \alpha)/\alpha]\Psi_{n_1,n_2}$, since in this case $k = n_1$, and once again we obtain the identity operator.

By the introduction of the indices j and k, we are able to apply one and the same set of operators to all of the wavefunctions. This choice may be compared with the results of Demkov,[3] who obtains two families of operators as well as the two complete sets of representations of $SU(2)$ in his treatment of the harmonic oscillator with a 2:1 frequency ratio. These family indices allow a unified treatment of the operators and become insignificant in the classical limit, but cannot be expressed simply in terms of coordinates and momentum operators.

Choosing the following bilinear combinations of the ladder operators, one obtains the constants of the motion, the first of which is a function of the Hamiltonian; the remainder generate a Lie algebra for $SU(2)$: $\mathcal{A}_x^-\mathcal{A}_x^+ + \mathcal{A}_y^-\mathcal{A}_y^-$, $\mathcal{A}_x^+\mathcal{A}_y^- + \mathcal{A}_x^-\mathcal{A}_y^+$, $i\mathcal{A}_x^+\mathcal{A}_y^- - i\mathcal{A}_x^-\mathcal{A}_y^+$, $\mathcal{A}_x^+\mathcal{A}_x^- - \mathcal{A}_y^+\mathcal{A}_y^-$. These constants of the motion are the well-known components of the tensor constant of the motion for the isotropic harmonic oscillator when $\alpha = \beta = 1$.

Although the derivation of an $SU(2)$ group is only possible when the frequency ratio is rational, an irrational frequency ratio may be approximated arbitrarily well by some nearby rational number. To the extent that this may be regarded as a perturbation in the Hamiltonian, we may obtain approximate symmetry groups for any anisotropic oscillator. We have already reconciled ourselves to the total lack of any degeneracy for incommensurable ratios on account of the frequency of appearence of one-dimensional and hence nondegenerate representations when the frequencies are not very commensurable. However, there is also an interesting complementary effect, that when an irrational ratio is close to a rational ratio, the nondegenerate states will cluster into nearly degenerate multiplets. Again if a slight perturbation is made, they will coalesce into a single degenerate state for the oscillator whose ratio is rational.

Figure 2 illustrates this effect for two examples. In Fig. 2(a), we compare the isotropic oscillator with an oscillator in which the frequency ratio is 10:11, while in Fig. 2(b), the comparison is made between the ratios 1:2 and 11:20. In both cases it is easy to identify the multiplets and the corresponding degenerate states. When the approximations are poor, and

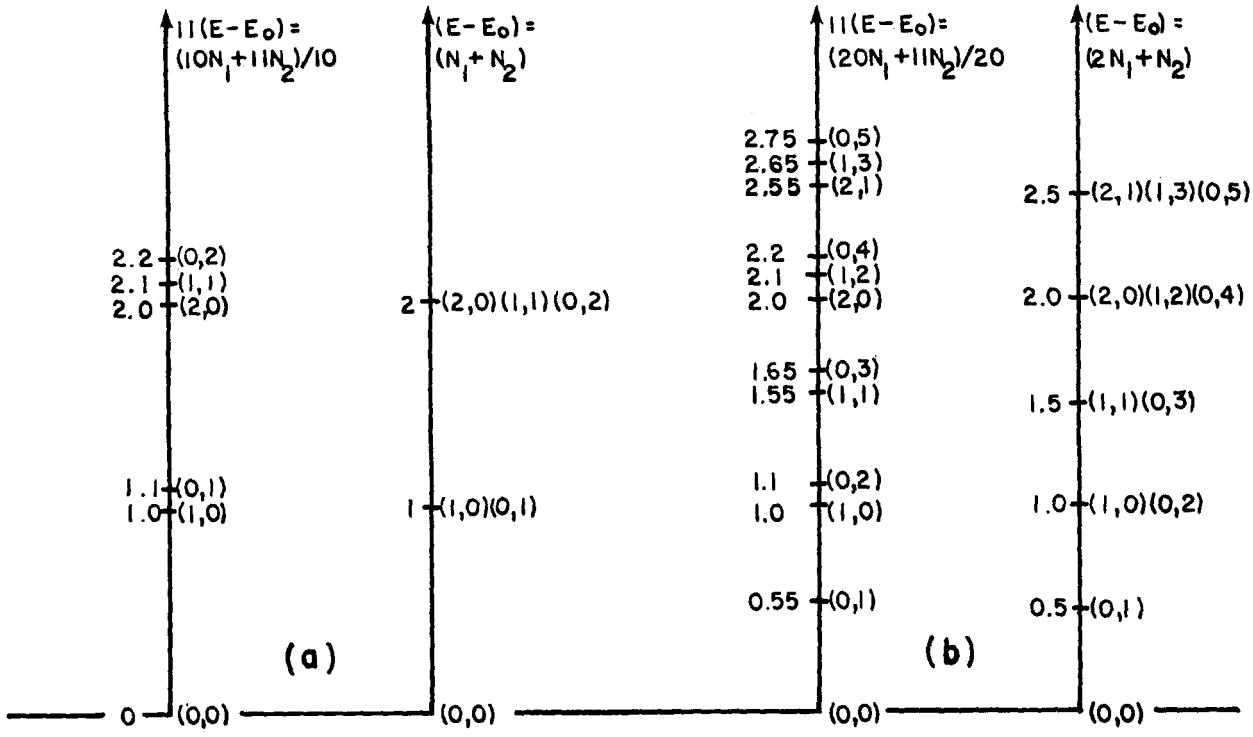


FIG. 2. Multiplet structure of the anisotropic oscillator with frequency ratio: (a) 11:10, compared with 1:1, and (b) 1:20, compared with 1:2.

hence the gap between multiplets is small, it is possible for levels between different multiplets to coincide. This corresponds to the occurrence of common divisors in the two frequencies. We should also not lose sight of the fact that even while the nondegenerate states are forming multiplets, there are still degenerate levels at higher energies, so that for very large energies, multiplets may be formed from degenerate as well as nondegenerate states. Since there is only a finite number of states with any given multiplicity of degeneracy, it is inevitable that the multiplets will eventually be formed from highly degenerate states. We should also note that the lower, singly degenerate levels, which appear to result from a broken symmetry are actually members of distinct families of states of degeneracy $1, 2, 3, \cdots$, each of whose families form bases for the irreducible representations of $SU(2)$.

We have assumed the numbers α and β appearing in the Hamiltonian (1) to have been integers, in order that one could define the operators (17). If the frequency ratio is rational this is always possible, because we can always absorb a common factor by our choice of units in defining the Hamiltonian. However, we should take care to ensure in so doing that α and β are relatively prime and not merely integers. If this is not done,we will not get the most fundamental ladder operators possible, and it will not always be possible to reach from a given state to all the others degenerate to it by the use of constants of the motion formed from the operators of Eq. (17). To illustrate the point, consider the isotropic harmonic oscillator in which we have chosen $\alpha = \beta = 2$. We will then obtain ladder operators which change the quantum numbers by two units; hence no combination of raising and lowering operators will be able to pass between the degenerate states $(0, n)$ and $(1, n-1)$. Indeed, there would be no way to change the parity of the individual quantum numbers.

III. POLAR CREATION AND ANNIHILATION OPERATORS FOR THE HARMONIC OSCILLATOR

Both the Hamilton–Jacobi and Schrödinger equations for the isotropic harmonic oscillator are separable in rectangular Cartesian coordinates as well as polar coordinates. We have seen in the last section that Dulock's procedure applied to the action-angle variables after separation in the Cartesian coordinates produces the well-known ladder operators for the harmonic oscillator, which in turn leads to the identification of $SU(n)$ as the symmetry group of the n-dimensional harmonic oscillator and the explanation of the observed degeneracies in the spectrum of the harmonic oscillator.

If Dulock's procedure is applied to the action-angle variables arising from the separation of the

Hamiltonian–Jacobi equation in polar coordinates, one obtains a collection of ladder operators for radial and angular quanta. Dulock did not discuss such operators, although it is clear from the general treatment that they could be formed and would have the desired properties. We will form them as a prelude to our attempt to construct quantum-mechanical operators with analogous properties. Again the study of two-dimensional problems is adequately representative of the general case, so we may thereby avoid the complications in the angular operators which arise in higher dimensions.

The Hamiltonian for the isotropic harmonic oscillator in two-dimensional polar coordinates is

$$H = (p_r^2 + r^2 + p_\theta^2/r^2)/2. \tag{20}$$

The corresponding Hamiltonian–Jacobi equation can be separated by writing the principal function in the form $S = S_r(r) + S_\theta(\theta)$. Since θ does not appear explicitly in the Hamiltonian, we have $S_\theta = \mu\theta$ where μ is a constant. S_r is given by

$$S_r = \int (2E - r^2 - \mu^2/r^2)^{\frac{1}{2}}\, dr. \tag{21}$$

The action variables are calculated using the separated Hamilton–Jacobi equations; the result is

$$J_r = \oint p_r\, dr = -\pi\mu + \pi E, \tag{22a}$$

$$J_\theta = \oint p_\theta\, d\theta = 2\pi\mu, \tag{22b}$$

from which the energy is given in terms of the action variables

$$E = (2J_r + J_\theta)/2\pi. \tag{23}$$

According to the procedure described in the preceding section, we will now construct "a" functions which satisfy commutators (7); in this problem they are given by

$$a_r^{\pm} = \mp i(J_r/2\pi)^{\frac{1}{2}} e^{\pm 2\pi i w_r}, \tag{24a}$$

$$a_\theta^{\pm} = -(J_\theta/4\pi)^{\frac{1}{2}} e^{\pm 4\pi i w_\theta}. \tag{24b}$$

The angle variables $w_j = \partial S/\partial J_j$, written in terms of polar coordinates and their conjugate momenta, are

$$w_r = \frac{1}{\pi}\int (2Er^2 - \mu^2 - r^4)^{-\frac{1}{2}} r\, dr,$$

$$w_\theta = \frac{w_r}{2} - \frac{\mu}{2\pi}\int (2Er^2 - \mu^2 - r^4)^{-\frac{1}{2}} \frac{dr}{r} + \frac{\theta}{2\pi}.$$

A straightforward evaluation of the integrals yields

$$w_r = \frac{1}{2\pi} \arcsin \frac{r^2 - E}{(E^2 - \mu^2)^{\frac{1}{2}}},$$

$$w_\theta = \frac{w_r}{2} - \frac{1}{4\pi} \arcsin \frac{Er^2 - \mu^2}{r^2(E^2 - \mu^2)^{\frac{1}{2}}} + \frac{\theta}{2\pi}.$$

When this result is inserted in (24) and momenta are substituted according to the separated Hamilton–Jacobi equations, we have

$$a_r^{\pm} = [(E - \mu)/2]^{\frac{1}{2}}[(E^2 - \mu^2)^{-\frac{1}{2}}(r^2 - E \mp irp_r)], \tag{25a}$$

$$a_\theta^{\pm} = [\mp ie^{\pm 2\pi i w_r}]\mu^{\frac{3}{2}}[2(E^2 - \mu^2)]^{-\frac{1}{2}} \times \left[\frac{\mu}{r^2} - \frac{E}{\mu} \mp i\frac{p_r}{r}\right] e^{\pm i2\theta}. \tag{25b}$$

The quantities

$$N = a_r^- a_r^+ + a_\theta^- a_\theta^+, \tag{26a}$$

$$L = a_r^- a_r^+ - a_\theta^- a_\theta^+, \tag{26b}$$

$$K = -ia_\theta^- a_r^+ + ia_r^- a_\theta^+, \tag{26c}$$

$$D = -a_r^- a_\theta^+ - a_\theta^- a_r^+ \tag{26d}$$

are constants of the motion. N commutes with the rest and is one-half the Hamiltonian; the other three satisfy an $SU(2)$ Poisson-bracket Lie algebra:

$$\{K, L\} = -2D, \quad \{L, D\} = -2K, \quad \{D, K\} = -2L.$$

If these quantities were to be written out explicitly in Cartesian coordinates, they would be fairly complicated with no apparent simple geometric or algebraic interpretation. Their form is nevertheless reminiscent of similar results which Dulock obtained in treating the hydrogen atom, for in both cases one can pick out the customary constants amongst some additional terms. He was able to prescribe a transformation which preserved the Poisson-bracket relations, and which reduced the constants of the motion obtained from the action-angle variable ladder operators to the components of the Runge vector and the angular momentum.

Dulock's transformation consisted in defining new constants as follows:

$$L' = N - L, \tag{27a}$$

$$K' = K[2 + (N + L)/(N - L)]^{\frac{1}{2}}, \tag{27b}$$

$$D' = D[2 + (N + L)/(N - L)]^{\frac{1}{2}}. \tag{27c}$$

The result of his transformation applied to the constants is

$$L' = p_\theta, \tag{28a}$$

$$K' = p_\theta\left[\frac{1}{r}\, p_r \cos 2\theta + \sin 2\theta\left(\frac{E}{p_\theta} - \frac{p_\theta}{r^2}\right)\right], \tag{28b}$$

$$D' = p_\theta\left[-\frac{1}{r}\, p_r \sin 2\theta + \cos 2\theta\left(\frac{E}{p_\theta} - \frac{p_\theta}{r^2}\right)\right]. \tag{28c}$$

The new constants are more easily recognized when written in Cartesian coordinates,

$$L' = xp_y - yp_x, \tag{29a}$$

$$K' = xy + p_x p_y, \tag{29b}$$

$$D' = (x^2 - y^2 + p_x^2 - p_y^2)/2, \tag{29c}$$

being nothing but the components of the well-known tensor constant of the motion.

We now turn to the quantum-mechanical treatment. The Schrödinger equation in polar coordinates is

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} - r^2 + 2E\right)\Psi = 0. \tag{30}$$

Its normalized solutions are given by[18]

$$\Psi_{n,m} = R_{n,|m|}(r)\Theta_m(\theta), \tag{31a}$$

$$R_{n,|m|}(r) = \left(\frac{2[(n-|m|)/2]!}{[\Gamma((n+|m|+2)/2)]^3}\right)^{\frac{1}{2}} \times r^{|m|}e^{-r^2/2}L^{|m|}_{(n+|m|)/2}(r)^2, \tag{31b}$$

$$\Theta_m(\theta) = e^{im\theta}/(2\pi)^{\frac{1}{2}}, \tag{31c}$$

where L is the associated Laguerre polynomial; the quantum numbers can take the values $n = 0, 1, 2, \cdots$, $m = 0, \pm 1, \pm 2, \cdots$, with the restriction $n - |m| = 0, 2, 4, \cdots$. Energy eigenvalues are given by

$$E = n + 1. \tag{32}$$

The derivation of ladder operators from their classical version is likely to be somewhat inconclusive when the classical form is a complicated function of noncommuting coordinate and momentum operators. Although the expressions are still relatively simple for the harmonic oscillator, we will content ourselves with obtaining them from other sources, and demonstrating that they approach in the classical limit the ladder operators obtained from the action-angle variables.

Accordingly, we note the recurrence relation given by Schrödinger[19] connecting the radial wavefunctions of the hydrogen atom belonging to different total quantum numbers. If we replace r by r^2, we get a recurrence relation connecting the radial wavefunctions for the present problem. These considerations, after including an appropriate normalizing factor, give us the operators

$$\mathcal{K}_r^+ = [(n+2)^2 - m^2]^{-\frac{1}{2}}\left[r^2 - (n+2) - r\frac{\partial}{\partial r}\right], \tag{33a}$$

$$\mathcal{K}_r^- = [n^2 - m^2]^{-\frac{1}{2}}\left[r^2 - n + r\frac{\partial}{\partial r}\right]. \tag{33b}$$

They are normalized ladder operators for the radial wavefunctions

$$\mathcal{K}_r^{\pm}R_{n,|m|} = R_{n\pm 2,|m|},$$

and hence are also ladder operators for the complete wavefunctions since the angular part does not depend on n;

$$\mathcal{K}_r^{\pm}\Psi_{n,m} = \Psi_{n\pm 2,m}.$$

Infeld and Hull[20] gave an operator that changes the orbital quantum number in the radial wavefunction for the hydrogen atom. We again replace r by r^2 to get a ladder operator that changes $|m|$ in the radial wavefunctions of the isotropic harmonic oscillator. After including appropriate normalizing factors, we get the operators

$$\mathcal{K}_\theta^{\pm} = \frac{|m| - 1}{[(n - |m| + 2)(n + |m|)]^{\frac{1}{2}}} \times \left[\frac{|m| - 1}{r^2} - \frac{n+1}{|m|-1} \mp \frac{1}{r}\frac{\partial}{\partial r}\right] \tag{34}$$

which are normalized m-changing ladder operators for the radial wavefunctions

$$\mathcal{K}_\theta^+ rR_{n,|m|-2} = rR_{n,|m|},$$
$$\mathcal{K}_\theta^- rR_{n,|m|} = rR_{n,|m|-2}.$$

Using the operators obtained, we define

$$\mathcal{A}_r^+ = [(n-m)/2 + 1]^{\frac{1}{2}}\mathcal{K}_r^+, \tag{35a}$$

$$\mathcal{A}_r^- = [(n-m)/2]^{\frac{1}{2}}\mathcal{K}_r^-, \tag{35b}$$

$$\mathcal{A}_\theta^+ = [(m - \tfrac{1}{2})/2 + 1]^{\frac{1}{2}}\mathcal{K}_\theta^+\mathcal{K}_r^+e^{i2\theta}, \quad m \geq 0,$$
$$\mathcal{A}_\theta^+ = [(m - \tfrac{1}{2})/2 + 1]^{\frac{1}{2}}\mathcal{K}_r^+\mathcal{K}_\theta^-e^{i2\theta}, \quad m < 0, \tag{35c}$$

$$\mathcal{A}_\theta^- = [(m - \tfrac{1}{2})/2]^{\frac{1}{2}}\mathcal{K}_r^-\mathcal{K}_\theta^-e^{-i2\theta}, \quad m > 0,$$
$$\mathcal{A}_\theta^- = [(m - \tfrac{1}{2})/2]^{\frac{1}{2}}\mathcal{K}_\theta^+\mathcal{K}_r^-e^{-i2\theta}, \quad m \leq 0. \tag{35d}$$

In the classical limit we have

$$\mathcal{A}_r^{\pm} \to a_r^{\pm},$$
$$\mathcal{A}_\theta^{\pm} \to a_\theta^{\pm}.$$

We will now see whether the commutators corresponding to (7) are satisfied for the $\mathcal{A}$'s. The first failure one notices is that $\mathcal{A}_\theta^-$ and $\mathcal{A}_r^+$ do not commute. It is easy to verify that $\mathcal{A}_\theta^-\mathcal{A}_r^+\Psi_{n,m}$ and $\mathcal{A}_r^+\mathcal{A}_\theta^-\Psi_{n,m}$ are equal if one excludes states such that $m = -(n-2)$. For these particular states, we have that $\mathcal{A}_\theta^-\mathcal{A}_r^+$ applied to $\Psi_{n,-(n-2)}$ is a constant times $\Psi_{n,-n}$, whereas $\mathcal{A}_r^+\mathcal{A}_\theta^-\Psi_{n,-(n-2)}$ is zero, since $\mathcal{A}_\theta^-$ annihilates the state before $\mathcal{A}_r^+$ gets a chance to operate on anything. This effect is illustrated in the diagram of states given in Fig. 3.

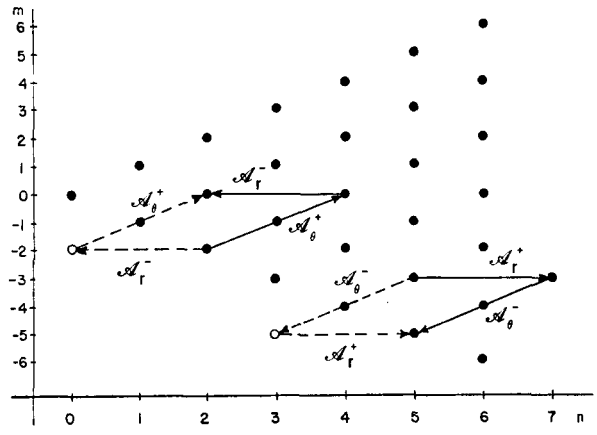


FIG. 3. Noncommutation of the operators $\mathcal{A}_\theta^+$, $\mathcal{A}_r^-$ and $\mathcal{A}_\theta^-$, $\mathcal{A}_r^+$ when applied to states such that $n = -m$ and $n = -(n-2)$, respectively.

Similar considerations apply to the commutator of $\mathcal{A}_\theta^+$ and $\mathcal{A}_r^-$; these commute when applied to states other than those for which $m = -n$ $(n > 0)$. For such states, we find that $\mathcal{A}_r^-\mathcal{A}_\theta^+$ applied to $\Psi_{n,-n}$ is a constant times $\Psi_{n,-(n-2)}$, whereas $\mathcal{A}_\theta^+\mathcal{A}_r^-\Psi_{n,-n}$ is zero. This is also illustrated in Fig. 3.

The commutator $[\mathcal{A}_r^-, \mathcal{A}_r^+]$ is found to be the identity operator except when applied to wavefunctions such that $m = -n$ $(n > 0)$, for which the commutator is different from the identity. The commutator $[\mathcal{A}_\theta^-, \mathcal{A}_\theta^+]$ fails to be the identity operator for states such that $m = -n$ and $m = -(n-2)$. But one finds that $\mathcal{A}_r^+$ and $\mathcal{A}_\theta^+$ do commute.

The result is that the supposedly universal procedure for obtaining constants of the motion satisfying $SU(2)$ Lie algebra commutation relations cannot be carried into quantum mechanics. At this point we have only seen that our ladder operators do not form the von Neumann algebra upon which we usually rely to form $SU(2)$. Our ladder operators nevertheless still form valid constants of the motion among their bilinear products.

In analogy to the classical constants (26) we define

$$\mathfrak{L} = \mathcal{A}_r^-\mathcal{A}_r^+ - \mathcal{A}_\theta^-\mathcal{A}_\theta^+, \tag{36a}$$

$$\mathfrak{K} = -i\mathcal{A}_\theta^-\mathcal{A}_r^+ + i\mathcal{A}_r^-\mathcal{A}_\theta^+, \tag{36b}$$

$$\mathfrak{D} = -\mathcal{A}_\theta^-\mathcal{A}_r^+ - \mathcal{A}_r^-\mathcal{A}_\theta^+. \tag{36c}$$

We find that

$$[\mathfrak{K}, \mathfrak{L}] = -2i\mathfrak{D}, \tag{37a}$$

$$[\mathfrak{L}, \mathfrak{D}] = -2i\mathfrak{K}, \tag{37b}$$

even though the $\mathcal{A}$'s do not satisfy the proper commutation relations, but

$$[\mathfrak{D}, \mathfrak{K}] \neq -2i\mathfrak{L}. \tag{37c}$$

By this final failure, we have to abandon the expectation that there will be a *universal unitary symmetry group constructed from ladder operators* in analogy to the classical universal groups. It is clear that the fault lies in the use of ladder operators, and not in any of the possible ambiguities which might exist in establishing them as the quantum analog of classical operators. More precisely, it appears that the difficulties arise from the states on the lower left boundary in the diagram of Fig. 3, where the expected commutation rules of our operators fail. Geometrically, we can attribute the difficulty to the fact that the ladder operators do not move states parallel to the boundaries of the region of acceptable wavefunctions. Thus a parallelogram representing the commutator of two operators may have one corner inside the region and the other outside. This nonclosure may in turn be traced to the fact that some ladder operators simultaneously change more than one quantum number, as well as that one quantum number may limit the values of another.

Although the construction of the universal unitary symmetry group has failed, we know that the harmonic oscillator does in fact have a unitary symmetry. We may disclose it by using a quantum version of Dulock's transformation (17), by which the tensor constants of the motion were recovered from those constructed from the polar ladder operators:

$$\mathfrak{L}' = 2\mathcal{A}_\theta^-\mathcal{A}_\theta^+ - \tfrac{3}{2}, \tag{38a}$$

$$\mathfrak{K}' = -i\mathcal{A}_\theta^-\mathcal{A}_r^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - \frac{1}{2}}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+ - 1}\right)^{\frac{1}{2}} + i\mathcal{A}_r^-\mathcal{A}_\theta^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - \frac{3}{2}}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+}\right)^{\frac{1}{2}}, \tag{38b}$$

$$\mathfrak{D}' = -\mathcal{A}_\theta^-\mathcal{A}_r^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - \frac{1}{2}}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+ - 1}\right)^{\frac{1}{2}} - \mathcal{A}_r^-\mathcal{A}_\theta^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - \frac{3}{2}}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+}\right)^{\frac{1}{2}}. \tag{38c}$$

The new constants do satisfy the commutators of $SU(2)$ generators,

$$[\mathfrak{K}', \mathfrak{L}'] = 2i\mathfrak{D}', \tag{39a}$$

$$[\mathfrak{L}', \mathfrak{D}'] = 2i\mathfrak{K}', \tag{39b}$$

$$[\mathfrak{D}', \mathfrak{K}'] = 2i\mathfrak{L}'. \tag{39c}$$

To complete the analogy with the classical results, one may verify the operator identities,

$$\mathfrak{L}' = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right), \tag{40a}$$

$$\mathfrak{K}' = xy - \frac{\partial^2}{\partial x \partial y}, \tag{40b}$$

$$\mathfrak{D}' = \frac{1}{2}\left(x^2 - y^2 - \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right). \tag{40c}$$

By writing the right-hand side in polar coordinates and applying the operators to a wavefunction, it will be found that the left-hand operator produces the same result.

A final detail that should be noticed is that although the isotropic harmonic oscillator written in polar coordinates has an energy expression

$$E = (2J_r + J_\theta)/2\pi$$

appropriate to an anisotropy of 2:1, only one family of $SU(2)$ representations occurs. Two such families occur for this anisotropy in Cartesian coordinates. That the polar oscillator exhibits only one family is due to the interdependence of the angular and radial quantum numbers. For example, singly degenerate states should arise from (n, m) quantum-number pairs $(0, 0)$ and $(0, 1)$. The restriction that $n \geq |m|$ prohibits the second state.

IV. THE HARMONIC OSCILLATOR WITH EXCESS ANGULAR MOMENTUM

We have now seen to what extent a transition is possible from classical to quantum mechanics in the case of the anisotropic harmonic oscillator. Classically, it is supposed that such oscillators always have a $SU(n)$ symmetry group, although the constants of the motion are in general transcendental and become algebraic only in the case where there are commensurable frequencies among the normal coordinate oscillators.

Even in the commensurable case, the assertion that $SU(n)$ is the quantum-mechanical symmetry group has to be treated with caution. The folklore of accidental degeneracy holds that when one has found the "hidden symmetry" of a system, the irreducible representations of the "true" symmetry group will occur once each. Only in this way do the symmetry-adapted wavefunctions uniquely diagonalize the Hamiltonian. No such thing applies to the anisotropic harmonic oscillator, and in fact the one-dimensional representation of $SU(n)$ occurs several times, as do also those of higher dimension. The different representations which occur, occur with the same multiplicity, forming families of $SU(n)$ representations.

Nevertheless, as we have already seen, the high incidence of one-dimensional representations allows us to reconcile the universal occurrence of $SU(n)$ as a symmetry group for harmonic oscillators with the lack of degeneracy which is observed for incommensurable frequency ratios.

Such was the state of affairs for a simple two-dimensional harmonic oscillator in Cartesian coordinates. However, as we know, the supposed universality of $SU(n)$ is not confined to harmonic oscillators in Cartesian coordinates. Rather, it extends at least to those systems which are classically degenerate. In the last section we examined in detail one such case, that of an isotropic harmonic oscillator written in polar coordinates. In terms of the radial and angular action-angle variables, it behaved rather much as an anisotropic harmonic oscillator with a 2:1 frequency ratio. The presence of non-Cartesian coordinates implies a constraint on the quantum numbers which was not present in the Cartesian case, and these constraints in turn influence the pattern of degeneracies and the commutation rules of the ladder operators. While the accustomed commutation rules may fail, we have seen that transformations might exist which would repair the failure.

The combination of the two effects—anisotropy and non-Cartesian coordinates—can be produced in a series of somewhat artificial Hamiltonians by considering an isotropic harmonic oscillator in which the moving particle is possessed of "excessive" angular momentum. The modification is wrought by adding a term proportional to L^2/r^2 to the harmonic oscillator Hamiltonian, where L is the angular momentum vector. The Cartesian Hamiltonian,

$$H = (p^2 + r^2 + \delta L^2/r^2)/2, \quad \delta \geq -1 , \qquad (41)$$

has the form

$$2H = p_r^2 + r^2 + \gamma^2 p_\theta^2/r^2, \quad \gamma = (1 + \delta)^{\frac{1}{2}} , \qquad (42)$$

in polar coordinates. Thus its interpretation is that it describes a particle for which, due to some anomaly, the angular momentum is γ times that computed from the classical formula $\mathbf{r} \times \mathbf{p}$. The restriction $\delta \geq -1$ ensures that γ is real. For example, with the factor $\gamma = \frac{1}{2}$, such a Hamiltonian arises in studying the magnetic monopole.

The Hamilton–Jacobi equation corresponding to the Hamiltonian (41),

$$2E = \left(\frac{\partial S}{\partial r}\right)^2 + \frac{\gamma^2}{r^2}\left(\frac{\partial S}{\partial \theta}\right)^2 + r^2, \qquad (43)$$

is separated by writing the principal function in the form $S = S_r(r) + S_\theta(\theta)$. We have immediately $S_\theta = \mu\theta$, where μ is a constant, since θ does not appear explicitly in the Hamiltonian. The resulting equation can than be solved for S_r:

$$S_r = \int (2E - r^2 - \gamma^2\mu^2/r^2)^{\frac{1}{2}}\, dr.$$

The action variables $J_j = \oint p_j\, dq_j$ are found by use of the relation $p_j = \partial S/\partial q_j$ and the separated

Hamilton–Jacobi equations

$$J_r = \pi(E - \gamma\mu), \tag{44a}$$

$$J_\theta = 2\pi\mu. \tag{44b}$$

The energy can then be expressed in terms of the action variables,

$$E = (2J_r + \gamma J_\theta)/2\pi, \tag{45}$$

an expression from which the "anisotropy" of the radial and angular action is evident.

Following the pattern of development to which we are now accustomed, we determine the classical ladder operators $a_r^{\pm}$ and $a_\theta^{\pm}$. Foreseeing the requirements to be imposed by the quantum-mechanical construction to be possible, the powers to which one has to raise $\exp(2\pi i w)$ have to be integral, which will require that γ be rational. Hence we set $\gamma = \alpha/\beta$, in which α and β are relatively prime integers. Since the factor of 2 multiplying the radial action will complicate this relationship, we prefer to write the Hamiltonian as a function of $(\beta' J_r + \alpha' J_\theta)$ in which α' and β' are relatively prime. When α and β are in least terms, we define $\beta' = 2\beta$, $\alpha' = \alpha$ when α is odd, and $\beta' = \beta$, $\alpha' = \alpha/2$ when α is even. Nevertheless,for the purposes of the classical construction α' and β' need not be restricted to integer values.

Our choice of the "a" functions will then be

$$a_r^{\pm} = (J_r/2\pi\alpha')^{\frac{1}{2}}(\mp i)^{\alpha'} e^{\pm 2\pi i \alpha' w_r}, \tag{46a}$$

$$a_\theta^{\pm} = (J_\theta/2\pi\beta')^{\frac{1}{2}}(-1)^{\alpha'} e^{\pm 2\pi i \beta' w_\theta}. \tag{46b}$$

The angle variables $w_j = \partial S/\partial J_j$ are evaluated using the separated Hamilton–Jacobi equations

$$w_r = \frac{1}{2\pi} \arcsin \frac{r^2 - E}{(E^2 - \gamma^2\mu^2)^{\frac{1}{2}}},$$

$$w_\theta = \frac{\gamma w_r}{2} - \frac{\gamma}{4\pi} \arcsin \frac{Er^2 - \gamma^2\mu^2}{r^2(E^2 - \gamma^2\mu^2)^{\frac{1}{2}}} + \frac{\theta}{2\pi};$$

hence,

$$\mp i e^{\pm 2\pi i w_r} = (E^2 - \gamma^2\mu^2)^{-\frac{1}{2}}(r^2 - E \mp i r p_r),$$

$$-e^{\pm 2\pi i w_\theta} = \mp i e^{\pm 2\pi i(\alpha'/\beta')w_r} \times \left[\frac{\gamma\mu}{(E^2 - \gamma^2\mu^2)^{\frac{1}{2}}}\left(\frac{\gamma\mu}{r^2} - \frac{E}{\gamma\mu} \mp i\,\frac{p_r}{r}\right)\right]^{\alpha'/\beta'} e^{\pm i\theta}$$

from which we find

$$a_r^{\pm} = \left(\frac{E}{2\alpha'} - \frac{\mu}{\beta'}\right)^{\frac{1}{2}}[(E^2 - \gamma^2\mu^2)^{-\frac{1}{2}}(r^2 - E \mp i r p_r)]^{\alpha'}, \tag{47a}$$

$$a_\theta^{\pm} = (\mp i)^{\alpha'} e^{\pm 2\pi i \alpha' w_r}\left[\frac{\mu}{\beta'}\right]^{\frac{1}{2}} \times \left[\frac{\gamma\mu}{(E^2 - \gamma^2\mu^2)^{\frac{1}{2}}}\left(\frac{\gamma\mu}{r^2} - \frac{E}{\gamma\mu} \mp i\,\frac{p_r}{r}\right)\right]^{\alpha'} e^{\pm i\beta'\theta}. \tag{47b}$$

The constants of the motion may then be given explicitly as bilinear combinations of the ladder operators,

$$L = a_r^- a_r^+ - a_\theta^- a_\theta^+, \tag{48a}$$

$$K = (-i)^{\alpha'} a_\theta^- a_r^+ + (i)^{\alpha'} a_r^- a_\theta^+, \tag{48b}$$

$$D = (-i)^{\alpha'+1} a_\theta^- a_r^+ + (i)^{\alpha'+1} a_r^- a_\theta^+. \tag{48c}$$

One can apply Dulock's transformation (27) to the constants given above, since the fact that the new constants also satisfy $SU(2)$ commutation relations only depends on the commutators among the a's; but only in the case of a pure harmonic oscillator potential do the constants have a simple interpretation, as considered in Sec. II.

The Schrödinger equation for the problem in polar coordinates is

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\gamma^2}{r^2}\frac{\partial^2}{\partial\theta^2} - r^2 + 2E\right)\Psi = 0, \tag{49}$$

whose normalized solutions are

$$\Psi_{u,m} = R_{u,|m|}(r)\Theta_m(\theta), \tag{50a}$$

$$R_{u,|m|}(r) = \left(\frac{2(u - \gamma|m|)!}{[\Gamma(u+1)]^3}\right)^{\frac{1}{2}} r^{\gamma|m|} e^{-r^2/2} L_u^{\gamma|m|}(r^2), \tag{50b}$$

$$\Theta_m(\theta) = e^{im\theta}/(2\pi)^{\frac{1}{2}}. \tag{50c}$$

The quantum numbers can take the values $m = 0, \pm 1, \pm 1, \cdots$ and $u - \gamma|m| = 0, 1, 2, \cdots$. Energy eigenvalues are given by the relation

$$(E - 1 + \gamma|m|)/2 = u$$

and hence we have

$$(E - 1) - \gamma|m| = 0, 2, 4, \cdots. \tag{51}$$

There will be no more than twofold degeneracy when γ is irrational, for the change in $\gamma|m|$ for different values of $|m|$ will not be integral and hence there will not be two values of $|m|$ such that $(E - 1) - \gamma|m|$ is an even integer for fixed E. Accidental degeneracy arises only for rational $\gamma = \alpha/\beta$ (written in least terms); in this case the energy eigenvalues have the form $E - 1 = n/\beta$, where n is a positive integer which obeys $(n - \alpha|m|)/\beta = 0, 2, 4, \cdots$.

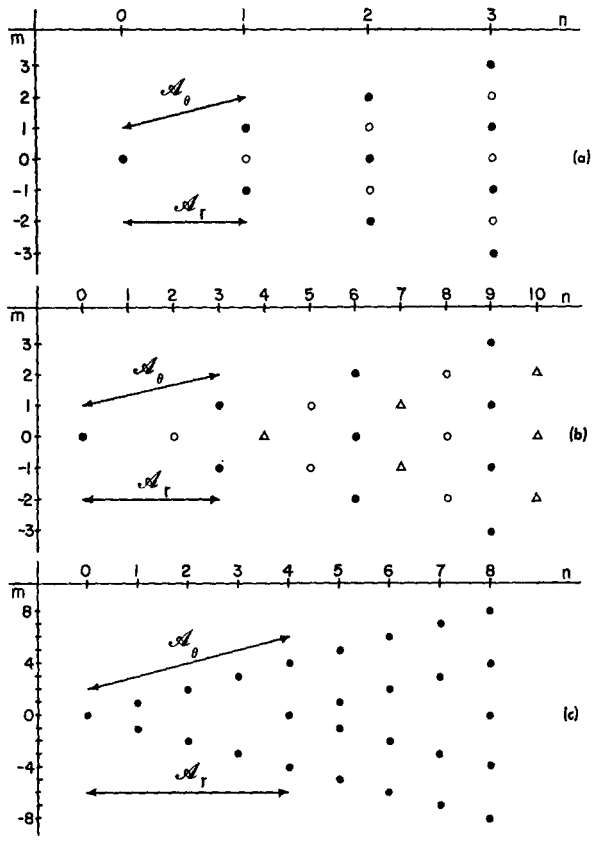


FIG. 4. Diagram of states for the harmonic oscillator with excess angular momentum showing ladder operators: (a) $\alpha = 2$, $\beta = 1$; two families of $SU(2)$ degeneracy interleave forming a family with $O(3)$ degeneracy. (b) $\alpha = 3$, $\beta = 1$; there are three independent families of $SU(2)$ degeneracy. (c) $\alpha = 1$, $\beta = 2$; uneven spacing of m for levels $n = 1 + 4p$ or $3 + 4p$ (p an integer) precludes the existence of an $SU(2)$ or $O(3)$ symmetry group.

Diagrams of states are given in Fig. 4 for several values of α and β. It will be convenient to give the wavefunctions in terms of the quantum numbers n and m for the case of rational γ:

$$\Psi_{n,m} = R_{n,|m|}(r)\,\Theta_m(\theta), \tag{52a}$$

$$R_{n,|m|}(r) = \left(\frac{2[(n-\alpha\,|m|)/2\beta]!}{[\Gamma[(n+\alpha\,|m|+2\beta)/2\beta]]^3}\right)^{\frac{1}{2}} \times r^{\alpha|m|/\beta} e^{-r_2/2} L^{\alpha|m|/\beta}_{(n+\alpha|m|)/2\beta}(r^2), \tag{52b}$$

$$\Theta(\theta) = e^{im\theta}/(2\pi)^{\frac{1}{2}}. \tag{52c}$$

The procedure for obtaining ladder operators is the same as that used in the preceding section; we find that the operators

$$\mathcal{K}_r^+ = \left[\left(\frac{n}{\beta}+2\right)^2 - \frac{\alpha^2 m^2}{\beta^2}\right]^{-\frac{1}{2}}\left[r^2 - \left(\frac{n}{\beta}+2\right) - r\frac{\partial}{\partial r}\right], \tag{53a}$$

$$\mathcal{K}_r^- = \left[\frac{n^2}{\beta^2} - \frac{\alpha^2 m^2}{\beta^2}\right]^{-\frac{1}{2}}\left[r^2 - \frac{n}{\beta} + r\frac{\partial}{\partial r}\right] \tag{53b}$$

are normalized ladder operators for the total quantum number

$$\mathcal{K}_r^{\pm} R_{n,|m|} = R_{n+2\beta,|m|}$$

and the operators

$$\mathcal{K}_\theta^+ = \left(\frac{\alpha\,|m|}{\beta}+1\right)\left[\left(\frac{n-\alpha\,|m|}{\beta}\right)\left(\frac{n+\alpha\,|m|}{\beta}+2\right)\right]^{-\frac{1}{2}} \times \left[\frac{\alpha\,|m|+\beta}{\beta r^2} - \frac{n+\beta}{\alpha\,|m|+\beta} - \frac{1}{r}\frac{\partial}{\partial r}\right], \tag{54a}$$

$$\mathcal{K}_\theta^- = \left(\frac{\alpha\,|m|}{\beta}-1\right)\left[\left(\frac{n-\alpha\,|m|}{\beta}+2\right)\left(\frac{n+\alpha\,|m|}{\beta}\right)\right]^{-\frac{1}{2}} \times \left[\frac{\alpha\,|m|-\beta}{\beta r^2} - \frac{n+\beta}{\alpha\,|m|-\beta} + \frac{1}{r}\frac{\partial}{\partial r}\right], \tag{54b}$$

which effect the following change in the radial wavefunctions:

$$\mathcal{K}_\theta^{\pm} r R_{n,|m|} = r R_{n,|m|\pm 2\beta/\alpha}.$$

$R_{n,|m|} \pm 2\beta/\alpha$ will not be the radial part of a wavefunction unless $\alpha = 1$ or 2; otherwise $(\mathcal{K}_\theta^{\pm})^\alpha$ or $(\mathcal{K}_\theta^{\pm})^{\alpha/2}$ are the lowest powers of $\mathcal{K}_\theta^{\pm}$ that will change $|m|$ by an integer for α odd or even, respectively, and hence will be ladder operators for the angular quantum number. It is just these powers (α') to which one raises the classical operators (47) according to the discussion preceding (46).

With the operators already obtained at hand we define

$$\mathcal{A}_r^+ = [((n-j)/\alpha - m)/\beta' + 1]^{\frac{1}{2}}(\mathcal{K}_r^+)^{\alpha'}, \tag{55a}$$

$$\mathcal{A}_r^- = [((n-j)/\alpha - m)/\beta']^{\frac{1}{2}}(\mathcal{K}_r^-)^{\alpha'}, \tag{55b}$$

$$\mathcal{A}_\theta^+ = [(m-\tfrac{1}{2})/\beta' + 1]^{\frac{1}{2}}(\mathcal{K}_\theta^+)^{\alpha'}(\mathcal{K}_r^+)^{\alpha'} e^{i\beta'\theta}, \quad m \geq 0,$$
$$\mathcal{A}_\theta^+ = [(m-\tfrac{1}{2})/\beta' + 1]^{\frac{1}{2}}(\mathcal{K}_r^+)^{\alpha'}(\mathcal{K}_\theta^-)^{\alpha'} e^{i\beta'\theta}, \quad m < 0, \tag{55c}$$

$$\mathcal{A}_\theta^- = [(m-\tfrac{1}{2})/\beta']^{\frac{1}{2}}(\mathcal{K}_r^-)^{\alpha'}(\mathcal{K}_\theta^-)^{\alpha'} e^{-i\beta'\theta}, \quad m > 0,$$
$$\mathcal{A}_\theta^- = [(m-\tfrac{1}{2})/\beta']^{\frac{1}{2}}(\mathcal{K}_\theta^+)^{\alpha'}(\mathcal{K}_r^-)^{\alpha'} e^{-i\beta'\theta}, \quad m \leq 0, \tag{55d}$$

where the value of j is determined from the values of n according to

n	j
$p\alpha$	0
$2\beta + p\alpha$	2β
$4\beta + p\alpha$	4β
.	.
.	.
.	.
$2\alpha\beta + p\alpha$	0

for integer values of p.

These quantum-mechanical ladder operators approach the classical ladder operators in the limit of large quantum numbers:

$$\mathcal{A}_r^{\pm} \to a_r^{\pm}, \quad \mathcal{A}_\theta^{\pm} \to a_\theta^{\pm}.$$

On the other hand, we have the same failure of the expected commutation relations among the "$\mathcal{A}$" operators that we have already seen in the more restricted problem of the isotropic harmonic oscillator expressed in polar coordinates which we treated in the last section. The reasons for the failure are again just the same ones, which arise from the fact that some of the ladder operators may change two quantum numbers simultaneously. Consequently, some products of ladder operators may allow us to pass limiting combinations of quantum numbers which cannot be done if they are applied in a different order.

Setting aside for the moment the question of their mutual commutation rules, we can use the ladder operators to define three constants of the motion independent of the Hamiltonian,

$$\mathfrak{L} = \mathcal{A}_r^- \mathcal{A}_r^+ - \mathcal{A}_\theta^- \mathcal{A}_\theta^+, \tag{56a}$$

$$\mathfrak{K} = (-i)^{\alpha'} \mathcal{A}_\theta^- \mathcal{A}_r^+ + (i)^{\alpha'} \mathcal{A}_r^- \mathcal{A}_\theta^+, \tag{56b}$$

$$\mathfrak{D} = (-i)^{\alpha'+1} \mathcal{A}_\theta^- \mathcal{A}_r^+ + (i)^{\alpha'+1} \mathcal{A}_r^- \mathcal{A}_\theta^+. \tag{56c}$$

The operators $\mathfrak{K}$ and $\mathfrak{D}$, when applied to the wavefunction with quantum numbers (n, m) gives a linear combination of those with quantum numbers $(n, m - \beta')$ and $(m, m + \beta')$. One can see that states with such quantum numbers do not exist for all n when β is greater than 1. For example, when $\alpha = 1$, $\beta = 2$, $\mathcal{A}_r^- \mathcal{A}_\theta^+$ applied to $(5, -1)$ gives a function which is not an acceptable eigenfunction, since there is no (5, 3) state [see Fig. 4(c)]. The reason that the function produced by this ladder operator is not a solution of Schrödinger's equation is because the corresponding radial function is not quadratically integrable. Infeld and Hull have noted that ladders may terminate in nonintegrable functions in certain circumstances.

In fact, when the total quantum number n is of the form $1 + 4p$ or $3 + 4p$, for $\alpha = 1$, $\beta = 2$, there is no way to combine the degenerate states of positive angular momentum with those of negative angular momentum by means of the above constants of the motion, since in each case there is a gap near zero angular momentum where the eigenfunctions are promoted into functions which are not acceptable as solutions of Schrödinger's equation. The result is that although the constants shown account for part of the degeneracy, they do not account for all of it, in the cases for $\beta > 1$. It seems that one must additionally make use of the reflective symmetry of the system to show that the negative angular momentum family is degenerate with the corresponding positive angular momentum family. It is also interesting to note that in those cases in which the given ladder operators generate the entire degenerate family, there may still appear several families of $SU(2)$ or $O(3)$ degenerate levels, the former occurring when α is odd, and the latter when α is even, but only when $\beta = 1$.

We see several effects at work in these spectra. From the anisotropy we find that there may be different groups of states with the same degeneracy belonging to different total quantum numbers. Thus we can expect to see several families of irreducible representations, if we succeed in finding a symmetry group. However, the effect of the non-Cartesian coordinates is to introduce constraints between different quantum numbers, which we have seen can interfere with the commutation relations of the ladder operators and prevent them from generating an $SU(n)$ group. Finally we have seen a new effect, that the ladder operators may not cause transformations which lie entirely within the Hilbert space of quadratically integrable solutions of our Schrödinger's equation.

Thus we expect to be prevented, in the most general case, from forming a symmetry group whose generators consist of products of our ladder operators. It is interesting to note that Dulock's regularization, which has previously served to correct the failure of the commutation rules on wavefunctions of extreme quantum numbers, is only applicable in the case of $\beta = 1$, since there is no way to overcome the unequal spacing of the angular momentum quantum numbers which occurs near zero angular momentum. But, when $\beta = 1$, the regularization produces

$$\mathfrak{L}' = 2\mathcal{A}_\theta^- \mathcal{A}_\theta^+ - 2 + 1/2\beta', \tag{57a}$$

$$\mathfrak{K}' = (-i)^{\alpha'} \mathcal{A}_\theta^- \mathcal{A}_r^+ \left[2 + \frac{\mathcal{A}_r^- \mathcal{A}_r^+ - 1 - 1/\beta'}{\mathcal{A}_\theta^- \mathcal{A}_\theta^+ - 1}\right]^{\frac{1}{2}} + (i)^{\alpha'} \mathcal{A}_r^- \mathcal{A}_\theta^+ \left[2 + \frac{\mathcal{A}_r^- \mathcal{A}_r^+ - 2 - 1/\beta'}{\mathcal{A}_\theta^- \mathcal{A}_\theta^+}\right]^{\frac{1}{2}}, \tag{57b}$$

$$\mathfrak{D}' = (-i)^{\alpha'+1} \mathcal{A}_\theta^- \mathcal{A}_r^+ \left[2 + \frac{\mathcal{A}_r^- \mathcal{A}_r^+ - 1 - 1/\beta'}{\mathcal{A}_\theta^- \mathcal{A}_\theta^+ - 1}\right]^{\frac{1}{2}} + (i)^{\alpha'+1} \mathcal{A}_r^- \mathcal{A}_\theta^+ \left[2 + \frac{\mathcal{A}_r^- \mathcal{A}_r^+ - 2 - 1/\beta'}{\mathcal{A}_\theta^- \mathcal{A}_\theta^+}\right]^{\frac{1}{2}}. \tag{57c}$$

These operators may be verified to generate the Lie algebra of $SU(2)$ or $O(3)$. We note that the constant j which was included in the definition of (55) was

required to ensure that these commutation rules would assume their proper form.

V. THE HYDROGEN ATOM WITH EXCESS ANGULAR MOMENTUM

The Hamiltonian for the hydrogen atom with excess angular momentum is given by

$$2H = p_r^2 + \frac{\gamma^2 p_\theta^2}{r^2} - \frac{2}{r} \tag{58}$$

in polar coordinates. An entirely analogous procedure to that used in Secs. III and IV yields, after separation of the Hamiltonian–Jacobi equation,

$$S_\theta = \mu\theta, \tag{59a}$$

$$S_r = \int \left(2E - \frac{\gamma^2\mu^2}{r^2} + \frac{2}{r}\right)^{\frac{1}{2}} dr, \tag{59b}$$

where μ is a separation constant. The action variables are given by

$$J_r = 2\pi[(-2E)^{-\frac{1}{2}} - \gamma\mu], \tag{60a}$$

$$J_\theta = 2\pi\mu, \tag{60b}$$

from which the energy is given in terms of the action variables

$$E = -2\pi^2/(J_r + \gamma J_\theta)^2. \tag{61}$$

Again, the quantum-mechanical construction of ladder operators will only be possible for rational γ. We set $\gamma = \alpha/\beta$ where α and β are relatively prime integers, the Hamiltonian is written as a function of $(\beta J_r + \alpha J_\theta)^n$; this allows construction of "a" operators with integral powers of $\exp(2\pi i w)$. Nevertheless, the classical construction does not depend on these considerations.

Our choice of "a" functions will be

$$a_r^\pm = (J_r/2\pi\alpha)^{\frac{1}{2}}(\mp i)^\alpha e^{\pm 2\pi i\alpha w_r}, \tag{62a}$$

$$a_\theta^\pm = (J_\theta/2\pi\beta)^{\frac{1}{2}}(-1)^\alpha e^{\pm 2\pi i\beta w_\theta}. \tag{62b}$$

For this problem we have

$$w_r = -\frac{1}{2\pi}(-2E)^{\frac{1}{2}} r p_r + \frac{1}{2\pi}\arcsin\frac{-2Er - 1}{(1 - 2E\gamma^2\mu^2)^{\frac{1}{2}}},$$

$$w_\theta = \gamma w_r - \frac{\gamma}{2\pi}\arcsin\frac{r - \gamma\mu}{r(1 + 2E\gamma^2\mu^2)^{\frac{1}{2}}} + \frac{\theta}{2\pi}.$$

Hence,

$$\mp i e^{\pm 2\pi i w_r} = e^{\mp i(-2E)^{\frac{1}{2}} r p_r}[(-2E)^{-1} - \gamma^2\mu^2]^{-\frac{1}{2}} \times [(-2E)^{\frac{1}{2}} r - (-2E)^{-\frac{1}{2}} \mp i r p_r],$$

$$-e^{\pm 2\pi i w_\theta} = \mp i e^{\pm 2\pi i\gamma w_r} \times \left[\frac{\gamma\mu}{(1 + 2E\gamma^2\mu^2)^{\frac{1}{2}}}\left(\frac{\gamma\mu}{r} - \frac{1}{\gamma\mu} \mp i p_r\right)\right]^\gamma e^{\pm i\theta},$$

from which we find

$$a_r^\pm = \left[\frac{(-2E)^{-\frac{1}{2}}}{\alpha} - \frac{\mu}{\beta}\right]^{\frac{1}{2}} \times \{e^{\mp i(-2E)^{\frac{1}{2}} r p_r}[(-2E)^{-1} - \gamma^2\mu^2]^{-\frac{1}{2}} \times [(-2E)^{\frac{1}{2}} r - (-2E)^{-\frac{1}{2}} \mp i r p_r]\}^\alpha, \tag{63a}$$

$$a_\theta^\pm = (\mp i)^\alpha e^{\pm 2\pi i\alpha w_r}\left(\frac{\mu}{\beta}\right)^{\frac{1}{2}} \times \left[\frac{\gamma\mu}{(1 + 2E\gamma^2\mu^2)^{\frac{1}{2}}}\left(\frac{\gamma\mu}{r} - \frac{1}{\gamma\mu} \mp i p_r\right)\right]^\alpha e^{\pm i\beta\theta}. \tag{63b}$$

One can construct the constants of the motion given by (48), setting $\alpha' = \alpha$. After Dulock's transformation (27), one gets another set of constants also satisfying $SU(2)$ commutation relations. The new constants have no simple interpretation for general γ, except in the case $\gamma = 1$, that is, for a pure $-1/r$ potential. The transformed constants are given by

$$L' = 2p_\theta, \tag{64a}$$

$$K' = 2(\sin\theta + p_r p_\theta\cos\theta - p_\theta^2\sin\theta/r)/(-2E)^{\frac{1}{2}}, \tag{64b}$$

$$D' = 2(\cos\theta - p_r p_\theta\sin\theta - p_\theta^2\cos\theta/r)/(-2E)^{\frac{1}{2}}; \tag{64c}$$

the quantities in parentheses in (64b) and (64c) are the y and x components, respectively, of the Runge vector. This result was given by Dulock.

The Schrödinger equation for the problem in polar coordinates is

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\gamma^2}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{2}{r} + 2E\right)\Psi = 0, \tag{65}$$

whose normalized solutions are

$$\Psi_{u,m} = R_{u,|m|}(r)\Theta_m(\theta), \tag{66a}$$

$$R_{u,|m|} = \left(\frac{2}{(u - \gamma|m| + \frac{1}{2})^3}\frac{(u - 2\gamma|m|)!}{[\Gamma(\mu + 1)]^3}\right)^{\frac{1}{2}} \times \rho^{\gamma|m|} e^{-\rho/2} L_u^{2\gamma|m|}(\rho), \tag{66b}$$

$$\Theta = e^{im\theta}/(2\pi)^{\frac{1}{2}}, \quad \rho = 2(-2E)^{\frac{1}{2}} r. \tag{66c}$$

The quantum numbers can take the values $m = 0, \pm 1, \pm 2, \pm\cdots$ and $u - 2\gamma|m| = 0, 1, 2, \cdots$. Energy eigenvalues are given by the relation

$$(-2E)^{-\frac{1}{2}} + \gamma|m| - \tfrac{1}{2} = u,$$

and we have

$$(-2E)^{-\frac{1}{2}} - \gamma\,|m| - \tfrac{1}{2} = 0, 1, 2, \cdots. \quad (67)$$

The same argument given after (51) leads to the conclusion that there will be no more than twofold degeneracy for irrational γ. For rational $\gamma = \alpha/\beta$ there is accidental degeneracy, and in this case the energy eigenvalues have the form $(-2E)^{-\frac{1}{2}} = n/\beta + \frac{1}{2}$, where n is a positive integer which obeys $(n - \alpha\,|m|)/\beta = 0, 1, 2, \cdots$. Diagrams of states are given in Fig. 5 for various values of α and β. It will be convenient to give the wavefunctions in terms of the quantum numbers n and m for the case of rational γ:

$$\Psi_{n,m} = R_{n,|m|}(r)\Theta_m(\theta), \quad (68a)$$

$$R_{n,|m|} = \left(\frac{2}{(n/\beta + \frac{1}{2})^3}\,\frac{[(n - \alpha\,|m|)/\beta]!}{[\Gamma((n + \alpha\,|m| + \beta)/\beta)]^3}\right)^{\frac{1}{2}} \times \rho^{\alpha|m|/\beta} e^{-\rho/2} L^{2\alpha|m|/\beta}_{(n+\alpha|m|)/\beta}(\rho); \quad (68b)$$

$$\Theta_m = e^{im\theta}/(2\pi)^{\frac{1}{2}}, \quad \rho = 2r/(n/\beta + \tfrac{1}{2}). \quad (68c)$$

The ladder operators given by Schrödinger for the total quantum number of the *three*-dimensional hydrogen atom can be adapted, by appropriate renaming of the indices and inclusion of an adequate normalizing factor, into operators appropriate for our two-dimensional case:

$$\mathcal{K}_r^+ = \exp\left(\ln\frac{n + \beta/2}{n + 3\beta/2}\right)\rho\,\frac{\partial}{\partial\rho}\left(\frac{n + \beta/2}{n + 3\beta/2}\right)^{\frac{3}{2}} \times \left[\left(\frac{n}{\beta} + 1\right)^2 - \frac{\alpha^2 m^2}{\beta^2}\right]^{-\frac{1}{2}}\left[\frac{\rho}{2} - \frac{n}{\beta} - 1 - \rho\,\frac{\partial}{\partial\rho}\right], \quad (69a)$$

$$\mathcal{K}_r^- = \exp\left(\ln\frac{n + \beta/2}{n - \beta/2}\right)\rho\,\frac{\partial}{\partial\rho}\left(\frac{n + \beta/2}{n - \beta/2}\right)^{\frac{3}{2}} \times \left[\frac{n^2}{\beta^2} - \frac{\alpha^2 m^2}{\beta^2}\right]^{-\frac{1}{2}}\left[\frac{\rho}{2} - \frac{n}{\beta} + \rho\,\frac{\partial}{\partial\rho}\right]. \quad (69b)$$

These operators are normalized ladder operators for the total quantum number n of the radial wavefunctions of the problem in consideration; the exponential part is included to change the multiplier $(n + \beta/2)/\beta$ in the argument ρ of the wavefunction to $(n \pm \beta + \beta/2)/\beta$. We have

$$\mathcal{K}_r^{\pm} R_{n,|m|} = R_{n\pm\beta,|m|}.$$

The ladder operators given by Infeld and Hull, for the orbital quantum number of the three-dimensional hydrogen atom, can similarly be transformed to get the operators

$$\mathcal{K}_\theta^+ = \frac{(n + \beta/2)(\alpha\,|m| + \beta/2)}{[(n - \alpha\,|m|)(n + \alpha\,|m| + \beta)]^{\frac{1}{2}}} \times \left(\frac{\alpha\,|m| + \beta/2}{\beta r} - \frac{\beta}{\alpha\,|m| + \beta/2} - \frac{\partial}{\partial r}\right), \quad (70a)$$

$$\mathcal{K}_\theta^- = \frac{(n + \beta/2)(\alpha\,|m| - \beta/2)}{[(n - \alpha\,|m| + \beta)(n + \alpha\,|m|)]^{\frac{1}{2}}} \times \left(\frac{\alpha\,|m| - \beta/2}{\beta r} - \frac{\beta}{\alpha\,|m| - \beta/2} + \frac{\partial}{\partial r}\right), \quad (70b)$$

which effect the following change in the radial wavefunction:

$$\mathcal{K}_\theta^{\pm} r^{\frac{1}{2}} R_{n,|m|} = r^{\frac{1}{2}} R_{n,|m|\pm\beta/\alpha}.$$

$R_{n,|m|} \pm \beta/\alpha$ does not belong to a wavefunction, unless $\alpha = 1$; otherwise $(\mathcal{K}_\theta^{\pm})^\alpha$ is the lowest power which will change from one wavefunction to another. This is the same power to which one raises the classical functions (63).

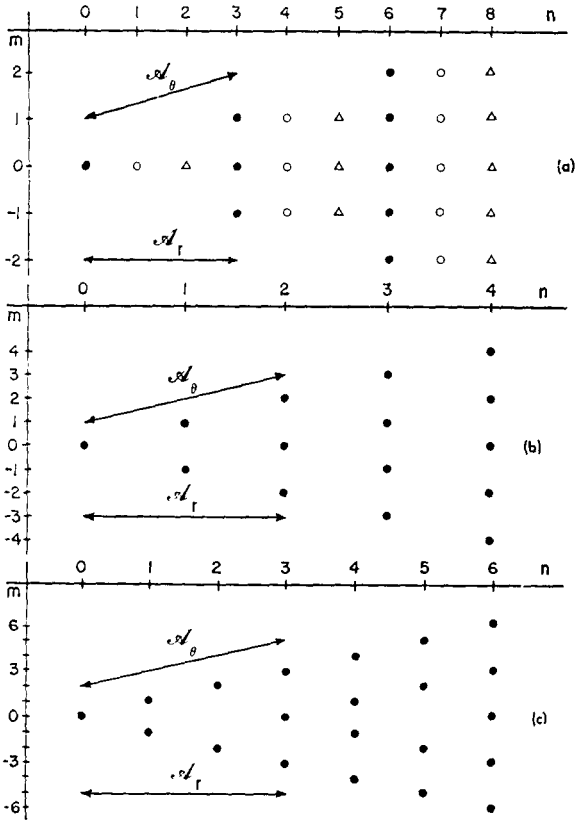


FIG. 5. Diagram of states for the hydrogen atom with excess angular momentum, showing ladder operators: (a) $\alpha = 3$, $\beta = 1$; there are three families with $O(3)$ degeneracy. (b) $\alpha = 1$, $\beta = 2$; there is one family with $SU(2)$ degeneracy. (c) $\alpha = 1$, $\beta = 3$; uneven spacing of m for level $n = 1 + 3p$ or $2 + 3p$ (p an integer) precludes the existence of $O(3)$ or $SU(2)$ symmetry group.

With these operators at hand we define

$$\mathcal{A}_r^+ = \{[(n-j)/\alpha - m]/\beta + 1\}^{\frac{1}{2}}(\mathcal{K}_r^+)^\alpha, \tag{71a}$$

$$\mathcal{A}_r^- = \{[(n-j)/\alpha - m]/\beta\}^{\frac{1}{2}}(\mathcal{K}_r^-)^\alpha, \tag{71b}$$

$$\mathcal{A}_\theta^+ = [(m-\tfrac{1}{2})/\beta + 1]^{\frac{1}{2}}(\mathcal{K}_\theta^+)^\alpha(\mathcal{K}_r^+)^\alpha e^{i\beta\theta}, \quad m \geq 0,$$
$$\mathcal{A}_\theta^+ = [(m-\tfrac{1}{2})/\beta + 1]^{\frac{1}{2}}(\mathcal{K}_r^+)^\alpha(\mathcal{K}_\theta^-)^\alpha e^{i\beta\theta}, \quad m < 0, \tag{71c}$$

$$\mathcal{A}_\theta^- = [(m-\tfrac{1}{2})/\beta]^{\frac{1}{2}}(\mathcal{K}_r^-)^\alpha(\mathcal{K}_\theta^-)^\alpha e^{-i\beta\theta}, \quad m > 0,$$
$$\mathcal{A}_\theta^- = [(m-\tfrac{1}{2})/\beta]^{\frac{1}{2}}(\mathcal{K}_\theta^+)^\alpha(\mathcal{K}_r^-)^\alpha e^{-i\beta\theta}, \quad m \leq 0, \tag{71d}$$

where the value of j is determined from the values of n according to

n	j
$p\alpha$	0
$\beta + p\alpha$	β
$2\beta + p\alpha$	2β
.	.
.	.
.	.
$(\alpha-1)\beta + p\alpha$	$(\alpha-1)\beta$.

These ladder operators approach the classical "a" functions in the limit of large quantum numbers:

$$\mathcal{A}_r^\pm \to a_r^\pm,$$
$$\mathcal{A}_\theta^\pm \to a_\theta^\pm.$$

Again there is a failure of the expected commutation relations among the $\mathcal{A}$ operators, for the same reasons as those considered in Sec. III.

The quantum-mechanical constants of the motion are given by (56) after setting $\alpha' = \alpha$. In this problem we find that the constants do not account for all the degeneracy in the cases $\beta > 2$, where uneven angular momentum spacing is found for some degenerate levels. In the cases where the constants do generate all of the degeneracy, we find that there may appear several families of $O(3)$ degeneracy, in the case $\beta = 1$; and several families of $SU(2)$ degeneracy in the case $\beta = 2$. One expects from the classical Poisson-bracket relations that the constants $\mathfrak{K}$, $\mathfrak{L}$, and $\mathfrak{D}$ should satisfy $SU(2)$ commutation relations, but there is a failure of the commutator of $\mathfrak{K}$ and $\mathfrak{D}$ as considered in Sec. III [Eqs.(37)]. In the case where the constants account for the degeneracy, we may apply a quantum-mechanical version of Dulock's transformation,

$$\mathfrak{L}' = 2\mathcal{A}_\theta^-\mathcal{A}_\theta^+ - 2 + 1/\beta, \tag{72a}$$

$$\mathfrak{K}' = (-i)^\alpha\mathcal{A}_\theta^-\mathcal{A}_r^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - 1 + 1/\beta}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+ - 1}\right)^{\frac{1}{2}} + (i)^\alpha\mathcal{A}_r^-\mathcal{A}_\theta^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - 2 + 1/\beta}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+}\right)^{\frac{1}{2}}, \tag{72b}$$

$$\mathfrak{D}' = (-i)^{\alpha+1}\mathcal{A}_\theta^-\mathcal{A}_r^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - 1 + 1/\beta}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+ - 1}\right)^{\frac{1}{2}} + (i)^{\alpha+1}\mathcal{A}_r^-\mathcal{A}_\theta^+\left(2 + \frac{\mathcal{A}_r^-\mathcal{A}_r^+ - 2 + 1/\beta}{\mathcal{A}_\theta^-\mathcal{A}_\theta^+}\right)^{\frac{1}{2}}, \tag{72c}$$

to obtain constants satisfying $SU(2)$ or $O(3)$ commutation relations. For a pure hydrogen-atom potential $\alpha = \beta = n$, we have the identities

$$\mathfrak{L}' = \frac{2}{i}\frac{\partial}{\partial\theta}, \tag{73a}$$

$$\mathfrak{K}' = 2\left(\cos\theta - \frac{\sin\theta}{2r}\frac{\partial}{\partial\theta} + \frac{\cos\theta}{2}\frac{\partial}{\partial r} + \sin\theta\frac{\partial^2}{\partial r\partial\theta} + \frac{\cos\theta}{r}\frac{\partial^2}{\partial\theta^2}\right)(-2\mathfrak{H})^{-\frac{1}{2}}, \tag{73b}$$

$$\mathfrak{D}' = 2\left(\sin\theta + \frac{\cos\theta}{2r}\frac{\partial}{\partial\theta} + \frac{\sin\theta}{2}\frac{\partial}{\partial r} - \cos\theta\frac{\partial^2}{\partial r\partial\theta} + \frac{\sin\theta}{r}\frac{\partial^2}{\partial\theta^2}\right)(-2\mathfrak{H})^{-\frac{1}{2}}, \tag{73c}$$

where $\mathfrak{H}$ is the Hamiltonian operator.

They are easily verified by applying the operators to any wavefunction, when transformed to Cartesian coordinates:

$$\mathfrak{L}' = \frac{2}{i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right), \tag{74a}$$

$$\mathfrak{K}' = 2\left(\frac{x}{(x^2+y^2)^{\frac{1}{2}}} + x\frac{\partial^2}{\partial y^2} - y\frac{\partial^2}{\partial x\partial y} - \frac{1}{2}\frac{\partial}{\partial x}\right)(-2\mathfrak{H})^{-\frac{1}{2}}, \tag{74b}$$

$$\mathfrak{D}' = 2\left(\frac{y}{(x^2+y^2)^{\frac{1}{2}}} + y\frac{\partial^2}{\partial x^2} - x\frac{\partial^2}{\partial y\partial x} - \frac{1}{2}\frac{\partial}{\partial y}\right)(-2\mathfrak{H})^{-\frac{1}{2}}; \tag{74c}$$

they can be recognized to be the angular momentum operator and the x and y components of the quantum-mechanical version of the Runge vector, respectively.

VI. THE RELATIONSHIP BETWEEN THE HYDROGEN ATOM AND THE HARMONIC OSCILLATOR

Historically there has always been considered to have been a very close relationship between the harmonic oscillator and the Kepler problem, which has persisted in the comparison of the quantum-mechanical versions of the two systems. The relation

is particularly apparent when polar coordinates are used to describe the motion, since it seems that the orbit is naturally expressed in terms of the polar radius in the Kepler problem, and in terms of the square of the polar radius for the harmonic oscillator. Moreover, where the polar angle enters into the equation of the Kepler orbits, the double angle appears in the orbital equation of the harmonic oscillator. Otherwise the equations of the orbits are functionally practically identical. A relationship can also be seen, due to the double angle, in the fact that the Kepler orbit has one point of perihelion and one aphelion, while these numbers are doubled in the harmonic-oscillator orbits. In both instances we have ellipses, but for the Kepler problem the origin lies at the focus, while for the harmonic oscillator it lies at the center of symmetry of the ellipse.

With such similarities it is hardly surprising that one might be able to transform one system into the other. However, the transformation cannot be made too directly, because one cannot simply substitute r^2 for r and double the angular coordinates. Actually this simple substitution of r^2 for r is possible in the radial equation of the separated Hamilton–Jacobi equation. In fact, substituting reciprocals or powers is a standard technique for obtaining solutions of problems with one central-force power law in terms of those of another. However, the radial equation will contain various coefficients, some of which will be constant parameters of the problem, and especially in higher dimension, others which will be separation constants from the Hamilton–Jacobi equation. Although they are constants of the motion, they still have a functional dependence on the coordinates and momenta. Hence, when a change of variable such as we have proposed is made, the significance of various constants appearing in the equation may be altered; for instance, the role of the energy in one equation may be played by the angular momentum after the substitution. So it is that the relationship between coefficients may not remain intact when one tries to extend a substitution such as of one power of the radius for another to a canonical transformation involving all the variables of the system.

Between the harmonic oscillator and the Kepler problem, the principal difficulty in trying to substitute r^2 for r and 2θ for θ is that while it will transform the orbits, it will not transform the energy properly. An energy transformation implies a conjugate transformation in the time variable, which may be understood from the fact that although the substitution mentioned will transform the orbits it will not transform properly the particle velocity within the orbit. However, in the context of Lagrangian mechanics, Wintner[21] has shown how, by introducing the eccentric anomaly of the Kepler problem as the new independent variable, the interchange of the two problems may be effected. It is also possible to interpret the use of parabolic coordinates, in which the Kepler problem is naturally expressed, as a means of realizing the square transformation, because the parabolic coordinates can be regarded as the real and imaginary parts of an analytic function and a canonical transformation based on the square of this function.

Quantum mechanically it has also been noticed that the wavefunctions of the harmonic oscillator and those of the hydrogen atom are very similar.[22] The angular wavefunctions are, of course, the same, so the only difference appears in the occurrence of r^2 as a variable in the former functions and of r in a corresponding role in the latter. Except for the additional appearance of a scale factor, the functional form of the wavefunctions in the two cases are quite similar. Originally regarded as a mere curiosity, this relationship has been exploited as a means of solving the one equation in terms of the other, and continues to receive more attention as other similarities between the two problems have been noted. This is particularly true when the full Schrödinger equation is taken into account, and not the radial part alone. Traditionally, the symmetry group $O(n+1)$ has been regarded as the exclusive province of the n-dimensional hydrogen atom, and $SU(n)$ that of the harmonic oscillator. Consequently, when it was found that certain hydrogen-atom wavefunctions transform according to the representations of a unitary group,[23] or while it was supposed that there existed universal symmetry groups, there was considerable interest in seeing what relation these two problems might have to one another.

An interesting approach, reminiscent of the transformations of the radial equation in the classical Hamilton–Jacobi theory, is the nonunitary transformation proposed by Plebanski.[24]

We feel, however, that the most productive point of view is to regard both the harmonic oscillator and Kepler problem as variants of the theme of an anisotropic classically degenerate action-angle system complicated somewhat by the presence of non-Cartesian coordinates. In developing this viewpoint, we consider first the classical orbits.

Given the radial principal functions of the two-dimensional harmonic oscillator and hydrogen atom, both with excess angular momentum [Eqs. (43) and (59b)], one obtains easily the equations of the orbits by a method described by Born.[25] The radial-angle

variable is given by

$$w_r = \int \left(2E - \frac{\gamma^2\mu^2}{r^2} + \frac{2}{r}\right)^{-\frac{1}{2}} \nu_r \, dr \qquad (75)$$

for the hydrogen-atom potential, where ν_r is the characteristic frequency in the radial direction. Since $dw_r = \nu_r \, dt$, we have

$$dt = \left(2E - \frac{\gamma^2\mu^2}{r^2} + \frac{2}{r}\right)^{-\frac{1}{2}} dr. \qquad (76)$$

Using the expression for the effective angular momentum,

$$\gamma\mu = \frac{r^2}{\gamma}\frac{d\theta}{dt},$$

to eliminate dt from (76), one gets the differential equation of the orbit. Integrating, one obtains the equation of the orbit

$$r_{\text{II}} = \frac{\gamma^2\mu^2}{1 + (1 + 2\gamma^2\mu^2 E)^{\frac{1}{2}} \cos(\theta/\gamma)}, \qquad (77)$$

where the polar axis is oriented to simplify the initial conditions.

Similarly for the harmonic oscillator one has

$$r_{\text{O}}^2 = \frac{\gamma^2\mu^2}{E + (E^2 - \gamma^2\mu^2)^{\frac{1}{2}} \cos(2\theta/\gamma)}. \qquad (78)$$

For $\gamma = 1$, these two orbits are well known to be an ellipse with focus at the center of force, and an ellipse with center of symmetry at the center of force, respectively.

One obtains an orbital precession when $\gamma \neq 1$. Examples of orbits for the harmonic oscillator with excess angular momentum are plotted in Fig. 6 for a

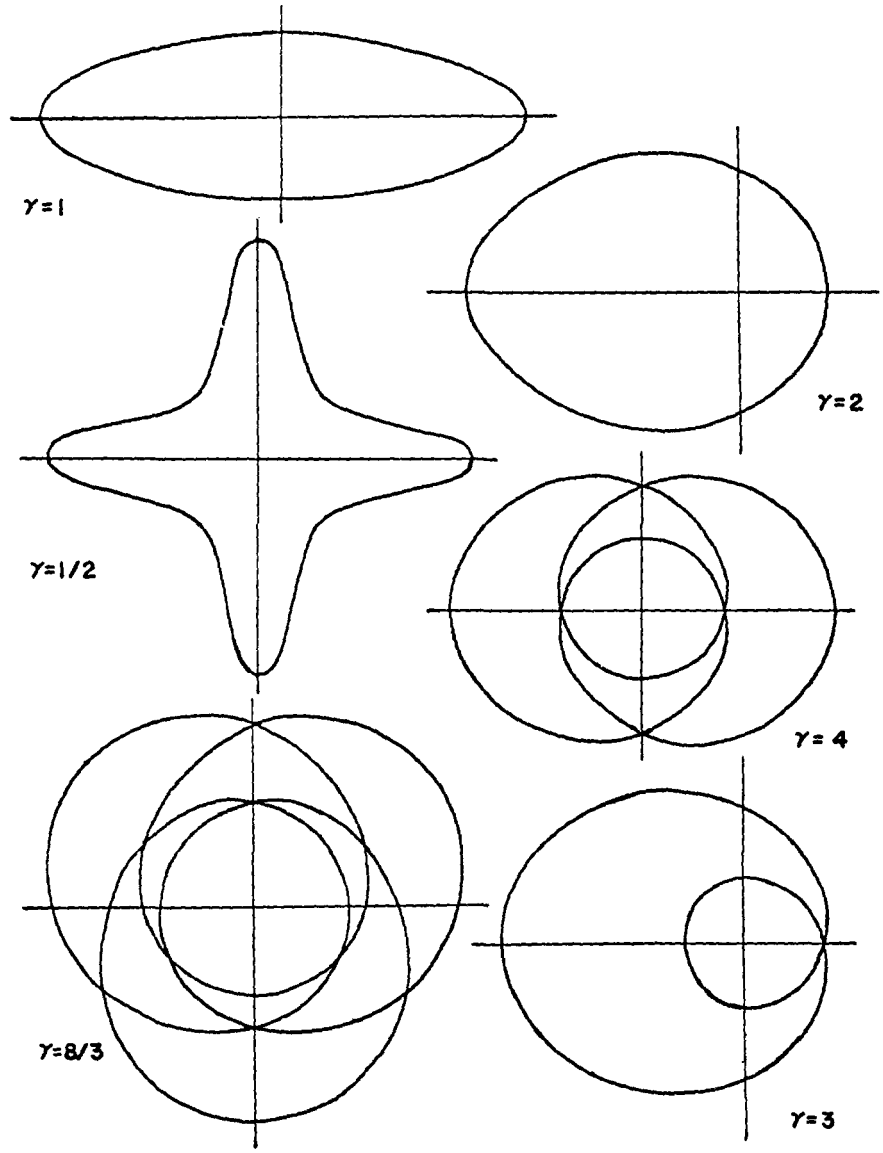


FIG. 6. Classical orbits for the harmonic oscillator with excess angular momentum.

few values of γ. The orbits for the hydrogen atom are qualitatively the same if one considers that the value of γ is half that given in the figure, and takes the square root in the radial direction. It is clear from inspection of the orbital equations that such a substitution transforms the functional form of one equation into that of the other. A complete transcription from one problem to the other is given by

$$(-2E_{\mathrm{H}})^{\frac{1}{2}} r_{\mathrm{H}} = r_{\mathrm{O}}^2, \tag{79a}$$

$$(-2E_{\mathrm{H}})^{-\frac{1}{2}} = E_{\mathrm{O}}, \tag{79b}$$

$$\gamma_{\mathrm{H}} = \gamma_{\mathrm{O}}/2, \tag{79c}$$

$$\mu_{\mathrm{H}} = 2\mu_{\mathrm{O}}. \tag{79d}$$

This transformation is valid for any number of dimensions if one considers θ to be the angular distance from the moving point to a fixed line in the plane of the motion. The angular components of the principal function contain the transformed variables in the combination $\gamma\mu$ and hence are unaltered by the transformation. When one considers the problem in Hamilton–Jacobi form, the energy for these systems is given in terms of the action variables by (45) and (61). One finds that there is a factor of 2 multiplying J_r in the case of the harmonic oscillator, which is due to the fact that one cycle of r is half a cycle of θ; when one applies transformation (79), this factor is corrected by the substitution $\gamma_{\mathrm{H}} = \gamma_{\mathrm{O}}/2$.

Another possible transcription would be

$$(-2E_{\mathrm{H}})^{\frac{1}{2}} r_{\mathrm{H}} = r_{\mathrm{O}}^2, \tag{80a}$$

$$(-2E_{\mathrm{H}})^{-\frac{1}{2}} = E_{\mathrm{O}}, \tag{80b}$$

$$\theta_{\mathrm{H}} = 2\theta_{\mathrm{O}}. \tag{80c}$$

Its extension to more than two dimensions is slightly more complicated, for θ is the angle measured in the plane of motion, and has to be properly described in terms of the angular variables in use in the particular system.

Quite generally, the n-dimensional Schrödinger equation for a particle moving in a central potential $V(r)$, with excess angular momentum, expressed in polar coordinates, is

$$\left(\frac{d^2}{dr^2} + \frac{N+1}{r}\frac{d}{dr} - \frac{\gamma^2 l(l+N-2)}{r^2} + 2(E-r)\right)\Psi = 0. \tag{81}$$

Here L^2, the square of the angular momentum operator, whose eigenfunctions are the n-dimensional hyperspherical harmonics Y_l^N has been replaced by its eigenvalue $l(l+N-2)$, $l = 0, 1, 2, \cdots$.

Equation (81) has solutions in terms of Laguerre polynomials for both the hydrogen-atom potential $V = -1/r$ and the harmonic oscillator potential $V = r^2/2$ in any number of dimensions.

For the Coulomb potential $V = -1/r$, the solution is given by[26]

$$^{\mathrm{H}}R_{u,l} = \left(\frac{2^{N-1}}{(u-\gamma l-(N-3)/2)^{N+1}} \times \frac{(u-(2\gamma l+N-2))!}{[\Gamma(u+1)]^3}\right)^{\frac{1}{2}} \rho^{\gamma l} e^{-\rho/2} L_u^{2\gamma l+N-2}(\rho), \qquad \rho = 2(-2E)^{\frac{1}{2}} r. \tag{82}$$

The quantum numbers satisfy $u - (2\gamma l + N - 2) = 0, 1, 2, \cdots$. Energy eigenvalues are given by the relation $(-2E)^{-\frac{1}{2}} + (2\gamma l + N - 3)/2 = u$. Hence we have

$$(-2E)^{-\frac{1}{2}} - \gamma l - (N-1)/2 = 0, 1, 2, \cdots. \tag{83}$$

For the harmonic potential $V = r^2/2$, the solution is

$$^{\mathrm{O}}R_{u,l} = \left(\frac{2(u-\gamma l-(N-2)/2)!}{[\Gamma(u+1)]^3}\right)^{\frac{1}{2}} \times r^{\gamma l} e^{-r^2/2} L_u^{\gamma l+(N-2)/2}(r^2); \tag{84}$$

the quantum numbers satisfy $u - \gamma l - (N-2)/2 = 0, 1, 2, \cdots$. Energy eigenvalues are given by the relation $2u \doteq (E - N/2) + (\gamma l + N - 2)$. From this we have

$$(E - N/2)/2 - \gamma l/2 = 0, 1, 2, \cdots. \tag{85}$$

Inspection of the relations (83) and (85) shows, that for the same number of dimensions, one can transform the degeneracy pattern of one problem into that of the other by simply choosing $\gamma_{\mathrm{H}} = \gamma_{\mathrm{O}}/2$. Let us suppose that $S_{\mathrm{H}} = (-2E)^{-\frac{1}{2}} - (N-1)/2$ and $S_{\mathrm{O}} = (E - N/2)/2$. To find degeneracy above that given by the possible combinations of the remaining angular quantum numbers, one has to determine the various values of l which make $S_{\mathrm{H}} - \gamma_{\mathrm{H}} l$ or $S_{\mathrm{O}} - \gamma_{\mathrm{O}} l/2$ a positive integer; hence, if $\gamma_{\mathrm{H}} = \gamma_{\mathrm{O}}/2$ we will find exactly the same degeneracy for the level $S_{\mathrm{H}} = S_{\mathrm{O}}$.

Now that we have refreshed our memory concerning the classical orbits, Schrödinger equations, wavefunctions, eigenvalues, and degeneracy patterns of these two problems, we may return to the question of the extent we may apply information pertaining to one of the systems to the other. In spite of their close similarities, one system is not a canonical transform of the other unless one wishes to tamper with the independent variable. Although such a transformation is feasible in classical mechanics, it is not

practical quantum mechanically. The closest reasonable canonical transformation, interchanging $(-2E)^{\frac{1}{2}}r$ with r^2 and θ with 2θ, transforms one Hamiltonian into a function of the other. However, in that manner one may still expect to find that the degeneracy patterns transform correctly, that there will be a functional relation between the eigenvalues of one problem and the transformed eigenvalues of the other, and that the symmetry group of the one system will pass over into the symmetry group of the other. It is also possible simply to interchange $(-2E)^{\frac{1}{2}}r$ with r^2, and instead of changing the angular coordinates, replace γ by $\gamma/2$. Effectively the anisotropy of the Hamiltonian has been changed, and will in turn influence the symmetry operators in the manner we have already seen.

Ample opportunity accordingly exists for comparing the wavefunctions, degeneracies, and symmetry groups of harmonic oscillators and hydrogen atoms, even of different dimensions. Somewhat more care is needed in comparing the symmetry groups than the wavefunctions, because of the possibility that anisotropy may produce two interleaving families, or it may result in some states being omitted.

It is in this way that one may more readily explain a recent paper of Ravenhall, Sharp and Pardee.[23] They were interested in a problem in solid state physics which involved solving the hydrogen atom with an impenetrable planar barrier passing through the nucleus. Such a system differs from an ordinary hydrogen atom in that all the states of even parity with respect to reflection in a plane are excluded. It was found that the remaining states showed the degeneracy patterns of, and transformed according to, the group $SU(3)$, rather than the $O(4)$ appropriate to the unrestricted problem. Motivated by this observation, they went on to investigate the bisected harmonic oscillator, and found that curiously enough its wavefunctions transformed according to $O(4)$ rather than the $SU(3)$ of the full oscillator.

It is interesting to note how some of their results are related to anisotropy and how they can be obtained and slightly generalized by our treatment. Returning to the remark at the end of Sec. II, we will now deliberately choose the anisotropy factors α and β of the harmonic oscillator as multiples of their least-term expression. For example, if we take $\alpha = \beta = 2$, we have the isotropic harmonic oscillator, but the ladder operators given by (17) will distinguish four families of states for the four possible combinations of the numbers k and j. Every family forms a basis for the irreducible representations of $SU(2)$, since the fact that one can obtain generators from the ladder operators (17) does not depend on α and β being relatively prime. Two of the four families contain all the states whose first quantum number is even. Eliminating the other two families we get a system equivalent, from the point of view of the group generators, to an isotropic harmonic oscillator with 2:1 frequency ratio. That this separation is possible is just the result given by Ravenhall, Sharp, and Pardee in Sec. A of their paper. Even more generally, if one chooses $\alpha = \beta = m$, the operators (17) will now distinguish m^2 families of states, each forming a basis for the irreducible representations of the group $SU(2)$. Moreover, if we select those states for which n_1 and n_2 are multiples of a and b, respectively, where a and b are relatively prime integers, and form the generators of $SU(2)$ for $\alpha = \beta = ab$ we find that the resulting system is equivalent to an anisotropic harmonic oscillator with $a:b$ frequency ratio. This means that such a separation of the isotropic harmonic-oscillator energy spectrum accounts for the degeneracy of every anisotropic harmonic oscillator with rational frequency relation.

The analysis of Sec. II can easily be generalized to n dimensions. We find that the degeneracy of the anisotropic harmonic oscillator with commensurable frequencies can be accounted for by an $SU(n)$ dynamical symmetry group. Each state will belong to one and only one of several families, each of which forms a basis for the symmetric tensor irreducible representations of $SU(n)$. For $n \geq 3$ it may not be possible to define the operators (17) in such a way that all the states of a given energy level belong to one and the same family. Again, if we consider the isotropic harmonic oscillator and choose $\alpha_1 = \alpha_2 = \cdots = \alpha_n = m$, we find that the n-dimensional version of the operators (17) distinguish m^n families of states, each of which forms a basis for the symmetric tensor irreducible representations of $SU(n)$. An interesting case is the three-dimensional anisotropic harmonic oscillator with 1:1:2 frequency relations, which is equivalent to choosing states from the isotropic harmonic oscillator one of whose quantum numbers is even. Ravenhall, Sharp, and Pardee considered such a system and showed that the states separated into two families forming a basis for the n^2 and $n(n+1)$-dimensional irreducible representations of $O(4)$. From the point of view of our treatment there appear four families of $SU(3)$ degeneracy, the lowest energy levels of which are (0, 0, 0); (1, 0, 0); (0, 1, 0); (1, 1, 0). A state of one of these four families will differ in n_1 and n_2 by an even integer and in n_3 by an integer from its corresponding lowest energy state.

They also showed how one could separate a basis

for the irreducible representations of $O(3)$ into two families forming a basis for the irreducible representations of $SU(2)$. We obtain the same result by constructing the operators (55) for $\gamma = \frac{2}{1}$ and choose $\alpha' = 2$ and $\beta' = 2$, double their least-term expression. This will give ladder operators changing the quantum numbers by 2 units, so that the arrows representing the operators in Fig. 4(a) will have double length. It is clear that the operators will distinguish two families, whose degeneracies correspond to $SU(2)$ irreducible representations, and for which one can get the two sets of generators by applying Dulock's transformation (57).

VII. CONCLUSION

With the consideration of a variety of examples we are able to understand more completely the limitations on the existence of a universal symmetry group. Although we have confined our attention exclusively to two-dimensional examples, the conclusions are quite generally applicable to a system with whatever number of degrees of freedom. To a certain extent, the limitations encountered depend upon the precise formulation of our concept of a "universal symmetry group." In its most restricted sense, one should think of a classically degenerate problem, which means that there exists some system of canonical coordinates in which the Hamilton–Jacobi equation is separable, action-angle variables can be introduced, and there exist rational relationships amongst the resulting frequencies.

Given the restriction to classical degeneracy, the ladder operators introduced by Dulock and McIntosh[9] demonstrably lead to an $SU(n)$ group formed from those constants of the motion which are bilinear combinations of the ladder operators. This group depends only on the number of degrees of freedom, and in that sense is universal for all such classical systems. Fradkin,[7] Mukunda,[8] and others have been more ambitious, claiming that operators very similar in appearance to those of Dulock and McIntosh should generate both $SU(n)$ and $O(n+1)$ groups for all potentials with spherical symmetry. Such arguments were based on the concept of vector or tensor constants of the motion.

The spherically symmetric potentials will, by definition, have an orthogonal group of the appropriate dimension as a symmetry group. Since vectors or tensors are defined by their transformation properties with respect to an orthogonal group, it is then a simple matter to set down a differential equation imposed by these transformation properties under an infinitesimal rotation. Solutions can be shown to exist,[27] involving generally some additional parameters which can be fixed by further considerations governing the nature of the orbits; for instance, the possible existence of circular or linear orbits may further restrict the constants possible. The exceptional role played by the classically degenerate systems—the harmonic oscillator and the Kepler problem—has made itself apparent in these analyses by the fact that they are the only systems for which the vector or tensor solutions are single valued. Presumably this would also mean that they are the only systems for which the additional constants would be acceptable quantum-mechanical operators.

This final and most extensive level of universality has never been carried out explicitly for any specific system, let us say for the two-center Coulomb problem, for which the classical results are known, much less for a system of the generality claimed. Nevertheless the theoretical arguments are sound; in principle any classical problem can be reduced to force-free motion in some appropriate space. The demonstration of this theorem is the culmination of the traditional classical mechanics course. However, two delicate points are generally left open. One relates to the topology of the force-free space, and indeed in the absence of forces, the topological nature of the space is of the utmost importance in determining the type of motion—whether it is bounded, periodic, and so on. The second, even more subtle, point concerns whether the required canonical mapping is a transcendental function or not, a point which has a bearing upon whether it could be realized by noncommuting quantum-mechanical operators.

Once these preliminaries have been conceded, one might for example follow the technique of Mukunda,[8] in which it is shown how the generators of a Euclidean group may be combined to form generators of other groups, such as the unitary or orthogonal groups, in a manner not unlike the constructions of Dulock. When the translation operators of the Euclidean group are the coordinates and momenta of the space of force-free motion, one has achieved his universal symmetry group, in principle.

So much, then, for the levels of universality which one could claim. Since the classical reasoning is valid, one has to seek an explanation of why one does not observe the degeneracy implied by the universal symmetry group in every quantum-mechanical system somewhere in the correspondence between quantum-mechanical and classical operators.

In our present discussion, we have only considered the first of these levels, treating systems which are classically degenerate. Difficulties which

are encountered on this level are bound to persist in the higher levels where more extensive universal groups are sought, and so their adequate understanding is a preliminary to any other investigation, even though the latter may unearth further obstacles to universal symmetry.

There seem to be two kinds of effects at work which are complicated by quantization. One is anisotropy, while the other is the interference between non-Cartesian coordinates. Both seem to imply constraints on possible quantum numbers which are incompatible with the formation of symmetry groups.

Anisotropy is conveniently studied, uncomplicated by other considerations, in the anisotropic plane harmonic oscillator. The degeneracy of the isotropic oscillator has been long known, and corresponds to the various possible ways by which a positive integer may be written as the sum of two others. Bilinear products of the ladder operators may be interpreted as transferring a quantum from one coordinate to the other. They generate the group $SU(2)$ on the one hand, and account for the degeneracy on the other.

In the case of anisotropy, quanta in the two coordinates do not carry the same amount of energy, and so such a transfer can only be understood if several of the one quanta have the same energy as several of the other. This implies a rational frequency ratio, but once again the degeneracy can be explained and an $SU(2)$ group generated. Apparently further details of the symmetry have been overlooked. One of these is that the multiplicity of degeneracy depends upon the residue class of the least common multiple of the two frequencies. In particular, there may be several singly degenerate states. Group theory does not preclude matrix elements in the Hamiltonian between such states, and so we might think that the group accounting for the accidental degeneracy has been incompletely determined. However, it is clear that no product of ladder operators could ever make a transition between these states, even though they have the same symmetry type.

In fact, if we think of the powers of the elementary ladder operators which produce the composite quantum corresponding to their least common multiple as forming derived ladder operators, it is these derived ladder operators which generate our $SU(2)$ group. It is therefore apparent that there exist a number of parallel families of degenerate states, because certain of the bilinear products of the derived ladder operators are the ladder operators which produce the different irreducible representations of $SU(2)$. It therefore makes sense to say that the wavefunctions of an anisotropic oscillator are grouped into k systems of irreducible representations of $SU(2)$, where k is the least common multiple of the frequency ratio.

Although we are unable to produce constants of the motion for an anisotropic harmonic oscillator with an incommensurable frequency ratio, we nevertheless can feel that there is a certain sort of "continuity" in the sense that the more and more discrepant the frequency ratio in an oscillator in which that ratio is rational, the more and more numerous the one-dimensional representations of $SU(n)$ become and the greater and greater becomes the energy gap to the first doubly degenerate state. If it is any reassurance, we may think that in the incommensurable case, the one-dimensional representation occurs with an infinite multiplicity, and that $SU(n)$ is still the symmetry group.

A complementary sort of "continuity" is afforded by the clustering of the nondegenerate states when the frequency ratio differs but little from a rational ratio, but is still incommensurable, because of the closeness of the numbers involved, some combinations of ladder operators may approximately fulfil the commutation rules of a unitary group, and so we may also think in terms of an approximate symmetry group with approximate degeneracies.

In any event it seems that ladder operators are a more fruitful concept than symmetry operators. They can be formed even when quantization conditions impose numerical constraints on the quantum numbers which make it difficult or impossible to form symmetry operators, and constants of the motion. Our first lesson is that when there is anisotropy, which means to say that we are forced to juggle quanta of different sizes, we must expect to find not one but several systems of irreducible representations of the group $SU(n)$, depending on the residue classes of the least compatible quantum.

When one is dealing with non-Cartesian coordinates, an additional complication arises. This comes about because of the structure of the Hamilton–Jacobi equation. In polar coordinates, for example, one finds that the z component of the angular momentum is a constant, which is essentially the content of the ϕ component of the Hamilton–Jacobi equation. One then uses this fact to simplify the theta equation, in three dimensions. In higher dimensions, one can usually obtain a chain of angular equations, such that fixing the variable in one as constant allows the separation of the next, and so on. In two dimensions one arrives at once to the radial equation, whose solution depends on the constant of angular momentum.

This series of separation is paralleled in the separation of the Schrödinger equation, with the result that the theta equation depends on the ϕ quantum number, and the radial equation will depend upon the angular quantum number. As a result there may be constraints on the value of one quantum number imposed by another. Moreover, and equally important, the ladder operator for one set of action-angle variables may involve more than one of the canonical coordinates, due to their involvement through the constants of the motion which allowed the separation.

Again, the simplest instance of this phenomenon uncomplicated by extraneous considerations is to be seen in the separation of the Kepler problem or hydrogen atom in polar coordinates. Since the ladder operators depend on the angle variables, which in turn depend upon which action variables involve the constants of separation, we find that the radial ladder operator only involves the principal quantum number, the theta operator involves both n and l, while the phi ladder operator involves all three, n, l, and m. At least, such is the case in three dimensions, with suitable modification for greater or lesser dimensionality. In our present case we have, for two dimensions,

$$\mathcal{A}_r^{\pm}\Psi_{n,m} = C\Psi_{n\pm1,m},$$

$$\mathcal{A}_\theta^{\pm}\Psi_{n,m} = C'\Psi_{n\pm1,m\pm1}.$$

If we expect to use these ladder operators directly to form an $SU(2)$ group, we must expect the radial ladder operators to commute with the angular ladder operators. However, we first have the restriction that $|m| \leq n$, with a zero wavefunction for those index combinations in which $|m| > n$. Then, we find that $\mathcal{A}_r$ changes n only, while $\mathcal{A}_\theta$ changes both n and m. Consequently, if we apply $\mathcal{A}_r^-\mathcal{A}_\theta^+$ to the state $(n, -n)$, and then apply $\mathcal{A}_\theta^+\mathcal{A}_r^-$ we find respectively

$$\mathcal{A}_r^-\mathcal{A}_\theta^+\Psi_{n,-n} = C''\Psi_{n,-(n-1)},$$

$$\mathcal{A}_\theta^+\mathcal{A}_r^-\Psi_{n,-n} = 0.$$

Consequently, our angular and radial ladder operators do not satisfy the commutation relations which their analogs do in classical mechanics, and thus there is no direct transcription of the $SU(2)$ symmetry group of classical mechanics into an $SU(2)$ quantum-mechanical group. The difficulty is a fundamental one, as can perhaps better be seen in the case of three dimensions, since the groups $SU(3)$ and $O(4)$ are very much more different than $SU(2)$ and $O(3)$ in two dimensions, and the patterns of degeneracies are also markedly different.

It is worth emphasizing that in this case, the pattern of degeneracies produced by the ladder operators is quite correct, and that it is only the commutation rules of $SU(2)$ which fail to be preserved in the quantum-mechanical transition. Moreover, in the classical limit, the commutation rules correspond, as the dimensions of the representations become larger and larger, and the effects at the edge of extreme quantum numbers become of lesser importance.

We have also seen that Dulock's nonlinear transformation, at least in two dimensions, will transform the ladder operators into a set which actually generate the orthogonal group instead of the unitary group.

Once having seen the effects of anisotropy and interference between coordinates, it is only natural to consider their combination as well. The simplest instance is the planar isotropic harmonic oscillator separated in polar coordinates, where we have a 2:1 frequency ratio between the radial and angular frequencies. The frequency ratio 2:1 is one which enjoys a number of exceptional properties, and in fact is responsible for the very close association between the hydrogen atom and harmonic oscillator which has been noted throughout the history of both classical and quantum mechanics. Other frequency ratios can be contrived by introducing a centrifugal potential proportional to the square of the angular momentum, both for the Kepler problem and the harmonic oscillator.

The first effect which we see is that there occur families of degenerate states characteristic of the anisotropy, so that even if a symmetry group were to appear, there would not be just a single assemblage of its irreducible representations. Then we discover that there is not just a single "key" state in the sense of Infeld and Hull, as would be characteristic of a spectrum belonging to $SU(2)$ generated by ladder operators, but that all the states of "extreme" quantum numbers are key states. This situation is connected with the anomalous commutation rules of the ladder operators at the extreme states. However, yet another effect makes its appearance, which is that not all the lattice points which could be generated by the ladder operators are admissible. Formally, the corresponding functions exist, but they may be defective as solutions of Schrödinger's equation in some way—by not being square integrable, or by being multiple valued.

The final results are rather unpredictable. Sometimes, the pattern will be such that another group than the unitary unimodular group will be a symmetry group. In the case of the 2:1 frequency ratio, the

degenerate families may interleave or separate in such an orderly manner, that one may discern representations of $SU(n)$ or $O(n+1)$; indeed in this way one may best understand the traditional relations between these groups, as well as some newly discovered relations. On other occasions, the pattern of degeneracies will be such that it appears to preclude any possible explanation in terms of symmetry groups.

In this investigation, we have not undertaken to understand the fate of the universal symmetry group in those problems which are not classically degenerate, an important example of which is the motion of a particle in the field of two fixed Coulombian centers. On the other hand, we have seen that it can be distorted considerably by the process of quantization in the classically degenerate problems. The general conclusion which we reach is that we might still expect to preserve the construction of ladder operators, and in this way obtain an operator calculus which will account for the observed degeneracies, at least for the familiar separable potentials. However, it does not appear that we can hope to preserve the commutation rules of these ladder operators with one another in different coordinates. The result may be that we could obtain a different symmetry group than the unimodular unitary group which we expect, as was the case for the Kepler problem; or it may be that we could not obtain any finite Lie group at all. Such a conclusion is rather disturbing for those who hope to give a group-theoretical account of all degeneracy and symmetry, but much more so for those of us who had hoped that $SU(n)$ would be that group.

* Based in part upon a professional thesis submitted to the Escuela Superior de Física y Matemáticas of the Instituto Politécnico Nacional in partial fulfilment of the requirements for the title "Licenciado en Física y Matemáticas."

[1] V. Fock, Z. Physik **98**, 145 (1935).

[2] J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).

[3] Yu. N. Demkov, Zh. Eksp. Teor. Fiz. **44**, 2007 (1963) [Sov. Phys.—JETP **17**, 1349 (1963)].

[4] V. A. Dulock and H. V. McIntosh, Am. J. Phys. **33**, 109 (1965).

[5] V. A. Dulock and H. V. McIntosh, J. Math. Phys. **7**, 1401 (1966).

[6] M. H. Johnson and B. A. Lippman, Phys. Rev. **76**, 828 (1949).

[7] D. M. Fradkin, Progr. Theoret. Phys. **37**, 798 (1967).

[8] N. Mukunda, Phys. Rev. **155**, 1383 (1967).

[9] V. A. Dulock and H. V. McIntosh, Pacific J. Math. **19**, 39 (1966).

[10] O. Laporte and G. Y. Rainich, Trans. Am. Math. Soc. **39**, 154 (1936).

[11] L. Susskind and J. Glogower, Physics **1**, 49 (1964).

[12] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, London, 1961), 4th ed., p. 83.

[13] R. Hermann, Commun. Math. Phys. **2**, 78 (1966).

[14] H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand, Princeton, N.J., 1956).

[15] We are indebted to V. A. Dulock for suggesting this representation.

[16] A. Cisneros and H. V. McIntosh, J. Math. Phys. **10**, 277 (1969).

[17] D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1951), pp. 298–304.

[18] S. Bell and P. A. Warsop, J. Mol. Spectry. **20**, 425 (1966).

[19] E. Schrödinger, Proc. Roy. Irish Acad. **46A**, 9 (1940).

[20] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).

[21] A. Wintner, *The Analytical Foundation of Celestial Mechanics* (Princeton University Press, Princeton, N.J., 1947).

[22] D. Bergman and Y. Frishman, J. Math. Phys. **6**, 1855 (1965); A. Giovannini and T. Tonietti, Nuovo Cimento **54A**, 1 (1968).

[23] D. G. Ravenhall, R. T. Sharp, and W. J. Pardee, Phys. Rev. **164**, 1950 (1967).

[24] J. Plebanski, Notes from Lectures on Elementary Quantum Mechanical Operators, Instituto Politécnico Nacional, Mexico (1964).

[25] M. Born, *The Mechanics of the Atom* (Frederick Ungar Publ. Co., New York, 1960), pp. 138–139.

[26] S. P. Alliluev, Zh. Eksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys.—JETP **6**, 156 (1958)].

[27] H. Bacry, H. Ruegg, and J. M. Souriau, Commun. Math. Phys. **3**, 323 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Degeneracy in the Presence of a Magnetic Monopole

HAROLD V. MCINTOSH AND ARTURO CISNEROS*
Escuela Superior de Fisica y Matemáticas, Instituto Politécnico Nacional, México 14, D.F., México

(Received 20 June 1968)

The symmetry of the force field of a magnetic monopole is comparable in its simplicity to that of the hydrogen atom or a harmonic oscillator. Both these latter systems possess a "hidden" symmetry which leads to an "accidental" degeneracy in the energy spectrum of their Schrödinger equations. Since the monopole field is derived from a vector potential which is not symmetric, but undergoes a gauge transformation under rotation, the concepts of symmetry and constants of the motion must be expressed properly in the presence of velocity-dependent forces. It is found that neither the mechanical nor the canonical angular momentum is conserved in the presence of a monopole field, but rather a total angular momentum which incorporates angular momentum resident in the magnetic field. The total angular momentum defines a cone, on whose surface the motion takes place, whatever central electrostatic potential may exist. Neither the harmonic oscillator nor the hydrogen atom retain their accidental degeneracy when the nucleus possesses a magnetic charge, but if a repulsive centrifugal potential proportional to the square of the pole strength is added, accidentally degenerate systems with a higher symmetry result. The symmetry of such a harmonic oscillator is still somewhat obscure, but the "charged Coulombic monopole" has an $O(4)$ symmetry group generated by the total angular momentum together with a Runge vector constructed from the total angular momentum. The irreducible representations of $O(4)$ which occur are not the n^2 representations of the hydrogen atom, but the $m \cdot n$ ($m - n = 2\epsilon$, twice the monopole charge) representations which cannot be realized by four-dimensional spherical harmonics. The magnetic pole strength must be quantized, if admissible solutions of Schrödinger's equation are to exist, and according to Schwinger's quantization ($\epsilon = nhc/e$) if the wavefunctions are to be single valued; in any event, the ground state will be degenerate.

I. INTRODUCTION

Problems involving the presence of accidental degeneracy are widely known and by now fairly well understood, insofar as it has always been possible to account for the additional degeneracy by "hidden symmetries." Thus we recognize that the generators of such symmetries form a larger dynamical group than the geometric symmetry group which was responsible for our belief that the degeneracy was accidental. Such considerations have always been of more concern quantum mechanically than classically because of the influence of a manifold of degenerate states on the appearance of the spectrum of a quantum-mechanical system, not to mention the complications which it produces in the perturbation treatment of degenerate systems. Moreover, as Stehle and Han[1] have shown, there is more to classical degeneracy than the simple existence of a manifold of orbits of equal energy, which is all that is implied by the existence of a symmetry group. It is additionally necessary to pay attention to quantization conditions as exemplified by the Bohr–Sommerfeld procedure in the old quantum mechanics.

This difference has tempered several recent results by diverse authors[2] to the effect that *both* $O(4)$ and $SU(3)$ are classical symmetry groups for *all* single-particle three-dimensional systems. Indeed they are, but the traditional criterion remains that the only degenerate systems are those for which all the bounded orbits are closed. In the case of a central force law in Euclidean space the only potentials showing such a degeneracy are $V(r) = r^2/2$ and $V(r) = -1/r$, corresponding, respectively, to an isotropic harmonic oscillator and motion in a Coulomb field of force. The degeneracy of the Coulomb, or Kepler, problem has been known since the 1935 work of Fock[3] to be due to the symmetry group $O(4)$ for the bound orbits, and by the work of Jauch and Hill[4] and others to be caused by the group $SU(3)$ for the harmonic oscillator.

These results are embodied in a theorem of Bertrand,[5] whose proof is obtained by perturbing one of the circular orbits which may be found in all central-force problems, to see if there will be a nearby closed orbit. Its hypotheses are therefore somewhat restrictive from a physical point of view; in particular, the Lorentz force which is exerted on a moving charged particle is a noncentral velocity-dependent force, to which the theorem does not apply.

The simplest spherically symmetric magnetic-field problem is doubtless that for a magnetic monopole which produces a radial **B** field whose magnitude diminishes with distance as the inverse square, that is,

$$\mathbf{B} = \epsilon \mathbf{r}/r^3.$$

Not only is the investigation of this problem interesting on account of its violating the hypotheses of Bertrand's theorem, but there is also another aspect—that it is necessary to take into account gauge transformations in conjunction with the geometric symmetry operations whenever there are velocity-dependent forces derived from a vector potential.

The theory of the magnetic monopole has a venerable history, perhaps the earliest published treatment being that of Poincaré[6] in 1896, who showed that the typical trajectory of an electron moving in a monopole field would be a geodesic on the surface of a circular cone whose vertex contained the monopole. However, because of its connection with the theory of the aurorae, most physical interest in magnetic problems of this type has centered on the motion of a charged particle in the field of a magnetic dipole, a very much more complicated problem which is not completely separable and which lacks the simple integrals possessed by motion in the monopole field. Nevertheless there have been some recent treatments of the classical monopole problem,[7,8] mostly from the point of view of vectorial or Lagrangian mechanics. Lehnert[9] cites the monopole as an especially straightforward instance of the magnetic mirror effect, including it in a recital of magnetic problems whose partial or complete solutions have been obtained analytically.

Quantum-mechanical interest dates from 1931, the year of appearance of Dirac's paper[10] speculating on the quantum-mechanical restrictions concerning the existence of monopoles, and the resultant necessity for the quantization of electric charge. The Schrödinger equation for the motion of a charged particle in the field of an isolated unchanged monopole was studied in a concomitant paper of Tamm,[11] published shortly thereafter.

The quantum-mechanical literature concerning monopoles can conveniently be divided into three categories. The first group of papers,[12–16] generally older, consists of investigations of the solutions of Schrödinger's equation or Dirac's equation for the monopole, to determine their dependence on the singularity in the vector potential, to determine whether the lack of bound states can be altered by considering the magnetic moment of the electron, and to determine the nature of the solutions for a charged monopole. The second group[17,18] is mostly experimental in nature, being concerned with attempts to discover monopoles in various surroundings—in meteorites, among cosmic rays, or produced by high-energy accelerators, of which all the results to date have been negative. Finally there is a more recent group of field-theoretic[19] papers which either reexamine Dirac's original considerations, or which try to predict scattering and other data which may aid the experimental efforts toward detection.

Since no bound states are expected for the electron–monopole system either classically or quantum mechanically, there has been little systematic discussion of either degeneracy or constants of the motion, with the exception that Fierz[13] has noticed that the mechanical angular momentum $\mathbf{r} \times \boldsymbol{\pi}$ is not conserved due to the presence of the magnetic field, and that instead, a more general quantity $\mathbf{L} - \epsilon\mathbf{r}/r$ is conserved. This quantity is a "total" angular momentum whose components still obey the usual commutation rules, and in terms of whose square the Hamiltonian may be defined. The classical treatments have rarely been given in Hamiltonian terms; hence, while integrals of the motion are known, little attention has been paid to finding symmetry properties or infinitesimal transformations to which they might correspond.

On the other hand, charged monopoles have been found to possess bound states.[16] Their radial wavefunctions are the same as for the ordinary Coulomb problem, with only the angular dependence on the quantum numbers being different with the result that the accidental degeneracy of the hydrogen atom is lost.

Since we are interested in the occurrence of accidental degeneracy, it is interesting to consider certain electrostatic potentials for a charged magnetic monopole which result in degeneracy, even if they might not correspond to a particularly realistic physical system. Two such potentials are

$$V(r) = r^2/2 + \epsilon^2/2r^2, \tag{1}$$

$$V(r) = -1/r + \epsilon^2/2r^2, \tag{2}$$

where ϵ is the pole strength of the monopole. The corresponding Hamiltonians are closely related to the harmonic oscillator and Coulomb problems in four dimensions, and in turn to the spherical top whose connection to the monopole problem has already been noted by several authors. This relationship is established by introducing quaternionic coordinates for the Hamiltonian equations, which in turn allow a Hopf mapping to be made which incorporates the magnetic pole strength as a conserved canonical momentum which must accordingly be quantized to permit integrable eigenfunctions of the Schrödinger equation. One is thereby led in an alternative way to Dirac's conclusion that magnetic pole strength must be quantized.

The most conspicuous feature of the motion of a particle in a monopole field is that as Poincaré showed, it is confined to the surface of a cone whose half-angle diminishes from $\pi/2$ with increasing pole strength, and whose axis is defined by a vector which takes the place of the angular momentum. This remark applies equally well whether there is a central electrical force field present or not. If the cone is allowed to roll without slipping on a plane, the orbital point in contact with the

plane will trace out an orbit corresponding to zero pole strength. Unless there are certain values of the half angle of the cone which correspond to certain combinations of initial conditions and pole strengths, one will find that closed orbits on the cone do not correspond to closed orbits on the plane. Hence potentials showing accidental degeneracy and closed orbits in the absence of a magnetic charge may no longer do so in the presence of the charge, or may do so only for certain combinations of initial conditions and pole strengths. It is therefore seen that the supplementary term $\epsilon^2/2r^2$ produces the proper environment for those potentials which we study, since its effect as seen from Newton's theory of revolving orbits,[20] is always to cause an orbital precession.

We are mostly interested in the symmetries and degeneracies which occur in the presence of a magnetic monopole, so we shall confine our investigation to classical mechanics and nonrelativistic quantum mechanics for which Schrödinger's equation will suffice. From a general point of view, we first study the meaning of symmetry in the presence of a vector potential, as is necessary for the study of any system in which magnetic fields occur. We find that infinitesimal generators of symmetry transformations in Hamiltonian mechanics must be altered in such a way that they will regauge the potentials correctly when they produce displacements in the phase space.

Many properties of the charged monopole can be seen quite clearly by means of ordinary vectorial mechanics, which we use to derive properties of the orbits and constants of the motion. We find that the general characteristic of all central force potentials in which a monopole is present is that the motion of a charged particle is confined to the surface of Poincaré's cone, that Kepler's second law still holds, and that there is a vector constant of the motion, the "total" angular momentum, which defines the axis of the cone. The total angular momentum differs from the mechanical angular momentum, $\mathbf{r} \times \boldsymbol{\pi}$, by just those terms needed to regauge the magnetic vector potential after a rotation.

The two electrostatic potentials of most interest are the harmonic oscillator potential and the Coulomb potential. Their orbits are not closed, but if a supplemental repulsive centrifugal potential is added, proportional to the square of the magnetic charge of the monopole, there result closed orbits. The charged monopole with an augmented Coulomb potential is particularly interesting, because the orbits are planar. Confined to the surface of a cone as well, they are conic sections, but the center of force is no longer at the focus, nor even in the plane of the orbits. However, a second vector constant of the motion entirely analogous to the Runge vector, with the exception that it is formed from the total angular momentum, can be exhibited.

The orbits for a harmonic oscillator potential with a monopole present at the attracting center resemble precessing bent ellipses, and will also be closed when the repulsive centrifugal potential is included. A tensor constant of the motion can be constructed, but eventually its quantum-mechanical properties will be seen to be much less convenient than those of the Runge vector for the charged monopole.

Once the vectorial treatment has made clear the nature of the classical orbits and the constants of the motion, we cast our problems in Hamiltonian form, verifying the relation of the constants of the motion to the requirements imposed by the vector potential. The principal advantage of the Hamiltonian formalism is the ease with which it allows us to formulate our problems in several important coordinate systems, namely polar and parabolic coordinates, ellipsoidal coordinates, and a particular kind of quaternionic coordinates. Since we are dealing with a classically degenerate problem, we can assuredly formulate the universal symmetry group of Dulock and McIntosh, Fradkin, and Mukunda in these coordinate systems, with the object of forming ladder operators to be used in their quantum-mechanical versions.

Again, the most interesting system is the charged monopole, whose symmetry group we find to be $O(4)$, exactly as for the Coulomb problem. However, different families of irreducible representations appear, each with their characteristic multiplicity of degeneracy, depending upon the magnitude of the magnetic pole strength, whose quantization was the result of Dirac's and Tamm's original investigation. None of these are representations realizable in terms of rotations of a hypersphere unless the magnetic charge is zero, so that Fock's stereographic transformation could no longer apply. Moreover, there is an absence of states of low angular-momentum quantum number; exactly how many again depending upon the magnetic charge and reflecting the intrinsic angular momentum of the magnetic field. This lack has its influence both on the spectrum of such a system—the ground state will be degenerate, for example—as well as on its scattering properties.

II. SYMMETRY IN THE PRESENCE OF A VECTOR POTENTIAL

Symmetry in a mechanical system, as it is ordinarily understood, means that if the motion of the system is followed from certain initial conditions, one should

expect to see similar motion develop from any other initial conditions symmetrically situated with respect to the original conditions and for which the same configuration of forces is to be found. To rephrase this consideration in a more precise mathematical form, let us suppose that T is a transformation of phase space, that I is a set of initial conditions, and that $O(I)$ is the solution to the equation of motion M with respect to these initial conditions. Symmetry prevails when $O(T(I)) = T(O(I))$ whenever $T(M) = M$. In more physical terms, symmetry prevails when image orbits proceed from image initial conditions, a result which we expect solely as a consequence of the invariance of the equations of motion. Thus in a central force problem, if we rotate our starting point 90° about the z axis, its trajectory will be just the rotated orbit. But, if we project a charged particle against a uniform electrical field we are likely to see an orbit entirely different from the reflected orbit if we project it along the field, since a reflection perpendicular to its direction is not a symmetry of the electric field.

In Hamiltonian mechanics, the kinetic energy expressed in Cartesian coordinates is invariant with respect to quite general transformations,[21] including translations and rotations. Therefore questions of symmetry depend upon the behavior of the potential energy with respect to various transformations. However, in the case of velocity-dependent forces arising from a vector potential, care must be taken in judging symmetry, on account of the nonuniqueness of the vector potential. In the case of the magnetic monopole, the **B** field is spherically symmetric, being radially directed inward or outward, according to the sign of the pole, with a magnitude proportional to $1/r^2$. Such is not the case for its corresponding vector potential, which in spherical polar coordinates has an azimuthal component proportional to $\tan \frac{1}{2}\theta$, in one commonly used gauge. However, upon rotation the vector potential differs by only a gauge transformation, and as a consequence the force field remains unchanged.

In the presence of a vector potential, the mechanical momentum $\boldsymbol{\pi}$ of a particle is related to its canonical momentum $\mathbf{p}$ by

$$\boldsymbol{\pi} = \mathbf{p} - \mathbf{A}, \tag{3}$$

with vector potential **A**. We shall choose units for the electronic charge e, velocity of light c, and other physical constants which will allow us to avoid distracting coefficients in our formulas. The Hamiltonian is then, using vectorial notation,

$$H = \tfrac{1}{2}(\mathbf{p} - \mathbf{A})^2 + V(r). \tag{4}$$

In transforming configuration space, the vector potential **A** is affected in two ways. First, its components are defined as functions of the coordinates and are therefore modified. Second, it itself is a vector and hence its components are subject to transformation. Properly transformed **A** becomes $T^{-1}(\mathbf{A}[T(r)]) = \mathbf{A}'(r)$. We say that the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ is symmetric if there exists a gauge function ϕ for which

$$\mathbf{A} - \mathbf{A}' = \nabla\phi. \tag{5}$$

Once it is known that there exists such a function ϕ, it is possible to define a generator of Goldstein's[22] type $F_2(q, P, t)$ for a canonical transformation:

$$F_2(q, P) = \sum_{i=1}^{n} T_i(q)P_i + \phi. \tag{6}$$

We then have

$$Q_j = T_j(q) \tag{7}$$

and

$$p_j = \sum_{i=1}^{n} \frac{\partial T_i}{\partial q_j} P_i + \frac{\partial \phi}{\partial q_j}. \tag{8}$$

Supposing that $V(T(r)) = V(r)$ and that T is an orthogonal transformation

$$\begin{aligned} H(q, p, t) &= \tfrac{1}{2}[T^{-1}(P) + \nabla\phi - (T^{-1}A(Q) + \nabla\phi)]^2 + V(r) \\ &= H(Q, P, t). \end{aligned} \tag{9}$$

Therefore, the Hamiltonian is functionally the same in the new coordinates as in the old, and one is assured both that he has carried out a symmetry transformation on the Hamiltonian and that he will see an image orbit in the image coordinates.

The important conclusion to be drawn from this derivation is that the canonical momenta must be modified to compensate for the gauge transformation. The physical significance of this compensation is not hard to discern, because as one might say, the magnetic field also contains momentum and it is the combined momentum which has to be scrutinized in discussing symmetry. Or alternatively, it is not potentials but potential differences which are important, and unless the reference point is also taken into consideration one can hardly expect correct results from a mere change of coordinates.

If we suppose that the transformation T, including the gauge transformation, is an infinitesimal transformation it may presumably be written in the form of a Poisson bracket transformation, and therefore there will be a corresponding constant of the motion. It will be the constant ordinarily associated with the transformation plus a function of the coordinates corresponding to the infinitesimal change of gauge resulting from the regular transformation.

To illustrate these remarks, let us consider the algebraically simpler example of a uniform magnetic field in the direction of the z axis,[23] whose vector potential in the symmetric gauge is

$$\mathbf{A} = \tfrac{1}{2}B(-y, x, 0),$$

yielding the Hamiltonian for a charged particle moving in this field,

$$H = \tfrac{1}{2}(\mathbf{p} - \mathbf{A})^2.$$

We expect translations in all three axial directions to be symmetries. An x translation δx causes a change in the y component of this vector potential equal to δx, and hence a gauge differential of $y\delta x$. To the generator p_x of the x translation we must then add the infinitesimal gauge y, to produce the constant of the motion $(p_x + \frac{1}{2}By)$. Likewise the generator of a y translation becomes $(p_y - \frac{1}{2}Bx)$, but the unchanged p_z generates a z translation. With respect to rotation about the z axis, it may be seen that the change in the components of $\mathbf{A}$ is exactly balanced by the necessary mixing of components, so that $L_z = xp_y - yp_x$, the z component of angular momentum is also a constant of the motion. No gauge transformation results from its usage. In considering the various constants of the motion which we have enumerated it should be borne in mind that the momenta which they contain are canonical momenta, which should be rewritten in terms of the mechanical momenta before comparing the Hamiltonian equations of motion with Lagrangian or vectorial equations.

Not only does the presence of the magnetic field distort the ordinary constants of the motion, both by the addition of the infinitesimal gauge and the usage of the canonical rather than mechanical momentum, but one may very well find that the commutation relations satisfied by the constants in the presence of a field is different from the ones pertaining to the field-free case. For example, the x and y translations no longer commute in our present example of a uniform $\mathbf{B}$ field in the z direction. Rather, the commutation rule is

$$[P_x + \tfrac{1}{2}By, P_y - \tfrac{1}{2}Bx] = B,$$

which depends upon the field strength B. It is readily verified that such a relation is true for any magnetic field $\mathbf{B}$.

Quantum mechanically the complications introduced by the vector potential can be treated somewhat differently, for the change of gauge can be absorbed into the phase of the wavefunction. This consideration actually was the cornerstone of Dirac's original investigation of the magnetic monopole. Given a symmetry group, if the gauge transformation is made without the compensating change in the canonical momenta, there results a projective representation of the symmetry group due to the general unobservability of the phase of a wavefunction, and thus a somewhat different situation from that which prevails classically.

Yet, there is not really such a difference, because representing a coordinate q and its conjugate momentum p_q by the operators q and $-i\partial/\partial q$, we may apply the transformed canonical momentum of Eq. (8) to a wavefunction Ψ to find

$$\left(-i\frac{\partial}{\partial q} - \frac{\partial \phi}{\partial q}\right)\Psi. \tag{10}$$

If we simply apply the differential operator to Ψ whose phase has been modified by the gauge ϕ, we obtain

$$-i\frac{\partial}{\partial q}e^{-i\phi}\Psi = e^{-i\phi}\left(-i\frac{\partial}{\partial q} - \frac{\partial \phi}{\partial q}\right)\Psi, \tag{11}$$

which amounts to the same thing.

The foregoing program gives us a definitive plan of action by which we may find a constant of the motion for the Hamiltonian whenever we are given a symmetry of the fields present, inclusive of magnetic fields. Its essential feature is that the constant of the motion generating the usual symmetry operation must be modified to include the infinitesimal gauge transformation thereby produced, which will result in a suitable change of the canonical momenta to accommodate the change in the vector potential.

These general precepts may be found in the literature, although not always so explicitly stated. For example, Fierz[13] is quite definite that his conserved angular momentum for the magnetic monopole consists of two parts, one mechanical and the other belonging to the field which he explicitly calculates from the Poynting vector. However, Goldhaber[19] is more inclined to treat the field angular momentum as an independent quantity, and seems to be unsure that the classical problem necessarily possesses a conserved total angular momentum. On the other extreme, Wilson[12] once proposed that the angular momentum of the field be arbitrarily quantized, in what would appear to be total disregard of the matter to which the field was coupled. However, the existence of a conserved angular-momentum vector remains a consequence of the spherical symmetry of the system to which it pertains, even though in the presence of magnetic fields it will consist of a mechanical and a gauge part, and not a mechanical part alone.

III. CLASSICAL TREATMENT OF THE MONOPOLE

One may use vectorial methods to obtain a very quick and direct understanding of the motion of a charged particle in the field of a monopole, which will considerably facilitate his understanding of the eventual Hamiltonian formulation of the problem. The constants of the motion can be obtained from vector identities, and in the case of the angular momentum will be the same as that derived from the symmetry of the Hamiltonian by the methods discussed in the last section. It only requires a slight adaptation of the methods applicable to an uncharged monopole to treat the charged monopole, or for that matter, one for which there is any kind of additional central force.

Constants of the monopole were determined already by Poincaré,[6] but we find it convenient to paraphrase Nadeau's[8] derivation in vectorial notation. The first constant is the total angular momentum, which we shall see may be found in the presence of whatever central force to which has been added the Lorentz force of the monopole field. We begin by writing the equation of motion in vector form

$$\frac{d^2\mathbf{r}}{dt^2} = \epsilon \frac{d\mathbf{r}}{dt} \times \frac{\mathbf{r}}{r^3} - \nabla V(r), \tag{12}$$

where ϵ is the monopole strength. Noting that

$$\nabla V(r) = \frac{\partial V}{\partial r}\hat{\mathbf{r}}$$

for unit radial vector $\hat{\mathbf{r}}$, and forming the cross product of $\hat{\mathbf{r}}$ with both members of the equation of motion (12), we obtain

$$\mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} = \epsilon \mathbf{r} \times \left[\frac{d\mathbf{r}}{dt} \times \frac{\mathbf{r}}{r^3}\right],$$

and by expanding the triple cross product, we find that

$$\mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} = \epsilon \frac{d\hat{\mathbf{r}}}{dt}. \tag{13}$$

In arriving at this stage we see that, on account of the cross product, a central force has no influence whatsoever on the subsequent derivation.

Now rewriting slightly the derivative of the cross product, we obtain

$$\mathbf{r} \times \frac{d\mathbf{r}}{dt} = \epsilon\hat{\mathbf{r}} + \mathbf{D}, \tag{14}$$

in which $\mathbf{D}$ is a vector constant, but is not equal to the angular momentum on account of the term $\epsilon\hat{\mathbf{r}}$. Nevertheless,

$$\mathbf{D} \cdot \hat{\mathbf{r}} = -\epsilon, \tag{15}$$

which shows that the position vector $\mathbf{r}$ always makes a constant angle with the vector $\mathbf{D}$, which means that the orbit lies on the surface of a right-circular cone. When the pole strength is zero, $\mathbf{L}$ becomes the angular momentum, and the cone, whose half angle becomes $\pi/2$, has degenerated into a plane. Thus, we have as a limiting case the familiar orbits of ordinary central-force problems.

By means of the two purely vectorial identities

$$\mathbf{D} = (\mathbf{D} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} + \hat{\mathbf{r}} \times (\mathbf{D} \times \hat{\mathbf{r}})$$

and

$$\mathbf{r} \times \frac{d\mathbf{r}}{dt} = r^2\hat{\mathbf{r}} \times \frac{d\hat{\mathbf{r}}}{dt},$$

we can rewrite the defining equation (15) for $\mathbf{D}$ in the form

$$\hat{\mathbf{r}} \times \left[\frac{d\hat{\mathbf{r}}}{dt} - \frac{\mathbf{D}}{r^2} \times \hat{\mathbf{r}}\right] = 0,$$

which implies that

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{\mathbf{D}}{r^2} \times \hat{\mathbf{r}}. \tag{16}$$

Since this is the formula describing the rate of change of a vector subject to rotation, we conclude that the orbital point rotates about the axis of its cone with an angular velocity $|\mathbf{D}|/r^2$, which depends upon the constant magnitude of $\mathbf{D}$ and varies inversely as the square of the radius. This is thus a version of Kepler's second law which applies in the presence of a monopole field, and reduces to the traditional result for zero monopole strength.

Since one has, setting $\mathbf{L} = \mathbf{r} \times \boldsymbol{\pi}$, the mechanical angular momentum

$$\mathbf{L} = \epsilon\hat{\mathbf{r}} + \mathbf{D}, \quad \mathbf{D} \cdot \hat{\mathbf{r}} = -\epsilon,$$

he finds

$$\mathbf{L}^2 = \mathbf{D}^2 - \epsilon^2, \tag{17}$$

which is a further general result for the monopole field, that the magnitude of the mechanical angular momentum is constant, and only its direction changes.

A further constant of the uncharged monopole, the magnitude of the particle velocity, is no longer a constant in the presence of electrical forces, its constancy having been due to the fact that the Lorentz force is always perpendicular to the velocity.

Further results may be obtained when one knows the specific form of the radial forces present; two cases will interest us, that of the charged monopole for which there is a Coulomb electric field, and that of a monopole with a harmonic potential. To these potentials it is convenient to add an inverse square potential, not because we know of any plausible physical origin for such a term, but because it provides

a system with closed orbits and hence one which is relevant to the study of accidental degeneracy. Mathematically we see that this term combines with the Lorentz force in the equation of motion to add a term $\epsilon\mathbf{D}/r^3$ to the remaining force; it therefore represents a contribution having a constant direction. Thus let us write $V(r) = U(r) + \epsilon^2/2r^2$. Including it in the equation of motion produces

$$\frac{d^2\mathbf{r}}{dt^2} = \epsilon \frac{d\mathbf{r}}{dt} \times \frac{\mathbf{r}}{r^3} + \epsilon^2 \frac{\hat{\mathbf{r}}}{r^3} - \nabla U(r), \tag{18}$$

and finally, making the combination which we have mentioned,

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\epsilon\mathbf{D}}{r^3} - \nabla U(r). \tag{19}$$

Let us first investigate in detail the Coulomb case, with $U(r) = -1/r$. The equation of motion is then

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\epsilon\mathbf{D}}{r^3} - \frac{\hat{\mathbf{r}}}{r^2}. \tag{20}$$

Were it not for the presence of the magnetic charge, it is known that the Coulomb potential has in addition to the angular momentum a second vector constant of the motion, the Runge vector

$$\mathbf{R} = \mathbf{D} \times \frac{d\mathbf{r}}{dt} + \hat{\mathbf{r}}. \tag{21}$$

We might surmise that there would be a similar constant even in the presence of a magnetic charge, but that the angular momentum entering into its definition would be the total rather than the mechanical angular momentum. To verify this supposition, we calculate

$$\frac{d}{dt}\left(\mathbf{D} \times \frac{d\mathbf{r}}{dt}\right) = \mathbf{D} \times \frac{d^2\mathbf{r}}{dt^2}$$

into which we substitute the equation of motion to find, after using (16),

$$\frac{d}{dt}\left(\mathbf{D} \times \frac{d\mathbf{r}}{dt}\right) = -\frac{d\hat{\mathbf{r}}}{dt}$$

and thus finally to conclude that

$$\frac{d}{dt}\left[\mathbf{D} \times \frac{d\mathbf{r}}{dt} + \hat{\mathbf{r}}\right] = 0, \tag{22}$$

which leaves us with the Runge vector as a constant of the motion when it is defined in terms of the *total* angular momentum.

We may obtain further insight into the motion of a charged particle in this field by computing $(d\mathbf{r}/dt) \times (d^2\mathbf{r}/dt^2)$, which is proportional to the binormal to the particle trajectory. We have

$$\frac{d\mathbf{r}}{dt} \times \frac{d^2\mathbf{r}}{dt^2} = -\frac{\epsilon}{r^3}\frac{d\mathbf{r}}{dt} \times \mathbf{D} - \frac{1}{r^2}\frac{d\mathbf{r}}{dt} \times \hat{\mathbf{r}}.$$

The two terms on the right each lack the same multiple of $\mathbf{r}$ to correspond to the constants $\mathbf{R}$ and $\mathbf{D}$, respectively, whereupon we write

$$\frac{d\mathbf{r}}{dt} \times \frac{d^2\mathbf{r}}{dt^2} = \frac{\mathbf{D} + \epsilon\mathbf{R}}{r^3}. \tag{23}$$

It therefore appears that the binormal has a constant direction, which shows that the orbit is confined to a plane, although it is not a plane which contains the monopole itself. Nevertheless it appears that the orbit is always a conic section, whatever the monopole strength.

If we compute the inner product $\mathbf{D} \cdot \mathbf{R}$, we obtain the result $-\epsilon$ which reduces to zero only in the case $\epsilon = 0$ or $e = 0$. Only in those two extreme cases will the Runge vector be perpendicular to the total angular momentum.

Although the plane formed by the total angular momentum and the Runge vector contains many vectors of interest, such as the focal point of the orbit, the vector to the perihelion, and so on, we have not found that any of these quantities has a simple algebraic expression in terms of the constants.

Now we turn our attention to the harmonic-oscillator-like case, in which we take $U(r) = \frac{1}{2}r^2$. Our equation of motion becomes

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\epsilon\mathbf{D}}{r^3} - \mathbf{r}. \tag{24}$$

In the absence of the monopole field, the harmonic oscillator possesses a "tensor" constant of the motion which can readily be expressed in dyadic form. With $\epsilon = 0$, the equation of motion of the harmonic oscillator is

$$\frac{d^2\mathbf{r}}{dt^2} = -\mathbf{r}. \tag{25}$$

We multiply this equation first by $d\mathbf{r}/dt$ on the left, and then on the right, to obtain

$$\frac{d\mathbf{r}}{dt}\frac{d^2\mathbf{r}}{dt^2} = -\frac{d\mathbf{r}}{dt}\mathbf{r},$$

$$\frac{d^2\mathbf{r}}{dt^2}\frac{d\mathbf{r}}{dt} = -\mathbf{r}\frac{d\mathbf{r}}{dt}.$$

Recognition of the derivative of a dyadic product means that if we add these two equations we obtain

$$\frac{d}{dt}\left(\frac{d\mathbf{r}}{dt}\frac{d\mathbf{r}}{dt} + \mathbf{r}\mathbf{r}\right) = 0, \tag{26}$$

which states that the dyadic

$$\frac{d\mathbf{r}}{dt}\frac{d\mathbf{r}}{dt} + \mathbf{rr}$$

is a constant. Moreover, it is easily seen that

$$\frac{d}{dt}\left(\mathbf{r}\frac{d\mathbf{r}}{dt} - \frac{d\mathbf{r}}{dt}\mathbf{r}\right) = 0 \tag{27}$$

since the components of this antisymmetric dyadic are just the components of angular momentum. As a result these two dyadic constants of the motion can be combined to form a single complex constant of the motion,

$$\left(\frac{d\mathbf{r}}{dt} + i\mathbf{r}\right)\left(\frac{d\mathbf{r}}{dt} - i\mathbf{r}\right) = K, \tag{28}$$

whose nine components include the total energy and the generators of the group $SU(3)$ which is the known symmetry group of the isotropic harmonic oscillator.

An attempt to repeat the derivation when the magnetic charge is not zero fails on account of the unsymmetrical position which the term $\epsilon\mathbf{D}/r^3$ occupies in the derivation. Nevertheless the constancy of $\mathbf{D}$ may be exploited to obtain a dyadic constant of the motion, since the equation of motion may be projected onto a plane perpendicular to $\mathbf{D}$. The equation of motion in this perpendicular plane is precisely that of a two-dimensional harmonic oscillator, and may be solved accordingly. This solution determines a unique trajectory on the surface of the cone to which the orbit is confined by counterprojection along the axis of the cone.

The simplest operator to be used for this purpose (which is a combination of a projection and a 90° rotation about the projection axis) is the dyadic corresponding to the vector product with respect to $\mathbf{D}$:

$$\frac{d^2}{dt^2}\mathbf{D} \times \mathbf{r} = -\mathbf{D} \times \mathbf{r}, \tag{29}$$

where $\mathbf{u} = \mathbf{D} \times \hat{\mathbf{r}}$. We may then apply our previous derivation to the vector $\mathbf{u}$ in place of the vector $\hat{\mathbf{r}}$, to obtain a dyadic constant of the motion

$$\left(\frac{d\mathbf{u}}{dt} + i\mathbf{u}\right)\left(\frac{d\mathbf{u}}{dt} - i\mathbf{u}\right). \tag{30}$$

These vectorial methods are somewhat more convenient for obtaining constants of the motion, since they are free from the complications imposed by the use of a vector potential and its accompanying gauge transformations. With this preparation, we may now formulate the Hamiltonian version of the theory, in order to then obtain the Schrödinger equation for the quantum-mechanical versions of the problem.

IV. THE MONOPOLE PROBLEM IN HAMILTONIAN FORM

In order to apply Hamiltonian mechanics to problems involving the field of a magnetic monopole, it is necessary to find a vector potential $\mathbf{A}$ from which its magnetic field $\mathbf{B}$ may be derived through the relation $\mathbf{B} = \nabla \times \mathbf{A}$. For this purpose, the vector potential,

$$\mathbf{A} = \epsilon\left(\frac{yz}{(x^2 + y^2)r}, -\frac{xz}{(x^2 + y^2)r}, 0\right), \tag{31}$$

is convenient, although it differs by the gauge ϵ arc tan $(y/x) = \epsilon\phi$ from the one most often used in the literature,

$$\mathbf{A}' = \epsilon\left(-\frac{y}{r(r + z)}, \frac{x}{r(r + z)}, 0\right).$$

This latter, in spherical polar coordinates, assumes the form

$$A_r' = 0, \quad A_\theta' = 0, \quad A_\phi' = (\epsilon/r)\tan\tfrac{1}{2}\theta.$$

If we write the Hamiltonian in its customary form

$$H = \tfrac{1}{2}(\mathbf{p} - \mathbf{A})^2 + V(r) \tag{32}$$

and select the vector potential $\mathbf{A}$ above, it becomes

$$H = \frac{1}{2}\left[\left(p_x - \frac{\epsilon xz}{(x^2 + y^2)r}\right)^2 + \left(p_y + \frac{\epsilon yz}{(x^2 + y^2)r}\right)^2 + p_z^2\right] + V(r)$$

or, on expansion,

$$H = \frac{p^2}{2} - \frac{\epsilon z}{(x^2 + y^2)r}(yp_x - xp_y) + \frac{\epsilon^2 z^2}{2(x^2 + y^2)r^2} + V(r). \tag{33}$$

This latter form suggests the introduction of polar coordinates. The required transformation equations are

$$\begin{aligned} x &= r\sin\theta\cos\phi, \\ y &= r\sin\theta\sin\phi, \\ z &= r\cos\theta, \end{aligned} \tag{34}$$

$$\begin{aligned} p_x &= \sin\theta\cos\phi p_r + \frac{\cos\theta\cos\phi}{r}p_\theta - \frac{\sin\phi}{r\sin\theta}p_\phi, \\ p_y &= \sin\theta\sin\phi p_r + \frac{\cos\theta\sin\phi}{r}p_\theta + \frac{\cos\phi}{r\sin\theta}p_\phi, \\ p_z &= \cos\theta p_r - \frac{\sin\theta}{r}p_\theta. \end{aligned} \tag{35}$$

Under this substitution the Hamiltonian becomes

$$H = \frac{1}{2}\left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{(p_\phi + \epsilon\cos\theta)^2}{r^2\sin^2\theta}\right) + V(r). \tag{36}$$

Even though this Hamiltonian is not spherically symmetrical, as can be seen most easily by examining its form in rectangular Cartesian coordinates, the force field from which it arises possesses such symmetry. We must therefore apply the procedures discussed in the second section. Knowing that the vector potential changes by a gauge transformation upon rotation, which cannot affect the equations of motion, the proper procedure is to add to the constant of the motion generating an infinitesimal rotation the infinitesimal gauge change produced by the rotation. The resulting generator will define a canonical transformation with an increment in the new canonical momentum just adequate to absorb the change in the vector potential due to the change in the gauge.

If that procedure is applied in the present instance, we first take the *canonical* angular-momentum components as the generators of infinitesimal rotations about the various coordinate axes. These components are then adjusted by adding a function whose gradient is the ensuing change in the vector potential. Calculation then produces the following constants of the motion:

$$D_x = l_x - \epsilon xr/(x^2 + y^2), \tag{37a}$$

$$D_y = l_y - \epsilon yr/(x^2 + y^2), \tag{37b}$$

$$D_z = l_z, \tag{37c}$$

where l_i is a component of the nonconserved *canonical* angular momentum.

It may be verified by explicit calculation that these three constants likewise obey the Poisson-bracket relations satisfied by the angular-momentum components: $\{D_x, D_y\} = D_z$, and so on, cyclically. They may legitimately be regarded as components of the *total* angular momentum, which is neither the same as the mechanical angular momentum nor the canonical angular momentum. Nevertheless, it is related to these quantities; if we recall the relationship between the mechanical linear momentum $\boldsymbol{\pi}$ and the canonical linear momentum $\mathbf{p}$, which is

$$\pi_x = p_x - \epsilon yz/(x^2 + y^2)r, \tag{38a}$$

$$\pi_y = p_y + \epsilon xz/(x^2 + y^2)r, \tag{38b}$$

$$\pi_z = p_z, \tag{38c}$$

it is again a matter of calculation to verify that

$$D_x = L_x - \epsilon x/r, \tag{39a}$$

$$D_y = L_y - \epsilon y/r, \tag{39b}$$

$$D_z = L_z, \tag{39c}$$

which is precisely the definition of the total angular momentum which we obtained from our vectorial analysis of the monopole problem.

Expressed in spherical coordinates, the components of the total angular momentum become

$$D_x = -\sin\phi p_\theta - \cos\phi(p_\theta \cot\theta + \epsilon/\sin\theta), \tag{40a}$$

$$D_y = \cos\phi p_\theta - \sin\phi(p_\phi \cot\theta + \epsilon/\sin\theta), \tag{40b}$$

$$D_z = p_\phi. \tag{40c}$$

The Hamiltonian itself may be expressed in terms of the total angular momentum, the result being

$$H = \tfrac{1}{2}p_r^2 + (\mathbf{D}^2 - \epsilon^2)/2r^2 + V(r). \tag{41}$$

Once again it appears to be convenient to write the potential energy in the form $V(r) = U(r) + \epsilon^2/2r^2$, giving the Hamiltonian the form

$$2H = p_r^2 + \mathbf{D}^2/r^2 + 2U(r). \tag{42}$$

In addition to the general spherical symmetry of all potentials of the form $V(r)$ in the presence of the monopole field, and to which the constancy of the total angular momentum is due, we are interested in the particular cases for which $U(r) = -1/r$ and $U(r) = r^2/2$. We have already seen that the Runge vector based on the total angular momentum in the case of the "Coulombic" potential $-1/r$,

$$\mathbf{R} = \mathbf{D} \times \boldsymbol{\pi} + \hat{\mathbf{r}}, \tag{43}$$

is a constant of the motion; but we need to express it in terms of the canonical coordinates and momenta. In those terms it becomes

$$\mathbf{R} = \mathbf{D} \times (\mathbf{p} - \mathbf{A}) + \hat{\mathbf{r}}. \tag{44}$$

With respect to the Poisson-bracket operation, the components of the vector $\mathbf{D}$ and $\mathbf{R}$ satisfy the commutation rules for the generators of the group $O(4)$, just as they do in the absence of a magnetic field, and hence we may conclude that we still have the same $O(4)$ group of symmetries even in the presence of the monopole field with repulsive centrifugal potential.

Turning to the harmonic-oscillator potential $r^2/2$, we have seen that the dyadic

$$\left(\frac{d\mathbf{u}}{dt} + i\mathbf{u}\right)\left(\frac{d\mathbf{u}}{dt} - i\mathbf{u}\right), \quad \text{where } \mathbf{u} = \left(1 - \frac{\mathbf{DD}}{\mathbf{D}\cdot\mathbf{D}}\right)\cdot\mathbf{r} \tag{45}$$

is a constant of the motion, and that we may even verify that its factors,

$$\frac{du_i}{dt} \pm iu_i$$

are eigenfunctions of the Hamiltonian, with respect to the operation of forming the Poisson bracket. Unfortunately the components of these vectors fail to satisfy the commutation rules of $SU(3)$, even when

$\epsilon = 0$. The difficulty is due to the presence of the projection operator in the definition of $\mathbf{u}$, whose functional dependence on the canonical coordinates complicates the Poisson brackets among the various vector or dyadic components. As it turns out, the analysis of the symmetry of the harmonic monopole is much more complex than that for the Coulombic monopole.

The monopole Hamiltonian written in spherical coordinates bears a very strong resemblance to the Hamiltonian of the spherical top, and suggests the introduction of quaternionic (or Euler-angle) coordinates. These are defined by[24]

$$\begin{aligned} \xi &= r \sin \tfrac{1}{2}\theta \cos \tfrac{1}{2}(\phi - \psi), \\ \eta &= r \sin \tfrac{1}{2}\theta \sin \tfrac{1}{2}(\phi - \psi), \\ \zeta &= r \cos \tfrac{1}{2}\theta \sin \tfrac{1}{2}(\phi + \psi), \\ \lambda &= r \cos \tfrac{1}{2}\theta \cos \tfrac{1}{2}(\phi + \psi) \end{aligned} \tag{46}$$

with the corresponding momenta

$$\begin{aligned} p_\xi &= \sin\frac{\theta}{2}\cos\frac{\phi-\psi}{2}\,p_r + \frac{2p_\theta}{r}\cos\frac{\theta}{2}\cos\frac{\phi-\psi}{2} \\ &\quad + \frac{(p_\psi - p_\phi)}{r\sin\theta/2}\sin\frac{\phi-\psi}{2}, \\ p_\eta &= \sin\frac{\theta}{2}\sin\frac{\phi-\psi}{2}\,p_r + \frac{2p_\theta}{r}\cos\frac{\theta}{2}\sin\frac{\phi-\psi}{2} \\ &\quad - \frac{(p_\psi - p_\phi)}{r\sin\theta/2}\cos\frac{\phi-\psi}{2}, \\ p_\zeta &= \cos\frac{\theta}{2}\sin\frac{\phi+\psi}{2}\,p_r - \frac{2p_\theta}{r}\sin\frac{\theta}{2}\sin\frac{\phi+\psi}{2} \\ &\quad + \frac{(p_\psi + p_\phi)}{r\cos\theta/2}\cos\frac{\phi+\psi}{2}, \\ p_\lambda &= \cos\frac{\theta}{2}\cos\frac{\phi+\psi}{2}\,p_r - \frac{2p_\theta}{r}\sin\frac{\theta}{2}\cos\frac{\phi+\psi}{2} \\ &\quad - \frac{(p_\psi + p_\phi)}{r\cos\theta/2}\sin\frac{\phi+\psi}{2}. \end{aligned} \tag{47}$$

Given these coordinates, which are appropriate to a four-dimensional space, and the Hamiltonian

$$H' = \tfrac{1}{2}(p_\xi^2 + p_\eta^2 + p_\zeta^2 + p_\lambda^2) + U(r), \tag{48}$$

we find that this Hamiltonian becomes, after substitution of the three-dimensional coordinates and momenta,

$$\begin{aligned} H' = \frac{1}{2}\Bigg(p_r^2 + \frac{4p_\theta^2}{r^2} \\ + \frac{4(p_\phi - p_\psi\cos\theta)^2}{r^2\sin^2\theta} + \frac{4p_\psi^2}{r^2}\Bigg) + U(r). \end{aligned} \tag{49}$$

Unfortunately, a discrepancy appears in the form of a factor of 4 multiplying that part of the Hamiltonian involving the angular coordinates and momenta but not the radial part, when this Hamiltonian is compared to that of the monopole problem. There is no clear way to remove this factor by means of a canonical transformation, and consequently the use of quaternionic coordinates is helpful only in treating the angular part of the Hamiltonian. The psi-momentum takes the place of the monopole charge, and becomes an argument in the quantum-mechanical treatment of the problem in favor of the quantization of the magnetic pole strength.

In the case of the spherical top, one imposes the constraint of constant radius, and so the factor of 4 in the angular parts of the three- or four-dimensional Hamiltonians causes no problem. However, in the present case the mapping between the two systems is a Hopf mapping, in which the ψ coordinate is lost, and the ψ momentum remains as a constant in the reduced Hamiltonian. This discrepancy is unfortunate, because we thereby lose the constants of the motion of the four-dimensional systems which do not depend on purely angular variables. This is even true if we try to select those constants (which are abundant since the four-dimensional systems of most interest to us are, respectively, those of force-free motion, the Kepler problem, or the harmonic oscillator) which commute in the Poisson-bracket sense with *both* the Hamiltonian and the ψ momentum.

In fact, the Hamiltonian in quaternionic coordinates to which the monopole problems correspond is

$$H = \frac{1}{2}\left(p^2 - \frac{3}{4}\frac{L^2}{r^2}\right) + U(r), \tag{50}$$

where

$$\begin{aligned} p^2 &= p_\xi^2 + p_\eta^2 + p_\zeta^2 + p_\lambda^2, \\ r^2 &= \xi^2 + \eta^2 + \zeta^2 + \lambda^2, \\ L^2 &= \sum_{i \neq j}(p_i j - i p_j)^2, \quad i, j = \xi, \eta, \zeta, \lambda. \end{aligned} \tag{51}$$

The appropriate four-dimensional problem to consider is then one with the same central force which interests us in three dimensions, but with "excess" angular momentum. Such systems differ from those with the conventional angular momentum in that the orbits precess in the plane of the angular momentum (which is a tensor in higher dimensions) at a rate depending on the angular-momentum surplus. We have discussed this effect at some length in an earlier paper.[25] From this vantage point we can already predict that the Coulombic monopole problem will possess a tractable symmetry group, but that the symmetry of the harmonic monopole will remain rather obscure.

The Coulombic monopole is so similar to the hydrogen atom, that it is worthwhile to exhibit its Hamiltonian in two more coordinate systems, in which it, as well as the hydrogen atom, is separable. These are parabolic and ellipsoidal coordinates. Treating first parabolic coordinates, we introduce their definitions[26]:

$$\begin{aligned} x &= (\mu\nu)^{\frac{1}{2}} \cos\phi, \\ y &= (\mu\nu)^{\frac{1}{2}} \sin\phi, \\ z &= (\mu - \nu)/2. \end{aligned} \tag{52}$$

Since we again deal with a canonical point transformation, the momenta may be readily obtained from a generating function of Goldstein's type F_2,[22] and are

$$\begin{aligned} p_x &= \frac{2(\mu\nu)^{\frac{1}{2}} \cos\phi}{\mu + \nu} p_\mu + \frac{2(\mu\nu)^{\frac{1}{2}} \cos\phi}{\mu + \nu} p_\nu - \frac{\sin\phi}{(\mu\nu)^{\frac{1}{2}}} p_\phi, \\ p_y &= \frac{2(\mu\nu)^{\frac{1}{2}} \sin\phi}{\mu + \nu} p_\mu + \frac{2(\mu\nu)^{\frac{1}{2}} \sin\phi}{\mu + \nu} p_\nu + \frac{\cos\phi}{(\mu\nu)^{\frac{1}{2}}} p_\phi, \\ p_z &= \frac{2\mu}{\mu + \nu} p_\mu - \frac{2\nu}{\mu + \nu} p_\nu. \end{aligned} \tag{53}$$

Substituting the parabolic coordinates and momenta, the Hamiltonian in parabolic coordinates is readily found to be

$$\begin{aligned} H = [&4\mu p_\mu^2 + 4\nu p_\nu^2 + (p_\phi^2 + \epsilon^2)(\mu + \nu)/\mu\nu \\ &+ 2\epsilon p_\phi(\mu - \nu)/\mu\nu - 4]/2(\mu + \nu). \end{aligned} \tag{54}$$

It can be rearranged in a form convenient for separation of the variables,

$$\begin{aligned} &[4\mu p_\mu^2 + (p_\phi - \epsilon)^2/\mu - 2\mu H] \\ &\quad = -[4\nu p_\nu^2 + (p_\phi + \epsilon)^2/\nu - 2\nu H] + 4. \end{aligned} \tag{55}$$

It is rather interesting to notice that the only trace of the magnetic pole strength in this Hamiltonian is in the constant added to the z component of the angular momentum.

Prolate ellipsoidal coordinates are the natural coordinate system to use in treating a two-center problem, because the two attracting centers may occupy the foci of the coordinate ellipses and hyperbolas. In this way we can treat the problem of the motion of a charged particle in the field of two charged magnetic monopoles. Again, to obtain a separable problem, it is necessary to include the repulsive centrifugal potential along with the attractive Coulomb potential. Since the vector potential is additive in the same manner as scalar potentials, we may introduce two vector potentials, one for each of two charged monopoles situated at the foci of an ellipsoidal coordinate system. We shall take the foci to be the points $(0, 0, d)$ and $(0, 0, -d)$, so that the internuclear separation is $2d$ along the z axis. We then have

$$\mathbf{A}_1 = \epsilon_1\left(\frac{y(z + d)}{(x^2 + y^2)\,|\mathbf{r} + \mathbf{d}|}, \frac{-x(z + d)}{(x^2 + y^2)\,|\mathbf{r} + \mathbf{d}|}, 0\right), \tag{56a}$$

$$\mathbf{A}_2 = \epsilon_2\left(\frac{y(z - d)}{(x^2 + y^2)\,|\mathbf{r} - \mathbf{d}|}, \frac{-x(z - d)}{(x^2 + y^2)\,|\mathbf{r} - \mathbf{d}|}, 0\right). \tag{56b}$$

The ellipsoidal coordinates are defined by[27]

$$\begin{aligned} &\xi = (r_1 + r_2)/2d, \quad r_1 = (x^2 + y^2 + (z + d)^2)^{\frac{1}{2}}, \\ &\eta = (r_1 - r_2)/2d, \quad r_2 = (x^2 + y^2 + (z - d)^2)^{\frac{1}{2}}, \\ &\phi = \arctan y/x. \end{aligned} \tag{57}$$

The kinetic energy expressed in ellipsoidal coordinates is

$$\begin{aligned} \frac{1}{2m}\mathbf{p}^2 = &\frac{1}{2m\,d^2}\frac{1}{\xi^2 + \eta^2} \\ &\times \left[(\xi^2 - 1)p_\xi^2 + (1 - \eta^2)p_\eta^2 \right. \\ &\left. + \left(\frac{1}{\xi^2 - 1} + \frac{1}{1 - \eta^2}\right)\right]p_\phi^2. \end{aligned} \tag{58}$$

The two vector potentials in ellipsoidal coordinates are given by

$$\begin{aligned} \mathbf{A}_1 = \epsilon_1\Bigg(&\frac{y\,d(\xi\eta + 1)}{d^2(\xi^2 - 1)(1 - \eta^2)\,d(\xi + \eta)}, \\ &\frac{-x\,d(\xi\eta + 1)}{d^2(\xi^2 - 1)(1 - \eta^2)\,d(\xi + \eta)}, 0\Bigg), \end{aligned} \tag{59a}$$

$$\begin{aligned} \mathbf{A}_2 = \epsilon_2\Bigg(&\frac{y\,d(\xi\eta - 1)}{d^2(\xi^2 - 1)(1 - \eta^2)\,d(\xi - \eta)}, \\ &\frac{-x\,d(\xi\eta - 1)}{d^2(\xi^2 - 1)(1 - \eta^2)\,d(\xi - \eta)}, 0\Bigg). \end{aligned} \tag{59b}$$

Calculating $\mathbf{p} \cdot \mathbf{A}$ first in Cartesian coordinates and then passing to ellipsoidal coordinates, we find that

$$\begin{aligned} \mathbf{p} \cdot \mathbf{A} = \frac{p_\phi}{d^2}\Bigg[&(\epsilon_1 + \epsilon_2)\frac{\eta}{(1 - \eta^2)(\xi^2 - \eta^2)} \\ &+ (\epsilon_1 - \epsilon_2)\frac{\xi}{(\xi^2 - 1)(\xi^2 - \eta^2)}\Bigg]. \end{aligned} \tag{60}$$

Finally, we have for $\mathbf{A}^2$,

$$\begin{aligned} \mathbf{A}^2 = &-\left[\frac{\epsilon_1^2}{r_1^2} + \frac{\epsilon_2^2}{r_2^2}\right] \\ &+ \frac{(\epsilon_1^2 + \epsilon_2^2)(\xi^2 - \eta^2) + 2\epsilon_1\epsilon_2(\xi^2\eta^2 - 1)}{d^2(\xi^2 - 1)(1 - \eta^2)(\xi^2 - \eta^2)}. \end{aligned} \tag{61}$$

The contribution of the two repulsive centrifugal potentials is

$$\frac{1}{2}\left[\frac{\epsilon_1^2}{r_1^2}+\frac{\epsilon_2^2}{r_2^2}\right], \tag{62}$$

and the Coulomb potential is

$$\begin{aligned} U &= \frac{Z_1 e}{d(\xi+\eta)}+\frac{Z_2 e}{d(\xi-\eta)} \\ &= \frac{1}{2m\,d^2}\frac{1}{\xi^2-\eta^2} \\ &\quad\times[\xi 2m\,de(Z_1+Z_2)+\eta 2m\,de(Z_2-Z_1)]. \end{aligned} \tag{63}$$

Finally, combining all these terms according to the formula

$$H=\frac{1}{2m}\mathbf{p}^2-\frac{e}{2mc}\mathbf{p}\cdot\mathbf{A}+\frac{e^2}{2mc^2}\mathbf{A}^2+V, \tag{64}$$

we obtain the Hamiltonian for a pair of charged Coulombic monopoles in ellipsoidal coordinates:

$$\begin{aligned} H &= \frac{1}{2m\,d^2}\frac{1}{\xi^2-\eta^2} \\ &\quad\times\left[(\xi^2-1)p_\xi^2+(1-\eta^2)p_\eta^2+\frac{(p_\phi+\epsilon_-\xi)^2}{\xi^2-1}\right. \\ &\quad\left.+\frac{(p_\phi+\epsilon_+\eta)^2}{1-\eta^2}+Z_+\xi-Z_-\eta+k\right]. \end{aligned} \tag{65}$$

It can be rearranged in a form convenient for separation of the variables:

$$\begin{aligned} &\xi^2 h-(\xi^2-1)p_\xi^2-\frac{(p_\phi+\epsilon_-\xi)^2}{\xi^2-1}-Z_+\xi \\ &\quad=k+\eta^2 h+(1-\eta^2)p_\eta+\frac{(p_\phi+\epsilon_+\eta)^2}{1-\eta^2}-Z_-\eta. \end{aligned} \tag{66}$$

We have used the abbreviations

$$\begin{aligned} h &= 2m\,d^2H, \\ \epsilon_+ &= e(\epsilon_1+\epsilon_2)/c, \\ \epsilon_- &= e(\epsilon_1-\epsilon_2)/c, \\ Z_+ &= 2m\,de(Z_1+Z_2), \\ Z_- &= 2m\,de(Z_1-Z_2), \end{aligned} \tag{67}$$

and

$$k=2e^2\epsilon_1\epsilon_2/c^2.$$

V. THE HAMILTON-JACOBI EQUATION, LADDER OPERATORS, AND SYMMETRY GROUPS

For those problems which are classically degenerate, there is a straightforward procedure by which ladder operators for the energy may be constructed. Due to the symplectic nature of the Hamiltonian equations of motion, these ladder operators will be Poisson-bracket eigenfunctions of the Hamiltonian belonging to negative pairs of eigenvalues. Accordingly, constants of the motion can be formed from products of pairs belonging to opposite eigenvalues, and because of the classical degeneracy, there will be many linearly independent such products to be formed. One may generate an $SU(3)$ classical symmetry group for systems such as the monopole with the central forces which we are studying. However, we generally find that only the ladder operators will carry over into quantum mechanics, and that the commutation rules of their bilinear products will not quite obey the commutation rules of an $SU(3)$ group, although in the case of the Coulombic systems, one can generally use the ladder operators to generate an $O(4)$ symmetry group. In previous papers, we have discussed both the technique to form the ladder operators and the "universal" symmetry group,[28] as well as the modifications which the universal symmetry group suffers in various typical circumstances.[25] A more extensive discussion of various details is found in these papers.

The technique is to solve the Hamilton–Jacobi equation in order to write the Hamiltonian in terms of the action-angle variables. Once the Hamiltonian has the functional form

$$H=H(\textstyle\sum \alpha_i^{-1}J_i),$$

in which the J_i are the action variables, and the α_i are integers, ladder operators may immediately be defined as follows:

$$a_i^{\pm}=(J_i/2\pi\alpha_i)^{\frac{1}{2}}e^{\pm 2\pi i\alpha_i w_i},$$

where the ω_i are the conjugate angle variables.

Although we have shown in the last section that the Coulombic monopole problem is separable in a great variety of coordinate systems—polar coordinates, parabolic coordinates, ellipsoidal coordinates, and quaternionic coordinates, we shall carry out the ladder-operator construction only in polar coordinates, both for the Coulombic and the harmonic monopole. The other coordinate systems give rise to some characteristic anomalies of their own, but the polar ladder operators will be entirely adequate for studying the degeneracy and symmetry of the quantum-mechanical monopoles.

The Hamilton–Jacobi equation corresponding to the Hamiltonian (36),

$$\begin{aligned} 2E &= \left(\frac{\partial S}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial S}{\partial\theta}\right)^2 \\ &\quad+\frac{1}{r^2\sin\theta}\left(\frac{\partial S}{\partial\phi}+\epsilon\cos\theta\right)^2+\frac{\epsilon^2}{r^2}+2U(r), \end{aligned} \tag{68}$$

is separated by writing the principal function in the form $S = S_r(r) + S_\theta(\theta) + S_\phi(\phi)$. We have immediately $S_\phi(\phi) = \beta\phi$, where β is a constant, since ϕ does not appear explicitly in the Hamiltonian. The other components of the principal function are

$$S_r = \int [2(E - U) - (\alpha^2 + \epsilon^2)/r^2]^{\frac{1}{2}}\, dr, \quad (69a)$$

$$S_\theta = \int [\alpha^2 - (\beta + \epsilon \cos\theta)^2/\sin^2\theta]^{\frac{1}{2}}\, d\theta, \quad (69b)$$

where α is a separation constant.

The action variables $J_j = \oint p_j\, dq_j$ are found by use of the relation $p_j = \partial S/\partial q$ and the separated Hamilton–Jacobi equations

$$J_\phi = \oint p_\phi\, d\phi = 2\pi\beta, \quad (70a)$$

$$J_\theta = \oint p_\theta\, d\theta = 2\pi[(\alpha^2 + \epsilon^2)^{\frac{1}{2}} - \beta], \quad (70b)$$

$$J_r = \oint p_r\, dr = 2\pi[(-2E)^{-\frac{1}{2}} - (\alpha^2 + \epsilon^2)^{\frac{1}{2}}], \quad U = -1/r, \quad (70c)$$

$$= \pi[E - (\alpha^2 + \epsilon^2)^{\frac{1}{2}}], \quad U = r^2/2. \quad (70d)$$

The energy can then be expressed in terms of the action variables

$$E = -2\pi^2(J_r + J_\theta + J_\phi)^{-2}, \quad U = -1/r, \quad (71a)$$

$$E = (2J_r + J_\theta + J_\phi)/2\pi, \quad U = r^2/2. \quad (71b)$$

Considering for the moment just the Kepler potential, we find that the form of the expression for the energy given above guarantees that bilinear combinations of the form a^-a^+, constructed from

$$a_r^\pm = \mp i(J_r/2\pi)^{\frac{1}{2}} e^{\pm 2\pi i w_r}, \quad (72a)$$

$$a_\theta^\pm = \mp i(J_\theta/2\pi)^{\frac{1}{2}} e^{\pm 2\pi i w_\theta}, \quad (72b)$$

$$a_\phi^\pm = -(J_\phi/2\pi)^{\frac{1}{2}} e^{\pm 2\pi i w_\phi}, \quad (72c)$$

are constants of the motion. The a-functions satisfy the Poisson-bracket relations

$$\{a_r^+, a_r^-\} = i, \quad (73a)$$

$$\{a_\theta^+, a_\theta^-\} = i, \quad (73b)$$

$$\{a_\phi^+, a_\phi^-\} = i; \quad (73c)$$

operators that belong to different coordinates commute.

The a functions will be given explicitly in terms of polar coordinates and momenta. The angle variables are evaluated using the separated Hamilton–Jacobi equations

$$w_r = -\frac{(-2E)^{\frac{1}{2}}}{2\pi} r p_r + \frac{1}{2\pi} \arcsin \frac{-2Er - 1}{[1 + 2E(\alpha^2 + \epsilon^2)]^{\frac{1}{2}}},$$

$$w_\theta = w_r - \frac{1}{2\pi} \arcsin \frac{r - (\alpha^2 - \epsilon^2)}{r[1 + 2E(\alpha^2 + \epsilon^2)]^{\frac{1}{2}}} - \frac{1}{2\pi} \arcsin \frac{(\alpha^2 + \epsilon^2)\cos\theta + \beta\epsilon}{[\beta^2\epsilon^2 + (\alpha^2 + \epsilon^2)(\alpha^2 - \beta^2)]^{\frac{1}{2}}},$$

$$w_\phi = w_\theta - \frac{1}{2\pi} \arcsin \frac{-\epsilon - \beta\cos\theta}{(\alpha^2 - \epsilon^2 - \beta^2)^{\frac{1}{2}} \sin\theta} + \frac{\phi}{2\pi},$$

from which the functions (72) are constructed:

$$a_r^\pm = [(-2E)^{-\frac{1}{2}} - (\alpha^2 + \epsilon^2)^{\frac{1}{2}}]^{\frac{1}{2}} \times [e^{\pm i(-2E)^{\frac{1}{2}} r p_r}((-2E)^{-1} - (\alpha^2 + \epsilon^2))^{-\frac{1}{2}} \times ((-2E)^{\frac{1}{2}} r - (-2E)^{-\frac{1}{2}} \mp i r p_r)], \quad (74a)$$

$$a_\theta^\pm = [(\alpha^2 + \epsilon^2)^{\frac{1}{2}} - \beta]^{\frac{1}{2}} (\mp i e^{\pm 2\pi i w_r}) \frac{(\alpha^2 + \epsilon^2)^{\frac{1}{2}}}{[1 + 2E(\alpha^2 + \epsilon^2)]^{\frac{1}{2}}} \times \left(\frac{-1}{(\alpha^2 + \epsilon^2)^{\frac{1}{2}}} + \frac{(\alpha^2 + \epsilon^2)^{\frac{1}{2}}}{r} \mp i p_r\right) \times \frac{(\alpha^2 + \epsilon^2)^{\frac{1}{2}}}{[\beta^2\epsilon^2 + (\alpha^2 + \epsilon^2)(\alpha^2 - \beta^2)]^{\frac{1}{2}}} \times \left[(\alpha^2 + \epsilon^2)^{\frac{1}{2}} \cos\theta + \frac{\beta\epsilon}{(\alpha^2 + \epsilon^2)^{\frac{1}{2}}} \pm i \sin\theta p_\theta\right], \quad (74b)$$

$$a_\phi^\pm = \beta^{\frac{1}{2}} (\mp i e^{\pm 2\pi i w_\theta}) \frac{1}{(\alpha^2 + \epsilon^2 - \beta^2)^{\frac{1}{2}}} \times \left(\frac{\epsilon}{\sin\theta} + \beta\cot\theta \mp i p_\theta\right) e^{\pm i\phi}. \quad (74c)$$

Similarly, for the harmonic-oscillator potential the form of the energy expression (7.1b) guarantees that bilinear combinations of the functions,

$$a_r^\pm = \mp i(J_r/2\pi)^{\frac{1}{2}} e^{\pm 2\pi i w_r}, \quad (75a)$$

$$a_\theta^\pm = (J_\theta/4\pi)^{\frac{1}{2}} e^{\pm 4\pi i w_\theta}, \quad (75b)$$

$$a_\phi^\pm = (J_\phi/4\pi)^{\frac{1}{2}} e^{\pm 4\pi i w_\phi}, \quad (75c)$$

are constants of the motion. Poisson-bracket relations (73) are also satisfied among the functions given above.

The angle variables are given in this case by

$$w_r = \frac{1}{2\pi} \arcsin \frac{r^2 - E}{[E^2 - (\alpha^2 + \epsilon^2)]^{\frac{1}{2}}},$$

$$w_\theta = \frac{w_r}{2} - \frac{1}{4\pi} \arcsin \frac{Er^2 - (\alpha^2 + \epsilon^2)}{r^2[E^2 - (\alpha^2 + \epsilon^2)]^{\frac{1}{2}}} - \frac{1}{2\pi} \arcsin \frac{(\alpha^2 + \epsilon^2)\cos\theta + \beta\epsilon}{[\beta^2\epsilon^2 + (\alpha^2 + \epsilon^2)(\alpha^2 - \beta^2)]^{\frac{1}{2}}},$$

$$w_\phi = w_\theta - \frac{1}{2\pi} \arcsin \frac{-\epsilon - \beta\cos\theta}{(\alpha^2 - \epsilon^2 - \beta^2)^{\frac{1}{2}} \sin\theta} + \frac{\phi}{2\pi},$$

from which functions (75) are constructed:

$$a_r^{\pm} = \left[-\frac{(\alpha^2+\epsilon^2)^{\frac{1}{2}}}{2} + \frac{E}{2}\right]^{\frac{1}{2}} \times \{(E-(\alpha^2+\epsilon^2))^{-\frac{1}{2}}[r^2 - E \mp irp_r]\}, \quad (76a)$$

$$a_\theta^{\pm} = \left[\frac{(\alpha^2+\epsilon^2)^{\frac{1}{2}}}{2} - \frac{\beta}{2}\right]^{\frac{1}{2}} (\mp ie^{\pm 2\pi i w_r}) \frac{(\alpha^2+\epsilon^2)^{\frac{1}{2}}}{[E^2-(\alpha^2+\epsilon^2)]^{\frac{1}{2}}} \times \left[\frac{-E}{(\alpha^2+\epsilon^2)^{\frac{1}{2}}} + \frac{(\alpha^2+\epsilon^2)^{\frac{1}{2}}}{r^2} \mp i\frac{p_r}{r}\right] \times \left\{\frac{(\alpha^2+\epsilon^2)^{\frac{1}{2}}}{[\beta^2\epsilon^2 + (\alpha^2+\epsilon^2)(\alpha^2-\beta^2)]^{\frac{1}{2}}} \times \left[(\alpha^2+\epsilon^2)\cos\theta + \frac{\beta\epsilon}{(\alpha^2+\epsilon^2)^{\frac{1}{2}}} \pm i\sin\theta p_\theta\right]\right\}^2 \quad (76b)$$

$$a_\phi^{\pm} = \left(\frac{\beta}{2}\right)^{\frac{1}{2}} (e^{\pm 4\pi i w_\theta}) \times \left\{\frac{1}{(\alpha^2+\epsilon^2-\beta^2)^{\frac{1}{2}}}\left[\frac{\epsilon}{\sin\theta} + \beta\cot\theta \mp ip_\theta\right]e^{\pm i\phi}\right\}^2. \quad (76c)$$

If one chooses the constants of the motion to be

$$M_{ij} = Q_iP_j - Q_jP_i,$$
$$N_{ij} = Q_jQ_j - P_iP_j,$$

where

$$Q_i = (a_i^+ + a_i^-)/2^{\frac{1}{2}},$$
$$P_i = (a_i^+ - a_i^-)/2^{\frac{1}{2}},$$

we find that M_{ij} and N_{ij} satisfy the commutation relations of $SU(3)$; this is valid for the two cases we have considered, since the conclusion depends only on the commutators (73) among the a's.

Dulock[2] gave a mapping from a set of functions satisfying $SU(3)$ Poisson-bracket relations to a new set satisfying the commutation relations of $O(4)$. This means that we should also have $O(4)$ as a dynamical symmetry for the two problems considered here, the generators given explicitly by using the "a" functions (74) for $U(r) = -1/r$ or those given by (76) for $U(r) = r^2/2$. Dulock's transformation is

$$L_1' = -\frac{1}{2}\left(\frac{b'b+2c'c}{c'c}\right)^{\frac{1}{2}}(c'b+b'c), \quad (77a)$$

$$L_2' = -\frac{i}{2}\left(\frac{b'b+2c'c}{c'c}\right)^{\frac{1}{2}}(c'b-b'c), \quad (77b)$$

$$L_3' = c'c, \quad (77c)$$

$$P_1 = \frac{-(a'a+2b'b'+2c'c')^{\frac{1}{2}}}{4b'b(c'c)^{\frac{1}{2}}(b'b+c'c)} \times [(b'a+a'b)(c'b+b'c)c'c + (b'a-a'b)(c'b-b'c)(b'b+c'c)], \quad (77d)$$

$$P_2 = \frac{-i(a'a+2b'b+2c'c)^{\frac{1}{2}}}{4b'b(c'c)^{\frac{1}{2}}(b'b+c'c)} \times [(b'a+a'b)(c'b-b'c)c'c + (b'a-a'b)(c'b+b'c)(b'b+c'c)], \quad (77e)$$

$$P_3 = -\frac{[(a'a+2b'b+2c'c)(b'b+2c'c)]^{\frac{1}{2}}}{2(b'b+c'c)} \times [b'a+a'b], \quad (77f)$$

where we have set $a = a_r^+$, $a' = a_r^-$, $b = a_\theta^+$, $b' = a_\theta^-$, $c = a_\phi^+$, $c' = a_\phi^-$.

If the a's satisfy (73), $\mathbf{L}'$ and $\mathbf{P}$ satisfy

$$[L_i', L_j'] = e_{ijk}L_k', \quad (78a)$$

$$[P_i, P_j] = e_{ijk}L_k', \quad (78b)$$

$$[P_i, L_j'] = e_{ijk}P_k. \quad (78c)$$

Only in the case of hydrogen-atom potential do the generators of $O(4)$ have a simple interpretation as constants of the motion, the result of the transformation applied to (74) being

$$\mathbf{L}' = \mathbf{D}, \quad (79a)$$

$$\mathbf{P} = (-2E)^{-\frac{1}{2}}\frac{(\alpha^2+\epsilon^2)^{\frac{1}{2}}}{\alpha}\left(\mathbf{R} + \frac{\epsilon}{\alpha^2+\epsilon^2}\mathbf{D}\right), \quad (79b)$$

where $\mathbf{D}$ is the total angular momentum and $\mathbf{R}$ is the Runge vector. One finds that $\mathbf{L}' \cdot \mathbf{P} = 0$; a quantum-mechanical version of these operators would predict energy levels in contradiction to those obtained by the solution of Schrödinger's equation. The difficulty is that the commutation relations among the L''s and P's are not preserved quantum mechanically. One can modify Dulock's transformation to get the vectors $\mathbf{D}$ and $\mathbf{R}$ which will be found to be the correct generators of the dynamical symmetry group for hydrogen-atom potential.

VI. THE SCHRÖDINGER EQUATION FOR THE MONOPOLE

Just as the Hamilton–Jacobi equation for the monopole in a central force field is separable in spherical polar coordinates, so are the corresponding Schrödinger equations. Making the usual substitutions for the kinetic energy operator in terms of the Laplacian in spherical polar coordinates, the Schrödinger equation for a charged particle in a magnetic field

$$[\tfrac{1}{2}(-i\nabla - \mathbf{A})^2 + V]\Psi = E\Psi, \quad (80)$$

becomes

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)-\frac{1}{\sin^2\theta}\left(-i\frac{\partial}{\partial\phi}+\epsilon\cos\theta\right)^2-\epsilon^2+2r^2(E-U)\right]\Psi=0 \quad (81)$$

for the vector potential (31) and the scalar potential given by $V(r) = U(r) + \epsilon^2/2r^2$. To separate the variables, we assume that the wavefunction Ψ has the form

$$\Psi(r,\theta,\phi) = R(r)\Theta(\theta)e^{im\phi}. \quad (82)$$

Under this assumption, we obtain the separated equations for the radial and colatitudinal variables,

$$\frac{1}{r^2}\left(\frac{\partial}{\partial r}r^2\frac{\partial R}{\partial r}\right)+2(E-U)R-\frac{\alpha^2+\epsilon^2}{r^2}R=0, \quad (83)$$

$$\left[\alpha^2+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)-\frac{(m+\epsilon\cos\theta)^2}{\sin^2\theta}\right]\Theta=0. \quad (84)$$

The second equation is well known to be the Schrödinger equation for the spherical top[24,29] when m and ϵ are positive or negative integers. Without imposing any conditions on m and ϵ, Fierz[13] showed that the equation may have physically acceptable solutions only for m and ϵ simultaneously integral or half integral, thus giving an alternative indicated in these special circumstances, of the correctness of Dirac's[10] conclusions regarding the quantization of the magnetic monopole strength. We will exclude half-integral values of pole strength, since it leads to half-integral values of orbital angular momentum. Reasons for rejecting such solutions of Schrödinger's equation have been extensively discussed in the literature.[30]

The normalized eigenfunctions of Eq. (84) are[31]

$$\Theta_{l,m}=(-1)^{l+\epsilon}\left[\frac{(l-\epsilon)!\,(l+m)!}{(l+\epsilon)!\,(l-m)!}\right]^{\frac{1}{2}}\frac{(2l+1)^{\frac{1}{2}}2^{-l-\frac{1}{2}}}{(2l)!}\times(1-t)^{-(m+\epsilon)/2}(1+t)^{(\epsilon-m)/2}\times\frac{d^{l-m}}{dt^{l-m}}(1-t)^{l+\epsilon}\frac{d^{l+\epsilon}}{dt^{l+\epsilon}}(1+t)^{2l},$$

with $t = \cos\theta$, (85)

and eigenvalues given by

$$\alpha^2+\epsilon^2=l(l+1),\quad l=0,1,2,\cdots,\quad m=0,\pm1,\pm2,\cdots,\pm l,\quad l\geq|\epsilon|. \quad (86)$$

As we see, these restrictions are identical to the usual quantization rules for angular momentum.

	ϵ	n	1	2	3	4	5	6	7	
	0		1	4	9	16	25	36	49	n^2
(a)	±1			3	8	15	24	35	48	n^2-1
	±2				5	12	21	32	45	n^2-4
	±3					7	16	27	40	n^2-9

	ϵ	n	0	1	2	3	4	5	6
	0		1	3	6	10	15	21	28
(b)	±1			3	5	10	14	21	27
	±2				5	7	14	18	27
	±3					7	9	18	22

FIG. 1. Table of degeneracies of the magnetic monopole, with supplemental centrifugal potential: (a) hydrogen atom, (b) harmonic oscillator.

However, we find the interesting effect that the lowest value of orbital angular momentum is limited by the magnetic pole strength; this means that for $\epsilon = \pm1$ there will be no s states and the lowest level will be degenerate; for $\epsilon = \pm2$ there will be no s and p states, etc. In general, the lowest level will be at least $(2|\epsilon| + 1)$-fold degenerate. It is clear that this result is independent of the potential U, as long as it depends only on the radius.

After substitution of the eigenvalue α^2 in the radial equation, we find it is identical to that of a particle with orbital angular momentum ℓ moving in the central potential $U(r)$; the solutions of this equation for $U(r) = -1/r$ and $r^2/2$ are well known, the only difference being that solutions for $l < |\epsilon|$ will not be present, and hence the over-all degeneracy pattern will be changed.

The radial wavefunctions for $-1/r$ potential are[32]

$${}^{H}R_{n,\ell}=\left[\frac{2}{n^4}\frac{(n-\ell-1)!}{[(n+\ell)!]^3}\right]^{\frac{1}{2}}e^{-\rho/2}\rho^{\ell}L_{n+l}^{2\ell+1}(\rho), \quad (87)$$

with energy eigenvalues given by

$$E=(2n^2)^{-1},\quad n=l+1,l+2,\cdots. \quad (88)$$

The degeneracies of the first energy levels for the first few values of magnetic monopole strength are given in Fig. 1(a).

In general the degeneracy of the level n, corresponding to pole strength $\pm\epsilon$, is given by

$$\sum_{l=|\epsilon|}^{n-1}(2l+1)=n^2-|\epsilon|^2. \quad (89)$$

As will be seen in the next section, this degeneracy is accounted for by an $O(4)$ dynamical symmetry group.

The radial wavefunctions for $r^2/2$ potential are given by[33]

$${}^{o}R_{n,l}=\left[\frac{2((n-l)/2)!}{[\Gamma((n+l+3)/2)]^3}\right]^{\frac{1}{2}}e^{-r^2/2}r^l L_{(n+l+1)/2}^{l+\frac{1}{2}}(r^2)$$

$$n=l,l+2,l+4,\cdots, \quad (90)$$

with energy eigenvalues

$$E = n + \tfrac{3}{2}. \tag{91}$$

The degeneracies of the first energy levels for the first few values of magnetic pole strength are given in Fig. 1(b).

Ladder operators for the wavefunctions (85) and (87) can be found in the literature. Musto[34] gives the operator

$${}^{\mathrm{H}}\mathcal{K}_r^{\pm} = e^{(\ln n/n\pm 1)\rho}\frac{\partial}{\partial\rho}\left(\frac{n}{n+1}\right)^2[n(n\pm 1) - l(l+1)]^{-\frac{1}{2}} \times \left(\frac{\rho}{2} - n \pm 1 \mp \rho\frac{\partial}{\partial\rho}\right), \tag{92}$$

which is a normalized ladder operator for the total quantum number of the radial wavefunction

$${}^{\mathrm{H}}\mathcal{K}_r^{\pm}\,{}^{\mathrm{H}}R_{n,l} = {}^{\mathrm{H}}R_{n\pm 1,l}.$$

Infeld and Hull[35] give the operators

$${}^{\mathrm{HR}}\mathcal{K}_\theta^{\pm} = nl(n^2 - l^2)^{-\frac{1}{2}}\left(\frac{l}{r} - \frac{1}{l} \mp \frac{\partial}{\partial r}\right) \tag{93}$$

which are normalized ladder operators for the orbital quantum number of the radial wavefunction:

$${}^{\mathrm{HR}}\mathcal{K}_\theta^{+} r\,{}^{\mathrm{H}}R_{n,l-1} = r\,{}^{\mathrm{H}}R_{n,l},$$
$${}^{\mathrm{HR}}\mathcal{K}_\theta^{-} r\,{}^{\mathrm{H}}R_{n,l} = r\,{}^{\mathrm{H}}R_{n,l-1}.$$

Burkhard[36] gives the operators

$${}^{\Theta}\mathcal{K}_\theta^{+} = \frac{l+1}{2l+1}\left[\frac{(2l+1)(2l+3)}{[(l+1)^2 - m^2][(l+1)^2 - \epsilon^2]}\right]^{\frac{1}{2}} \times \left[(l+1)\cos\theta + \frac{m\epsilon}{l+1} + \sin\theta\frac{\partial}{\partial\theta}\right], \tag{94a}$$

$${}^{\Theta}\mathcal{K}_\theta^{-} = \frac{l}{2l+1}\left[\frac{(2l-1)(2l+3)}{[l^2 - m^2][l^2 - \epsilon^2]}\right]^{\frac{1}{2}} \times \left[l\cos\theta + \frac{m\epsilon}{l} - \sin\theta\frac{\partial}{\partial\theta}\right], \tag{94b}$$

which are normalized ladder operators for the orbital quantum number of the angular wavefunction,

$${}^{\Theta}\mathcal{K}_\theta^{\pm}\Theta_{l,m} = \Theta_{l\pm 1,m}.$$

Infeld and Hull[35] give the operators

$$\mathcal{K}_\phi^{\pm} = [(l \pm m + 1)(l \pm m)]^{-\frac{1}{2}} \times \left[(m - \tfrac{1}{2})\cot\theta + \frac{\epsilon}{\sin\theta} \mp \frac{\partial}{\partial\theta}\right]e^{\pm i\phi}, \tag{95}$$

which are normalized ladder operators for the magnetic quantum number of the angular wavefunctions,

$$\mathcal{K}_\phi^{\pm}\sin^{\frac{1}{2}}\theta\Theta_{l,m}\Phi_m = \sin^{\frac{1}{2}}\theta\Theta_{l,m\pm 1}\Phi_{m\pm 1}.$$

In the limit of large quantum numbers we have

$${}^{\mathrm{H}}\mathcal{K}_r^{\pm} \to \mp ie^{\pm 2\pi i w_r}, \tag{96a}$$
$${}^{\mathrm{HR}}\mathcal{K}_\theta^{\pm\Theta}\mathcal{K}_\theta^{\pm} \to e^{\pm 2\pi i(w_\theta - w_r)}, \tag{96b}$$
$$\mathcal{K}_\phi^{\pm} \to \mp ie^{\pm 2\pi i(w_\phi - w_\theta)}; \tag{96c}$$

this is seen by comparing the ladder operators with (74).

With the ladder operators given, one can construct the quantum-mechanical version of the functions "a" given by (74) for Kepler potential. We would expect to find an $SU(3)$ dynamical symmetry group in the quantum-mechanical problem also. Nevertheless, the operators fail to satisfy the expected commutation relations when applied to states having extreme quantum numbers, as found in the simpler two-dimensional cases.[28]

We now turn to the harmonic-oscillator potential; the recurrence relation given by Schrödinger[37] connecting radial wavefunctions of different total quantum number, for the hydrogen atom, can be adapted by changing r into r^2 to get ladder operators for the total quantum number of the wavefunction (90). After inclusion of a proper normalizing factor we have

$${}^{\mathrm{o}}\mathcal{K}_r^{+} = [(n + \tfrac{5}{2})^2 - (l + \tfrac{1}{2})^2]^{-\frac{1}{2}}\left[r^2 - (n+3) - r\frac{\partial}{\partial r}\right], \tag{97a}$$

$${}^{\mathrm{o}}\mathcal{K}_r^{-} = [(n + \tfrac{1}{2})^2 - (l + \tfrac{1}{2})^2]^{-\frac{1}{2}}\left[r^2 - n + r\frac{\partial}{\partial r}\right]; \tag{97b}$$

$${}^{\mathrm{o}}\mathcal{K}_r^{\pm}\,{}^{\mathrm{o}}R_{n,l} = {}^{\mathrm{o}}R_{n\pm 2,l}.$$

The same transformation used above can be applied to the operator (93); after adjusting the normalizing factor we find

$${}^{\mathrm{oR}}\mathcal{K}_\theta^{\pm} = \frac{l - \frac{1}{2}}{[(n - l + 2)(n + l + 1)]^{\frac{1}{2}}} \times \left[\frac{l - \frac{1}{2}}{r^2} - \frac{n + \frac{3}{2}}{l - \frac{1}{2}} \mp \frac{1}{r}\frac{\partial}{r\partial}\right]; \tag{98}$$
$${}^{\mathrm{oR}}\mathcal{K}_\theta^{+} r^{\frac{3}{2}}\,{}^{\mathrm{o}}R_{n,l-2} = r^{\frac{3}{2}}\,{}^{\mathrm{o}}R_{n,l},$$
$${}^{\mathrm{oR}}\mathcal{K}_\theta^{-} r^{\frac{3}{2}}\,{}^{\mathrm{o}}R_{n,l} = r^{\frac{3}{2}}\,{}^{\mathrm{o}}R_{n,l-2}.$$

Ladder operators for the angular wavefunctions are the same [(94) and (95)] as those given for hydrogen-atom potential. We find that in the limit of large quantum numbers,

$${}^{\mathrm{o}}\mathcal{K}_r^{\pm} \to \mp ie^{\pm 2\pi i w_r}, \tag{99a}$$
$${}^{\mathrm{oR}}\mathcal{K}_\theta^{\pm}({}^{\Theta}\mathcal{K}_\theta^{\pm})^2 \to \pm ie^{\pm 4\pi i(w_\theta - w_r/2)}, \tag{99b}$$
$$(\mathcal{K}_\phi^{\pm})^2 \to -e^{\pm 4\pi i(w_\phi - w_\theta)}. \tag{99c}$$

In this case we also find that an attempt to construct

the operators corresponding to (76) leads to a failure of the expected commutators.

It is instructive to examine the actual wavefunctions for the degenerate ground state of the monopole. If we consider Schwinger's[19] monopole, with the quantized magnetic charge equal to one, we find that the half angle of the orbital cone in a state of zero mechanical angular momentum is $\gamma = \arccos(\frac{1}{2}^{\frac{1}{2}})$, or 45°. Due to the mixture of repulsive centrifugal potential and attractive Coulomb potential, the classical orbits in such a case will consist of straight lines in which the particle oscillates back and forth in a radial direction on the surface of the cone. Generally speaking, however, there will be two orbits possible for each combination of energy and mechanical angular momentum, one lying on the upper branch of the cone and circulating, if at all, in one direction, while the other lies on the lower branch and circulates in the opposite direction.

When we think of the quantum-mechanical wavefunction, we have to remember that the probability distribution will be spread out in the general region of the classical orbit, and that the orientation of the orbit will not be entirely definite. There is another aspect to be considered. Classically, the conserved total angular momentum and the mechanical angular momentum are related by the formula $\mathbf{D} = \mathbf{L} - \epsilon\mathbf{r}/r$ and their lengths by the formula $\mathbf{D}^2 = \mathbf{L}^2 + \epsilon^2$. Consequently, when $\epsilon = 1$ and $\mathbf{D}^2 = 1$, the mechanical angular momentum is necessarily zero, and only the states of radial oscillation are admissible. However, the quantum-mechanical "length" of $\mathbf{D}^2$ is of the form $l(l+1)$, which means that the corresponding quantum-mechanical formula is

$$\mathfrak{L}^2 + \epsilon^2 = l(l+1),$$

and this in turn implies that $\mathfrak{L}^2$ has eigenvalue 1. However $\mathfrak{L} \cdot \mathfrak{D}$ is a conserved constant, so that the component of the mechanical angular momentum along the axis of the orbital cone is a constant, which can take one of the three values $+1$, 0, or -1. Hence, a circulation of the particle is admissible in the ground state, something which would not be possible classically. If the circulation is clockwise, the particle will favor the upper branch of the cone, when it is counterclockwise it will favor the lower branch, and when it is zero, presumably the particle may divide its time equally between the two branches, but still avoid the axis of the cone. Hence we may understand the wavefunctions shown in the sketch of Fig. 2. The degeneracy of the ground state then seems to be a consequence of the fact that the particle, by being confined to the surface of a cone and not to a plane, never occupies a spherically symmetrical probability distribution, and so singly degenerate s states can never occur.

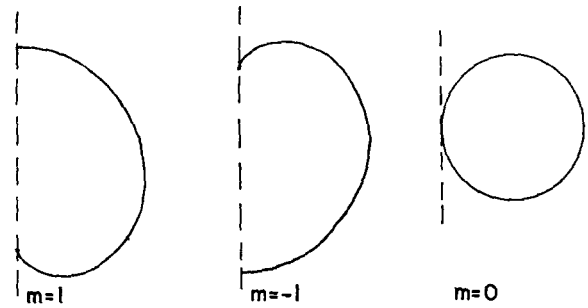


FIG. 2. Angular wavefunctions of the degenerate ground state for the first-level magnetic monopole charge.

VII. SYMMETRY OF THE MONOPOLE

In this section we show that the quantum-mechanical symmetry group of the monopole problem with the potential $\epsilon^2/2r^2 - 1/r$ is the rotation group in four dimensions; its generators are given by the quantum-mechanical operators corresponding to the vectors $\mathbf{D}$ and $\mathbf{R}$ considered in the preceding sections. The total angular momentum operator is given by

$$\mathfrak{D} = \mathbf{r} \times \boldsymbol{\pi} - \epsilon\mathbf{r}/r, \tag{100}$$

where $\boldsymbol{\pi}$ is the mechanical momentum operator; substituting $\boldsymbol{\pi} = (-i\nabla - \mathbf{A})$, where $\mathbf{A}$ is the vector potential (31), we find

$$\mathfrak{D}_x = \mathfrak{L}_x - \epsilon xr/(x^2 + y^2), \tag{101a}$$

$$\mathfrak{D}_y = \mathfrak{L}_y - \epsilon yr/(x^2 + y^2), \tag{101b}$$

$$\mathfrak{D}_z = \mathfrak{L}_z, \tag{101c}$$

where

$$\mathfrak{L} = -i\mathbf{r} \times \nabla. \tag{102}$$

The Runge vector operator is given by

$$\mathfrak{R} = \tfrac{1}{2}(\mathfrak{D} \times \boldsymbol{\pi} - \boldsymbol{\pi} \times \mathfrak{D}) + \mathbf{r}/r \tag{103}$$

after symmetrization of the noncommuting operators $\mathfrak{D}$ and $\boldsymbol{\pi}$. Its Cartesian components are

$$\mathfrak{R}_x = \pi_z\mathfrak{D}_y - \mathfrak{D}_z\pi_y + x/r, \tag{104a}$$

$$\mathfrak{R}_y = \pi_x\mathfrak{D}_z - \mathfrak{D}_x\pi_z + y/r, \tag{104b}$$

$$\mathfrak{R}_z = \pi_y\mathfrak{D}_x - \mathfrak{D}_y\pi_x + z/r. \tag{104c}$$

The operators $\mathfrak{D}$ and $\mathfrak{R}' = (-2\mathfrak{K})^{-\frac{1}{2}}\mathfrak{R}$ commute with the Hamiltonian $\mathfrak{K}$ and satisfy the commutation relations of $O(4)$:

$$[\mathfrak{D}_i, \mathfrak{D}_j] = ie_{ijk}\mathfrak{D}_k, \tag{105a}$$

$$[\mathfrak{R}'_i, \mathfrak{R}'_j] = ie_{ijk}\mathfrak{D}_k, \tag{105b}$$

$$[\mathfrak{R}'_i, \mathfrak{D}_j] = ie_{ijk}\mathfrak{R}'_k. \tag{105c}$$

We will now apply Pauli's[38] derivation of the formula for the energy levels of the hydrogen atom to this problem. By doing so, the degeneracy given by (89) will be accounted for completely. We define the operators

$$^1\mathfrak{J} = (\mathfrak{D} + \mathfrak{R}')/2, \tag{106a}$$

$$^2\mathfrak{J} = (\mathfrak{D} - \mathfrak{R}')/2. \tag{106b}$$

As a direct consequence of the commutators (105), we have

$$[^1\mathfrak{J}_i, {}^1\mathfrak{J}_j] = ie_{ijk}{}^1\mathfrak{J}_k, \tag{107a}$$

$$[^2\mathfrak{J}_i, {}^2\mathfrak{J}_j] = ie_{ijk}{}^2\mathfrak{J}_k, \tag{107b}$$

$$[^1\mathfrak{J}_i, {}^2\mathfrak{J}_j] = 0, \tag{107c}$$

A straightforward calculation yields

$$\mathfrak{R}^{12} = -(\mathfrak{D}^2 - \epsilon^2 + 1) + (-2\mathfrak{H})^{-1}, \tag{108}$$

$$\mathfrak{D} \cdot \mathfrak{R} = \mathfrak{R} \cdot \mathfrak{D} = -\epsilon, \tag{109}$$

which are used to obtain the relations,

$$^1\mathfrak{J}^2 = [(-2\mathfrak{H})^{-\frac{1}{2}} - \epsilon - 1)][(-2\mathfrak{H})^{-\frac{1}{2}} - \epsilon + 1]/4, \tag{110a}$$

$$^2\mathfrak{J}^2 = [(-2\mathfrak{H})^{-\frac{1}{2}} + \epsilon - 1)][(-2\mathfrak{H})^{-\frac{1}{2}} + \epsilon + 1]/4. \tag{110b}$$

$^1\mathfrak{J}^2$, $^2\mathfrak{J}^2$, and $\mathfrak{H}$ can be diagonalized simultaneously, since they commute. From the commutation relation (107a), satisfied by the components of $^1\mathfrak{J}$, we conclude that $^1\mathfrak{J}^2$ has eigenvalues given by $i(i + 1)$, where $i = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$. The dimension of the representation is $2i + 1 = s$, and hence $i(i + 1) = (s - 1) \times (s + 1)/4$. Comparing with (110a) we find that the energy level E_s is given by

$$E_s = -\tfrac{1}{2}(s + \epsilon)^2. \tag{111}$$

The eigenvalues of $^2\mathfrak{J}^2$ are given by $k(k + 1)$, where k is integer or half-integer. The dimension of the representation is $2k + 1 = q$ and hence $k(k + 1) = (q - 1)(q + 1)/4$. From (111) and (110b) we see that for the representation corresponding to the energy level E_s, the eigenvalue of $^2\mathfrak{J}^2$ is $(s + 2\epsilon - 1) \times (s + 2\epsilon + 1)/2$, and we have $q = s + 2\epsilon$. The dimensionalities of the representations of $^1\mathfrak{J}$ and $^2\mathfrak{J}$ differ by 2ϵ. This originates from the fact that $\mathfrak{D} \cdot \mathfrak{R}$ is $-\epsilon$ and is not zero as in the case of the hydrogen atom. The degeneracy of the level s is simply the product of the dimensions of the representations of $^1\mathfrak{J}$ and $^2\mathfrak{J}$ which is given by $s(s + 2\epsilon)$, and which is the same as formula (89) since $S = n - \epsilon$. It is also clear that the lowest level will be degenerate. Since s and q differ by 2ϵ, we will not have simultaneous one-dimensional representations of $^1\mathfrak{J}$ and $^2\mathfrak{J}$, as for the hydrogen atom. The lowest degeneracy is given by $2|\epsilon| + 1$.

One gets the quantization of magnetic pole strength from these considerations. 2ϵ is integral simply because it is the difference of the dimensions of two representations. Nevertheless, this derivation is not as general as that of Dirac, since it depends on a particular potential.

It is interesting to note the role played by the identity $\mathfrak{D} \cdot \mathfrak{R} = -\epsilon$ in the determination of the symmetry of the Coulombic monopole, since it is responsible for the fact that one does not obtain the usual representations of $O(4)$ of dimension n^2, but rather those of dimension mn, where $m - n = 2\epsilon$.

Bacry, Ruegg, and Souriau,[39] in a recent paper investigating the possible vector constants of the motion which might exist for a problem with a spherically symmetrical potential, found that such a requirement neither fixed the angle between the angular momentum and the second vector constant, nor their relative magnitudes. The familiar relationship $\mathbf{L} \cdot \mathbf{R} = 0$ which holds for the ordinary Coulomb potential, stating that the Runge vector lies in the plane of the orbit and is consequently perpendicular to the angular momentum, is a consequence of the existence of linear orbits passing through the origin. These orbits carry zero angular momentum and allow the evaluation of $\mathbf{L} \cdot \mathbf{R}$, which is otherwise only known to be a constant. However, in the case of the monopole, the total angular momentum always has a minimum value due to the angular momentum resident in the field and distinct from the mechanical angular momentum. Thus the linear orbits fix the value of $\mathbf{D} \cdot \mathbf{R}$ at this minimum value, $-\epsilon$.

The other relationship, $L^2 + R^2 = (-2E)^{-1}$, was determined from the possible circular orbits, and yields similar results for the monopole.

The failure of Dulock's mapping from the classical generators of $SU(3)$ to the generators of $O(4)$ as applied to the quantum-mechanical ladder operators of the monopole problem is evidenced by the foregoing discussion. However, we have not investigated the possibility that other transformations of a similar nature might not exist, and allow a successful construction of the generators of the symmetry group.

The question of the symmetry of the harmonic oscillator is considerably more complicated than for the Coulombic monopole, and we have not yet succeeded in relating the observed pattern of degeneracies to a symmetry group, even though the degenerate multiplicities strongly suggest the existence of a $SU(2) \times O(3)$ symmetry group.

VIII. CONCLUSION

Perhaps the most important point which we have investigated is the discussion of the proper formulation

in Hamiltonian mechanics of the treatment of geometrical symmetry operations in the presence of a magnetic field. We have seen that the important concept is the symmetry of the field, or more generally the force system present in a mechanical problem. The potentials describing a symmetrical force field will not necessarily show the same symmetry, but rather will be invariant up to a gauge transformation when the symmetry operation is performed. As a consequence, the generators of the symmetry transformation which would be conserved quantities in the absence of magnetic forces must be modified to incorporate the necessary gauge transformations to compensate, in the definition of the canonical momentum, the change which takes place in the vector potential. The conserved quantities in the presence of a magnetic field will then be somewhat different from the familiar constants of the motion when there are purely electrostatic forces, or generally forces derived from a scalar potential alone.

A case in point is the conservation of angular momentum in the presence of a spherically symmetric magnetic field, of which the simplest example is the field due to a point magnetic monopole. It is not the angular momentum $\mathbf{L} = \mathbf{r} \times \boldsymbol{\pi}$ which is conserved, but rather a "total angular momentum" $\mathbf{D}$,

$$D = \mathbf{r} \times \frac{d\mathbf{r}}{dt} - \epsilon \hat{\mathbf{r}}.$$

Although $\mathbf{D}$ is also a vector constant of the motion, it does not satisfy the same algebraic identities as $\mathbf{L}$, and in particular, we have $\mathbf{D} \cdot \mathbf{r} = -\epsilon r$ rather than $\mathbf{D} \cdot \mathbf{r} = 0$. Instead of being confined to a plane perpendicular to the vector $\mathbf{D}$, the radius vector always lies in the surface of a cone whose axis is defined by $\mathbf{D}$. As the magnetic pole strength becomes weaker, we find that in the limit of zero pole strength, $\mathbf{D}$ becomes equal to $\mathbf{L}$, and the half angle of the orbital cone becomes $\pi/2$. Nevertheless, the law of equal areas in equal times, the second law of Kepler, still retains its validity, even though the area is swept out on the surface of the cone.

Accidental degeneracies fare somewhat differently in the presence of the magnetic field, and we have seen that both the harmonic oscillator and the Kepler problem lose their degeneracies when a monopole is located at their attraction centers. In their place other potentials become more degenerate, particularly these same two potentials when they are supplemented by the centrifugal potential $\epsilon^2/2r^2$. There are probably no such potentials occurring in nature, but their study is nevertheless instructive. The behavior of a particle in their presence can be understood qualitatively if we understand that when the particle motion is warped onto the surface of a cone by the monopole field, the circumference of a circle is less than $2\pi r$ and therefore a closed planar orbit will precess when it is transferred to the conical surface.

Taking account of this effect, we have seen that there is a modified Runge vector which is a constant of the motion for the charged monopole, and a dyadic constant of the motion for the harmonic monopole. Moreover, both problems possess the same energy formula which they have in the absence of the monopole, and in fact the only difference in the pertinent Schrödinger equations is that there are somewhat different theta functions composing the separated wavefunctions and certain restrictions on the range of the quantum numbers: otherwise both the solutions and spectra are identical.

These restrictions, at first sight innocuous, in fact radically change the pattern of degeneracies. In the case of the Coulomb monopole, the symmetry group is still $O(4)$. However, instead of the irreducible representations of dimension n^2 which are typical of orbital angular momentum problems, we find that the representations which occur are of type mn, in which $m - n = 2\epsilon$ i.e., twice the quantized magnetic pole strength. The alteration in the degeneracy pattern of the harmonic monopole is even more profound, since the resulting degeneracies correspond to no possible irreducible representation of $SU(3)$, although it is not precluded that some other symmetry group might serve to account for the degeneracies.

Yet another byproduct of the modifications is the fact that the ground states of neither problem are any longer singly degenerate; the actual degeneracy occurring depending on the quantized monopole strength, and hence on the residual angular momentum of the magnetic field of the monopole.

It was known since the earliest work of Dirac[10] and Tamm[11] that the monopole strength had to be quantized if one was to obtain a solution for the Schrödinger equation of the monopole. However, it seems to have escaped notice that assuming Dirac's quantization would in fact lead to double-valued wavefunctions. Dirac's treatment was based entirely on arguments of phase change resulting from gauge transformations, and so did not encounter problems with the actual wavefunctions. Tamm used a gauge which obscured the single or double-valued nature of his solutions. Fierz[13] used the same gauge while Banderet[14] only treated scattering solutions of the Schrödinger equation. Eliezer and Roy[16] apparently abused an absolute value sign, and missed the multiple degeneracy of the ground state, which was implicit in

the restrictions of the range of the quantum numbers mentioned in the other papers. However, only Eliezer and Roy treated the charged monopole, which in contrast to the uncharged monopole possesses bound states in which the multiple degeneracy would be obvious, and for which doubly degenerate ground state would have a readily apparent double valuedness. Malkus,[18] who also treated a charged monopole, did so in the framework of Dirac's equation, where spin also plays a role. As Peres[40] has recently confirmed, one must use Schwinger's[19] quantization, double that of Dirac, if wavefunctions which are single valued under spatial rotations are to result.

The charged monopole with repulsive centrifugal potential shares with the hydrogen atom separability in a variety of coordinate systems, including parabolic and ellipsoidal coordinates. In fact, the latter admit the more general problem of two such monopoles with arbitrary electric and magnetic charges. It is unfortunate that the limit in which two opposite magnetic poles move into coincidence does not produce the magnetic dipole as a limiting case. The obstacle is the fact that the repulsive potential necessary to render the system separable depends on the square of the pole strength and thus does not pass over into a derivative in the limit as do the pole strengths. In fact, the repulsive potential is so strong that there are no particle trajectories in which the particle spirals around a line of magnetic force, as is characteristic of the magnetic dipole problem.[41]

We have not elaborated the solution of the monopole problem in these coordinate systems, but it is perhaps interesting to note a remark of Schweiger[42] in his monumental treatment of the ordinary Kepler problem in parabolic coordinates, that the confinement of the attracting center to the orbital plane was only a consequence of the fact that the same constant p_ϕ appeared in both separated parts of the Hamilton–Jacobi equation. In the presence of the monopole, the constants which appear are $(p_\phi + \epsilon)$ and $(p_\phi - \epsilon)$, and as we have seen, the orbital plane moves away from the origin.

One coordinate system which is of especial utility, however, is formed by the quaternionic coordinates. Although these are four-dimensional coordinates, the magnetic pole strength enters naturally as a constrained coordinate; it is the constant momentum conjugate to one of the angular quaternionic coordinates. The angular part of the equation reduces immediately to the Schrödinger equation for the spherical top, and it is in fact this interpretation which leads most naturally to the quantization of magnetic pole strength in such problems. However, there is an anomalous factor of 4 multiplying the angular momentum which prevents the monopole with a certain electrostatic potential in three-dimensional space from passing over directly into the analogous problem in four dimensions, and so one obtains the analog with "excess angular momentum." The result is that the symmetry of the Coulombic monopole can be extracted from the four-dimensional symmetry, whereas that of the harmonic monopole cannot be extracted from a symmetry group even though a calculus of ladder operators exists accounting for the degeneracy.

The monopole with a Coulombic charge, even though the repulsive centrifugal potential provides it with the symmetry group $O(4)$, contrasts in another interesting way with the ordinary hydrogen atom. Fock's explanation of the accidental degeneracy involved a stereographic projection which reduced the problem to that of force-free motion on the surface of a hypersphere. It was this simple geometric version of the problem which gave the concept of "hidden symmetry" its strong intuitive appeal. However, it is known that only the family of irreducible representations of $O(4)$ of dimension n^2 may be realized by transformations of spherical harmonics over a geometrical sphere. Since it is the family mn with $m - n = 2\epsilon$ which occurs among the wavefunctions of the monopole problem, we must conclude that there is no mapping of Fock's type to the surface of a hypersphere which will provide us with a simple geometrically symmetric version of the monopole problem.

ACKNOWLEDGMENTS

It is a pleasure to recollect a very interesting discussion with Dr. H. Kolm concerning the hunting of monopoles. M. Berrondo has been of great assistance in preparing the Hamiltonian treatment of the monopole. We have made many computations of the classical trajectories for the one- and two-center problems using the IBM 1130 computer of the Centro Nacional de Cálculo, Instituto Politécnico Nacional, whose cooperation we gratefully acknowledge.

* Based in part upon a professional thesis submitted to the Escuela Superior de Física y Matemáticas of the Instituto Politécnico Nacional in partial fulfillment of the requirements for the title "Licenciado en Física y Matemáticas."

[1] P. Stehle and M. Y. Han, Phys. Rev. **159**, 1076 (1967).
[2] D. M. Fradkin, Progr. Theoret. Phys. (Kyoto) **37**, 798 (1967); N. Mukunda, Phys. Rev. **155**, 1383 (1967); V. A. Dulock and H. V. McIntosh, Pacific J. Math. **19**, 39 (1966).
[3] V. Fock, Z. Physik **98**, 145 (1935).
[4] J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).
[5] J. Bertrand, Compt. Rend. **77**, 849 (1873).
[6] H. Poincaré, Compt. Rend. **123**, 530 (1896).
[7] J. R. Lapidus and J. L. Pietenpol, Am. J. Phys. **28**, 17 (1960).
[8] G. Nadeau, Am. J. Phys. **28**, 566 (L) (1960).

[9] B. Lehnert, *Dynamics of Charged Particles* (North-Holland Publ. Co., Amsterdam, 1964), pp. 32–34.
[10] P. A. M. Dirac, Proc. Roy. Soc. (London) **A133**, 60 (1931).
[11] Ig. Tamm, Z. Physik **71**, 141 (1931).
[12] B. O. Grönblom, Z. Physik **98**, 283 (1935); M. N. Saha, Indian J. Math. **10**, 141 (1936); P. Jordan, Ann. Physik **32**, 66 (1938); H. A. Wilson, Phys. Rev. **75**, 309 (L) (1949); J. A. Eldridge, *ibid.*, 1614 (L) (1949); M. N. Saha, *ibid.*, 1968 (L) (1949); N. F. Ramsey, *ibid.* **109**, 225 (L) (1958).
[13] M. Fierz, Helv. Phys. Acta **17**, 27 (1944).
[14] P. P. Banderet, Helv. Phys. Acta **19**, 503 (1946).
[15] Harish-Chandra, Phys. Rev. **74**, 883 (1948).
[16] C. J. Eliezer and S. K. Roy, Proc. Cambridge Phil. Soc. **58**, 401 (1962).
[17] M. A. Tuve, Phys. Rev. **43**, 770 (L) (1933); H. J. D. Cole, Proc. Cambridge Phil. Soc. **47**, 196 (1951); E. Bauer, Proc. Cambridge Phil. Soc. **47**, 777 (1951); N. C. Fitz, Jr., W. B. Good, J. L. Kasner, Jr., and A. R. Ruark, Phys. Rev. **111**, 1406 (1958); H. Bradner and W. M. Isbell, Phys. Rev. **114**, 603 (1959); R. Katz and D. R. Parnell, *ibid.* **116**, 236 (1959); M. Fidecaro, G. Finnochiaro, and G. Giacomelli, Nuovo Cimento **22**, 657 (1961); E. Goto, H. H. Kolm, and K. W. Ford, Phys. Rev. **132**, 387 (1963); E. Amaldi, G. Baroni, A, Manfredini, H. Bradner, L. Hoffman, and G. Vanderhaeghe, Nuovo Cimento **28**, 773 (1963); E. M. Purcell, G. B. Collins, T. Fuji, J. Hornbostel, and F. Turkot, Phys. Rev. **129**, 2326 (1963); W. C. Carithers, R. Stefanski, and R. K. Adair, *ibid.* **149**, 1070 (1966).
[18] W. V. R. Malkus, Phys. Rev. **83**, 899 (1951).
[19] P. A. M. Dirac, Phys. Rev. **74**, 817 (1948); N. Cabbibo and E. Ferrari, Nuovo Cimento **23**, 1147 (1962); R. A. Ferrel and J. S. Hopfield, Physics **1**, 1 (1964); D. Zwanziger, Phys. Rev. **137**, B647 (1965); D. R. Tompkins, *ibid.* **138**, B248 (1965); C. R. Hagen, *ibid.* **140**, B804 (1965); S. Weinberg, *ibid.* **138**, B988 (1965); S. Goldhaber, *ibid.* **140**, B1407 (1965); J. Schwinger, *ibid.* **144**, 1087 (1966); R. V. Tevikyan, Zh. Eksp. Teor. Fiz. **50**, 911 (1966) [Sov. Phys.—JETP **23**, 606 (1966)]; L. I. Schiff, Phys. Rev. **160**, 1257 (1967).
[20] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, London, 1961), 4th ed., p. 83.
[21] N. Ya. Vilenkin, G. J. Kunetsov, and Ya. A. Smorodinskii, Yad. Fiz. **2**, 906 (1965) [Sov. J. Nucl. Phys. **2**, 645 (1966)].
[22] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publ. Co., Reading, Mass., 1950), Chap. 8.
[23] E. Daltabuit and H. V. McIntosh, Rev. Mex. Física **16**, 105 (1967).
[24] F. Hund, Z. Physik **51**, 1 (1928).
[25] A. Cisneros and H. V. McIntosh, J. Math. Phys. **11**, 870 (1970).
[26] M. Born, *The Mechanics of the Atom* (Frederick Ungar Publ. Co., New York, 1960), Chap. 35.
[27] Ref. 26, Chap. 39.
[28] A. Cisneros and H. V. McIntosh, J. Math. Phys. **10**, 277 (1969).
[29] F. Reiche, Z. Physik **39**, 444 (1926); H. Rademacher and F. Reiche, *ibid.* **41**, 453 (1927); R. de L. Kronig and I. I. Rabi, Phys. Rev. **29**, 262 (1927); O. Klein, Z. Physik **58**, 730 (1929).
[30] W. Pauli, Helv. Phys. Acta **12**, 147 (1939); M. L. Whippman, Am. J. Phys. **34**, 656 (1966).
[31] W. H. Shaffer, J. Mol. Spectry. **1**, 69 (1957).
[32] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Co., New York, 1955), 2nd ed.
[33] S. Bell and P. A. Warsop, J. Mol. Spectry. **22**, 360 (1967).
[34] R. Musto, Phys. Rev. **148**, 1274 (1966).
[35] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).
[36] D. G. Burkhard, J. Mol. Spectry. **2**, 187 (1958).
[37] E. Schrödinger, Proc. Roy. Irish Acad. **46A**, 9 (1940).
[38] W. Pauli, Z. Physik **36**, 336 (1926).
[39] H. Bacry, H. Ruegg, and J. M. Souriau, Commun. Math. Phys. **3**, 323 (1966).
[40] A. Peres, Phys. Rev. **167**, 1449 (1968).
[41] M. Sandoval-Vallarta, "Theory of the Geomagnetic Effects of Cosmic Radiation" in *Handbuch der Physik* (Springer-Verlag, Berlin, 1961), Vol. 46/1.
[42] F. Schweiger, Acta Phys. Austriaca **19**, 138 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On Heat Conduction and Wave Propagation in Rigid Solids

D. B. BOGY AND P. M. NAGHDI
Department of Mechanical Engineering, University of California, Berkeley 94720

(Received 25 March 1969)

This paper is concerned with conduction of heat and the related problem of propagation of thermal waves in stationary rigid solids. Special attention is given to rate-dependent response and the ensuing conditions of propagation in the conducting medium. The case in which the response of the medium is characterized by functions that are nonlinear in temperature and temperature gradient but are linear in the rate of temperature is studied in detail. The conclusions reached are also discussed in the context of the infinitesimal theory.

1. INTRODUCTION

This paper is concerned with heat conduction and propagation of thermal waves in rigid solids. Our approach and point of view is that of recent developments in continuum thermodynamics. We employ the balance of energy and the (Clausius–Duhem) entropy-production inequality as well as constitutive assumptions characterizing the local thermal behavior of the material. For example, if temperature, temperature gradient, and rate of temperature are taken as independent variables, then the thermal variables such as specific energy, specific entropy, and heat flux are determined by constitutive equations. When the constitutive assumptions include rate of temperature as an independent variable, there arises a local production of entropy [see Eq. (2.2)], in addition to the entropy production caused by conduction, which is inherent in the usual rate-independent response. The rate-dependent thermal response studied here and the corresponding generalization of the nonlinear heat conduction equation leads to a prediction of finite-speed thermal waves, as well as to other novel phenomena associated with local thermal dissipation.

Before describing the scope of the paper in detail, it is desirable to recall for background information certain other developments which bear on conduction of heat in rigid solids. Within the framework of modern developments in continuum physics, attention was largely confined until recently to purely mechanical constitutive equations. Arguments for obtaining necessary and sufficient conditions for the validity of the entropy production inequality, limited to nonlinear elastic materials with heat conduction and viscosity, were given by Coleman and Noll.[1] In a paper by Coleman and Mizel[2] there is a more extensive discussion of generalizations of the classical theory of linearly viscous fluids with linear heat conduction, in which the independent thermodynamic variables are temperature and temperature gradient. Among subsequent developments, we mention Coleman's[3] work on thermodynamics of materials with fading memory, in which the thermomechanical constitutive equations are assumed to be functionals over the time histories of the chosen independent variables. Similar constitutive assumptions are utilized by Coleman and Gurtin[4] in their investigation of the thermal behavior of rigid heat conductors.

Certain types of well-known mechanical behavior—typical of materials of rate type—such as that for linear (Newtonian) viscous fluids are appropriately characterized by rate-dependent functions rather than by memory functionals.[5] Likewise, we may expect certain thermal behavior of rigid conductors to be best described by rate-dependent functions, and this is the underlying premise of the generalization of the classical heat-conduction equation sought here.

That a generalization of the classical heat conduction equation for solids is desirable has been recognized for several years. This view is based primarily on the fact that the classical equation is parabolic and does not admit the possibility of finite speeds of propagation of thermal pulses. Most of the previous attempts to alter the heat conduction equation so as to predict thermal waves, often referred to in the literature as "second sound," have been based on an *ad hoc* generalization of Fourier's heat conduction law; see, for example, Chester[6] and Ulbrich.[7] This generalization, referred to by Ulbrich[7] as Vernotte's hypothesis,[8] appears as

$$\mathbf{q} + \tau\dot{\mathbf{q}} = k \operatorname{grad} \theta \tag{1.1}$$

for rigid conductors, where $\mathbf{q}$, $\dot{\mathbf{q}}$, θ represent the heat flux, the time rate of heat flux, and temperature, while τ, k are material constants. With reference to (1.1) Ulbrich states[7]: "Although Vernotte's proposed revision of Fourier's hypothesis adequately circumvents the paradox of infinite velocity, no apparent physical justification can be offered for the addition of the second term . . . ."

A different approach from that based on (1.1), but

with the same objective, is taken by Kaliski.[9] He starts with the assumption that the heat conduction equation should be a second-order hyperbolic equation, and then generalizes Onsager's reciprocal relations and the associated form of the entropy inequality so as to accommodate this assumption.

We also draw attention to a very recent paper by Gurtin and Pipkin,[10] whose work is motivated by (1.1). They develop a nonlinear theory for rigid heat conductors in which the constitutive assumptions are in the form of functionals over the temperature history. They show that an equation of the form of (1.1) is a special case of their linearized constitutive equation for the heat flux $\mathbf{q}$. Their associated generalized heat-conduction equation in the linear theory is given by

$$c\ddot{\theta}(\mathbf{x}, t) + \beta(0)\dot{\theta}(\mathbf{x}, t) + \int_0^\infty \beta'(s)\dot{\theta}(\mathbf{x}, t - s)\, ds = a(0)\Delta\theta(\mathbf{x}, t) + \int_0^\infty a'(s)\Delta\theta(\mathbf{x}, t - s)\, ds + \dot{r}(\mathbf{x}, t), \tag{1.2}$$

where $a(s)$ and $\beta(s)$ are called the heat-flux relaxation function and the energy relaxation function. This equation is in a form that predicts finite speeds of propagation of thermal waves. An unusual feature of (1.2), however, is the presence of $\dot{r}$ (rather than r) which, evidently, is due to the fact that (1.2) is obtained from the first time derivative of the energy equation rather than from the energy equation itself.

In the present paper, which is concerned with rigid stationary conductors, we use, in general, *direct* (coordinate-free) notation. Thus, vectors and points in 3-dimensional Euclidean space are denoted by boldface Latin lower case letters, while boldface Latin capital letters are used to designate tensors of order two; also, Greek lower case letters are used (although not exclusively) for scalar thermodynamic variables. An exception to the above notation occurs when we write the energy equation as a partial differential equation in the temperature. Then, for ease of comparison and analysis, all quantities in the energy equation are expressed in terms of their Cartesian components.

In Sec. 2, following some preliminaries, we include, for clarity and later comparison, a discussion of nonlinear constitutive equations for rate-independent response,[11] when temperature and temperature gradient are taken as independent variables. A generalization, in which the independent variables include the rate of temperature, is introduced and developed in Sec. 3. This leads to a local thermal dissipation of energy and to the possibility of temperature changes without heat flow. The appropriate "heat-conduction equation" is then derived and is used in Sec. 4 to investigate the existence of finite speeds or propagation of second-order discontinuities in the temperature.

In order to arrive at more explicit conclusions, we specialize in Sec. 5 the results of Sec. 3 by assuming that the dependence on the temperature rate is linear. We then re-examine the energy equation for the case of isotropic materials with a center of symmetry and, in particular, make certain observations regarding the speed of propagation of thermal waves in one space dimension. Of special significance is the result in Sec. 5 that real wave speeds are possible only if there is a sufficiently large temperature rate present in the medium through which the wave is to propagate.

Finally, motivated partially by the results of Sec. 5, we discuss in Sec. 6 the possibility of wave propagation in an infinitesimal theory, as well as a theory governing infinitesimal time-dependent temperature variations superposed on a finite (nonuniform) equilibrium temperature change from the reference temperature. We show that the entropy production inequality precludes the possibility of thermal waves in such linearized theories which are obtained by systematic linearization of the results of Sec. 5.

2. PRELIMINARIES: RATE-INDEPENDENT RESPONSE

Let $\mathbf{x}$ be the (3-dimensional) position vector relative to a fixed rectangular Cartesian coordinate system, and let t denote time. Since we shall be concerned only with rigid stationary conductors, no distinction between material particles and their positions in space is necessary. Let ϵ, η, θ, and $\mathbf{q}$ denote the internal energy per unit mass, the entropy per unit mass, the temperature (assumed positive), and the heat-flux vector, all functions of $\mathbf{x}$, t in a prescribed domain. These quantities, together with the heat supply per unit mass r, must satisfy for each $\mathbf{x}$, t the energy equation[12]

$$-\operatorname{div}\mathbf{q} + \rho r = \rho\dot{\epsilon}, \tag{2.1}$$

and the entropy production inequality which, for later convenience, we write as

$$\gamma = \gamma_{\text{loc}} + \gamma_{\text{con}} \geq 0, \tag{2.2a}$$

$$\rho\gamma_{\text{loc}} = \rho\dot{\eta} + \theta^{-1}\operatorname{div}\mathbf{q} - \rho r\theta^{-1}, \tag{2.2b}$$

$$\rho\gamma_{\text{con}} = -\theta^{-2}\mathbf{q}\cdot\mathbf{g}, \tag{2.2c}$$

where γ_{loc} and γ_{con} represent the local entropy production and the production of entropy due to conduction, the abbreviation $\mathbf{g} = \operatorname{grad}\theta$ stands for the temperature gradient, a superposed dot denotes partial differentiation with respect to t, and ρ is the

mass density, a function of $\mathbf{x}$ only for rigid stationary bodies. Combining (2.1) and (2.2), we obtain the inequality

$$-\rho\dot{\epsilon} + \rho\theta\dot{\eta} - \theta^{-1}\mathbf{q}\cdot\mathbf{g} \geq 0. \tag{2.3}$$

In terms of the free energy function ψ, defined by

$$\psi = \epsilon - \eta\theta, \tag{2.4}$$

the energy equation (2.1) and the inequality (2.3) can be expressed in the alternative forms

$$-\operatorname{div}\mathbf{q} + \rho r = \rho(\dot{\psi} + \eta\dot{\theta} + \dot{\eta}\theta) \tag{2.5}$$

and

$$-\rho\dot{\psi} - \rho\eta\dot{\theta} - \theta^{-1}\mathbf{q}\cdot\mathbf{g} \geq 0. \tag{2.6}$$

Suppose θ and $\mathbf{g}$ are taken as independent variables. Then, the rate-independent response of a rigid conductor is characterized by constitutive assumptions of the form[13]

$$\psi = \psi(\theta, \mathbf{g}), \tag{2.7a}$$

$$\eta = \eta(\theta, \mathbf{g}), \tag{2.7b}$$

$$\mathbf{q} = \mathbf{q}(\theta, \mathbf{g}). \tag{2.7c}$$

From (2.7) and the inequality (2.6), it can be shown that ψ must be independent of $\mathbf{g}$ and that[14]

$$\psi = \psi(\theta), \tag{2.8a}$$

$$\eta = -\frac{\partial\psi}{\partial\theta} = \eta(\theta), \tag{2.8b}$$

$$-\theta^{-1}\mathbf{q}\cdot\mathbf{g} \geq 0. \tag{2.8c}$$

Thus, the entropy is determined by the free energy, a function of temperature only, and, according to (2.8c), heat cannot flow in a direction of increasing temperature.

Let the symmetry group of the material be defined by the set of all time-dependent orthogonal second-order tensors $\mathbf{A}$. Then, under a change of reference frame characterized by $\mathbf{A}$, the scalars ϵ, η, θ, ψ are unaffected, but the vectors $\mathbf{q}$, $\mathbf{g}$ transform according to

$$\mathbf{q} \to \mathbf{A}\mathbf{q}, \quad \mathbf{g} \to \mathbf{A}\mathbf{g}. \tag{2.9}$$

The constitutive equations (2.7) are invariant under such a change of frame if and only if, for each $\mathbf{A}$ and all values of $\mathbf{q}$ and $\mathbf{g}$,

$$\mathbf{A}\mathbf{q}(\theta, \mathbf{g}) = \mathbf{q}(\theta, \mathbf{A}\mathbf{g}). \tag{2.10}$$

If, in particular, the material has a center of symmetry, the identity (2.10) must be satisfied also for $\mathbf{A} = -\mathbf{1}$, where $\mathbf{1}$ is the unit tensor, so that

$$-\mathbf{q}(\theta, \mathbf{g}) = \mathbf{q}(\theta, -\mathbf{g}). \tag{2.11}$$

It follows that if $\mathbf{q}$ is a continuous function of $\mathbf{g}$ at $\mathbf{g} = \mathbf{0}$, then

$$\mathbf{q}(\theta, \mathbf{0}) = \mathbf{0}, \tag{2.12}$$

i.e., there can be no heat flow in the absence of a temperature gradient.[15]

From the energy equation (2.5), with (2.7) and (2.8), there follows in component form

$$\left(\frac{\partial q_k}{\partial g_l}\right)\theta_{,lk} + \left(\frac{\partial q_k}{\partial\theta}\right)\theta_{,k} - \rho\theta\left(\frac{\partial^2\psi}{\partial\theta^2}\right)\dot{\theta} = \rho r, \tag{2.13}$$

where q_k, g_l ($\equiv \theta_{,l}$), are the Cartesian components of $\mathbf{q}$, $\mathbf{g}$, a comma preceding an index denotes partial differentiation with respect to rectangular Cartesian coordinates x_k, $k = 1, 2, 3$, and the usual summation is implied by repeated indices. When $\mathbf{q}$ and ψ in (2.7c) and (2.8a) are given explicit representations in terms of θ and $\mathbf{g}$, (2.13) provides a quasilinear partial differential equation to be satisfied by θ. From (2.13), together with (2.12), it follows that when the heat supply r vanishes in rigid conductors having the rate-independent-type thermal response (2.7), there can be no time dependence in the temperature field without a corresponding spatial dependence,[16] i.e., there can be no change in temperature except that resulting from the flow of heat. Correspondingly, in view of (2.8c) and (2.2c), the entropy production is caused by conduction alone, and there is no local entropy production.

The residual energy equation (2.13) is second order in its spatial derivatives and only first order in its time derivative and, therefore, clearly cannot predict thermal waves (propagating second-order discontinuities) with real finite wave speeds. This will be elaborated upon in Secs. 4–6.

We close this section by noting that the classical linear Fourier heat conduction equation results from the linearization of (2.13), when (2.7c) takes the special form represented by the Fourier law

$$\mathbf{q}(\theta, \mathbf{g}) = -\mathbf{K}(\theta)\mathbf{g} \tag{2.14}$$

and $\psi(\theta)$ in (2.8a) is assumed to be polynomial.

3. DEPENDENCE ON TEMPERATURE RATE

We consider here a generalization of the constitutive assumptions (2.7) by including the first time derivative of temperature, as well as the temperature and its gradient, as independent variables. Thus, instead of (2.7), we now assume that

$$\psi = \psi(\theta, \mathbf{g}, \dot{\theta}), \tag{3.1a}$$

$$\eta = \eta(\theta, \mathbf{g}, \dot{\theta}), \tag{3.1b}$$

$$\mathbf{q} = \mathbf{q}(\theta, \mathbf{g}, \dot{\theta}), \tag{3.1c}$$

so that the inequality (2.6) becomes

$$-\rho\left(\frac{\partial\psi}{\partial\theta}+\eta\right)\dot{\theta}-\rho\left(\frac{\partial\psi}{\partial\mathbf{g}}\right)\cdot\dot{\mathbf{g}}-\rho\left(\frac{\partial\psi}{\partial\dot{\theta}}\right)\ddot{\theta}-\frac{1}{\theta}\mathbf{q}\cdot\mathbf{g}\geq 0. \quad (3.2)$$

With θ, $\mathbf{g}$, and $\dot{\theta}$ fixed, it follows from (3.1) that all terms in (3.2) are determined except $\ddot{\theta}$, which may assume arbitrary values. In order for (3.2) to be satisfied for all $\ddot{\theta}$, its coefficient $(\partial\psi/\partial\dot{\theta})$ must vanish, i.e., ψ must be independent of $\dot{\theta}$. Similarly, ψ cannot depend on $\mathbf{g}$, so that (3.1a) and (3.2) reduce to

$$\psi=\psi(\theta) \quad (3.3)$$

and

$$-\rho\left(\frac{\partial\psi}{\partial\theta}+\eta\right)\dot{\theta}-\frac{1}{\theta}\mathbf{q}\cdot\mathbf{g}\geq 0, \quad (3.4)$$

respectively.

Next, we decompose η into two parts $\eta^{(0)}$, $\eta^{(e)}$, defined by[17]

$$\eta(\theta,\mathbf{g},\dot{\theta})=\eta^{(0)}(\theta)+\eta^{(e)}(\theta,\mathbf{g},\dot{\theta}),$$
$$\eta^{(0)}(\theta)=\eta(\theta,\mathbf{0},0). \quad (3.5)$$

Assuming that η is continuous in $\dot{\theta}$ at $\dot{\theta}=0$, we have,[18] in the limit as $\alpha\to 0$,

$$\eta(\theta,\mathbf{g},\alpha\dot{\theta})=\eta(\theta,\mathbf{g},0)+o(1), \quad (3.6)$$

so that

$$\eta^{(e)}(\theta,\mathbf{0},\alpha\dot{\theta})=o(1), \quad (3.7)$$

where α is a real number. Now, in (3.4) put $\mathbf{g}=\mathbf{0}$, replace $\dot{\theta}$ by $\alpha\dot{\theta}$, and use (3.5) and (3.7) to obtain

$$-\rho\left(\frac{\partial\psi}{\partial\theta}+\eta^{(0)}\right)\alpha\dot{\theta}+o(\alpha)\geq 0, \quad (3.8)$$

where $\lim o(\alpha)/\alpha=0$, as $\alpha\to 0$. It then follows from (3.8) that

$$\eta^{(0)}=-\frac{\partial\psi}{\partial\theta}, \quad (3.9)$$

and the inequality (3.4) assumes the form

$$-\rho\eta^{(e)}\dot{\theta}-\theta^{-1}\mathbf{q}\cdot\mathbf{g}\geq 0. \quad (3.10)$$

In view of (3.1), (3.3), (3.5), and (3.9), the energy equation (2.5) becomes (in component form)

$$\left(\frac{\partial q_k}{g_i}\right)\theta_{,ik}+\left(\frac{\partial q_k}{\partial\dot{\theta}}+\rho\theta\frac{\partial\eta^{(e)}}{\partial g_k}\right)\dot{\theta}_{,k}+\rho\theta\left(\frac{\partial\eta^{(e)}}{\partial\dot{\theta}}\right)\ddot{\theta}+\left(\frac{\partial q_k}{\partial\theta}\right)\theta_{,k}+\rho\left(\eta^{(e)}+\theta\frac{\partial\eta^{(e)}}{\partial\theta}-\theta\frac{\partial^2\psi}{\partial\theta^2}\right)\dot{\theta}=\rho r, \quad (3.11)$$

which is the counterpart of (2.13) under the assumptions (3.1).[19]

Many of the conclusions drawn in Sec. 2 are altered for the rate-dependent response under consideration. In particular, even though ψ still reduces to a function only of θ, we cannot conclude the same about η which remains dependent on $\mathbf{g}$ and $\dot{\theta}$. Furthermore, it is clear from (2.2) and (3.10) that the inclusion of $\dot{\theta}$ as an independent variable in constitutive assumptions (3.1) has led to a local entropy production. In addition, the total entropy is no longer expressible in terms of the free energy as in the case not only of the rate-independent-type response but also of the more general response characterized by a functional over the time history of temperature as investigated by Coleman and Gurtin.[20]

The requirements that the constitutive equations (3.3), and (3.1b), and (3.1c) satisfy the appropriate transformation relations under a change of reference frame (imposed by the symmetry group of the material), that the heat flux function $\mathbf{q}$ be continuous in $\mathbf{g}$ at $\mathbf{g}=\mathbf{0}$, and that the material possess a center of symmetry imply [with an argument parallel to that used to arrive at (2.12)] that $\mathbf{q}$ must be an odd, and η an even, function of $\mathbf{g}$ with

$$\mathbf{q}(\theta,\mathbf{0},\dot{\theta})=\mathbf{0}. \quad (3.12)$$

Therefore, as in the rate-independent case, there can be no heat flux corresponding to zero temperature gradient. We also observe the inequalities

$$\mathbf{q}(\theta,\mathbf{g},0)\cdot\mathbf{g}\leq 0, \quad \eta^{(e)}(\theta,\mathbf{0},\dot{\theta})\dot{\theta}\leq 0, \quad (3.13)$$

which follow from (3.10), but we note that, in general, we cannot conclude that $\mathbf{q}\cdot\mathbf{g}\leq 0$.

4. PROPAGATION OF THERMAL WAVES

We investigate here the possibilities of the existence of propagating second order discontinuities, i.e., discontinuities in the second derivatives, in the temperature field. For convenience, however, we confine our analysis to one space dimension x. Writing q and g for the x components of $\mathbf{q}$ and $\mathbf{g}$, we obtain from (3.11) its 1-dimensional counterpart

$$\left(\frac{\partial q}{\partial g}\right)\theta_{,xx}+\left(\frac{\partial q}{\partial\dot{\theta}}+\rho\theta\frac{\partial\eta^{(e)}}{\partial g}\right)\dot{\theta}_{,x}+\rho\left(\theta\frac{\partial\eta^{(e)}}{\partial\dot{\theta}}\right)\ddot{\theta}+\left(\frac{\partial q}{\partial\theta}\right)\theta_{,x}+\rho\left(\eta^{(e)}+\theta\frac{\partial\eta^{(e)}}{\partial\theta}-\theta\frac{\partial^2\psi}{\partial\theta^2}\right)\dot{\theta}=\rho r. \quad (4.1)$$

We assume that θ, g, $\dot{\theta}$, and r are continuous functions of x, t and denote by brackets the value at the propagating discontinuity of the jump of whatever

quantity it brackets. If we allow discontinuities in second order derivatives of θ and assume that the first derivatives are continuous,

$$\left(\frac{\partial q}{\partial g}\right)[\theta_{,xx}] + \left(\frac{\partial q}{\partial \dot\theta} + \rho\theta\frac{\partial \eta^{(e)}}{\partial g}\right)[\dot\theta_{,x}] + \rho\left(\theta\frac{\partial \eta^{(e)}}{\partial \dot\theta}\right)[\ddot\theta] = 0 \quad (4.2)$$

follows from (4.1). Using standard jump conditions,[21] we have

$$[\dot\theta_{,x}] = -U[\theta_{,xx}], \quad [\ddot\theta] = U^2[\theta_{,xx}], \quad (4.3)$$

where U is the speed of propagation of the discontinuity. Substitution of (4.3) into (4.2) gives

$$U = \left\{\frac{\partial q}{\partial \dot\theta} + \rho\theta\frac{\partial \eta^{(e)}}{\partial g} \pm \left[\left(\frac{\partial q}{\partial \dot\theta} + \rho\theta\frac{\partial \eta^{(e)}}{\partial g}\right)^2 - 4\rho\theta\frac{\partial q}{\partial g}\frac{\partial \eta^{(e)}}{\partial \dot\theta}\right]^{\frac{1}{2}}\right\} \Big/ 2\rho\theta\frac{\partial \eta^{(e)}}{\partial \dot\theta}, \quad (4.4)$$

provided $[\theta_{,xx}]$ and $\partial\eta^{(e)}/\partial\dot\theta$ are not zero. In the event $\partial\eta^{(e)}/\partial\dot\theta$ does vanish, (4.2) and (4.3) yield

$$U = \frac{\partial q}{\partial g}\Big/\left(\frac{\partial q}{\partial \dot\theta} + \rho\theta\frac{\partial \eta^{(e)}}{\partial g}\right). \quad (4.5)$$

The terms on the right-hand side of (4.4) and (4.5) are evaluated at the propagating discontinuity, so that U can be considered alternatively as a function only of x or t.[22]

The following observations are apparent from (4.4) and (4.5):

(i) Equation (4.4) shows that second order discontinuities in temperature, i.e., thermal waves, may possibly occur with two distinct real wave speeds. This depends, of course, on the relative magnitudes and signs of the terms in (4.4) and will be investigated further for particular representations of $\eta^{(e)}$, ψ, and q in the following sections.

(ii) It follows from (4.5) that if $\eta^{(e)}$ is independent of $\dot\theta$, then only one wave-speed is possible. In fact, we see that thermal waves may occur when $\eta^{(e)}$ vanishes altogether. Thus, thermal dissipation caused by a local entropy production, represented by the first term of the inequality in (3.10), may not be necessary for thermal waves, provided q remains dependent on $\dot\theta$ as well as g. Equation (4.5) also shows that when $\eta^{(e)}$ vanishes and q does not depend on $\dot\theta$, as in the rate-independent response discussed in Sec. 2, no finite-speed thermal wave can occur.

5. LINEARITY IN TEMPERATURE RATE

Recalling (3.1b), (3.1c), and (3.3) we introduce in this section specific assumptions regarding the dependence of the response functions on the temperature rate. Before doing so, we decompose $\mathbf{q}$ [as was done with η in (3.5)] into two parts $\mathbf{q}^{(0)}$, $\mathbf{q}^{(e)}$ defined by

$$\mathbf{q}(\theta, \mathbf{g}, \dot\theta) = \mathbf{q}^{(0)}(\theta, \mathbf{g}) + \mathbf{q}^{(e)}(\theta, \mathbf{g}, \dot\theta), \quad \mathbf{q}^{(0)}(\theta, \mathbf{g}) = \mathbf{q}(\theta, \mathbf{g}, 0). \quad (5.1)$$

Then, (3.12) implies

$$\mathbf{q}(\theta, \mathbf{0}, \dot\theta) = \mathbf{q}^{(0)}(\theta, \mathbf{0}) + \mathbf{q}^{(e)}(\theta, \mathbf{0}, \dot\theta) = \mathbf{0}. \quad (5.2)$$

Since $\mathbf{q}^{(e)}(\theta, \mathbf{g}, 0)$ vanishes by (5.1), it follows that

$$\mathbf{q}(\theta, \mathbf{0}, 0) = \mathbf{q}^{(0)}(\theta, \mathbf{0}) = \mathbf{0}. \quad (5.3)$$

From (3.5), (5.2), and (5.3) we have

$$\eta^{(e)}(\theta, \mathbf{0}, 0) = 0, \quad (5.4a)$$

$$\mathbf{q}^{(0)}(\theta, \mathbf{0}) = \mathbf{0}, \quad (5.4b)$$

$$\mathbf{q}^{(e)}(\theta, \mathbf{0}, \dot\theta) = \mathbf{0}. \quad (5.4c)$$

Now ψ, and consequently $\eta^{(0)}$, are arbitrary functions of θ, while $\mathbf{q}^{(0)}$ is any continuous function of θ and $\mathbf{g}$ which is odd in $\mathbf{g}$ and satisfies (5.4b). We assume that the "extra" parts of entropy and heat flux, namely $\eta^{(e)}$ and $\mathbf{q}^{(e)}$, are nonlinear functions of θ and $\mathbf{g}$, but are linear functions of degree one in $\dot\theta$ so that

$$\eta^{(e)}(\theta, \mathbf{g}, \dot\theta) = -n(\theta, \mathbf{g})\dot\theta, \quad \mathbf{q}^{(e)}(\theta, \mathbf{g}, \dot\theta) = \mathbf{I}(\theta, \mathbf{g})\dot\theta. \quad (5.5)$$

Assuming further that $\mathbf{q}^{(0)}$ and $\mathbf{q}^{(e)}$ are polynomials in $\mathbf{g}$ and recalling (5.4), we arrive at the forms

$$\mathbf{q}^{(0)}(\theta, \mathbf{g}) = -\mathbf{K}(\theta, \mathbf{g})\mathbf{g}, \quad \mathbf{q}^{(e)}(\theta, \mathbf{g}, \dot\theta) = \mathbf{H}(\theta, \mathbf{g})\mathbf{g}\dot\theta. \quad (5.6)$$

By (3.13), (5.1), and (5.2), the functions $n(\theta, \mathbf{g})$ $\mathbf{K}(\theta, \mathbf{g})$ in (5.5), (5.6) must satisfy

$$n(\theta, \mathbf{0}) \geq 0, \quad \mathbf{K}(\theta, \mathbf{g})\mathbf{g} \cdot \mathbf{g} \geq 0. \quad (5.7)$$

In view of (5.6), the energy equation (3.11) now assumes the form

$$\left(-\frac{\partial}{\partial g_l}(K_{kj}g_j) + \frac{\partial}{\partial g_l}(H_{kj}g_j)\dot\theta\right)\theta_{,kl} + \left(H_{kj}g_j - \rho\theta\dot\theta\frac{\partial n}{\partial g_k}\right)\dot\theta_{,k} - \rho n\theta\ddot\theta + \left[-\left(\frac{\partial K_{kj}}{\partial\theta}\right)g_j + \left(\frac{\partial H_{kj}}{\partial\theta}\right)\dot\theta g_j\right]\theta_{,k} - \rho\left(n\dot\theta + \theta\dot\theta\frac{\partial n}{\partial\theta} + \theta\frac{\partial^2\psi}{\partial\theta^2}\right)\dot\theta = \rho r, \quad (5.8)$$

which is a quasilinear second-order partial-differential equation to be satisfied by the temperature. The classification and type of initial-boundary-value

problem appropriate to such an equation has been extensively discussed in the literature.[23]

The essential character of Eq. (5.8) will not be altered if we make some further simplifying assumptions. In particular, we assume, henceforth, that the material is isotropic with a center of symmetry so that n, $\mathbf{K}$, and $\mathbf{H}$ in (5.5) and (5.6) are isotropic functions of θ and $g_i g_i$. It is tempting at this stage to introduce a further simplification and to consider the case in which n, $\mathbf{K}$, and $\mathbf{H}$ are functions of θ only. However, for the reason that will become apparent later, we limit the discussion to a slightly more general case in which n, $\mathbf{K}$, and $\mathbf{H}$ have the forms[24]

$$n(\theta, \mathbf{g}) = \nu(\theta) + \mu(\theta) g_i g_i, \tag{5.9a}$$

$$K_{kj}(\theta, \mathbf{g}) = [\kappa(\theta) + \beta(\theta) g_i g_i]\delta_{kj}, \tag{5.9b}$$

$$H_{kj}(\theta, \mathbf{g}) = [h(\theta) + \gamma(\theta) g_i g_i]\delta_{kj}. \tag{5.9c}$$

The inequality (3.10) now becomes

$$\rho\theta[\nu + \mu g_i g_i]\dot{\theta}^2 + \kappa g_i g_i + \beta g_i g_i g_k g_k - h g_i g_i \dot{\theta} - \gamma g_i g_i g_k g_k \dot{\theta} \geq 0, \tag{5.10}$$

from which we deduce the restrictions

$$\gamma = 0, \quad \nu \geq 0, \quad \kappa \geq 0, \quad \beta \geq 0, \quad \mu \geq 0. \tag{5.11}$$

With the help of (5.11), we can also deduce the necessary condition

$$h^2 g_i g_i \leq 4\rho\theta(\kappa + \beta g_i g_i)(\nu + \mu g_k g_k). \tag{5.12}$$

Although the inequality (5.12) depends on $g_i g_i$, it places no restriction on h in the limits as $g_i g_i \to 0, \infty$. It is not difficult to see that (5.12) permits nonzero values of h. For example, a sufficient condition (independent of $g_i g_i$) for the satisfaction of (5.12) is

$$h^2 \leq 4\rho\theta(\nu\beta + \mu\kappa). \tag{5.13}$$

Substituting (5.9) and (5.11) into (5.8), we obtain

$$\begin{aligned}(-\kappa - \beta\theta_{,i}\theta_{,i} + h\dot{\theta})\theta_{,kk} - 2\beta\theta_{,k}\theta_{,l}\theta_{,kl} \\ + (h - 2\rho\theta\dot{\theta}\mu)\theta_{,k}\dot{\theta}_{,k} - \rho\theta(\nu + \mu\theta_{,i}\theta_{,i})\ddot{\theta} \\ + \left(-\frac{\partial\kappa}{\partial\theta} - \frac{\partial\beta}{\partial\theta}\theta_{,i}\theta_{,i} + \frac{\partial h}{\partial\theta}\dot{\theta}\right)\theta_{,k}\theta_{,k} \\ - \rho\left(\nu\dot{\theta} + \mu\theta_{,i}\theta_{,i}\dot{\theta} + \theta\dot{\theta}\frac{\partial\nu}{\partial\theta}\right. \\ \left. + \theta\dot{\theta}\frac{\partial\mu}{\partial\theta}\theta_{,i}\theta_{,i} + \theta\frac{\partial^2\psi}{\partial\theta^2}\right)\dot{\theta} = \rho r.\end{aligned} \tag{5.14}$$

We now use the 1-dimensional form of (5.14) to examine more closely the propagation of thermal waves. Following the procedure that led to (4.4) and (4.5), we obtain for the speed of propagation U of second-order discontinuities the expression

$$\begin{aligned}U = \{-(h - 2\rho\theta\mu\dot{\theta}) \pm (h - 2\rho\theta\mu\dot{\theta})^2(\theta_{,x})^2 \\ - 4\rho\theta[\nu + \mu(\theta_{,x})^2][\kappa + 3\beta(\theta_{,x})^2 - h\dot{\theta}]\}^{\frac{1}{2}}\}/ \\ 2\rho\theta[\nu + \mu(\theta_{,x})^2],\end{aligned} \tag{5.15}$$

provided that $\nu + \mu(\theta_{,x})^2$ does not vanish. If this expression does vanish, no finite wave speed is possible corresponding to (4.5), since by (5.12) the vanishing of $\nu + \mu(\theta_{,x})^2$ implies $(h - 2\rho\theta\mu\dot{\theta})\theta_{,x}$ is also zero.

Equation (5.15) reveals an interesting phenomenon. No real wave speed can result if both $\theta_{,x}$ and $\dot{\theta}$ are zero "in front of the wave," that is, if the conductor into which the wave is to propagate is in a uniform equilibrium (time-independent) thermal state. On the other hand, it appears that real wave speeds can occur provided the temperature gradient and the temperature rate in the conductor are large enough; this will be further discussed in Sec. 6.

Before closing this section, we return to (5.9) and observe that if we had assumed the coefficient functions n, $\mathbf{K}$, and $\mathbf{H}$ to depend on θ only [or, equivalently, if μ, β, and γ were absent from (5.9)], then from the inequality (3.10) we would have found $h = 0$ and no propagation of thermal waves would be possible. Moreover, within the limitation of the special forms (5.9) and recalling the inequality (5.10), it becomes evident that of the terms in (5.9) which contain $g_i g_i$ only the one in $\mathbf{K}$ seems to be essential for possible existence of a real wave speed.

6. FURTHER REMARKS: PROPAGATION CONDITIONS IN A LINEARIZED THEORY

In this section we discuss the possibilities of propagation of thermal waves in the context of a linearized theory governing infinitesimal time-dependent temperature variations. Since the propagation condition in (5.15) depends on the temperature gradient and the temperature rate, we permit the temperature field upon which the infinitesimal time-dependent field is superposed to be nonuniform but time-independent, and we place no restriction on the magnitude of this equilibrium temperature variation from the reference temperature. Therefore, we set

$$\theta(\mathbf{x}, t) = \tilde{\theta}(\mathbf{x}) + \epsilon\theta'(\mathbf{x}, t), \tag{6.1}$$

and consider temperature variations from equilibrium that are $O(\epsilon)$ in the limit as $\epsilon \to 0$. It follows from (6.1) that $\dot{\theta}$ is $O(\epsilon)$, whereas $\theta_{,i}$ is $O(1)$ in general. In the special case when $\tilde{\theta}$ is uniform, $\theta_{,i}$ is also $O(\epsilon)$.

It is not necessary to develop here the complete linearized theory in order to point out the significant

results concerning the possibility of wave propagation, as predicted by the linear equation for $\theta'(\mathbf{x}, t)$, which emerges from the systematic linearization of (5.14). For brevity we omit such a development and proceed directly to a discussion of the results by a further examination of (5.15).

The classification of the 1-dimensional form of the differential equation (5.14) is determined by the sign of the expression within the braces in (5.15). This expression can be written as the sum of two terms, namely

$$T_1 = (h\theta_{,x})^2 - 4\rho\theta[\nu + \mu(\theta_{,x})^2][\kappa + \beta(\theta_{,x})^2],$$

$$T_2 = 4\rho\theta\{\mu(\theta_{,x})^2[\rho\theta\mu\dot{\theta}^2 - 2\beta(\theta_{,x})^2] + \nu[h\dot{\theta} - 2\beta(\theta_{,x})^2]\}, \quad (6.2)$$

and the condition for propagation of finite-speed thermal waves becomes

$$T_1 + T_2 > 0. \quad (6.3)$$

Comparing T_1 in (6.2) with the inequality (5.12) when $g_i g_i = (\theta_{,x})^2$, we see that $T_1 \leq 0$. Therefore, in order for waves to be possible, it is necessary that

$$T_2 > -T_1 \geq 0. \quad (6.4)$$

This condition is possible provided that $\nu h\dot{\theta}$ is large enough, since $\dot{\theta}$ appears in T_2 but not in T_1. Hence, (5.14) can predict waves, as already noted at the end of the previous section.

Consider now the propagation condition (6.4) subject to (6.1) which restricts the time-dependent temperature variation to $O(\epsilon)$. In the case when $\tilde{\theta}$ is merely a uniform reference temperature, the expression T_1 in (6.2) when linearized is $O(1)$, while T_2 is only $O(\epsilon)$. Clearly, the propagation condition (6.4) cannot be satisfied when $\tilde{\theta}$ is uniform in the linear theory. In the more general case,when $\tilde{\theta}$ is nonuniform and $\theta_{,i}$ is $O(1)$, it follows that both T_1, T_2 are $O(1)$ in the linearization; however, the $O(1)$ term in T_2 is $-8\rho\tilde{\theta}\tilde{\beta}[\tilde{\nu} + \tilde{\mu}(\tilde{\theta}_{,x})^2](\tilde{\theta}_{,x})^2$, where $\tilde{\beta} = \beta(\tilde{\theta})$, etc., and is nonpositive in view of (5.11). Thus, we again conclude that the propagation condition (6.4) cannot be satisfied.

In summary, we have found that within the framework of the rate-dependent constitutive assumptions which led to Eq. (5.14) for the temperature in a rigid conductor, thermal waves can occur in the finite theory but not in the corresponding linearized theory governing infinitesimal time-dependent temperature variation.

ACKNOWLEDGMENTS

We are grateful to Professor A. E. Green for discussions concerning the implication of the entropy inequality in special cases, which led to the remarks made at the end of Sec. 5. The results reported here were supported by the U.S. Office of Naval Research under Contract N00014-67-A-0114-0021 with the University of California, Berkeley.

[1] B. D. Coleman and W. Noll, Arch. Ratl. Mech. Anal. **13,** 167 (1963).

[2] B. D. Coleman and V. J. Mizel, J. Chem. Phys. **40,** 1116 (1964).

[3] B. D. Coleman, Arch. Ratl. Mech. Anal. **17,** 1 (1964).

[4] B. D. Coleman and M. E. Gurtin, Z. Angew. Math. Phys. **18,** 199 (1967).

[5] We may recall here that constitutive equations of linear viscous fluids can be regarded as exact constitutive relations which describe the behavior of a class of fluids in all motions. On the other hand, in a functional theory (e.g., of the type developed by Coleman, Ref. 3) a linearly viscous material approximates a general material with fading memory only in the limit of "slow" motions.

[6] M. Chester, Phys. Rev. **131,** 2013 (1963).

[7] C. W. Ulbrich, Phys. Rev. **123,** 2001 (1961).

[8] P. Vernotte, Compt. Rend. **246,** 3154 (1958).

[9] S. Kaliski, Bull. Acad. Polon. Sci., Ser. Sci. Tech., **13,** 253 (1965).

[10] M. E. Gurtin and A. C. Pipkin, Arch. Ratl. Mech. Anal. **31,** 113 (1969).

[11] These constitutive equations when properly linearized reduce to those appropriate to the classical theory of heat conduction.

[12] These local forms can be deduced from the corresponding statements in integral form. See, for example, C. Truesdell and W. Noll, in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1965), Vol. III/3, p. 294; our $\mathbf{q}$, r correspond to their $-\mathbf{h}$, q.

[13] The same symbols can be used here for a function and its value without confusion.

[14] The results of this section follow as a special case of those obtained in another context by B. D. Coleman and V. J. Mizel, Arch. Ratl. Mech. Anal. **13,** 245 (1963). They have studied heat conduction based on nonlinear constitutive equations for rate-independent response, in which the independent variables are the temperature and the first n spatial gradients of temperature .

[15] Similar arguments have appeared many times in the literature. See, for example, Ref. 1.

[16] Assuming, of course, that $\partial^2\psi/\partial\theta^2$ does not vanish.

[17] It should be noted that the decomposition of η in (3.5) does not imply that $\eta^{(0)}$ is independent of t, since θ, $\mathbf{g}$, and $\dot{\theta}$ are considered as independent variables.

[18] A function of α defined in a neighborhood of α_0 is $o(1)$ or $O(1)$ as $\alpha \to \alpha_0$, according to whether it is zero or bounded in this limit, respectively.

[19] It is worth noting that if in place of (3.1) we had assumed $\psi = \psi(\theta, \mathbf{g}, \dot{\theta}, \dot{\mathbf{g}})$, etc., that is, if the rate of the temperature gradient had been included as an independent variable also, then (3.3) would be replaced by $\psi = \psi(\theta, \mathbf{g})$. This should be contrasted with the results obtained by Coleman and Mizel (Ref. 14), where it is shown that ψ cannot depend on temperature gradients when a dependence on rates is excluded. See also, in this regard, A. C. Eringen [Intern. J. Eng. Sci. **4,** 179 (1966)], where the rate of temperature and rates of temperature gradients are included as independent variables in the context of a general thermomechanical investigation.

[20] See Ref. 4.

[21] See T. Y. Thomas, *Plastic Flow and Fracture in Solids* (Academic Press, New York, 1961), Vol. 2, p. 47. Alternatively, see C. Truesdell and R. A. Toupin, "The Classical Field Theories," in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1960), Vol. III/1, Sec. 181.

[22] An alternate derivation of (4.4) results from the characteristic relation appropriate to the quasilinear partial differential equation (4.1). It is, thereby, seen that discontinuities of the type under consideration can occur only along a characteristic. See R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. II: Partial Differential Equations* (Interscience Publishers, Wiley, New York, 1962), p. 416.

[23] See Footnote 22.

[24] Here we regard (5.9) merely as special cases of (5.5) and (5.6). In the context of rate-independent response, where only $\mathbf{K}(\theta, \mathbf{g})$ arises, it follows from the work of Coleman and Mizel (Ref. 14) that a function of the form (5.9b) can be regarded also as a special case of a certain higher-order approximate theory for which Fourier's law is a complete first-order theory.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Infilling for NUT Space

J. C. JACKSON
*King's College, London, England**

(Received 9 June 1969)

A method of generating interior NUT solutions from spherically symmetric ones is presented. It is similar to the well-known "derivation" of the corresponding vacuum metric from the Schwarzschild solution. These sources turn out to consist of material bridges, joining otherwise independent vacuum regions through a "wormhole." It is shown that this matter cannot have an equation of state which is physically reasonable everywhere.

1. INTRODUCTION

In recent years, much effort has been expended upon the production of algebraically special vacuum solutions of Einstein's field equations. Many of these space–times possess singularities, which in principle should be removable by introducing suitable matter-filled regions. In this paper, the interior extension of such a space–time, namely the NUT[1,2] solution, is presented. (This is also called the Taub–NUT solution. However, we are interested in the stationary part of this manifold, which is usually referred to as NUT space.)

Our starting point is the curious "derivation" of this solution, given by Demianski and Newman,[3] in which the exterior Schwarzschild metric is first complexified, and then a certain complex coordinate transformation performed, in such a way that the result is a "new" real vacuum metric. It is shown that a similar procedure can in fact be applied to any spherically symmetric static solution. In particular, given a static interior Schwarzschild solution joined smoothly onto the exterior one, the process can be chosen to yield a metric which is still smoothly joined onto the "new" exterior metric. The vacuum NUT manifold has no singularities in the elementary sense of the word; however, it is not geodesically complete,[4] and we shall show that it can be made complete by introducing matter in this way. The interior NUT spaces so obtained have several features in common with the vacuum solution, namely their Petrov classification and symmetries. It turns out that they inevitably possess regions where the matter is unphysical. This behavior seems to be associated with certain unusual topological properties; the matter distributions are bridges joining two otherwise independent vacuum regions through a "wormhole."

A brief discussion of the spherically symmetric situation is necessary. In this, a suitable physically reasonable interior Schwarzschild solution is given, from which an explicit interior NUT solution will be generated later. We write the spherically symmetric metric in the form

$$ds^2 = Y^2Z\,dt^2 - Z^{-1}\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \quad (1)$$

where Y and Z are arbitrary functions of r. The Einstein tensor has eigenvalues such that

$$\begin{aligned} 8\pi\rho &= \frac{1}{r^2} - \frac{Z}{r^2} - \frac{Z_1}{r}, \\ 8\pi p_1 &= 2\frac{Z}{r}\frac{Y_1}{Y} - 8\pi\rho, \\ 8\pi p_2 &= 8\pi p_3 \\ &= Z\frac{Y_{11}}{Y} + \tfrac{1}{2}Z_{11} + \tfrac{3}{2}Z_1\frac{Y_1}{Y} + \frac{Z_1}{r} + \frac{Z}{r}\frac{Y_1}{Y}, \end{aligned} \quad (2)$$

where ρ and the p_a are the principal density and pressures, and $Z_1 \equiv \partial Z/\partial r$, etc. (Greek indices take values $0, \cdots, 3$, and Roman ones $1, \cdots, 3$. The usual relativistic units, i.e., $G = c = 1$, are being used.)

Let the boundary of the source be at $r = d$. It will prove necessary to work in an admissible coordinate system, so that Y and Z and their first derivatives must be continuous here. The first of Eqs. (2) shows that in this case the density vanishes at the surface. (It is for this reason that the well-known incompressible perfect-fluid solution cannot be used.) With any reasonable equation of state, it is then essential for all the pressures to be zero at $r = d$. The Lichnerowicz conditions ensure that p_1 vanishes at the surface, and a simple way of making p_2 and p_3 do so is to introduce the additional requirement that Z_{11} and Y_{11} be continuous.

In view of the regularity requirement $Z = 1$ at the center $r = 0$, we assume for Z an expansion in positive-integral powers of r:

$$Z = 1 + \sum_{n=1}^{\infty} A_n r^n. \quad (3)$$

Equations (2) show that there will be singularities at $r = 0$ unless $A_1 = 0$. The three boundary conditions are

$$Z - 1 = -2m/d, \quad Z_1 = 2m/d^2, \quad Z_{11} = -4m/d^3, \quad (4)$$

at $r = d$, and (3) will be terminated when it contains enough constants to satisfy them, i.e., after $n = 4$. Equations (4) then give

$$Z = 1 - \frac{2m}{d}\frac{r^2}{d^2}\left(10 - 15\frac{r}{d} + 6\frac{r^2}{d^2}\right),$$

and from (2) we find

$$8\pi\rho = 60\frac{m}{d^3}\left(1 - \frac{r}{d}\right)^2.$$

The continuity requirements on Y at $r = d$ are

$$Y = 1, \quad Y_1 = Y_{11} = 0. \tag{5}$$

It is ensured that p_1 is proportional to m^2 by setting

$$\frac{2}{r}\frac{Y_1}{Y} = 8\pi\rho, \quad \text{i.e.,} \quad \frac{Y_1}{Y} = 30\frac{m}{d^2}\frac{r}{d}\left(1 - \frac{r}{d}\right)^2,$$

which integrates to

$$Y = \exp\left[-\frac{5}{2}\frac{m}{d}\left(1 - \frac{r}{d}\right)^3\left(1 + 3\frac{r}{d}\right)\right],$$

and clearly Eqs. (5) are satisfied. Equations (2) lead to

$$8\pi p_1 = -1200\frac{m^2}{d^4}\frac{r^2}{d^2}\left(1 - \frac{r}{d}\right)^2\left(1 - \frac{3}{2}\frac{r}{d} + \frac{3}{5}\frac{r^2}{d^2}\right).$$

We shall not bother to show p_2 and p_3 explicitly, but note that the last of (2) can be written

$$p_2 = p_3 = p_1 + \tfrac{1}{2}r\frac{dp_1}{dr} + \frac{1}{2}\frac{r}{Z}(\rho + p_1) \times [4\pi r(\rho + p_1) + \tfrac{1}{2}Z_1].$$

It is easily seen that with Z, ρ, and p_1 given above, this expression contains no powers of m lower than the second.

This matter distribution can be made physically reasonable because:

(1) there are no regions of infinite density or pressure;

(2) the density is everywhere positive if $m > 0$;

(3) the pressures can be made everywhere arbitrarily smaller than the density by making m sufficiently small.

2. THE METRIC

Introduce a null coordinate

$$u = t - \int (YZ)^{-1}\, dr,$$

in which case (1) becomes

$$ds^2 = Y^2Z\, du^2 + 2Y\, du\, dr - r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

A contravariant null tetrad for this is (with x^0, x^1, x^2, $x^3 = u, r, \theta, \phi$, respectively)

$$\begin{aligned} l^\mu &= \delta^\mu_1, \quad n^\mu = Y^{-1}\delta^\mu_0 - \tfrac{1}{2}Z\delta^\mu_1, \\ m^\mu &= [(\sqrt{2})r]^{-1}(\delta^\mu_2 + i\csc\theta\delta^\mu_3), \\ \bar{m}^\mu &= [(\sqrt{2})r]^{-1}(\delta^\mu_2 - i\csc\theta\delta^\mu_3), \end{aligned} \tag{6}$$

where

$$g^{\mu\nu} = l^\mu n^\nu + l^\nu n^\nu - m^\mu\bar{m}^\nu - m^\nu\bar{m}^\mu.$$

Following Demianski and Newman,[3] the coordinate r is now allowed to take complex values, and (6) is rewritten as

$$\begin{aligned} l^\mu = \delta^\mu_1, \quad n^\mu &= Y^{-1}\delta^\mu_0 - \tfrac{1}{2}Z\delta^\mu_1, \\ m^\mu &= [(\sqrt{2})\bar{r}]^{-1}(\delta^\mu_2 + i\csc\theta\delta^\mu_3), \\ \bar{m}^\mu &= [(\sqrt{2})r]^{-1}(\delta^\mu_2 - i\csc\theta\delta^\mu_3), \end{aligned}$$

which have been chosen so that

(1) l^μ and n^μ are still real, i.e., Y and Z are real functions of r and $\bar{r}$, which reduce to the original functions in (1) when r is real.

(2) $\bar{m}^\mu$ is still the complex conjugate of m^μ.

Y and Z are, of course, not uniquely determined by condition (1). This tetrad is now subjected to the complex-coordinate transformation

$$\begin{aligned} u' &= u + 4ib\log(\tfrac{1}{2}\sec\tfrac{1}{2}\theta), \\ r' &= r + ib, \quad \theta' = \theta, \quad \phi' = \phi, \end{aligned}$$

where b is a real constant. The tetrad components with respect to the new coordinate basis are

$$\begin{aligned} l'^\mu &= \delta^\mu_1, \quad n'^\mu = Y^{-1}\delta^\mu_0 - \tfrac{1}{2}Z\delta^\mu_1, \\ m'^\mu &= [(\sqrt{2})(r' + ib)]^{-1}(2ib\tan\tfrac{1}{2}\theta + \delta^\mu_2 + i\csc\theta\delta^\mu_3). \end{aligned} \tag{7}$$

The "new" metric is defined by letting r' be real, and defining $\bar{m}'^\mu$ to be the complex conjugate of m'^μ. Y and Z are now real functions of r'. The metric is obtained from the null tetrad in the usual manner. Its covariant form is, dropping the primes,

$$\begin{aligned} ds^2 = {} & Y^2Z(du + 4b\sin^2\tfrac{1}{2}\theta\, d\phi)^2 \\ & + 2Y(du - 4b\sin^2\tfrac{1}{2}\theta\, d\phi)\, dr \\ & - (r^2 + b^2)(d\theta^2 + \sin^2\theta\, d\phi^2), \end{aligned}$$

which can be simplified by introducing a coordinate

$$t = u + \int (YZ)^{-1}\, dr$$

to give

$$\begin{aligned} ds^2 = {} & Y^2Z(dt - 4b\sin^2\tfrac{1}{2}\theta\, d\phi)^2 - Z^{-1}\, dr^2 \\ & - (r^2 + b^2)(d\theta^2 + \sin^2\theta\, d\phi^2). \end{aligned} \tag{8}$$

In the future, this coordinate system will be used.

We shall have more to say about the way Y and Z must be "complexified" later. The empty-space Schwarzschild values of these quantities are

$$Y = 1, \quad Z = 1 - (2m/r),$$

from which Demianski and Newman[3] obtain the NUT vacuum solution, using the following complexification:

$$Y \Rightarrow 1, \quad m \Rightarrow m - ib,$$
$$Z \Rightarrow 1 - \frac{m}{r} - \frac{\bar{m}}{\bar{r}} = 1 - \frac{2(mr' + b^2)}{r'^2 + b^2}.$$

3. BOUNDARY CONDITIONS

Suppose we start with a spherically symmetric interior solution having Z of the form (3), except that for later convenience we let $A_n \Rightarrow mA_n$:

$$Z = 1 + \sum_{n=1}^{\infty} A_n m r^n, \quad Z_1 = \sum_{n=1}^{\infty} n A_n m r^{n-1}.$$

No assumptions are made about the form of Y. Assume that the Lichnerowicz boundary conditions are satisfied at $r = d$, i.e.,

$$\sum_{n=1}^{\infty} A_n m d^n = -\frac{2m}{d}, \quad \sum_{n=1}^{\infty} n A_n m d^n = \frac{2m}{d},$$
$$Y(d) = 1, \quad Y_1(d) = 0. \tag{9}$$

We now show that Y and Z can be complexified so that they and their first derivatives are equal to the respective vacuum NUT quantities at a certain new boundary, namely $r^2 + b^2 = d^2$. (It is assumed that $b^2 < d^2$.) Externally we have

$$Z = 1 - \frac{2(mr + b^2)}{y^2},$$
$$Z_1 = -\frac{2m}{y^2} + \frac{4(mr + b^2)r}{y^4},$$
$$Y = 1, \quad Y_1 = 0,$$

where $y^2 = r^2 + b^2$. Internally, let

$$m \Rightarrow m - ib,$$
$$Z(r) \Rightarrow 1 + \sum_{n=1}^{\infty} \tfrac{1}{2} A_n \left(\frac{m}{r} + \frac{\bar{m}}{\bar{r}}\right)(r\bar{r})^{(n+1)/2},$$
$$Y(r) \Rightarrow Y[(r\bar{r})^{\frac{1}{2}}].$$

Remembering that the new real radial coordinate is $r' = r + ib$, we see that these become, dropping the prime,

$$Z(r) \Rightarrow 1 + \sum_{n=1}^{\infty} A_n (mr + b^2) y^{n-1},$$
$$Y(r) \Rightarrow Y(y),$$

so that internally, we now have

$$Z_1 = \sum_{n=1}^{\infty} A_n m y^{n-1} + \sum_{n=1}^{\infty} A_n (mr + b^2)(n-1) y^{n-3} r,$$
$$Y_1 = \frac{\partial Y}{\partial y}\frac{r}{y}.$$

The required continuity conditions at $y = d$ are, therefore,

$$\sum_{n=1}^{\infty} A_n (mr + b^2) d^{n-1} = -\frac{2(mr + b^2)}{d^2},$$
$$\sum_{n=1}^{\infty} A_n m d^{n-1} + \sum_{n=1}^{\infty} A_n (mr + b^2)(n-1) d^{n-3} r = -\frac{2m}{d^2} + \frac{4(mr + b^2)r}{d^4},$$
$$Y(d) = 1, \quad \left(\frac{\partial Y}{\partial y}\right)_{y=d} \frac{r}{d} = 0,$$

which are automatically satisfied if Eqs. (9) are satisfied.

In fact, it can be shown by induction that C^n boundary conditions are preserved under this complexification. However, it has been verified that if an admissible coordinate system is not being used, and the O'Brien–Synge conditions[5] adopted, these are not preserved.

4. EINSTEIN AND WEYL TENSORS

Using the exterior-calculus techniques developed by Misner,[6] we can easily evaluate the components $G_{\alpha\beta}$ and $C_{\alpha\beta\gamma\delta}$ of these tensors (the Riemann and Ricci tensor conventions are given by $\zeta_{\nu;[\rho\sigma]} = \frac{1}{2} R^{\mu}_{\nu\rho\sigma}\zeta_{\mu}$, $R_{\nu\rho} = R^{\mu}_{\nu\rho\mu}$), with respect to the orthonormal frame

$$\overset{0}{\omega} = YZ^{\frac{1}{2}}(dt - 4b \sin^2 \tfrac{1}{2}\theta \, d\phi),$$
$$\overset{1}{\omega} = Z^{-\frac{1}{2}} dr, \quad \overset{2}{\omega} = y \, d\theta, \quad \overset{3}{\omega} = y \sin\theta \, d\phi, \tag{10}$$

the dual of which is

$$\underset{0}{e} = \left(\frac{1}{Y^2 Z}\right)^{\frac{1}{2}} \frac{\partial}{\partial t}, \quad \underset{1}{e} = Z^{\frac{1}{2}} \frac{\partial}{\partial r},$$
$$\underset{2}{e} = \frac{1}{y}\frac{\partial}{\partial \theta}, \quad \underset{3}{e} = 2\frac{b}{y}\tan\frac{\theta}{2}\frac{\partial}{\partial t} + \frac{1}{y \sin\theta}\frac{\partial}{\partial \phi}, \tag{11}$$

The conformal tensor is of Petrov type D. Its double principal null vectors, satisfying

$$C_{\alpha\beta\gamma[\delta} k_{\epsilon]} k^{\beta} k^{\gamma} = 0,$$

are, in terms of (11),

$$l^{\alpha} = Z^{-\frac{1}{2}}(\delta^{\alpha}_0 + \delta^{\alpha}_1), \quad n^{\alpha} = \tfrac{1}{2} Z^{\frac{1}{2}}(\delta^{\alpha}_0 - \delta^{\alpha}_1),$$

and are, in fact, the real members of the null tetrad (7). In Newman–Penrose language,[7] therefore, this tensor is characterized by one complex invariant, namely

$$\Psi_2 = -C_{\alpha\beta\gamma\delta} l^\alpha m^\beta n^\gamma \bar{m}^\delta,$$

which becomes

$$\Psi_2 = -\tfrac{1}{6}(2A + 2B + C + E) + iD,$$

where

$$A = \tfrac{1}{2}Z\left(\frac{Y_{11}}{Y} + \frac{3}{2}\frac{Y_1}{Y}\frac{Z_1}{Z} + \frac{1}{2}\frac{Z_{11}}{Z}\right),$$

$$B = -\tfrac{1}{2}y^{-2}[1 - (r^2 + Y^2b^2)y^{-2}Z + 4b^2y^{-2}Y^2Z],$$

$$C = -ry^{-2}Z\left(\frac{Y_1}{Y} + \frac{1}{2}\frac{Z_1}{Z}\right) - b^2y^{-4}Y^2Z,$$

$$D = by^{-2}YZ\left(\frac{Y_1}{Y} + \frac{1}{2}\frac{Z_1}{Z} - ry^{-2}\right),$$

$$E = -y^{-2}(b^2y^{-2}Z + \tfrac{1}{2}rZ_1).$$

By substituting the empty-space values of Y and Z, the well-known expression[2]

$$\Psi_2 = (m - ib)/(r - ib)^3 \tag{12}$$

is regained.

We also find the nonvanishing components

$$G_{00} = 2(B - E),$$
$$G_{11} = 2(C - B), \tag{13}$$
$$G_{22} = G_{33} = C + E - 2A,$$

so that tetrad (10) is the eigenframe of the Einstein tensor.

The physical significance of these results will be discussed later.

5. SYMMETRIES AND TOPOLOGY

Metric (8) has the following Killing fields:

$$\zeta_0 = \frac{\partial}{\partial t},$$

$$\zeta_1 = -2b\tan\frac{\theta}{2}\cos\phi\frac{\partial}{\partial t} + \sin\phi\frac{\partial}{\partial\theta} + \cos\phi\cot\theta\frac{\partial}{\partial\phi},$$

$$\zeta_2 = 2b\tan\frac{\theta}{2}\sin\phi\frac{\partial}{\partial t} + \cos\phi\frac{\partial}{\partial\theta} - \sin\phi\cot\theta\frac{\partial}{\partial\phi},$$

$$\zeta_3 = 2b\frac{\partial}{\partial t} + \frac{\partial}{\partial\phi}.$$

Thus, our general manifolds admit the same four-parameter group of motions as the vacuum solution. The fact that the latter possesses this group of motions has consequences which have been worked out by Misner[6] and Misner and Taub,[4] and their results are applicable to our situation. The relevant ones are summarized in the following paragraph.

The group, acting on a given point, generates the whole of a hypersurface of constant r. These hypersurfaces have a universal covering space with topology S^3. If one makes the assumption that leads to the smallest number of identifications in them, namely that their topology is just S^3, the coordinates t, θ, and ϕ all behave as "angles," with

$$t \equiv t + 8\pi b,$$
$$\theta \equiv \theta + 4\pi,$$
$$\phi \equiv \phi + 4\pi,$$

so that even this assumption leaves the space with closed timelike curves. Although a given point is regained only if θ or ϕ is increased by 4π, values of these coordinates outside the ranges $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$ are redundant. The structure constants of the group are given by

$$[\zeta_0, \zeta_a] = 0,$$
$$[\zeta_a, \zeta_b] = \epsilon_{abc}\zeta_c,$$

from which we see that the space–time admits a 3-parameter subgroup whose Lie algebra is isomorphic to that of the rotation group $SO(3)$. However, the subgroup must be simply connected, as it generates by itself the constant-r hypersurfaces, which we have taken to be simply connected. Thus it is, in fact, $SU(2)$. (These results are, of course, invalid when $b = 0$.)

Situations with this symmetry might be called spherically symmetric. However, perhaps this name should be reserved for cases where the "rotation" group acts on 2-surfaces. Invariantly defined quantities, like the eigenvalues of the Einstein tensor, are functions only of r; points in a given r = const surface are indistinguishable, i.e., these surfaces are homogeneous.

Another aspect of the manifold's topology will now be examined, namely, whether or not it possesses a center. In the fully spherically symmetric case, such a point is defined[8,9] as one where the area of the spheres which are the rotation group orbits tends to zero. If it exists, the usual coordinate r is introduced, and the metric becomes singular at $r = 0$, restricting r to positive values. In our case, the metric remains regular at $r = 0$, suggesting that now this coordinate can be continued to negative values. The "rotation" group has 3-dimensional orbits, and a center can reasonably be defined as a point where their volume vanishes.

This volume is

$$V = \int_{\text{orbit}} \overset{0}{\omega} \wedge \overset{2}{\omega} \wedge \overset{3}{\omega} = \int_0^{8\pi b} \int_0^{\pi} \int_0^{2\pi} YZ^{\frac{1}{2}} y^2 \sin\theta \, dt \, d\theta \, d\phi,$$

i.e.,

$$V = 32\pi^2 b Y Z^{\frac{1}{2}} y^2,$$

where $\overset{0}{\omega}$, $\overset{2}{\omega}$, $\overset{3}{\omega}$ are defined in (10). It never vanishes if Y and Z are never zero, which in the next section turns out to be a condition for the manifold $0 \leq t < 8\pi b$, $-\infty < r < +\infty$, $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$ to be geodesically complete.

6. GEODESIC COMPLETENESS

This treatment is similar to that given for the vacuum solution in Ref. 4. We employ the Killing fields, in the usual way, to give us the linear first integrals of the geodesic equations. If $u^\mu = dx^\mu/d\lambda$ is the unit tangent vector to the geodesic, λ being an affine parameter, and ζ^μ a Killing vector, then $u_\mu \zeta^\mu$ is a constant of the motion. Thus, writing

$$u_\mu \zeta^\mu_0 = q_0, \quad u_\mu \zeta^\mu_a = q_a$$

gives

$$u_0 = q_0, \quad 2bu_0 + u_3 = q_3,$$

$$-2b \tan \tfrac{1}{2}\theta \cos\phi u_0 + \sin\phi u_2 + \cos\phi \cot\theta u_3 = q_1,$$

$$2b \tan \tfrac{1}{2}\theta \sin\phi u_0 + \cos\phi u_2 - \sin\phi \cot\theta u_3 = q_2.$$

The symmetry can also be used to find a coordinate system in which $q_1 = q_2 = 0$, when

$$u_0 = q_0, \quad u_2 = 0, \quad u_3 = q_3 - 2bq_0.$$

The vanishing of $u_2 \Rightarrow d\theta/d\lambda = 0$, so that θ is constant, given by

$$\cos\theta = 2bq_0/q_3,$$

which reduces when b is zero to the familiar motion in the plane $\theta = \frac{1}{2}\pi$. The contravariant components u^μ are

$$u^0 = \frac{dt}{d\lambda} = \frac{q_0}{Y^2 Z} - 4b \frac{q_3}{y^2} \sin^2 \tfrac{1}{2}\theta,$$

$$u^3 = \frac{d\phi}{d\lambda} = -\frac{q_3}{y^2},$$

and u^1 is found from

$$g_{\mu\nu} u^\mu u^\nu = \epsilon = +1, 0, -1$$

for timelike, null, and spacelike geodesics, respectively. It is

$$u^1 = \frac{dr}{d\lambda} = \pm \left(\frac{q_0^2}{y^2} - \frac{q_3^2}{y^2} Z \sin^2\theta - \epsilon Z \right)^{\frac{1}{2}}.$$

Further integration of these equations is not possible unless we know Y and Z explicitly. However, if the latter possess no zeros, and Z has no infinities, $dt/d\lambda$, $dr/d\lambda$, and $d\phi/d\lambda$ are bounded, so that every geodesic can be continued to all values of its affine parameter, and the manifold is geodesically complete. The vacuum manifold $0 \leq t < 8\pi b$, $-\infty < r < +\infty$, $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$ is not geodesically complete, since $Z = 0$ at

$$r = m \pm (m^2 + b^2)^{\frac{1}{2}},$$

and Misner and Taub[4] show that it cannot be extended to one that is complete.

7. EXPLICIT SOLUTION

We are now in a position to examine the physical features of our solutions. A typical one is obtained from the Schwarzschild solution given in the introduction, and is

$$Z = 1 - \frac{2(mr + b^2)}{d^2} \frac{y}{d} \left(10 - 15 \frac{y}{d} + 6 \frac{y^2}{d^2} \right),$$

$$Y = \exp\left[-\frac{5}{2} \frac{m}{d} \left(1 - \frac{y}{d} \right)^3 \left(1 + 3 \frac{y}{d} \right) \right]. \qquad (14)$$

It was shown in general that the continuity conditions are satisfied at

$$y^2 = r^2 + b^2 = d^2,$$

i.e., at the two distinct surfaces

$$r = +(d^2 - b^2)^{\frac{1}{2}} \quad \text{and} \quad r = -(d^2 - b^2)^{\frac{1}{2}}.$$

As r diminishes from large positive values, so does the volume V of the surfaces $r = \text{const}$. This reaches a positive minimum somewhere near $r = 0$, and then opens out when r assumes large negative values. The manifold has no center, and the two vacuum regions $r > (d^2 - b^2)^{\frac{1}{2}}$ and $r < -(d^2 - b^2)^{\frac{1}{2}}$ are joined through a "wormhole" by a bridge of matter.

If m is positive, say, Y and Z of (14) fulfill the conditions for geodesic completeness if

$$d^2 > 2(m^2 + b^2)^{\frac{1}{2}} [(m^2 + b^2)^{\frac{1}{2}} + m].$$

This completeness is a definition of singularity free.[10–13] We have therefore removed the singularities from NUT space by introducing a matter-filled region, which can justifiably be regarded as a source for the exterior field.

The field equations obtained from (13) are

$$8\pi y^2\rho = 1 - (rZ)_{,1} + 3\frac{b^2}{y^2}Y^2Z - \frac{b^2}{y^2}Z,$$

$$8\pi y^2(\rho + p_1) = 2rZ\frac{Y_1}{Y} - 2\frac{b^2}{y^2}Z(1 - Y^2),$$

$$8\pi y^2 p_2 = 8\pi y^2 p_3 \tag{15}$$

$$= y^2Z\left(\frac{Y_{11}}{Y} + \frac{3}{2}\frac{Y_1}{Y}\frac{Z_1}{Z} + \frac{1}{2}\frac{Z_{11}}{Z}\right)$$

$$+ rZ\left(\frac{Y_1}{Y} + \frac{Z_1}{Z}\right) + \frac{b^2}{y^2}Z(1 + Y^2).$$

At $r = 0$, we find

$$8\pi b^2(\rho + p_1) = -2(1 - Y^2). \tag{16}$$

If the original spherically symmetric matter distribution has a physically reasonable equation of state, it can be shown that $Y(y) < 1$ for $y^2 < d^2$. (When $b = 0$, $4\pi(\rho + p_1) = (Z/r)(Y_1/Y) > 0$ if physically reasonable. Then $Y_1 > 0$ for $0 < r < d$, and as $Y = 1$ at $r = d$, it follows that $Y(r) < 1$ in this region.) Thus (16) shows that at $r = 0$, $\rho + p_1 < 0$, and the equation of state must now be unphysical in places. (This behavior is unreasonable because it implies the existence of observers for whom the energy density is negative.) The possibility that an interior NUT solution, consisting entirely of realistic matter, might be generated from a spherically symmetric one that does not, is excluded by the following argument. Such a solution would certainly have $Y(y) > 1$ at $r = 0$. Consider the region $0 < r < r_0$, where r_0 is the smallest positive value of r for which $Y(y)$ is unity, i.e., $r_0 \leq d$. Imposing the "reasonable" condition on (15) in this region gives

$$(Y^2 - 1)^{-1}Y^{-1}Y_1 > -b^2r^{-1}y^{-2},$$

which, on integrating between the limits ϵ, r, where $0 < \epsilon < r < r_0$, leads to

$$\left[\log\left(\frac{Y^2 - 1}{Y^2}\right)\right]_\epsilon^r > -\left[\log\left(\frac{r^2}{r^2 + b^2}\right)\right]_\epsilon^r.$$

There exists a $\delta > 0$ such that this condition is violated when $r > r_0 - \delta$.

The vacuum regions are characterized by (12), i.e.,

$$\Psi_2 = (m - ib)/(r - ib)^3.$$

As in the case $b = 0$, m is the "electric"-type mass of the source; b has been identified[3] as the source's "magnetic"-type mass, i.e., as the gravitational equivalent of a magnetic monopole. If $m > 0$, test particles dropped in vacuum move in the direction of decreasing r. This means that an observer in one vacuum part sees a source whose mass is the negative of that seen by an observer in the other.

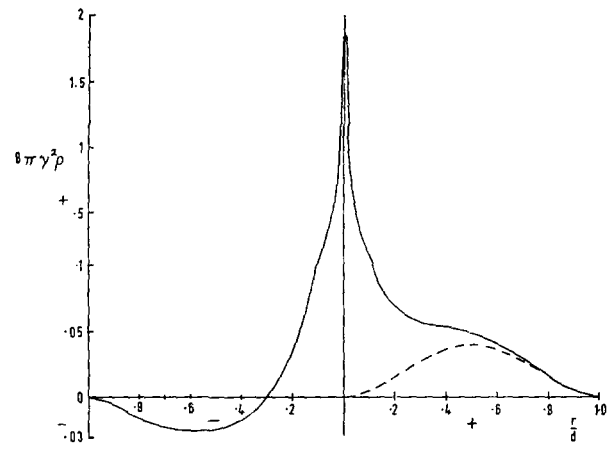


FIG. 1. Behavior of the density for $m/d = 0.01$. $b^2/d^2 = 0.001$ is represented by ———; $b^2/d^2 = 0$ is represented by – – –.

The particular example (14) gives cumbersome expressions when substituted into the field equations, and therefore ρ and p_1 have been given graphically (Figs. 1 and 2) for typical values of m/d and b/d, namely $m/d = 0.01$, $b^2/d^2 = 0.001$. The behavior of the other pressures is similar to that of p_1. For example, they are zero at the two boundaries, because the second derivatives of the metric components are continuous there. We see that these quantities behave in a wildly unphysical manner in regions of negative r; even negative principal densities are encountered. However, in view of the fact that for positive m, an observer on the $r < -(d^2 - b^2)^{\frac{1}{2}}$ side of the "wormhole" sees a source with negative mass, this is not surprising.

We shall now consider the way in which the matter is rotating. The vorticity tensor is defined by $\omega_{\mu\nu} = h_\mu^\rho h_\nu^\sigma u_{[\rho;\sigma]}$, where u_ρ is the velocity of the matter,

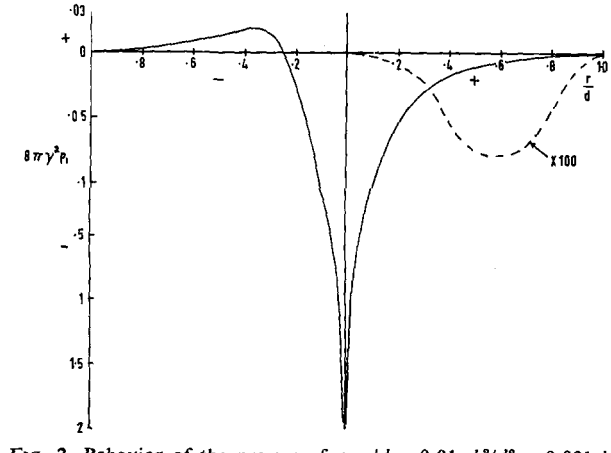
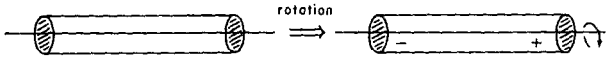


FIG. 2. Behavior of the pressure for $m/d = 0.01$. $b^2/d^2 = 0.001$ is represented by ———; $b^2/d^2 = 0$ is represented by – – –.

i.e., the timelike eigenvector given in (10), and $h_\mu^\rho = \delta_\mu^\rho - u_\mu u^\rho$ is the usual operator projecting into the instantaneous rest space of an observer moving with this velocity. The components $\omega_{\alpha\beta}$ of the vorticity tensor with respect to the eigentriad given in (11) are

$$\omega_{12} = \omega_{13} = 0,$$
$$\omega_{23} = -by^{-2}YZ^{\frac{1}{2}}.$$

The matter is therefore rotating everywhere about the normals to the surfaces of constant r. This provides a heuristic explanation of why the source appears to have a magnetic-type mass. Consider a rod of matter:



which, when rotating about its axis, develops a "magnetic" dipole moment, with magnetic monopoles at each end. In our source, rotating rods of this nature are threaded through the "wormhole," with their axes along the r-lines. Their ends, carrying the monopoles, coincide with its surfaces.

8. CONCLUSION

We have given a prescription for generating solutions which are, in every sense, interior extensions of the vacuum metric considered. However, the method unavoidably gives rise to matter distributions having unphysical regions, and to "wormhole" topologies.

The question of whether this behavior is due just to the particular method used has not been answered. However, there are very strong reasons for believing that the NUT field cannot have a physically reasonable, isolated source. This is because a coordinate system in which the metric is asymptotically Minkowskian cannot be found, essentially because the constant-r hypersurfaces have topology S^3 for arbitrarily large values of this coordinate. Indeed, in view of the peculiarities of vacuum NUT space, such a source would be an embarrassment, so that the one presented here is probably as reasonable as is possible under the circumstances.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Professor F. A. E. Pirani, for the advice, encouragement, and constructive criticism given during the course of this work. It was carried out while the author held a research studentship, given by the Science Research Council of Great Britain.

* Present address: Astronomy Centre, University of Sussex, Brighton, Sussex, England.

[1] A. H. Taub, Ann. Math. **53**, 472 (1951).
[2] E. T. Newman, L. Tamburino, and T. J. Unti, J. Math. Phys. **4**, 915 (1963).
[3] M. Demianski and E. T. Newman, Bull. Acad. Polon. Sci., Ser. Sci. Math. Astron. Phys. **11**, 653 (1966).
[4] C. W. Misner and A. H. Taub, "A Singularity Free Empty Universe" (to be published).
[5] S. O'Brien and J. L. Synge, Commun. Dublin Inst. Advan. Studies **A9** (1952).
[6] C. W. Misner, J. Math. Phys. **4**, 924 (1963).
[7] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[8] H. Bondi, Proc. Roy. Soc. (London) **A282**, 303 (1964).
[9] H. P. Künzle, Proc. Roy. Soc. (London) **A297**, 244 (1967).
[10] R. Penrose, Phys. Rev. Letters **14**, 57 (1965).
[11] S. W. Hawking, Proc. Roy. Soc. (London) **A294**, 511 (1966).
[12] S. W. Hawking, Proc. Roy. Soc. (London) **A295**, 490 (1966).
[13] S. W. Hawking, Proc. Roy. Soc. (London) **A300**, 187 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Transfer Matrix Solution of Some One- and Two-Medium Transport Problems in Slab Geometry*

RAPHAEL ARONSON
Department of Nuclear Engineering, New York University, New York, N.Y. 10453

(Received 27 June 1969)

The transfer-matrix method is used to solve several one- and two-medium problems in plane geometry. The formal solutions are given in terms of operators which make no reference to the interactions or to the nature of the flux approximations. This formulation points up the structural similarities between solutions of various problems.

1. INTRODUCTION

The transfer-matrix method[1] offers a simple perspective on radiation transport in slab geometries for both single- and multiple-medium problems. The reason is that the physics—that is, the interactions and the form of the approximation for the flux—is isolated from the specific geometry and boundary conditions that characterize the problem under consideration.

The group property of transfer matrices implies that the transfer matrix for any configuration of slabs is given simply in terms of the transfer matrices for an infinitesimal slab. The latter is characterized by two operators, which we call α and β, which are essentially the transmission and reflection operators for an infinitesimal slab. The flux in a given configuration can be expressed in terms of α and β for each medium, the geometry, the boundary conditions, and nothing else. This expression is exact, so it can be used for exact computations.[2] There is no explicit dependence on either the cross sections or on the type and order of the flux approximation.

The operators α and β, since they give the transmission and reflection properties of an infinitesimal slab, depend respectively on the forward and backward parts of the differential cross section. In a given finite approximation scheme for the flux, one uses the corresponding finite-matrix representation of α and β. The algebra by which one computes the flux from α and β remains the same, however, as for the exact calculation.

The formulation of the problems in terms of operators derived from α and β permits one to write down the solutions explicitly in operator form. The results derived in this paper apply whether one is dealing with isotropic or anisotropic scattering. They hold for polarized radiation and for energy-dependent interactions (so long as the diagonalization in Sec. 2 can be carried out, as it always can in a multigroup scheme). They are valid whether the fluxes are computed exactly or in a P_N, double-P_N, S_N, or any of the other standard representations.

One finds in solving various problems of interest that the same few combinations of operators arise from problem to problem. The simple form of the results shows up certain structural similarities which are obscured by other calculational methods.[3] (The simplicities occur in the operator formulation, and are lost, or at least are hard to find, when the kernels corresponding to all the operators are substituted explicitly.) The results fall into certain patterns which have not been observed before. In particular:

(1) The operators appropriate to a half-space problem are generalized in a very specific way in going to the corresponding slab problem.

(2) The operators appropriate to one-medium problems are generalized in a very specific way when reflectors are added.

(3) The operators appropriate to one-medium problems are generalized in a closely related, very specific way when the slab becomes one component of a two-component lattice.

By "very specific way," we mean that the modification is the same regardless of the specific problem being considered, e.g., whether one is dealing with a critical slab or a medium with sources.

It can be shown[4] that the transfer matrix formulation is completely equivalent to a formulation in terms of Case eigenfunctions[5] for arbitrary interaction. Every step in a Case-type calculation is paralleled by a step in the transfer-matrix development. This has already been demonstrated for one-speed isotropic scattering,[2] but it is true quite generally.[4] It follows that all the expressions to be obtained have analogs in the Case method. Many of these have already been noted for one-speed isotropic scattering.[2] The results thus have meaning even outside of the transfer-matrix formulation. The operator notation makes it possible both to derive the expressions in a simple way and to

recognize recurrent patterns which have physical significance independent of the particular formulation.

Further, since ultimately every problem must be solved numerically, it is useful to isolate and recognize the modules out of which solutions for various problems are built. In this method they become the actual computational intermediates, and they must appear in some form, however hard to recognize, for any method.

In deriving the results, we have found it convenient to put everything in a form directly suitable for computation once α and β are known. In the original formulation of the transfer-matrix method, both increasing and decreasing exponentials occur. The increasing exponentials arise because the transfer matrix gives both incoming and outgoing fluxes on the right-hand surface of the slab in terms of those on the left. That is, it gives part of the input in terms of the output, as if the output were known. In any actual problem, the inputs are given as boundary conditions and the outputs determined from them. The increasing exponentials then cancel. In doing numerical calculations, however, it is preferable to avoid the numerical round-off difficulties arising from the computation and cancellation of large numbers.[6] For this reason, we reformulate the equations to avoid any explicit appearance of increasing exponentials in the final results.

We consider as examples of the use of the method three single-medium problems and five two-medium problems. These are:

(1) Milne problem
(2) Slab with sources, with extension to half-space with sources
(3) Symmetric slab
(4) Reflected critical slab
(5) Critical lattice
(6) Reflected slab with sources
(7) Lattice with alternating source and source-free regions
(8) Two-medium Milne problem.

In probs. (3–7), we restrict the discussion to symmetric slabs. In regions with sources, the source distributions will be assumed to be uniform. These are not necessary restrictions, but represent situations of intrinsic interest and permit an interesting and useful simplification.

In Sec. 2, we summarize the transfer-matrix method. Section 3 discusses the meaning of the operators and states certain conventions employed in the subsequent presentation. In Sec. 4, the method is reformulated to eliminate the increasing exponentials. The one-medium problems are discussed in Secs. 5 and 6. Section 7 contains a treatment of a single symmetric slab. The two medium problems are treated, in the order indicated, in Secs. 8–12. The results are discussed in Sec. 13.

2. THE TRANSFER MATRIX FOR A SOURCELESS SLAB

The transfer-matrix method for a sourceless homogeneous medium was treated in detail by Aronson and Yarmush.[1] We give here a brief summary of the pertinent results.

Consider a sourceless slab of material normal to the x axis, extending from $x = 0$ to $x = t$ (Fig. 1). Let the right-directed distributions be given at $x = 0$ and $x = t$ by $\psi_+(0)$ and $\psi_+(t)$, respectively, and the left-directed distributions by $\psi_-(0)$ and $\psi_-(t)$. In other words, the incident distributions from the right and left are, respectively, $\psi_+(0)$ and $\psi_-(t)$, and the exit distributions are $\psi_-(0)$ and $\psi_+(t)$.

The fluxes obey the matrix equation

$$\begin{pmatrix}\psi_+(t)\\ \psi_-(t)\end{pmatrix} = \mathbf{H}(t)\begin{pmatrix}\psi_+(0)\\ \psi_-(0)\end{pmatrix}, \tag{2.1}$$

where $\mathbf{H}$ is a 2×2 matrix of operators,

$$\mathbf{H} = \begin{pmatrix}\mathbf{T} - \mathbf{R}^*\mathbf{T}^{*-1}\mathbf{R} & \mathbf{R}^*\mathbf{T}^{*-1}\\ -\mathbf{T}^{*-1}\mathbf{R} & \mathbf{T}^{*-1}\end{pmatrix}. \tag{2.2}$$

Here $\mathbf{T}$ and $\mathbf{T}^*$ are transmission operators for radiation incident from the left and right, respectively; $\mathbf{R}$ and $\mathbf{R}^*$ are the respective reflection operators. For a symmetric slab, $\mathbf{T} = \mathbf{T}^*$ and $\mathbf{R} = \mathbf{R}^*$. The central idea is that a knowledge of $\mathbf{H}$ is sufficient to obtain any quantity of interest.

If the slab is thought of as being made up of a sequence of n slabs whose individual $\mathbf{H}$ matrices are given by $\mathbf{H}_i$, where i designates the ith slab from the left, then the $\mathbf{H}$ matrix for the entire configuration is given by

$$\mathbf{H} = \mathbf{H}_n \cdots \mathbf{H}_2\mathbf{H}_1. \tag{2.3}$$

For homogeneous slabs, it is convenient to introduce the operators α and β, defined by the linear terms in

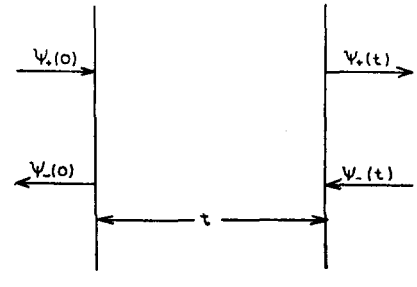


FIG. 1. Schematic slab geometry.

the expansions in t of $\mathbf{T}$ and $\mathbf{R}$, respectively:

$$\mathbf{T}(t) = \mathbf{I} - \boldsymbol{\alpha} t + \cdots, \tag{2.4}$$

$$\mathbf{R}(t) = \boldsymbol{\beta} t + \cdots. \tag{2.5}$$

Here $\mathbf{I}$ is the unit operator. Then if a constant matrix $\mathbf{W}$ is defined by

$$\mathbf{W} = \begin{pmatrix} \boldsymbol{\alpha} & -\boldsymbol{\beta} \\ \boldsymbol{\beta} & -\boldsymbol{\alpha} \end{pmatrix}, \tag{2.6}$$

it is easy to show that

$$\mathbf{H} = \exp(-\mathbf{W}t). \tag{2.7}$$

It is most convenient to determine $\mathbf{H}$ by a diagonalization procedure carried out on an operator related to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ rather than directly on a 2×2 matrix operator. Define

$$\boldsymbol{\sigma} = \boldsymbol{\alpha} + \boldsymbol{\beta}, \tag{2.8}$$

$$\boldsymbol{\delta} = \boldsymbol{\alpha} - \boldsymbol{\beta}, \tag{2.9}$$

and let the diagonalization of $\boldsymbol{\sigma\delta}$ be given by

$$\boldsymbol{\Lambda}^2 = \mathbf{X}^{-1}\boldsymbol{\sigma\delta}\mathbf{X}, \tag{2.10}$$

where $\boldsymbol{\Lambda}^2$ is diagonal. Then it can be shown that

$$\mathbf{H} = \tfrac{1}{4}\begin{pmatrix} \mathbf{B}_+ & \mathbf{B}_- \\ \mathbf{B}_- & \mathbf{B}_+ \end{pmatrix}\begin{pmatrix} \exp(-\boldsymbol{\Lambda}t) & 0 \\ 0 & \exp(\boldsymbol{\Lambda}t) \end{pmatrix}\begin{pmatrix} \mathbf{C}_+ & \mathbf{C}_- \\ \mathbf{C}_- & \mathbf{C}_+ \end{pmatrix}, \tag{2.11}$$

where

$$\mathbf{B}_\pm = \mathbf{X} \pm \boldsymbol{\xi}, \tag{2.12}$$

$$\mathbf{C}_\pm = \mathbf{X}^{-1} \pm \boldsymbol{\xi}^{-1}, \tag{2.13}$$

and

$$\boldsymbol{\xi} = \boldsymbol{\delta}\mathbf{X}\boldsymbol{\Lambda}^{-1} = \boldsymbol{\sigma}^{-1}\mathbf{X}\boldsymbol{\Lambda}. \tag{2.14}$$

The last equality comes from Eq. (2.10). Note that

$$\mathbf{B}_+\mathbf{C}_+ + \mathbf{B}_-\mathbf{C}_- = \mathbf{C}_+\mathbf{B}_+ + \mathbf{C}_-\mathbf{B}_- = 4\mathbf{I}, \tag{2.15}$$

$$\mathbf{B}_+\mathbf{C}_- + \mathbf{B}_-\mathbf{C}_+ = \mathbf{C}_+\mathbf{B}_- + \mathbf{C}_-\mathbf{B}_+ = 0, \tag{2.16}$$

where $\mathbf{I}$ is the unit operator.

If $\psi_+(0)$ and $\psi_-(0)$ are known, the flux at an interior point x can also be computed. One considers a slab bounded by 0 and x and replaces t in Eq. (2.11) by x.

It is convenient to define the matrices

$$\mathbf{F} = \mathbf{C}_-\mathbf{C}_+^{-1} = -\mathbf{B}_+^{-1}\mathbf{B}_-, \tag{2.17}$$

$$\mathbf{K}_\pm(x) = \mathbf{B}_\pm + \mathbf{B}_\mp \exp(-\boldsymbol{\Lambda}x)\mathbf{F}\exp(-\boldsymbol{\Lambda}x). \tag{2.18}$$

It will become apparent that $\mathbf{K}_\pm(t)$ is the generalization of $\mathbf{B}_\pm$ for a finite slab of thickness t.

From Eqs. (2.2) and (2.11), one finds for the transmission and reflection operators:

$$\mathbf{T}(t) = 4\mathbf{C}_+^{-1}\exp(-\boldsymbol{\Lambda}t)\mathbf{K}_+^{-1}(t), \tag{2.19}$$

$$\mathbf{R}(t) = \mathbf{K}_-(t)\mathbf{K}_+^{-1}(t). \tag{2.20}$$

3. MEANING OF THE OPERATORS

The operators $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are integral operators in direction, energy, polarization state, and any other variables describing the cross sections. The operands $\psi_\pm$ are functions of these variables. Since, in the transfer-matrix approach, the flux is split up into its forward and backward components, it is necessary to adopt some convention about direction. In the following, we take μ as the cosine of the angle the particle direction makes with the slab normal, which we take to be in the x direction, and when convenient, we exhibit the μ dependence explicitly. We describe $\psi_+(x, \mu)$ with respect to the positive x axis and $\psi_-(x, \mu)$ with respect to the negative x axis, so that μ is always in the interval $(0, 1)$.

We denote the eigenvalues of $(\boldsymbol{\sigma\delta})^{\frac{1}{2}}$, or of the diagonal operator $\boldsymbol{\Lambda}$, by λ. By Eq. (2.10), if λ is an eigenvalue of $(\boldsymbol{\sigma\delta})^{\frac{1}{2}}$, so is $-\lambda$. We choose λ so that $\operatorname{Re}\lambda > 0$, and if $\operatorname{Re}\lambda = 0$, so that $\operatorname{Im}\lambda > 0$. The pathological case with a double root $\pm\lambda = 0$ occurs in a system without absorption and must be considered separately. Note that no terms with increasing exponentials $\exp \lambda t$ appear in Eqs. (2.19) and (2.20).

We designate the eigenvalue with the smallest real part by λ_0. The eigenvalues λ are ordered according to increasing real part, so that $\operatorname{Re}\lambda_i \geq \operatorname{Re}\lambda_{i-1}$.

In the following, we assume that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are known. They are computed from the cross sections by methods discussed elsewhere.[1]

An operator such as $\boldsymbol{\alpha}$ operates in angle. Its kernel is a function of two angle variables, so we may think of $\boldsymbol{\alpha}$ as a matrix with elements $\alpha(\mu, \mu')$, indexed by μ and μ'. Similarly, the elements of operators like $\mathbf{X}$ are $X(\mu, \lambda)$, those of $\mathbf{B}_+^{-1}$ are $B_+^{-1}(\lambda, \mu)$, and those of $\mathbf{F}$ are $F(\lambda, \lambda')$. We denote the first column of a matrix such as $\mathbf{X}$ or $\mathbf{F}$ by a subscript zero. For instance, the elements of the vector $\mathbf{X}_0$ are $X(\mu, \lambda_0)$. Similarly, $F_{00} = F(\lambda_0, \lambda_0)$. For convenience of description, we think of the eigenvalues as being discrete, even though we recognize that they may form a continuum.[7] Thus, for instance, we may write Eq. (2.15) explicitly as

$$\sum_i B_+(\mu, \lambda_i)C_+(\lambda_i, \mu') + \sum_i B_-(\mu, \lambda_i)C_-(\lambda_i, \mu') = 4\delta(\mu - \mu'), \tag{3.1}$$

$$\int_0^1 C_+(\lambda_i, \mu)B_+(\mu, \lambda_j)\,d\mu + \int_0^1 C_-(\lambda_i, \mu)B_-(\mu, \lambda_j)\,d\mu = 4\delta_{ij}, \tag{3.2}$$

where δ_{ij} is to be interpreted as a Kronecker delta if both λ_i and λ_j are discrete and as a Dirac delta function if λ_i and λ_j are in the continuum. It is, of course, zero otherwise.

For a finite-matrix representation of α and β, all the operators have corresponding finite representations and all the eigenvalues are discrete. In an exact calculation, there will, in general, be a continuum.[8]

4. REFORMULATION OF THE METHOD

We want to reformulate the general equations for $\psi_{\pm}(x)$ at an arbitrary point x to avoid the increasing exponential in Eq. (2.11). We recognize that the increasing exponential occurs only because the problem is formulated in terms of $\psi_{+}(0)$ and $\psi_{-}(0)$. The natural boundary conditions involve not $\psi_{-}(0)$, an output flux, which must first be computed, but $\psi_{-}(t)$, which is presumed given. Since the problem is linear and the slab is uniform, we can write

$$\psi_{\pm}(x) = \mathcal{K}_{\pm}(t, x)\psi_{+}(0) + \mathcal{K}_{\mp}(t, t - x)\psi_{-}(t), \quad (4.1)$$

where $\mathcal{K}_{\pm}$ are a pair of operators to be determined. The notation indicates explicitly that $\mathcal{K}_{\pm}$ are functions of both t, the thickness of the slab, and the position x. If there are sources in the interior, there is an additional term due to the sources, but this does not affect the calculation of $\mathcal{K}_{\pm}$.

To compute $\mathcal{K}_{\pm}$, consider a slab of thickness t, with a flux $\psi_{+}(0)$ incident from the left and nothing incident from the right, i.e., $\psi_{-}(t) = 0$. Then

$$\psi_{-}(0) = \mathbf{R}(t)\psi_{+}(0), \quad (4.2)$$

and from Eqs. (2.1) and (2.11), one finds

$$\psi_{\pm}(x) = \tfrac{1}{4}\mathbf{B}_{\pm} \exp(-\mathbf{\Lambda}x)(\mathbf{C}_{+} + \mathbf{C}_{-}\mathbf{R})\psi_{+}(0) + \tfrac{1}{4}\mathbf{B}_{\mp} \exp(\mathbf{\Lambda}x)(\mathbf{C}_{-} + \mathbf{C}_{+}R)\psi_{+}(0). \quad (4.3)$$

But from Eqs. (2.18) and (2.20),

$$\mathbf{C}_{-} + \mathbf{C}_{+}\mathbf{R}(t) = 4 \exp(-\mathbf{\Lambda}t)\mathbf{F} \exp(-\mathbf{\Lambda}t)\mathbf{K}_{+}^{-1}(t) \quad (4.4)$$

$$\mathbf{C}_{+} + \mathbf{C}_{-}\mathbf{R}(t) = 4\mathbf{K}_{+}^{-1}(t), \quad (4.5)$$

so

$$\psi_{\pm}(x) = \{\mathbf{B}_{\pm} \exp(-\mathbf{\Lambda}x) + \mathbf{B}_{\mp} \exp[-\mathbf{\Lambda}(t - x)]\mathbf{F} \times \exp(-\mathbf{\Lambda}t)\}\mathbf{K}_{+}^{-1}(t)\psi_{+}(0). \quad (4.6)$$

Thus, finally, comparison of Eqs. (4.1) and (4.6) gives

$$\mathcal{K}_{\pm}(t, x) = \mathbf{K}_{\pm}(t - x) \exp(-\mathbf{\Lambda}x)\mathbf{K}_{+}^{-1}(t). \quad (4.7)$$

Equation (4.7) has several useful properties. It nowhere contains increasing exponentials, so the purpose of introducing this formulation is satisfied. Further, it shows that the two-parameter operators $\mathcal{K}_{\pm}(t, x)$ can be decomposed into a product of one-parameter operators. Perhaps most interesting; the flux everywhere depends only on the exponential and on a particular pair of functions $\mathbf{K}_{\pm}$ of the basic α and β operators. All the information about the system is contained in these three operators, which thus have a universal significance in slab transport problems.

It is of interest to note the limiting forms

$$\mathbf{K}_{+}(0) = 4\mathbf{C}_{+}^{-1} \quad (4.8)$$

and

$$\mathbf{K}_{-}(0) = 0. \quad (4.9)$$

It follows, as one must require, that

$$\mathcal{K}_{+}(t, 0) = \mathbf{I}, \quad (4.10)$$

$$\mathcal{K}_{-}(t, 0) = \mathbf{R}(t), \quad (4.11)$$

$$\mathcal{K}_{+}(t, t) = \mathbf{T}(t), \quad (4.12)$$

and

$$\mathcal{K}_{-}(t, t) = 0. \quad (4.13)$$

Equations (4.11) and (4.12) are obtained by comparing Eq. (4.7) with Eqs. (2.19) and (2.20).

5. THE MILNE PROBLEM

The Milne problem has been formulated previously by the transfer-matrix method.[2] In the usual way, one considered a half-space $x > 0$ with $\psi_{+}(0) = 0$ and imposed a condition that the flux was bounded asymptotically by $\exp \lambda_0 x$. We reformulate the problem here without assuming the boundedness condition *ab initio*. Rather it comes out naturally.

In the present formulation, we consider a slab between 0 and t, with $\psi_{+}(0) = 0$ but $\psi_{-}(t) \neq 0$. We let $t \to \infty$ and let $\psi_{-}(t)$ increase so that the flux at any finite x is finite and nonzero.

From Eqs. (4.1) and (4.7),

$$\psi_{\pm}(x) = \lim_{t\to\infty} \mathcal{K}_{\mp}(t, t - x)\psi_{-}(t) = \mathbf{K}_{\mp}(x) \exp(\mathbf{\Lambda}x) \lim_{t\to\infty} \exp(-\mathbf{\Lambda}t)\mathbf{K}_{+}^{-1}(t)\psi_{-}(t). \quad (5.1)$$

Because of the factor $\exp(-\mathbf{\Lambda}t)$, each element of the vector

$$\phi(t) = \exp(-\mathbf{\Lambda}t)\mathbf{K}_{+}^{-1}(t)\psi_{-}(t) \quad (5.2)$$

would decrease roughly exponentially with t, if $\psi_{-}(t)$ were to remain finite. The dominant element for large t is clearly that corresponding to λ_0. The flux remains finite and nonzero for finite x if and only if $\psi_{-}(t) \sim \exp \lambda_0 t$. Then $\phi(t)$ has only one finite element in the limit:

$$\phi(\infty) = \lim_{t\to\infty} (-\mathbf{\Lambda}t)\mathbf{K}_{+}^{-1}(t)\psi_{-}(t) = A\delta_{i0}, \quad (5.3)$$

where δ_{i0} is the Kronecker delta. Here $i = 0$ corresponds to the eigenvalue λ_0, and A is a normalization constant.

Inserting Eq. (5.3) into Eq. (5.1) gives

$$\psi_{\pm}(x) = A\mathbf{K}_{\mp 0}(x) \exp \lambda_0 x, \tag{5.4}$$

where in accord with Sec. 3, $\mathbf{K}_{\mp 0}$ is the first column of $\mathbf{K}_{\mp}$, corresponding to λ_0. Equation (5.4) may be rewritten

$$\psi_{\pm}(x) = A[\mathbf{B}_{\mp} \exp (\mathbf{\Lambda} x) + \mathbf{B}_{\pm} \exp (-\mathbf{\Lambda} x)\mathbf{F}]_0, \tag{5.5}$$

so that

$$\psi_{+}(x) + \psi_{-}(x) = 2A\mathbf{X}[\exp (\mathbf{\Lambda} x) + \exp (-\mathbf{\Lambda} x)\mathbf{F}]_0. \tag{5.6}$$

We use this result below.

The flux distribution emitted from the free surface is

$$\psi_{-}(0) = 4A(\mathbf{C}_{+}^{-1})_0, \tag{5.7}$$

where we have used Eqs. (2.15) and (2.17).

By Eq. (2.12),

$$\mathbf{B}_{+} + \mathbf{B}_{-} = 2\mathbf{X}. \tag{5.8}$$

Now the columns of $\mathbf{X}$ are the eigenvectors of $\boldsymbol{\sigma\delta}$ and nothing said previously has fixed the normalization. It is convenient to normalize so that

$$\int_0^1 X(\mu, \lambda)\, d\mu = 1. \tag{5.9}$$

The total flux at x is

$$\rho(x) = 2\pi \int_0^1 [\psi_{+}(x, \mu) + \psi_{-}(x, \mu)]\, d\mu. \tag{5.10}$$

The asymptotic flux is that part of the flux which does not fall off more rapidly than $\exp (-\lambda_0 x)$. Thus

$$\psi_{+}^{\mathrm{as}}(x) + \psi_{-}^{\mathrm{as}}(x) = 2A\mathbf{X}_0[\exp (\lambda_0 x) + F_{00} \exp (-\lambda_0 x)]. \tag{5.11}$$

The total asymptotic flux is then

$$\rho^{\mathrm{as}}(x) = 4\pi A[\exp (\lambda_0 x) + F_{00} \exp (-\lambda_0 x)]. \tag{5.12}$$

The extrapolated end point Z_0 is defined by

$$\rho^{\mathrm{as}}(-z_0) = 0, \tag{5.13}$$

from which

$$z_0 = \frac{1}{2\lambda_0} \ln (-1/F_{00}). \tag{5.14}$$

6. SLAB WITH SOURCES

The transfer-matrix method has been used by Gelbard[6] to solve the problem of the leakage from a slab with sources. He too was concerned with eliminating the increasing exponentials. The reformulation in this section in terms of the $\psi_{\pm}$ operators enables one to obtain the interior flux as well.

Equations (2.1) and (2.7) are equivalent to the differential equation

$$\mathbf{\Psi}'(x) = -\mathbf{W}\mathbf{\Psi}(x), \tag{6.1}$$

where $\mathbf{\Psi}(x)$ is the 2-component vector

$$\mathbf{\Psi}(x) = \begin{pmatrix} \psi_{+}(x) \\ \psi_{-}(x) \end{pmatrix}. \tag{6.2}$$

The prime denotes a derivative with respect to x. Equation (6.1) holds when no sources are present. In general, there may be an additional source term, so that

$$\mathbf{\Psi}'(x) = -\mathbf{W}\mathbf{\Psi}(x) + \mathbf{Q}(x), \tag{6.3}$$

where $\mathbf{Q}(x)$ is also a 2-component vector.

Let the source distribution be $S(x, \mu)$, $-1 \leq \mu \leq 1$, where μ is the cosine of the angle with the $+x$ direction. It is evident, either by comparison with the Boltzmann equation or by considering the relation between the sources and the discontinuity in the current, that the components $Q_{+}(x)$ and $Q_{-}(x)$ of $\mathbf{Q}(x)$ are given by

$$Q_{\pm}(x, \mu) = \pm\mu^{-1} S(x, \pm\mu), \quad 0 < \mu \leq 1. \tag{6.4}$$

The minus signs arise from our convention about the sign of μ.

The solution of Eq. (6.2) is

$$\mathbf{\Psi}(x) = \mathbf{H}(x)\mathbf{\Psi}(0) + \int_0^x \mathbf{H}(x - x')\mathbf{Q}(x')\, dx', \tag{6.5}$$

where we have used Eq. (2.7).

We restrict the discussion to a uniform source distribution for simplicity. In that case, $\mathbf{Q}$ is independent of position and the integration in Eq. (6.5) can be performed to give

$$\mathbf{\Psi}(x) = \mathbf{H}(x)\mathbf{\Psi}(0) + [\mathbf{I} - \exp (-\mathbf{W}x)]\mathbf{W}^{-1}\mathbf{Q}. \tag{6.6}$$

Define

$$\mathbf{\Phi}(x) = \mathbf{\Psi}(x) - \mathbf{W}^{-1}\mathbf{Q}. \tag{6.7}$$

Then

$$\mathbf{\Phi}(x) = \mathbf{H}(x)\mathbf{\Phi}(0). \tag{6.8}$$

Note that for a source distribution constant over all space, Eq. (6.3) degenerates to

$$-\mathbf{W}\mathbf{\Psi}(x) + \mathbf{Q} = 0, \tag{6.9}$$

or

$$\mathbf{\Psi} = \mathbf{W}^{-1}\mathbf{Q}. \tag{6.10}$$

Thus $\mathbf{\Phi}(x)$ is the change in the flux vector for $x > 0$ due to the removal of sources and scatterers in the half-space $x < 0$.

Equation (6.8) is of the same form as the transfer-matrix equation with no sources present. Further,

we know $\phi_+(0)$ and $\phi_-(t)$, where ϕ_+ and ϕ_- are the components of $\mathbf{\Phi}$:

$$\phi_+(0) = \psi_+(0) - (W^{-1}Q)_+, \tag{6.11}$$

$$\phi_-(t) = \psi_-(t) - (W^{-1}Q)_-. \tag{6.12}$$

It follows from Eq. (4.1) that

$$\phi_\pm(x) = \mathcal{H}_\pm(t, x)[\psi_+(0) - (W^{-1}Q)_+] + \mathcal{H}_\mp(t, t - x)[\psi_-(t) - (W^{-1}Q)_-]. \tag{6.13}$$

Thus

$$\psi_\pm(x) = \psi_\pm^{(0)}(x) + (\mathbf{W}^{-1}\mathbf{Q})_\pm - \mathcal{H}_\pm(t, x)(\mathbf{W}^{-1}\mathbf{Q})_+ - \mathcal{H}_\mp(t, t - x)(\mathbf{W}^{-1}\mathbf{Q})_-, \tag{6.14}$$

where $\psi_\pm^{(0)}(x)$ is the solution for the problem with the same incoming distributions but no sources.

If we introduce the symmetric and antisymmetric parts of the source distribution:

$$Q_s(u) = \frac{1}{2\mu}[S(\mu) + S(-\mu)] = \tfrac{1}{2}[Q_+(\mu) - Q_-(\mu)], \tag{6.15}$$

$$Q_a(u) = \frac{1}{2\mu}[S(\mu) - S(-\mu)] = \tfrac{1}{2}[Q_+(\mu) + Q_-(\mu)], \tag{6.16}$$

and note that

$$\mathbf{W}^{-1} = \tfrac{1}{2}\begin{pmatrix} \boldsymbol{\sigma}^{-1} + \boldsymbol{\delta}^{-1} & \boldsymbol{\sigma}^{-1} - \boldsymbol{\delta}^{-1} \\ -(\boldsymbol{\sigma}^{-1} - \boldsymbol{\delta}^{-1}) & -(\boldsymbol{\sigma}^{-1} + \boldsymbol{\delta}^{-1}) \end{pmatrix}, \tag{6.17}$$

we obtain

$$\mathbf{W}^{-1}\mathbf{Q} = \begin{pmatrix} \boldsymbol{\sigma}^{-1}Q_a + \boldsymbol{\delta}^{-1}Q_s \\ -\boldsymbol{\sigma}^{-1}Q_a + \boldsymbol{\delta}^{-1}Q_s \end{pmatrix}. \tag{6.18}$$

Then Eq. (6.14) becomes

$$\psi_\pm(x) = \psi_\pm^{(0)}(x) + [\mathbf{K}_+(t) - \mathbf{K}_\pm(t - x)e^{-\mathbf{\Lambda}x} - \mathbf{K}_\mp(x)e^{-\mathbf{\Lambda}(t-x)}]\mathbf{K}_+^{-1}(t)\boldsymbol{\delta}^{-1}Q_s + [\pm\mathbf{K}_+(t) - \mathbf{K}_\pm(t - x)e^{-\mathbf{\Lambda}x} + \mathbf{K}_\mp(x)e^{-\mathbf{\Lambda}(t-x)}]\mathbf{K}_+^{-1}(t)\boldsymbol{\sigma}^{-1}Q_a. \tag{6.19}$$

This gives the flux in the interior in terms of the incident fluxes and the angular source distribution, again with no increasing exponentials. The exiting fluxes are then found to be

$$\psi_-(0) = \psi_-^{(0)}(0) + [\mathbf{I} - \mathbf{R}(t) - \mathbf{T}(t)]\boldsymbol{\delta}^{-1}Q_s - [\mathbf{I} + \mathbf{R}(t) - \mathbf{T}(t)]\boldsymbol{\sigma}^{-1}Q_a \tag{6.20}$$

$$\psi_+(t) = \psi_+^{(0)}(t) + [\mathbf{I} - \mathbf{R}(t) - \mathbf{T}(t)]\boldsymbol{\delta}^{-1}Q_s + [\mathbf{I} + \mathbf{R}(t) - \mathbf{T}(t)]\boldsymbol{\sigma}^{-1}Q_a. \tag{6.21}$$

Define also $\psi_\pm^{(s)}(x)$ as the part of $\psi_\pm(x)$ arising from the symmetric part of the source distribution and $\psi_\pm^{(a)}(x)$ as that arising from the antisymmetric part. In general,

$$\psi_\pm(x) = \psi_\pm^{(0)}(x) + \psi_\pm^{(s)}(x) + \psi_\pm^{(a)}(x). \tag{6.22}$$

Equation (6.19) implies, as expected, that

$$\psi_+^{(s)}(x) = \psi_-^{(s)}(t - x), \tag{6.23}$$

$$\psi_+^{(a)}(x) = -\psi_-^{(a)}(t - x). \tag{6.24}$$

Thus $\psi^{(s)}$ and $\psi^{(a)}$ are themselves symmetric and antisymmetric, respectively.

Two properties of these equations are worth noting. First, the symmetric and antisymmetric parts of the source appear in essentially different ways; Q_s is multiplied by $\boldsymbol{\delta}^{-1}$ and Q_a by $\boldsymbol{\sigma}^{-1}$, with Q_s producing a symmetric flux and Q_a an antisymmetric flux. Second, because $\sigma(\mu, \mu')$ and $\delta(\mu, \mu')$ are of the form of μ^{-1} times something that has no singularities, except for a delta function at $\mu = \mu'$, the expressions $\boldsymbol{\delta}^{-1}Q_s$ and $\boldsymbol{\sigma}^{-1}Q_a$ have no singularities at $\mu = 0$, even though, according to Eqs. (6.15) and (6.16), Q_s and Q_a by themselves do. As a special case of Eqs. (6.23) and (6.24), we have

$$\psi_+^{(s)}(t) = \psi_-^{(s)}(0), \tag{6.25}$$

$$\psi_+^{(a)}(t) = -\psi_-^{(a)}(0). \tag{6.26}$$

Define

$$\mathbf{L}_\pm(t) = \mathbf{B}_\pm + \mathbf{B}_\mp \exp(-\mathbf{\Lambda}t), \tag{6.27}$$

$$\mathbf{J}_\pm(t) = \mathbf{B}_\pm - \mathbf{B}_\mp \exp(-\mathbf{\Lambda}t). \tag{6.28}$$

From Eqs. (4.7), (4.11), and (4.12), one has that

$$\mathbf{R}(t) + \mathbf{T}(t) = \mathbf{L}_-(t)\mathbf{L}_+^{-1}(t), \tag{6.29}$$

$$\mathbf{R}(t) - \mathbf{T}(t) = \mathbf{J}_-(t)\mathbf{J}_+^{-1}(t). \tag{6.30}$$

One obtains, from Eq. (6.19), after some algebra

$$\psi_+^{(s)}(x) = \mathbf{J}_+(t - x)[\mathbf{I} - \exp(-\mathbf{\Lambda}x)]\mathbf{L}_+^{-1}(t)\boldsymbol{\delta}^{-1}Q_s, \tag{6.31}$$

$$\psi_+^{(a)}(x) = \mathbf{L}_+(t - x)[\mathbf{I} - \exp(-\mathbf{\Lambda}x)]\mathbf{J}_+^{-1}(t)\boldsymbol{\sigma}^{-1}Q_a. \tag{6.32}$$

The fluxes $\psi_-^{(s)}$ and $\psi_-^{(a)}$ can be determined by Eqs. (6.23) and (6.24). The exit fluxes can be written

$$\psi_+^{(s)}(t) = (\mathbf{B}_+ - \mathbf{B}_-)[\mathbf{I} - \exp(-\mathbf{\Lambda}t)]\mathbf{L}_+^{-1}(t)\boldsymbol{\delta}^{-1}Q_s, \tag{6.33}$$

$$\psi_+^{(a)}(t) = (\mathbf{B}_+ + \mathbf{B}_-)[\mathbf{I} - \exp(-\mathbf{\Lambda}t)]\mathbf{J}_+^{-1}(t)\boldsymbol{\sigma}^{-1}Q_a. \tag{6.34}$$

For a half-space, we let $t \to \infty$. From Eq. (6.19) one has

$$\psi_\pm^{(s)}(x) = [\mathbf{I} - \mathbf{B}_\pm \exp(-\mathbf{\Lambda}x)\mathbf{B}_+^{-1}]\boldsymbol{\delta}^{-1}Q_s, \tag{6.35}$$

$$\psi_\pm^{(a)}(x) = [\pm\mathbf{I} - \mathbf{B}_\pm \exp(-\mathbf{\Lambda}x)\mathbf{B}_+^{-1}]\boldsymbol{\sigma}^{-1}Q_a. \tag{6.36}$$

The emergent flux at the free surface is then

$$\psi_-(0) = \mathbf{R}_\infty \psi_+(0) + [\mathbf{I} - \mathbf{R}_\infty]\boldsymbol{\delta}^{-1}Q_s - [\mathbf{I} + \mathbf{R}_\infty]\boldsymbol{\sigma}^{-1}Q_a, \quad (6.37)$$

where $\mathbf{R}_\infty$ is the reflection operator for a half-space. According to Eqs. (2.18) and (2.20),

$$\mathbf{R}_\infty = \mathbf{B}_-\mathbf{B}_+^{-1}. \quad (6.38)$$

Note also that the solution, as given in Eqs. (6.35) and (6.36), approaches the infinite medium solutions as $x \to \infty$.

The form of the half-space solution is closely related to the form of the solution of the half-space albedo problem, which is[2]

$$[\psi_\pm(x)]_{\text{albedo}} = \mathbf{B}_\pm \exp(-\boldsymbol{\Lambda}x)\mathbf{B}_+^{-1}\psi_+(0). \quad (6.39)$$

This similarity is not fortuitous. Consider the complementary half-space problem with a constant source $S(\mu)$ in the half-space $x < 0$, and a vacuum for $x > 0$. Let the solution of this problem be denoted by $\psi_\pm^{(2)}(x)$. Then

$$\psi_\pm^{(2)}(0) = \psi_\mp^{(s)}(0) - \psi_\mp^{(a)}(0), \quad (6.40)$$

where $\psi_\pm^{(s)}$ and $\psi_\pm^{(a)}$ still refer to the solutions of the original problem, for the half-space $x > 0$.

Let $\chi_\pm(x)$ denote the solution of the uniform source problem in the entire space. From Fig. 2, one sees that

$$\chi_+(0) = \mathbf{R}_\infty \chi_-(0) + \psi_+^{(2)}(0), \quad (6.41)$$

$$\chi_-(0) = \mathbf{R}_\infty \chi_+(0) + \psi_-^{(1)}(0). \quad (6.42)$$

In Eq. (6.41), the first term on the right represents radiation reflected from the $x < 0$ region, due to the radiation incident from $x > 0$, and the second term represents radiation born in $x < 0$ crossing at $x = 0$ for the first time. Equation (6.42) has a similar interpretation. The infinite medium solution is, from Eq. (6.18),

$$\chi_\pm(x) = \boldsymbol{\delta}^{-1}Q_s \pm \boldsymbol{\sigma}^{-1}Q_a, \quad (6.43)$$

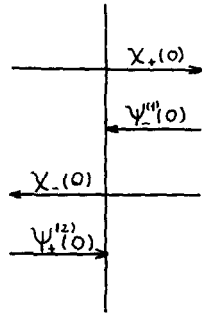


FIG. 2. Two half-spaces.

independent of x. Inserting this into Eqs. (6.41) and (6.42) and using Eq. (6.40) gives

$$\psi_\mp^{(s)}(0) \mp \psi_\mp^{(a)}(0) = \boldsymbol{\delta}^{-1}Q_s \pm \boldsymbol{\sigma}^{-1}Q_a - \mathbf{R}_\infty(\boldsymbol{\delta}^{-1}Q_s \mp \boldsymbol{\sigma}^{-1}Q_a) = (\mathbf{I} - \mathbf{R}_\infty)\boldsymbol{\delta}^{-1}Q_s \pm (\mathbf{I} + \mathbf{R}_\infty)\boldsymbol{\sigma}^{-1}Q_a, \quad (6.44)$$

in agreement with Eq. (6.37).

For $x > 0$, the half-space and full-space problems are related by

$$\chi_\pm(x) = \mathbf{B}_\pm \exp(-\boldsymbol{\Lambda}x)\mathbf{B}_+^{-1}\chi_+(0) + \psi_\pm^{(s)}(x) + \psi_\pm^{(a)}(x). \quad (6.45)$$

The first term on the right gives the flux at x due to the incident flux $\chi_+(0)$ at $x = 0$. With Eq. (6.43), this equation is equivalent to Eqs. (6.35) and (6.36).

7. THE SYMMETRIC SLAB

A considerable simplification in the treatment of slab problems results in the case of the symmetric slab, i.e., when the inputs and sources are symmetric.

Consider a sourceless symmetric slab with inputs ψ_i on both sides [Fig. 1 with $\psi_+(0) = \psi_-(t) = \psi_i$]. The outputs will then also be symmetric. Let us call the output on either side ψ_0. Clearly,

$$\psi_0 = (\mathbf{R} + \mathbf{T})\psi_i. \quad (7.1)$$

We can rewrite this as

$$\psi_0 = \mathbf{M}(t)\psi_i, \quad (7.2)$$

where, by Eq. (6.29),

$$\mathbf{M}(t) = \mathbf{L}_-(t)\mathbf{L}_+^{-1}(t). \quad (7.3)$$

According to Eqs. (4.1) and (4.7), the interior flux is given by

$$\begin{aligned}\psi_\pm(x) &= \{\mathbf{K}_\pm(t-x)\exp(-\boldsymbol{\Lambda}x) + \mathbf{K}_\mp(x)\exp[-\boldsymbol{\Lambda}(t-x)]\}\mathbf{K}_+^{-1}(t)\psi_i \\ &= \{\mathbf{B}_\pm \exp(-\boldsymbol{\Lambda}x) + \mathbf{B}_\mp \exp[-\boldsymbol{\Lambda}(t-x)]\}\mathbf{L}_+^{-1}(t)\psi_i \\ &= \mathbf{L}_\pm(t-2x)\exp(-\boldsymbol{\Lambda}x)\mathbf{L}_+^{-1}(t)\psi_i \\ &= \mathbf{L}_\mp(2x-t)\exp[-\boldsymbol{\Lambda}(t-x)]\mathbf{L}_+^{-1}(t)\psi_i. \end{aligned} \quad (7.4)$$

The symmetry implies that

$$\psi_+(t-x) = \psi_-(x), \quad (7.5)$$

so that it is necessary to compute the flux only for $x \leq t/2$, using the expression involving $\mathbf{L}_\pm(t-2x)$.

If the slab has a uniform symmetric source distribution in the interior, then we have

$$\psi_0 = \mathbf{M}(t)\psi_i + \psi_Q, \quad (7.6)$$

where ψ_Q is the output flux due to the sources. It is given by Eq. (6.33):

$$\psi_Q = (\mathbf{B}_+ - \mathbf{B}_-)[\mathbf{I} - \exp(-\mathbf{\Lambda}t)]\mathbf{L}_+^{-1}(t)\boldsymbol{\delta}^{-1}Q_s. \quad (7.7)$$

In the interior of the slab,

$$\psi_+(x) = \{\mathbf{B}_+ \exp(-\mathbf{\Lambda}x) + \mathbf{B}_- \exp[-\mathbf{\Lambda}(t-x)]\}\mathbf{L}_+^{-1}(t)\psi_i + \mathbf{J}_+(t-x)[\mathbf{I} - \exp(-\mathbf{\Lambda}x)]\mathbf{L}_+^{-1}(t)\boldsymbol{\delta}^{-1}Q_s. \quad (7.8)$$

It is sufficient to calculate ψ_+, since ψ_- is given by the symmetry condition, Eq. (7.5).

8. REFLECTED CRITICAL SLAB

Consider a multiplying slab of thickness a surrounded on both sides by identical reflectors of thickness b (Fig. 3). We ask for the conditions under which the system is critical, and in particular, for the critical core thickness a for given materials and reflector thickness b. There is no restriction that the outer material be nonmultiplying, but just that a thickness $2b$ of that material be subcritical.

We denote operators for the core by unprimed symbols and operators for the reflector by primed symbols in this and the following sections. Let ψ_i be the flux into the core from the reflector and ψ_0 be the flux out of the core into the reflector. By the discussion in the previous section,

$$\psi_0 = \mathbf{M}(a)\psi_i. \quad (8.1)$$

But also

$$\psi_i = \mathbf{R}'(b)\psi_0. \quad (8.2)$$

These equations have a solution if and only if

$$[\mathbf{M}^{-1}(a) - \mathbf{R}'(b)]\psi_0 = 0. \quad (8.3)$$

With the aid of Eq. (7.3), this becomes

$$[\mathbf{L}_+(a) - \mathbf{R}'(b)\mathbf{L}_-(a)]\mathbf{L}_-^{-1}(a)\psi_0 = 0, \quad (8.4)$$

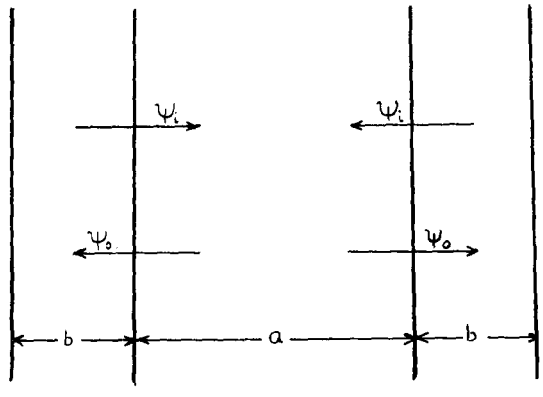


FIG. 3. Symmetrically reflected slab.

or

$$\{[\mathbf{B}_+ - \mathbf{R}'(b)\mathbf{B}_-] + [\mathbf{B}_- - \mathbf{R}'(b)\mathbf{B}_+]\exp(-\mathbf{\Lambda}a)\}\mathbf{L}_-^{-1}(a)\psi_0 = 0. \quad (8.5)$$

This is an eigenvalue equation in which the smallest value for a for which there is a solution is the critical thickness. The eigensolution ψ_0 is the critical output flux. In the absence of a reflector, Eq. (8.4) degenerates to

$$\mathbf{L}_+(a)\mathbf{L}_-^{-1}(a)\psi_0 = 0. \quad (8.6)$$

The reason we write the equation in terms of ψ_0 instead of ψ_i is that ψ_i is zero for the unreflected slab.

The flux in the core is, by Eqs. (7.3), (7.4), and (8.1),

$$\psi_\pm(x) = \mathbf{L}_\pm(a - 2x)\exp(-\mathbf{\Lambda}x)\mathbf{L}_-^{-1}(a)\psi_0. \quad (8.7)$$

Here the core is assumed to extend from 0 to a. The flux in the reflector is given by

$$\psi_\pm(x) = \mathcal{K}'_\mp(-x)\psi_0, \quad (8.8)$$

where $-b \leq x \leq 0$, or

$$\psi_\pm(x) = \mathbf{K}'_\mp(b + x)\exp(\mathbf{\Lambda}'x)\mathbf{K}_+'^{-1}(b)\psi_0. \quad (8.9)$$

In the right-hand reflector, where $a \leq x \leq a + b$

$$\psi_\pm(x) = \mathcal{K}'_\pm(x - a)\psi_0, \quad (8.10)$$

or

$$\psi_\pm(x) = \mathbf{K}'_\pm(a + b - x) \times \exp[-\mathbf{\Lambda}'(x - a)]\mathbf{K}_+'^{-1}(b)\psi_0. \quad (8.11)$$

9. CRITICAL LATTICE

Consider a lattice made up of an infinite array of alternating layers of multiplying material of thickness a and nonmultiplying material of thickness b. Again, let ψ_i and ψ_0 be respectively the fluxes into and out of a multiplying slab. In every slab, the flux distributions are symmetric. From Fig. 3, one sees that

$$\psi_0 = \mathbf{M}(a)\psi_i, \quad (9.1)$$

$$\psi_i = \mathbf{M}'(b)\psi_0. \quad (9.2)$$

The formulas of the last section still apply, with $\mathbf{M}'(b)$ replacing $\mathbf{R}'(b)$. The critical equation becomes

$$\{[\mathbf{B}_+ - \mathbf{M}'(b)\mathbf{B}_-] + [\mathbf{B}_- - \mathbf{M}'(b)\mathbf{B}_+]\exp(-\mathbf{\Lambda}a)\}\mathbf{L}_-^{-1}(a)\psi_0 = 0. \quad (9.3)$$

The flux in the multiplying medium, $0 \leq x \leq a$, is still given by Eq. (8.7). In the nonmultiplying slab, where $-b < x < 0$, Eq. (8.8) must be replaced by a form of Eq. (7.4) translated through a distance b.

Thus, for $-(b/2) \leq x \leq 0$,

$$\psi_{\pm}(x) = \mathbf{L}'_{\mp}(b + 2x) \exp(\mathbf{\Lambda}'x)\mathbf{L}'^{-1}_{+}(b)\psi_0. \quad (9.4)$$

The symmetry property can be used to obtain the flux for $-b \leq x \leq b/2$. Since the cell thickness is $(a + b)$, we have for all x

$$\psi_{\pm}(x) = \psi_{\pm}[x + n(a + b)], \quad (9.5)$$

where n is any integer.

10. SYMMETRICALLY REFLECTED SLAB WITH SYMMETRIC SOURCES

Consider a slab of thickness a with a uniform symmetric source distribution. The slab is reflected on both sides by identical reflectors of thickness b. Again, reflector operators are denoted by primes.

From Eq. (7.6),

$$\psi_0 = \mathbf{M}(a)\psi_i + \psi_Q. \quad (10.1)$$

Also,

$$\psi_i = \mathbf{R}'(b)\psi_0. \quad (10.2)$$

Putting these two equations together, one has

$$\psi_0 = [\mathbf{I} - \mathbf{M}(a)\mathbf{R}'(b)]^{-1}\psi_Q. \quad (10.3)$$

From Eqs. (6.20) and (6.29), one can write ψ_Q in the alternate form

$$\psi_Q = [\mathbf{I} - \mathbf{M}(a)]\mathbf{\delta}^{-1}Q_s. \quad (10.4)$$

Thus,

$$\begin{aligned}\psi_0 &= [\mathbf{I} - \mathbf{M}(a)\mathbf{R}'(b)]^{-1}[\mathbf{I} - \mathbf{M}(a)]\mathbf{\delta}^{-1}Q_s \\ &= \mathbf{L}_-(a)[\mathbf{L}_+(a) - \mathbf{R}'(b)\mathbf{L}_-(a)]^{-1} \\ &\quad \times [\mathbf{L}_+(a)\mathbf{L}_-^{-1}(a) - \mathbf{I}]\mathbf{\delta}^{-1}Q_s. \quad (10.5)\end{aligned}$$

Note the appearance of the same operator, $\mathbf{L}_+ - \mathbf{R}'\mathbf{L}_-$, that appears in the critical problem.

Putting together Eqs. (7.8), (10.2), and (10.5), one obtains the interior flux in the source medium, $0 \leq x \leq a$:

$$\begin{aligned}\psi_+(x) &= \{[\mathbf{B}_+ \exp(-\mathbf{\Lambda}x) + \mathbf{B}_- \exp(-\mathbf{\Lambda}(a - x))] \\ &\quad \times [\mathbf{L}_+(a) - \mathbf{R}'(b)\mathbf{L}_-(a)]^{-1}\mathbf{R}'(b)[\mathbf{L}_+(a) - \mathbf{L}_-(a)] \\ &\quad + \mathbf{J}_+(a - x)[\mathbf{I} - \exp(-\mathbf{\Lambda}x)]\}\mathbf{L}_+^{-1}(a)\mathbf{\delta}^{-1}Q_s. \quad (10.6)\end{aligned}$$

One could obtain an algebraically neater form for ψ_+, but Eq. (10.6) separates explicitly the part that vanishes as $b \to 0$. The flux $\psi_-(x)$ can be found from the symmetry condition, Eq. (7.5). The fluxes in the left-hand and right-hand reflectors are given by Eqs. (8.9) and (8.11), respectively.

11. LATTICE WITH ALTERNATING SOURCE AND SOURCE-FREE REGIONS

Consider a lattice made up of regions of thickness a with uniform symmetric sources alternating with sourceless regions of thickness b. In place of Eq. (10.2), one has

$$\psi_i = \mathbf{M}'(b)\psi_0. \quad (11.1)$$

The expressions for the fluxes in the core are therefore the same as for the symmetrically reflected slab, but with $\mathbf{M}'(b)$ replacing $\mathbf{R}'(b)$. That is, Eqs. (10.3)–(10.6) hold with $\mathbf{M}'(b)$ instead of $\mathbf{R}'(b)$. The fluxes in the sourceless medium are given by Eq. (9.4). The lattice condition, Eq. (9.5), determines the fluxes when x is not in the interval $(-b, a)$.

12. TWO-MEDIUM MILNE PROBLEM

Let the half-space $x > 0$ be filled with Medium 1 and the half-space $x < 0$ with Medium 2. We will use unprimed symbols for operators for Medium 1 and primed symbols for operators for Medium 2. We assume an infinite source, at $+\infty$, of such a strength that the flux at any finite value of x is finite and nonzero. The flux coming from infinity is given by Eq. (5.4). In this problem, unlike in the simple half-space Milne problem, there is in addition a flux $\psi_+(0)$ incident on Medium 1 from the left, reflected from Medium 2. The solution for $x \geq 0$ is then the sum of the one-medium Milne solution and the solution of the albedo problem for the half-space $x > 0$. The latter is given elsewhere.[2] Thus, for $x \geq 0$,

$$\psi_{\pm}(x) = \mathbf{B}_{\pm} \exp(-\mathbf{\Lambda}x)\mathbf{B}_+^{-1}\psi_+(0) + A\mathbf{K}_{\mp 0}(x) \exp \lambda_0 x. \quad (12.1)$$

The interface flux distributions $\psi_+(0)$ and $\psi_-(0)$ are connected by

$$\psi_+(0) = \mathbf{R}'_\infty \psi_-(0). \quad (12.2)$$

Since $\mathbf{K}_+(0) = 4\mathbf{C}_+^{-1}$, we can solve Eq. (12.1) for $\psi_-(0)$:

$$\begin{aligned}\psi_-(0) &= 4A(\mathbf{I} - \mathbf{B}_-\mathbf{B}_+^{-1}\mathbf{R}'_\infty)^{-1}(\mathbf{C}_+^{-1})_0 \\ &= 4A[(\mathbf{C}_+ + \mathbf{C}_-\mathbf{R}'_\infty)^{-1}]_0, \quad (12.3)\end{aligned}$$

where we have used Eq. (2.16) to obtain the last equality. Inserting this result in Eq. (12.1), we get for the flux distribution in Medium 1, after some algebra involving Eqs. (2.15) and (2.16):

$$\psi_{\pm}(x) = A[\mathbf{B}_{\mp} \exp(\mathbf{\Lambda}x) + \mathbf{B}_{\pm} \exp(-\mathbf{\Lambda}x)\mathcal{F}]_0, \quad (12.4)$$

where $\mathcal{F}$ is the generalization of $\mathbf{F}$ in the presence of a medium for $x < 0$:

$$\begin{aligned}\mathcal{F} &= (\mathbf{C}_- + \mathbf{C}_+\mathbf{R}'_\infty)(\mathbf{C}_+ + \mathbf{C}_-\mathbf{R}'_\infty)^{-1} \\ &= -(\mathbf{B}_+ - \mathbf{R}'_\infty\mathbf{B}_-)^{-1}(\mathbf{B}_- - \mathbf{R}'_\infty\mathbf{B}_+). \quad (12.5)\end{aligned}$$

Equation (12.4) is identical to Eq. (5.5) except that $\mathcal{F}$ replaces $\mathbf{F}$. Thus the expression for the extrapolated end point in the two-medium problem is

$$z_0 = (1/2\lambda_0) \ln (-1/\mathcal{F}_{00}). \tag{12.6}$$

When $\mathbf{R}'_\infty \to 0$, $\mathcal{F} \to \mathbf{F}$ and all the results approach the one-medium results.

For $x < 0$ we have a simple albedo problem and the flux is

$$\psi_\pm(x) = \mathbf{B}'_\mp \exp(\mathbf{\Lambda}'x)\mathbf{B}'^{-1}_+ \psi_-(0). \tag{12.7}$$

In obtaining Eq. (12.7) we have made the appropriate changes in the known results for the albedo problem for the half-space $x > 0$.[2]

If we use Eq. (12.3) for $\psi_-(0)$, we obtain for $x < 0$

$$\psi_\pm(x) = [4A\mathbf{B}'_\mp \exp(\mathbf{\Lambda}'x)(\mathbf{C}_+\mathbf{B}'_+ + \mathbf{C}_-\mathbf{B}'_-)^{-1}]_0. \tag{12.8}$$

Thus, if all space is filled with Medium 1, i.e., if Media 1 and 2 are identical, Eq. (2.15) implies that for $x < 0$,

$$\psi_\pm(x) = A\mathbf{B}_{\mp 0} \exp(\lambda_0 x). \tag{12.9}$$

Since in this case,

$$\mathcal{F} = 0, \tag{12.10}$$

Eq. (12.9) then holds for all x. Note that Eq. (12.9) is obtained from Eq. (5.4) by letting $\mathbf{K}_\pm \to \mathbf{B}_\pm$. This is in accord with our ideas that $\mathbf{K}_\pm$ is the generalization of $\mathbf{B}_\pm$ that takes into account boundary effects.

13. CONCLUSION

The convenience in notation and the generality offered by the transfer-matrix method point up strong similarities between apparently very different problems. We note, for instance:

(1) The ubiquitous appearance of $\mathbf{K}_\pm$, $\mathbf{L}_\pm$, and $\mathbf{J}_\pm$.

(2) The generalization of $\mathbf{B}_\pm$ in semi-infinite-medium problems to $\mathbf{K}_\pm(t)$ for slabs of thickness t.

(3) The generalization of $\mathbf{B}_\pm$ in problems involving an unreflected medium to $(\mathbf{B}_\pm - \mathbf{R}'\mathbf{B}_\mp)$ in problems for reflected media.

(4) The identical formulation of lattice problems and symmetric reflector problems, with $\mathbf{M}'$ in the lattice problems replacing $\mathbf{R}'$ in the reflector problems.

These generalizations are summarized in Table I.

These features hold completely generally, for every type of interaction, and for every approximation method. They can be used to predict results for problems not yet solved. The simplest such extension is given by the last point. A given lattice problem can be formulated immediately in terms of the corresponding reflected slab problem and vice versa.

It has been shown in the literature for isotropic scattering[2] (and it is true in general[4]) that to within a normalization factor, the kernel of the integral operator $\mathbf{B}_+$ is the forward part of the Case eigenfunction and the kernel of $\mathbf{B}_-$ is the backward part. Thus the generalizations of $\mathbf{B}_\pm$ describe the effective modification of the modes by reflections at the boundaries. The comparison with Case's method will be described in detail in a future article.

TABLE I. Generalizations of $B_\pm$ operators.

Geometry	Operator
Half-space	$\mathbf{B}_\pm$
Slab	$\mathbf{K}_\pm(t)$
Symmetric slab	$\mathbf{L}_\pm(t)$
Reflected symmetric slab	$\mathbf{B}_\pm - \mathbf{R}'\mathbf{B}_\mp$ $\mathbf{L}_\pm(t) - \mathbf{R}'\mathbf{L}_\mp(t)$
Lattice	$\mathbf{B}_\pm - \mathbf{M}'\mathbf{B}_\mp$ $\mathbf{L}_\pm(t) - \mathbf{M}'\mathbf{L}_\mp(t)$

Finally, when the operators are reduced to matrices in some finite approximation scheme, all the indicated algebra can actually be carried out. The generalized operators are useful intermediate results. In effect, they enable one to decouple various parts of the calculation.

Extensive numerical calculations have been performed for several of the problems listed, for a variety of differential cross sections for one speed. Some of the results have already been given[9,10] and others will be presented elsewhere. The speed and accuracy of the computations compare favorably with those by other methods. Some calculations for polarized photons, as well as for energy-dependent neutron interactions in a multigroup approximation, have been done.

* This work was supported by the National Science Foundation under Grants NSF-GK-148. and NSF-GK-4037.

[1] R. Aronson and D. L. Yarmush, J. Math. Phys. **7**, 221 (1966).

[2] R. Aronson, Nucl. Sci. Eng. **27**, 271 (1967).

[3] See, for instance, M. R. Mendelsohn, J. Math. Phys. **7**, 345 (1966), where the problems of Secs. 6, 10, and 11 of this paper are solved by Case's method.

[4] R. Aronson, in *Proceedings of Neutron Transport Theory Conference*, Virginia Polytechnic Institute, Blacksburg, Virginia, Report ORO-3858-1, January, 1969. This work will be reported in detail in a subsequent paper.

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

[6] E. M. Gelbard, *Spherical Harmonics Methods: P_L and Double-P_L Approximations*, in *Computing Methods in Reactor Physics*, H. Greenspan, C. N. Kelber, and D. Okrent, Eds. (Gordon & Breach, Sci. Publishers, Inc., New York, 1968).

[7] In fact, λ_0 is always discrete in one speed problems.

[8] We do not discuss here the conditions under which the diagonalization is possible. It is sufficient that it is always possible in a multigroup treatment of energy. The question is equivalent to whether the eigenfunctions in Case's method are complete.

[9] G. Caroll and R. Aronson, Trans. Am. Nucl. Soc. **10**, 606 (1967).

[10] G. Caroll, Ph.D. thesis, New York University, New York, 1969.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Nonlinear Electrodynamics: Lagrangians and Equations of Motion

GUY BOILLAT
Département de Mathématiques, Université de Clermont, Clermont-Ferrand, France

(Received 17 January 1969)

After a brief discussion of well-known classical fields we formulate two principles: When the field equations are hyperbolic, particles move along rays like disturbances of the field; the waves associated with stable particles are exceptional. This means that these waves will not transform into shock waves. Both principles are applied to nonlinear electrodynamics. The starting point of the theory is a Lagrangian which is an arbitrary nonlinear function of the two electromagnetic invariants. We obtain the laws of propagation of photons and of charged particles, along with an anisotropic propagation of the wavefronts. The general "exceptional" Lagrangian is found. It reduces to the Lagrangian of Born and Infeld when some constant (probably simply connected with the Planck constant) vanishes. A nonsymmetric tensor is introduced in analogy to the Born–Infeld theory, and finally, electromagnetic waves are compared with those of Einstein–Schrödinger theory.

1. WAVE DYNAMICS

A. Waves, Rays, and Exceptional Waves

Quite generally we have seen that, as far as wave propagation is concerned, the field equations can always be written as a quasilinear system of partial differential equations that is a system of equations which are linear with respect to the highest derivatives of the dependent field variables.[1] We assume that across the hypersurface S,

$$\varphi(x^\alpha) = 0, \quad \alpha = 0, 1, \cdots, n,$$

these derivatives are discontinuous. Precisely, if q is the highest order of the derivatives of the field component u, we assume that, after a change of variables,

$$x^\alpha \to \varphi(x^\alpha), \xi^i(x^\alpha), \quad i = 1, 2, \cdots, n,$$

the jump

$$\left[\frac{\partial^q u}{\partial \varphi^q}\right] = \left(\frac{\partial^q u}{\partial \varphi^q}\right)_{\varphi=+0} - \left(\frac{\partial^q u}{\partial \varphi^q}\right)_{\varphi=-0} = \delta^q u$$

is finite, while

$$\delta^r u = 0, \quad 0 \leq r < q.$$

S is called the wave surface, and since Hadamard[2] it is well known that

$$[\partial_{\alpha_1\alpha_2\cdots\alpha_q} u] = \varphi_{\alpha_1}\varphi_{\alpha_2}\cdots\varphi_{\alpha_q}\delta^q u,$$

where a subscript α denotes partial differentiation with respect to x^α.

To allow for the above-mentioned discontinuities, φ must be a solution of some characteristic equation of the form

$$\psi = G^{\alpha\beta\cdots\nu}\varphi_\alpha\varphi_\beta\cdots\varphi_\nu, \quad \psi = 0. \tag{1.1}$$

When the field is nonlinear, the completely symmetric tensor G may depend on the field and all its continuous derivatives.

The discontinuities—or disturbances—propagate themselves along the rays:

$$\frac{dx^\alpha}{d\sigma} = \frac{\partial\psi}{\partial\varphi_\alpha}, \quad \frac{d\varphi_\alpha}{d\sigma} = -\frac{\partial\psi}{\partial x^\alpha}. \tag{1.2}$$

Thus in general relativity the ray velocity $\partial\psi/\partial\varphi_\alpha$ must be a timelike (or possibly null) vector, i.e.,

$$\mathcal{N} = \left(g^{\alpha\beta}\frac{\partial\psi}{\partial\varphi_\alpha}\frac{\partial\psi}{\partial\varphi_\beta}\right)^{\frac{1}{2}}$$

must exist (be real). We normalize (if $\mathcal{N} \neq 0$) the ray velocity

$$u^\alpha = \frac{1}{\mathcal{N}}\frac{\partial\psi}{\partial\varphi_\alpha}, \quad u_\alpha u^\alpha = 1.$$

Equation (1.1) shows that

$$u^\alpha\varphi_\alpha = 0, \tag{1.3}$$

which now implies

$$g^{\alpha\beta}\varphi_\alpha\varphi_\beta \leq 0.$$

Now what about the growth of the disturbances of the field (such as $\delta^q u$)? Though we cannot here enter into details, the following results generally hold[3]: It is possible (because of the nonlinearity of the field) by a suitable choice of the (initial) disturbance to produce an accelerated wave (e.g., in fluid mechanics, push the piston into the cylinder). Consequently, the disturbances will cease to be finite after some critical time, thus tending to a shock. From this moment on, the field equations are no longer valid and must be replaced by others (e.g., Rankine–Hugoniot conditions). However, the form of the field equations might be such that this phenomenon does not appear on some wavefront; then we say with Lax[4] that this wave is exceptional. If this is true for all the waves, the system of field equations is called completely exceptional.[4,5]

The condition for (1.1) to be exceptional is[1,3]

$$\delta\psi = 0, \quad \varphi_\alpha\varphi_\beta \cdots \varphi_\nu \delta G^{\alpha\beta\cdots\nu} = 0, \tag{1.4}$$

or, equivalently,

$$\varphi_\alpha \delta u^\alpha = 0. \tag{1.5}$$

For instance, gravitational waves,

$$g^{\alpha\beta}\varphi_\alpha\varphi_\beta = 0,$$

are exceptional, for $g^{\alpha\beta}$ only depends on the metric tensor $g_{\alpha\beta}$ whose first-order derivatives are continuous[6] (hence $\delta g^{\alpha\beta} = 0$).

Alfvén waves of magnetohydrodynamics are exceptional too[3,5]; Alfvén shocks cannot be produced directly from a continuous initial state.[7]

B. Wave Dynamics: Two Principles

In this section we want to show briefly how particles, sources of hyperbolic fields, may be considered as disturbances of the field (in the sense of Sec. 1A) and therefore move along the rays of the associated waves. Let us distinguish several cases.

(1) The field equations contain the four-dimensional velocity u^α. They give the trajectories and, in order to be able to identify them with a family of rays, u^α must be a ray velocity. Hence, by virtue of (1.3), the characteristic equation

$$u^\alpha\varphi_\alpha = 0 \tag{1.6}$$

must exist. Let us examine the equations derived from the conservation laws

$$\nabla_\alpha T^{\alpha\beta} = 0$$

of two classical energy tensors.

i. Incoherent Matter

$$T^{\alpha\beta} = \rho u^\alpha u^\beta,$$

$$\nabla_\alpha(\rho u^\alpha) = 0, \tag{1.7}$$

$$u^\alpha\nabla_\alpha u^\beta = 0. \tag{1.8}$$

The discontinuities are of first order and a convenient way to obtain them is to make the replacement

$$\partial_\alpha, \nabla_\alpha \to \varphi_\alpha\delta. \tag{1.9}$$

(The Christoffel symbols are continuous; see the end of the previous section.) From the very definition of δ it is obvious that it operates as does the operator of differentiation, i.e., $f(u_1, u_2, \cdots)$ being a continuously differentiable function of the continuous field variables $u_1, u_2, \cdots$:

$$\delta f = \frac{\partial f}{\partial u_1}\delta u_1 + \frac{\partial f}{\partial u_2}\delta u_2 + \cdots.$$

From

$$\varphi_\alpha\delta(\rho u^\alpha) = 0, \quad u^\alpha\varphi_\alpha\delta u^\beta = 0,$$

we immediately get (1.6) and (1.5); the wave is exceptional. Equation (1.8) describes the time track of a test particle (geodesic).

ii. Perfect Fluids

The entropy satisfies the adiabatic condition

$$u^\alpha\partial_\alpha S = 0;$$

otherwise the motion of the fluid would not be determined.[8] On the other hand, an equation similar to (1.7) holds and it is easy to see that we obtain again (1.6) and (1.5) for the entropy wave,

$$\delta S \neq 0.$$

Similar results are valid for magnetohydrodynamics.[9] Thus we have seen on these examples that particles move in the same way as the disturbances of the wave (1.6) and that this wave is exceptional. This, of course, does not give us more information about the particle paths; they are what they are, already given by the field equations. But let us go further.

(2) From the Einstein equations

$$S_{\alpha\beta} = \chi T_{\alpha\beta},$$

we get

$$g^{\alpha\beta}\varphi_\alpha\varphi_\beta = 0, \quad \delta^2 g_{\alpha\beta} \neq 0,$$

no matter what the continuous energy–momentum tensor is. This is the equation of a null surface propagating with the speed of light. The wave is exceptional (Sec. 1A) and it is well known that the rays are null geodesics and the trajectories of particles with vanishing mass.

(3) The field equations do not give (directly) the laws of motion of particles. This is the case of nonlinear electrodynamics which we investigate in Sec. 2.

We propose now the following principles[10]:

Ray Principle: When the field equations are hyperbolic, particles move along rays like disturbances of the field.

Principle of Exception: The waves associated with stable particles are exceptional.

Generally other waves—that is, waves which are not associated with particles (e.g., fast or slow waves

in magnetohydrodynamics) or which are associated with unstable particles—are not exceptional. In the latter case the critical time might be connected with the lifetime of the particle: if the instability lies in the nature of the particle, it must also lie in the nature of the disturbance that represents it.

However, if all the waves of the field can represent stable particles, the system of field equations must be completely exceptional.

As yet we have only checked the above principles on the examples of (1) and (2). We are now going to see what comes out of them when applied to nonlinear electrodynamics.

2. NONLINEAR ELECTRODYNAMICS

A. Field Equations

It is well known that from the electromagnetic tensor $F_{\alpha\beta}$ and its dual,

$$F^{*\gamma\delta} = \tfrac{1}{2}\eta^{\alpha\beta\gamma\delta}F_{\alpha\beta}, \quad \eta^{\alpha\beta\gamma\delta} = -\frac{1}{(-g)^{\frac{1}{2}}}\epsilon^{\alpha\beta\gamma\delta}, \quad (2.1)$$

one can construct the two invariants

$$Q = \tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta}, \quad R = \tfrac{1}{4}F^{*}{}_{\alpha\beta}F^{\alpha\beta}. \quad (2.2)$$

[In a pseudo-Cartesian frame at some point, $Q = \frac{1}{2}(\mathbf{H}^2 - \mathbf{E}^2)$, $R = \mathbf{E}\cdot\mathbf{H}$, where $\mathbf{E}$ and $\mathbf{H}$ are the ordinary electric and magnetic 3-vectors.]

Let us introduce the 4-potential ϕ_α through

$$F_{\alpha\beta} = \partial_\alpha\phi_\beta - \partial_\beta\phi_\alpha \quad (2.3)$$

and the Lagrangian $L(Q, R)$ arbitrary function of the invariants (2.2).

Then the Euler equations for the variational principle applied to

$$\int (-g)^{\frac{1}{2}} L\, dx$$

read as

$$\nabla_\alpha L^{\alpha\beta} = 0, \quad L^{\alpha\beta} = L_Q F^{\alpha\beta} + L_R F^{*\alpha\beta}, \quad L_Q = \partial L/\partial Q, \quad (2.4)$$

which, together with

$$\nabla_\alpha F^{*\alpha\beta} = 0, \quad (2.5)$$

locally satisfied by (2.3), and the Einstein equations (cf. Ref. 11),

$$\begin{aligned} S_{\alpha\beta} &= \chi T_{\alpha\beta}, \\ T^{\alpha\beta} &= Lg^{\alpha\beta} - L^{\alpha\rho}F^{\beta}{}_{\rho} \\ &= L_Q\tau^{\alpha\beta} + (L - QL_Q - RL_R)g^{\alpha\beta}, \end{aligned} \quad (2.6)$$

form the system of the field equations. We recall that

$$\tau^{\alpha\beta} = Qg^{\alpha\beta} - F^{\alpha\rho}F^{\beta}{}_{\rho} \quad (2.7)$$

is the Maxwellian energy tensor.

Remark: Equation (2.4) shows that L needs to be defined only up to multiplicative and additive constants [more precisely, because of (2.5), to an additive linear function of R]. The first constant can be absorbed by χ and the second one can cancel a "cosmological" constant of the Einstein tensor, so we shall not bother here ensuring the conditions

$$L_Q(0, 0) = 1, \quad L(0, 0) = 0.$$

The following relations are useful too:

$$F^{*}{}_{\alpha\rho}F^{\beta\rho} = Rg^{\beta}_{\alpha}, \quad F^{*}{}_{\alpha\rho}F^{*\beta\rho} = F_{\alpha\rho}F^{\beta\rho} - 2Qg^{\beta}_{\alpha}, \quad (2.8)$$

$$\tau_{\alpha\rho}\tau^{\beta\rho} = P^2 g^{\beta}_{\alpha}, \quad P = (Q^2 + R^2)^{\frac{1}{2}}. \quad (2.9)$$

We shall suppose, of course, that L is nonlinear, and more precisely that

$$L_Q(L_{QQ} + L_{RR}) \neq 0.$$

B. Wave Surface

We assume that, across the wavefront $\varphi(x^\alpha) = 0$, ϕ_α has a discontinuity of the second order,

$$\pi_\alpha = \delta^2\phi_\alpha,$$

so that, from (2.3),

$$\delta F_{\alpha\beta} = \varphi_\alpha\pi_\beta - \varphi_\beta\pi_\alpha, \quad \varphi_\alpha = \partial_\alpha\varphi. \quad (2.10)$$

Equations (2.4) and (2.5) give [see Eq. (1.9)]

$$\begin{aligned} \varphi_\alpha(F^{\alpha\beta}\delta L_Q + F^{*\alpha\beta}\delta L_R + L_Q\delta F^{\alpha\beta}) = 0, \\ \varphi_\alpha\delta F^{*\alpha\beta} = 0, \end{aligned} \quad (2.11)$$

or by introducing the vectors

$$U^\beta = F^{\alpha\beta}\varphi_\alpha, \quad V^\beta = F^{*\alpha\beta}\varphi_\alpha$$

and taking account of (2.10),

$$U_\alpha\delta L_Q + V_\alpha\delta L_R + L_Q\{\mathfrak{G}\pi_\alpha - (\pi^\beta\varphi_\beta)\varphi_\alpha\} = 0. \quad (2.12)$$

If $\mathfrak{G} = g^{\alpha\beta}\varphi_\alpha\varphi_\beta \neq 0$, this equation shows that

$$\pi_\alpha = aU_\alpha + bV_\alpha + c\varphi_\alpha, \quad (2.13)$$

with

$$a\mathfrak{G}L_Q + \delta L_Q = 0, \quad b\mathfrak{G}L_Q + \delta L_R = 0. \quad (2.14)$$

The coefficient c is undetermined, but the last term of (2.13) has no effect on the disturbance (2.10) of the electromagnetic tensor; the Lorentz condition ($\nabla_\alpha\phi^\alpha = 0$) will make it vanish (provided $\mathfrak{G} \neq 0$).

From (2.2) and (2.10), we have

$$\delta Q = U^\alpha\pi_\alpha, \quad \delta R = V^\alpha\pi_\alpha; \quad (2.15)$$

inserting (2.13) and taking account of (2.7) and (2.8),

we obtain

$$\delta Q = a(Q\mathcal{G} - \mathcal{C}) + bR\mathcal{G}, \quad \delta R = aR\mathcal{G} - b(Q\mathcal{G} + \mathcal{C}); \qquad \mathcal{C} = \tau^{\alpha\beta}\varphi_\alpha\varphi_\beta. \quad (2.16)$$

Now Eqs. (2.14) give

$$a\{\mathcal{G}(L_Q + QL_{QQ} + RL_{QR}) - \mathcal{C}L_{QQ}\} + b\{\mathcal{G}(RL_{QQ} - QL_{QR}) - \mathcal{C}L_{QR}\} = 0,$$

$$a\{\mathcal{G}(QL_{QR} + RL_{RR}) - \mathcal{C}L_{QR}\} + b\{\mathcal{G}(L_Q + RL_{QR} - QL_{RR}) - \mathcal{C}L_{RR}\} = 0, \quad (2.17)$$

wherefrom we get the characteristic equations,[12]

$$\mathcal{C} + \mu\mathcal{G} = 0; \quad (2.18)$$

$\mu(Q, R)$ being a solution of

$$\varpi\mu^2 + \mu + \omega - \varpi P^2 = 0; \quad (2.19)$$

$$\varpi = \frac{L_{QQ}L_{RR} - L_{QR}^2}{L_Q(L_{QQ} + L_{RR})},$$

$$\omega = \frac{L_Q + Q(L_{QQ} - L_{RR}) + 2RL_{QR}}{L_{QQ} + L_{RR}}.$$

The roots of (2.19) always exist since the discriminant can be written[13]

$$\Delta = \left(\frac{L_{QQ} - L_{RR}}{L_{QQ} + L_{RR}} - 2Q\varpi\right)^2 + 4\left(\frac{L_{QR}}{L_{QQ} + L_{RR}} - R\varpi\right)^2. \quad (2.20)$$

(We disregard the case $\varpi = 0$ that does not seem of physical interest.)

Let $G^{\alpha\beta}$ be a normal tensor,[6]

$$G^{\alpha\beta} = s_0 V^\alpha_{(0)} V^\beta_{(0)} - \sum_{i=1}^{3} s_i V^\alpha_{(i)} V^\beta_{(i)}, \quad V_{(\mu)\alpha} V^\alpha_{(\nu)} = \eta_{\mu\nu},$$

and let the characteristic equation be

$$G^{\alpha\beta}\varphi_\alpha\varphi_\beta = 0.$$

Then the following inequalities[10,14] must hold according to the definition of the ray velocity (see Sec. 1A):

$$0 < \frac{s_i}{s_0} \leq 1, \quad i = 1, 2, 3. \quad (2.21)$$

If one equality is satisfied, light velocity is reached in two directions:

$$s_1 = s_0 \to \mathcal{G} = 0, \quad l_\alpha l^\alpha = 0:$$

$$\frac{dx^\alpha}{d\sigma} = l^\alpha \propto \varphi^\alpha \propto V^\alpha_{(0)} \pm V^\alpha_{(1)}. \quad (2.22)$$

[When $\mathcal{N} = 0$ (Sec. 1A), we note l^α instead of u^α in order to avoid confusion.]

Here

$$G^{\alpha\beta} = \tau^{\alpha\beta} + \mu g^{\alpha\beta}$$

and the eigenvalues are

$$s_0 = \mu + P; \quad s_i: \mu + P, \quad \mu - P, \quad \mu - P.$$

We must have (2.21)

$$\mu > P. \quad (2.23)$$

This inequality must be true for both values of μ. Hence,[13]

$$\varpi < 0, \quad \varpi P + \tfrac{1}{2} > 0, \quad \omega + P < 0. \quad (2.24)$$

Propagation with the speed of light [(2.22) is satisfied] will be studied in Sec. 2.D.

C. Double Root: The Born–Infeld Lagrangian

The discriminant (2.20) is positive and vanishes identically only for a solution of the system of partial differential equations:

$$p(r - t) - 2x(rt - s^2) = 0, \qquad ps - y(rt - s^2) = 0. \quad (2.25)$$

Here we have replaced $L(Q, R)$ by $z(x, y)$ and used p, q, r, s, t to denote the partial derivatives. Since we assume that $\varpi \neq 0$, the system is equivalent to

$$p(y_q - x_p) - 2x = 0, \qquad py_p + y = 0. \quad (2.26)$$

By integration this equation gives

$$y = -F'(q)/p.$$

Then

$$x_q = y_p = F'/p^2, \quad x = F(q)/p^2 + G(p).$$

Inserting these expressions into (2.26) results in

$$F''(q) + pG'(p) + 2G(p) = 0,$$

or (since $\varpi \neq 0$)

$$F''(q) = \text{const} = k, \quad F(q) = \tfrac{1}{2}kq^2 + a$$

(the coefficient of q in F can be taken equal to zero by changing q into q + const; cf. Remark, Sec. 2.A).

$$G(p) = (b/p^2) - \tfrac{1}{2}k.$$

p and q are easily obtained and so is z,

$$z = \text{const}[-y^2 + k(2x + k)]^{\frac{1}{2}}.$$

By (2.19) and (2.25) we have

$$\mu = -\frac{1}{2\varpi} = \frac{p}{rt - s^2}[\tfrac{1}{2}(r - t) - r]$$
$$= x - py_q = x + F'' = x + k.$$

Therefore,[13]

$$L = [-R^2 + k(2Q + k)]^{\frac{1}{2}}, \tag{2.27}$$
$$\mu = Q + k, \tag{2.28}$$

and according to (2.23) k must be positive. This is the Lagrangian of Born and Infeld.[11] [Note that it can also be written $L = (\mu^2 - P^2)^{\frac{1}{2}}$.] We shall see the fundamental role it plays in the theory.

The spherically symmetric solution obtained with this Lagrangian is well known[11,15]:

$$F_{01} = k^{\frac{1}{2}}\Big/\left[1 + \left(\frac{r}{r_0}\right)^4\right]^{\frac{1}{2}}, \quad x^0 = t, x^1 = r. \tag{2.29}$$

It removes the singularity at the origin that appeared in the Maxwellian field:

$$r = 0: \quad F_{01} = k^{\frac{1}{2}}.$$

This finite value is called, after Born and Infeld, the "absolute field."

D. Propagation with the Speed of Light

At the end of Sec. 2B we have seen that, owing to (2.22), the speed of light is reached in two directions that we are going to determine. Equations (2.7), (2.8), and (2.18) give

$$U_\alpha U^\alpha = (\mu + Q)\mathcal{G},$$
$$U_\alpha V^\alpha = R\mathcal{G}, \quad V_\alpha V^\alpha = U_\alpha U^\alpha - 2Q\mathcal{G}. \tag{2.30}$$

Thus when $\mathcal{G} = 0$, U_α, V_α, φ_α are null vectors. Since on the other hand

$$U^\alpha \varphi_\alpha = V^\alpha \varphi_\alpha = 0,$$

it is necessary that

$$U_\alpha = u\varphi_\alpha, \quad V_\alpha = v\varphi_\alpha. \tag{2.31}$$

Multiplying (2.8) by φ_β results in

$$uv = -R, \quad v^2 - u^2 = 2Q$$

or

$$u = \pm(P - Q)^{\frac{1}{2}}, \quad v = \mp(P + Q)^{\frac{1}{2}} \operatorname{sgn} R.$$

The equations [cf. (2.22) and (2.31)]

$$\frac{dx^\alpha}{d\sigma} = l^\alpha, \quad [F^{\alpha\beta} \pm (P - Q)^{\frac{1}{2}} g^{\alpha\beta}]l_\beta = 0, \quad l_\alpha l^\alpha = 0$$

give the trajectories of photons.[10]

It is worthwhile to note that for a null field ($P \equiv 0$) these equations—together with (2.4)–(2.6)—coincide with those of the Maxwellian case[16]:

$$P = 0, \quad F^{\alpha\beta}l_\beta = 0, \quad l_\alpha l^\alpha = 0,$$

implying[17,18]

$$F_{\alpha\beta} = l_\alpha p_\beta - l_\beta p_\alpha, \quad p_\alpha p^\alpha = -1,$$
$$p_\alpha l^\alpha = 0, \quad \tau_{\alpha\beta} = l_\alpha l_\beta, \quad l^\alpha \nabla_\alpha l^\beta = 0.$$

The null field still describes a fluid of photons[6] that still follow null geodesics.

In Sec. 2F we shall show that the results of this paragraph are compatible with Eq. (2.12) by calculating the associated disturbances.

E. Rays[19]

Let us define the tensors

$$G^{\alpha\beta} = (\tau^{\alpha\beta} + \mu g^{\alpha\beta})/(\mu^2 - P^2)^{\frac{1}{2}},$$
$$H_{\alpha\beta} = -(\tau_{\alpha\beta} - \mu g_{\alpha\beta})/(\mu^2 - P^2)^{\frac{1}{2}},$$

such that, by virtue of (2.9),

$$H_{\alpha\beta}G^{\beta\gamma} = g_\alpha^\gamma.$$

Let us introduce the relative spacelike 4-vectors,

$$e_\alpha = F_{\rho\alpha}u^\rho, \quad h_\alpha = F^*_{\rho\alpha}u^\rho, \quad S^\alpha = \eta^{\alpha\beta\gamma\delta}e_\beta h_\gamma u_\delta \qquad (u_\alpha u^\alpha = 1),$$

which coincide in the rest frame ($u^i = 0$, $u^0 = 1$) with the electric, magnetic, and Poynting vectors, respectively. When expressed with these quantities, we have

$$F_{\alpha\beta} = u_\alpha e_\beta - u_\beta e_\alpha + \eta_{\alpha\beta\gamma\delta}h^\gamma u^\delta,$$
$$F^{*\alpha\beta} = u^\alpha h^\beta - u^\beta h^\alpha - \eta^{\alpha\beta\gamma\delta}e_\gamma u_\delta,$$
$$\tau_{\alpha\beta} = (e^2 + h^2)(u_\alpha u_\beta - \tfrac{1}{2}g_{\alpha\beta}) - (e_\alpha e_\beta + h_\alpha h_\beta + u_\alpha S_\beta + u_\beta S_\alpha), \tag{2.32}$$
$$e^2 = -e_\alpha e^\alpha, \cdots.$$

From Sec. 1A and Eq. (2.18), the rays are given by

$$\frac{dx^\alpha}{d\tau} = \frac{1}{(-\mathcal{G})^{\frac{1}{2}}} G^{\alpha\beta}\varphi_\beta = u^\alpha, \tag{2.33}$$

$$\frac{d\varphi_\alpha}{d\tau} = -\frac{1}{2(-\mathcal{G})^{\frac{1}{2}}} \partial_\alpha G^{\beta\gamma}\varphi_\beta\varphi_\gamma. \tag{2.34}$$

Equation (2.33) yields

$$\frac{\varphi_\alpha}{(-\mathcal{G})^{\frac{1}{2}}} = H_{\alpha\beta}u^\beta = \frac{1}{(\mu^2 - P^2)^{\frac{1}{2}}}\{(\zeta - e^2)u_\alpha + S_\alpha\},$$
$$\zeta = \mu - Q, \tag{2.35}$$

while

$$u^\alpha \varphi_\alpha = 0 \rightarrow H_{\alpha\beta} u^\alpha u^\beta = 0,$$

i.e.,

$$e^2 = \zeta, \tag{2.36}$$

or equivalently,

$$\tfrac{1}{2}(e^2 + h^2) = \mu, \quad S^2 = \mu^2 - P^2$$

[because $Q = \frac{1}{2}(h^2 - e^2)$, $R = e_\alpha h^\alpha$, $S^2 = e^2 h^2 - R^2$].

Thus,

$$\frac{\varphi_\alpha}{(-\mathfrak{G})^{\frac{1}{2}}} = \frac{S_\alpha}{(\mu^2 - P^2)^{\frac{1}{2}}}, \tag{2.37}$$

the Poynting vector is orthogonal to the wavefront, ζ (or rather $\zeta^{\frac{1}{2}}$) is the absolute field of the charged particle which moves along the rays according to the first principle of Sec. 1B. μ is the absolute density of energy.

According to (2.33) and (2.34), particle paths are null geodesics for the metric

$$d\bar{s}^2 \propto H_{\alpha\beta}\, dx^\alpha\, dx^\beta.$$

Consequently,

$$u^\alpha \nabla_\alpha u^\beta = (g_\alpha^\beta - u_\alpha u^\beta)\gamma^\alpha,$$

where

$$\gamma^\lambda = -\tfrac{1}{2} G^{\lambda\rho}(\nabla_\mu H_{\nu\rho} + \nabla_\nu H_{\mu\rho} - \nabla_\rho H_{\mu\nu}) u^\mu u^\nu.$$

[As for the uncharged particles, they move along geodesics according to (1.8).]

Now it is easy to show that the Born–Infeld Lagrangian is well suited to describe a particle in the spherically symmetric approximation. We have (2.29)

$$e^2 = -g^{00} g^{11} (F_{01})^2 = F_{01}^2, \quad u^i = 0.$$

On the other hand, (2.28) and (2.35),

$$\zeta = k, \tag{2.38}$$

so that at the origin $r = 0$, where the particle is at rest, the equality (2.36) is true.

To end this section, we shall say a few words about the ray velocity diagram. This diagram gives us the shape of the wavefront resulting from a point disturbance (that is, a disturbance initially localized within a small sphere) and propagating into a constant state (specified by $g_{\alpha\beta} = \eta_{\alpha\beta}$, **E**, **H**). A proper choice of the reference frame in the three-dimensional physical space (the z axis lies in the direction of the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$) allows one to write the equation of the wavefront at time t[20]:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{1}{c^2}(z - Vt)^2 = t^2, \quad \mathbf{V} = \frac{\mathbf{S}}{\mu + w}, \tag{2.39}$$

with

$$V = |\mathbf{V}|, \quad w = \tfrac{1}{2}(E^2 + H^2),$$

$$a = \left(\frac{\mu + P}{\mu + w}\right)^{\frac{1}{2}}, \quad b = \left(\frac{\mu - P}{\mu + w}\right)^{\frac{1}{2}}, \quad c = ab.$$

It is immediately seen that

$$1 > a \geq b > c, \quad V^2 = (1 - a^2)(1 - b^2).$$

The propagation is anisotropic, and besides there is a draught effect of speed **V**. In the limit of an infinite absolute energy density $\mu \rightarrow \infty$, the ellipsoid (2.39) becomes the usual sphere of the linear theory of Maxwell.

F. Disturbances

i. Disturbances Associated with Photons[12,16]

Inserting (2.31) into (2.15) and (2.12) results in

$$\delta Q = u\pi^\alpha \varphi_\alpha, \quad \delta R = v\pi^\alpha \varphi_\alpha, \quad (\omega - P)\pi^\alpha \varphi_\alpha = 0,$$

or [see (2.24)]

$$\pi^\alpha \varphi_\alpha = 0, \tag{2.40}$$

$$\delta Q = \delta R = 0. \tag{2.41}$$

Furthermore we have from Eqs. (2.10) and (2.11)

$$\varphi_\alpha \delta F^{\alpha\beta} = 0, \quad \varphi_\alpha \delta F^{*\alpha\beta} = 0,$$

which means that the electric and magnetic disturbances are transverse, orthogonal to each other, and of equal strength, i.e., the usual results of the linear theory. But here Eqs. (2.41), $\delta Q = \delta R = 0$, replace the null field conditions (fluid of photons) $Q = R = 0$.

It is even possible to show that (2.41) characterizes photons. Let us assume that (2.41) is satisfied. Then, by (2.15),

$$U^\alpha \pi_\alpha = V^\alpha \pi_\alpha = 0, \tag{2.42}$$

and, by (2.12),

$$\mathfrak{G}\pi_\alpha = (\pi^\beta \varphi_\beta)\varphi_\alpha,$$

which means

$$\mathfrak{G} = 0, \quad \pi^\alpha \varphi_\alpha = 0 \tag{2.43}$$

for a wave (i.e., a disturbance) to exist: $\delta F_{\alpha\beta} \neq 0$ (2.10). Then, taking account of (2.30),

$$U_\alpha V^\alpha = 0, \quad V_\alpha V^\alpha = U_\alpha U^\alpha. \tag{2.44}$$

π_α is certainly a spacelike vector; otherwise it would be colinear to φ_α, which is just what we wanted to avoid when writing (2.43). U_α cannot be timelike for $U^\alpha \varphi_\alpha = 0$; if U_α were spacelike, then according to (2.42)–(2.44) φ_α would be orthogonal to the three linearly independent spacelike vectors π_α, U_α, V_α and would be timelike (and not null). Therefore, U_α is a

null vector and we are led back to the conclusions of Sec. 2D: light rays are eigendirections of the electromagnetic tensor.

ii. Disturbances Associated with Charged Particles[12]

The first equation (2.17) and (2.18) yield

$$a = -\pi\{R\eta_Q + (\mu - Q)\eta_R\}, \quad \eta = \log|L_Q|,$$
$$b = \pi\{1 + (\mu + Q)\eta_Q + R\eta_R\}; \tag{2.45}$$

$\pi(x^\alpha)$ is a function of the coordinates whose growth in the course of time has been briefly discussed in Sec. 1A. Equations (2.16) give

$$\delta Q = \mathfrak{G}\{a(2Q + \zeta) + bR\}, \quad \delta R = \mathfrak{G}(aR + b\zeta), \tag{2.46}$$

or

$$\delta Q = \pi\mathfrak{G}(R - f^2\eta_R), \quad \delta R = \pi\mathfrak{G}(\zeta + f^2\eta_Q), \tag{2.47}$$

with the definition

$$f(Q, R; \zeta) = [-R^2 + \zeta(2Q + \zeta)]^{\frac{1}{2}} \; [= (\mu^2 - P^2)^{\frac{1}{2}}].$$

The expressions (2.45) and (2.47) are not valid for the Born–Infeld Lagrangian: the coefficients of a and b in (2.17) all vanish; a and b are arbitrary. It had to be so because of the double root in Sec. 2C.

Making use of the relation

$$-\eta^{\alpha\beta\gamma\delta}S_\alpha = \epsilon^{\beta\gamma\delta}_{\lambda\mu\nu}e^\lambda h^\mu u^\nu = (e^\beta h^\gamma - e^\gamma h^\beta)u^\delta + (e^\gamma h^\delta - e^\delta h^\gamma)u^\beta + (e^\delta h^\beta - e^\beta h^\delta)u^\gamma$$

and of Eqs. (2.32) and (2.37), we can write (2.13) as

$$\pi_\alpha = -\frac{1}{f(-\mathfrak{G})^{\frac{1}{2}}}(e_\alpha\delta Q + h_\alpha\delta R + c\mathfrak{G}S_\alpha),$$

while (2.10) takes the form

$$\delta F_{\alpha\beta} = \frac{1}{f^2}\{(e_\alpha S_\beta - e_\beta S_\alpha)\delta Q + (h_\alpha S_\beta - h_\beta S_\alpha)\delta R\}; \tag{2.48}$$

the disturbance of the dual tensor is easily obtained[14]:

$$\delta F^{*\gamma\delta} = \frac{1}{f^2}(-\eta^{\alpha\beta\gamma\delta}S_\alpha)(e_\beta\delta Q + h_\beta\delta R),$$

i.e.,

$$\delta F^{*\gamma\delta} = \mathfrak{G}\{a(u^\gamma h^\delta - u^\delta h^\gamma) - b(u^\gamma e^\delta - u^\delta e^\gamma)\}. \tag{2.49}$$

G. Exceptional Waves

We apply now the second principle. According to (2.7), (2.10), and (2.15), we have

$$\varphi_\alpha\varphi_\beta\delta(\tau^{\alpha\beta} + \mu g^{\alpha\beta}) = \mathfrak{G}\delta(\mu + Q) - 2U_\rho(\varphi^\alpha\pi^\rho - \varphi^\rho\pi^\alpha)\varphi_\alpha = \mathfrak{G}\delta\zeta.$$

Thus the wave will be exceptional [cf. (1.4)] if [12]

$$\delta\zeta = 0. \tag{2.50}$$

From (2.38) it is already obvious that the Born–Infeld waves are exceptional. When the absolute field is not constant, this gives

$$\zeta_Q\delta Q + \zeta_R\delta R = 0,$$

or, by (2.47),

$$f^2J = R\zeta_Q + \zeta\zeta_R, \quad J = \zeta_Q\eta_R - \zeta_R\eta_Q. \tag{2.51}$$

If we assume the Jacobian $J \neq 0$,

$$\zeta_Q = JR_\eta, \quad \zeta_R = -JQ_\eta,$$

and (2.51) become

$$f + f_\eta = 0,$$

whose integration is easy,[13]

$$fL_Q = \mathrm{Z}(\zeta), \tag{2.52}$$

Z being an arbitrary function of ζ. When $J = 0$ this equation still holds:

$$J = 0 \quad \text{and} \quad \varpi \neq 0 \to \zeta = \zeta(\eta), \qquad RL_{QQ} + \zeta L_{QR} = 0.$$

With the change of variables of Sec. 2C, this last equation is equivalent to

$$yy_q - \zeta(p)x_q = 0, \quad \tfrac{1}{2}y^2 - \zeta(p)x = \text{funct}\,(p),$$

which is analogous to (2.52) if ζ is not constant.

Now Eq. (2.19) can be written

$$(2x + \zeta)r + 2ys + \zeta t + p + \frac{f^2}{p}(rt - s^2) = 0, \tag{2.53}$$

which is an equation of Monge–Ampère since, by virtue of (2.52), ζ can be considered as a function of x, y, p:

$$pf(x, y; \zeta) = \mathrm{Z}(\zeta). \tag{2.54}$$

It is then easily seen that, along the characteristic curves of (2.53) (see, for instance, Ref. 21),

$$\begin{cases} f^2\dfrac{dp}{p} + \zeta\,dx - y\,dy = 0, \\ f^2\dfrac{dq}{p} - y\,dx + (2x + \zeta)\,dy = 0, \end{cases}$$

the quantities

$$\zeta, \quad q + y\frac{p}{\zeta}$$

are constant. As a result z is obtained by eliminating

ζ between the equations

$$z = F(\zeta)f + yG(\zeta) + H(\zeta) \quad (F = Z/\zeta),$$
$$\frac{\partial z}{\partial \zeta} = 0: \quad F'f^2 + (yG' + H')f + (x + \zeta)F = 0,$$

where the prime denotes differentiation with respect to ζ.

Therefore the Lagrangian is of the form[22]

$$L = F(\zeta)f(Q, R; \zeta) + RG(\zeta) + H(\zeta), \quad (2.55)$$

where the absolute field $\zeta(Q, R)$ is obtained by solving

$$F'f^2 + (RG' + H')f + (Q + \zeta)F = 0, \quad (2.56)$$

F, G, H being arbitrary functions of ζ.

Moreover,

$$L_Q = \frac{F\zeta}{f}, \quad L_R = -R\frac{F}{f} + G. \quad (2.57)$$

H. Exceptional Lagrangians[22]

To each value μ_1, μ_2 of μ solution of (2.19) corresponds a wave (2.18). If both families of waves can describe trajectories of stable particles,they must both be exceptional. Then the field equations will be completely exceptional (Sec. 1A) and we say that the Lagrangian is exceptional. It has already been made clear that the Born–Infeld Lagrangian is exceptional ($\zeta_1 = \zeta_2 = \text{const} = k$).

In the general case when neither ζ_1 nor ζ_2 is constant, (2.50) must be true for each value; the equations

$$\begin{cases} \delta_1\zeta_1 = 0, \\ \delta_2\zeta_2 = 0 \end{cases}$$

must be satisfied simultaneously, where the subscript 1 in δ_1 means that we must put $\zeta = \zeta_1$ in the expressions (2.47) (and also, of course, $\pi = \pi_1$, $\mathfrak{S} = \mathfrak{S}_1$).

Consequently, the Eqs. (2.55)–(2.57) must be valid no matter which value of ζ is used. In other words, the following relations hold:

$$F_1f_1 + RG_1 + H_1 = F_2f_2 + RG_2 + H_2, \quad (2.58)$$

$$\frac{F_1\zeta_1}{f_1} = \frac{F_2\zeta_2}{f_2}, \quad -R\frac{F_1}{f_1} + G_1 = -R\frac{F_2}{f_2} + G_2. \quad (2.59)$$

We eliminate Q and R between these equations to obtain the identity

$$\zeta_1\zeta_2(G_1 - G_2)^2 - (\zeta_1 - \zeta_2)(F_1^2\zeta_1 - F_2^2\zeta_2) + (H_1 - H_2)^2 \equiv 0. \quad (2.60)$$

(ζ_1, ζ_2 cannot be linked by any relation, for Q and R would then be functions of each other.)

This identity can also be written in the form

$$\psi_{12} + H_1^2 - F_1^2\zeta_1^2 + H_2^2 - F_2^2\zeta_2^2 \equiv 0 \quad (2.61)$$

with

$$\psi_{12} = \zeta_1\theta_2 + \zeta_2\theta_1 - 2\phi_1\phi_2 - 2H_1H_2,$$
$$\theta = \zeta(G^2 + F^2), \quad \phi = G\zeta. \quad (2.62)$$

According to the Remark of Sec. 2A, we shall make use of the possibility to change G and H into G + const and H + const, respectively, thus avoiding unnecessary constants of integration. We shall mention in parentheses the function to which the constant is added.

By virtue of (2.61), ψ_{12} is equal to the sum of a function of ζ_1 and of a function of ζ_2; hence,

$$\frac{\partial^2\psi_{12}}{\partial\zeta_1\partial\zeta_2} = 0,$$

i.e.,

$$\theta_1' + \theta_2' - 2\phi_1'\phi_2' - 2H_1'H_2' = 0. \quad (2.63)$$

It is only a matter of simple calculations to see that if both (ϕ')'s were constant, the same would be true of the absolute fields. Hence we assume $\phi_2'' \neq 0$. Differentiating (2.63) with respect to ζ_2, we get

$$(\theta_2''/\phi_2'') - 2\phi_1' - 2H_1'(H_2''/\phi_2'') = 0. \quad (2.64)$$

The assumption $H_1' = \text{const}$ would lead to conclusions analogous to the above ones. Thus we are left with

$$\frac{H_2''}{\phi_2''} = \text{const} = \gamma \neq 0,$$
$$H_2 = \gamma\phi_2. \quad (G_2, H_2)$$

Equations (2.64) and (2.63) give

$$\phi_1 + \gamma H_1 = 0, \quad (H_1)$$
$$\theta_1 = -a\zeta_1 + b, \quad \theta_2 = a\zeta_2 + c.$$

By (2.62) we have

$$\psi_{12} = c\zeta_1 + b\zeta_2,$$

and taking account of (2.61) we finally obtain the systems (I) and (II):

$$\text{(I)} \quad \begin{cases} G_1\zeta_1 + \gamma H_1 = 0, \\ \zeta_1(G_1^2 + F_1^2) = -a\zeta_1 + b, \\ H_1^2 - F_1^2\zeta_1^2 + c\zeta_1 + l = 0; \end{cases}$$

$$\text{(II)} \quad \begin{cases} \gamma G_2\zeta_2 - H_2 = 0, \\ \zeta_2(G_2^2 + F_2^2) = a\zeta_2 + c, \\ H_2^2 - F_2^2\zeta_2^2 + b\zeta_2 - l = 0. \end{cases}$$

Any one of these systems may be used with (2.55) and (2.56) to determine the Lagrangian. We note that, on passing from (I) to (II), that is,from the value ζ_1 to

the value ζ_2 of the absolute field, we make the following changes (C):

$$(C)\quad (\zeta_1 \to \zeta_2):\ \gamma \to -\frac{1}{\gamma},\ a \to -a,\ b \rightleftarrows c,\ l \to -l.$$

From (I) and (II) we get

$$H_1^2 = \frac{-1}{1+\gamma^2}\{a\zeta_1^2 - (b-c)\zeta_1 + l\},$$

$$H_2^2 = \frac{\gamma^2}{1+\gamma^2}\{a\zeta_2^2 - (b-c)\zeta_2 + l\}.$$

If $a \neq 0$,we can always assume $a > 0$. The discriminant of the polynomial inside brackets is necessarily positive, so we can put

$$k = \frac{1}{2}\frac{b-c}{a},\quad l = a(k^2 - h^2).$$

The new constants h and k are C invariant, i.e., invariant by the transformations (C). Then,

$$H_1^2 = [a/(1+\gamma^2)]\{h^2 - (\zeta_1 - k)^2\};$$

the Lagrangian (2.55) reduces to the Born–Infeld Lagrangian when $h = 0$ ($\zeta_1 = k$), provided we assume $k > 0$. We have for $h = 0$

$$F_1^2(k) = \frac{1}{2}\frac{b+c}{k},\quad G_1(k) = H_1(k) = 0,$$

which shows that the C invariant

$$\frac{1}{2}\frac{b+c}{k} \tag{2.65}$$

must be positive.

We introduce the dimensionless constant [23]

$$s = \frac{1}{2}\frac{b+c}{k}\frac{1}{a},$$

and,assuming that L has been divided by the square root of (2.65), the solution of (I) can be taken equal to

$$F_1^2 = \frac{1}{s(1+\gamma^2)}\left\{\frac{\gamma^2}{\zeta_1^2}(k^2 - h^2) + 2K\frac{\gamma}{\zeta_1} - 1\right\},$$

$$G_1 = -\frac{\gamma}{\zeta_1}H_1,\quad H_1^2 = \frac{1}{s(1+\gamma^2)}\{h^2 - (\zeta_1 - k)^2\},$$

where the C-invariant constant K is defined by

$$2K = \frac{1}{a\gamma}(b + c\gamma^2) = \frac{k}{\gamma}\{s + 1 + \gamma^2(s-1)\}.$$

The law of transformation (C) of γ suggests to one to put

$$\gamma = tg\tfrac{1}{2}\theta.$$

The absolute field depends on the new constants in the following way:

$$\zeta_1 = \zeta(Q, R; k, h, s, \theta),$$
$$\zeta_2 = \zeta(Q, R; k, h, -s, \theta + \pi);$$
$$h = 0:\quad \zeta_1 = \zeta_2 = k.$$

The constant h might well be simply connected with the Planck constant. When it vanishes, both values of ζ reduce to the constant absolute field k of Born–Infeld, thus allowing for a spherical approximation of the particle (Sec. 2E). The spin of the stable particle seems to be represented by s and θ is some constant angle.

To illustrate the role played by θ we consider the first-order approximation of the absolute field when the parameter h/k is small (which is probably the case). Let Q_0, R_0 be the values of the invariants built from a solution of the Born–Infeld field equations and let

$$f_0 = f(Q_0, R_0; k).$$

We introduce the three coplanar dimensionless vectors **A**, **B**, **C**:

$$|\mathbf{A}| = \frac{f_0^2}{k^2},\quad |\mathbf{B}| = s\,|1 - (R_0^2/k^2)|,$$

$$|\mathbf{C}| = 2s(|R_0|/k),$$

the last ones being orthogonal to each other ($\mathbf{B}\cdot\mathbf{C} = 0$) and making with **A** the constant angles

$$(\mathbf{A}, \mathbf{B}) = \left|\theta + \frac{\pi}{2}[1 - \operatorname{sgn}(k^2 - R_0^2)]\right|,$$

$$(\mathbf{A}, \mathbf{C}) = \left|\frac{\pi}{2} + \theta \operatorname{sgn} R_0\right|;$$

then, using (2.56), we find[23]

$$\zeta_1 \cong k \pm h\frac{|\mathbf{A}| - |\mathbf{B}+\mathbf{C}|}{|\mathbf{A}+\mathbf{B}+\mathbf{C}|},$$

$$\zeta_2 \cong k \pm h\frac{|\mathbf{A}| + |\mathbf{B}+\mathbf{C}|}{|\mathbf{A}+\mathbf{B}+\mathbf{C}|}.$$

Remarks: (1) Quite generally in one-dimensional propagation when a field has only two components and when the field equations are completely exceptional,[3] the characteristics (wave surfaces in the x, t plane) belong to two families of isocline curves. This means that the slope of curves of one family is constant along each curve of the other family; in other words, curves of the same family cannot intersect. This phenomenon had already been noted for the Born–Infeld one-dimensional field equations.[24]

(2) It has also been predicted[25] that shocks might arise when the Lagrangian of Heisenberg–Euler[26] is used. This is true since this Lagrangian is not exceptional. This Lagrangian, however, as it is given,appears as the first terms of an expansion with respect to a small parameter. As we ignore the exact form of this Lagrangian, we cannot even calculate the absolute field. Nevertheless it is probable that the exact form is not exceptional since it is used in connection with the production of unstable particles.

I. Disturbance of the Ray Velocity

It is given by the formula[14,1]

$$\delta u^\alpha = [1/f(-\mathfrak{G})^{\frac{1}{2}}](\delta^\alpha_\beta - u^\alpha u_\beta)\varphi_\gamma \delta(\tau^{\beta\gamma} + \mu g^{\beta\gamma})$$

and calculations show that it is equal to

$$\delta u^\alpha = (S^\alpha/f^2)\delta\zeta.$$

When the wave is exceptional,

$$\delta u^\alpha = 0,$$

from (2.48) and (2.49) we have

$$\delta e^\alpha = u_\rho \delta F^{\rho\alpha} = 0, \quad \delta h^\alpha = u_\rho \delta F^{*\rho\alpha} = \mathfrak{G}(ah^\alpha - be^\alpha),$$

$$\delta S_\alpha = \eta_{\alpha\lambda\mu\nu} e^\lambda \delta h^\mu u^\nu = a\mathfrak{G} S_\alpha.$$

Thus a stable particle appears in its proper frame as a disturbance of the sole magnetic field.

On the other hand, we obtain[10], by solving Eqs. (2.46),

$$a\mathfrak{G} = \frac{\delta f}{f} - \frac{\mu}{f^2}\delta\zeta, \quad b\mathfrak{G} = \frac{f}{\zeta}\delta\left(\frac{R}{f}\right) + \frac{R}{\zeta}\frac{\mu}{f^2}\delta\zeta.$$

Inserting these expressions into the above relations,we immediately see that in the exceptional case ($\delta\zeta = 0$), the tetrad of vectors

$$V^\alpha_{(\gamma)}: u^\alpha, \quad \frac{e^\alpha}{\zeta^{\frac{1}{2}}}, \quad \frac{Re^\alpha + \zeta h^\alpha}{f(\zeta)^{\frac{1}{2}}}, \quad \frac{S^\alpha}{f}; \tag{2.66}$$

$$V_{(\mu)\alpha}V^\alpha_{(\nu)} = \eta_{\mu\nu},$$

is not disturbed by the stable particle wave:

$$\delta V^\alpha_{(\mu)} = 0. \tag{2.67}$$

J. Introduction of a Nonsymmetric Tensor

In their famous paper,[11] Born and Infeld introduce the nonsymmetric tensor

$$a_{\alpha\beta} = g_{\alpha\beta} + F_{\alpha\beta}, \tag{2.68}$$

i.e., identify the symmetric and antisymmetric part with the gravitational and electromagnetic field tensors, respectively, and use the Lagrangian density,

$$\mathfrak{L} = A(-|a_{\alpha\beta}|)^{\frac{1}{2}} \pm B(|F_{\alpha\beta}|)^{\frac{1}{2}} + C(-|g_{\alpha\beta}|)^{\frac{1}{2}}. \tag{2.69}$$

If a positive constant k is introduced for the sake of dimension, we rewrite (2.68) as

$$a_{\alpha\beta} = k^{\frac{1}{2}}g_{\alpha\beta} + F_{\alpha\beta}, \tag{2.70}$$

while (2.69) gives the Born–Infeld Lagrangian,

$$L = Af(Q, R; k) + BR + C,$$
$$\mathfrak{L} = (-g)^{\frac{1}{2}}L, \quad g = |g_{\alpha\beta}|. \tag{2.71}$$

Now an obvious generalization of (2.70) and (2.71) consists of replacing the constant k by the function $\zeta(Q, R)$ of Q and R,

$$a_{\alpha\beta} = \zeta^{\frac{1}{2}}g_{\alpha\beta} + F_{\alpha\beta}, \tag{2.72}$$

$$L = F(\zeta)f(Q, R; \zeta) + G(\zeta)R + H(\zeta). \tag{2.73}$$

This is nothing but the Lagrangian (2.55),and,from the results of Sec. 2G, we know that ζ is indeed an absolute field if it is defined by

$$\partial L/\partial\zeta = 0,$$

a result which can also be established by direct calculation.[16,27]

Let us compare waves of nonlinear electrodynamics with waves of the Einstein–Schrödinger theory. To this aim we introduce the adjoint tensor of (2.72),[23]

$$a_{\alpha\beta}\dot{a}^{\alpha\gamma} = a_{\beta\alpha}\dot{a}^{\gamma\alpha} = \delta^\gamma_\beta,$$

i.e.,

$$\dot{a}^{\alpha\beta} = \frac{\zeta^{\frac{1}{2}}}{f^2}(\tau^{\alpha\beta} + \mu g^{\alpha\beta}) + \frac{1}{f^2}(\zeta F^{\alpha\beta} - RF^{*\alpha\beta}).$$

We see at once that the quantity

$$\partial_\alpha[(-a)^{\frac{1}{2}}\dot{a}^{[\alpha\beta]}], \quad a = |a_{\alpha\beta}| = gf^2,$$

vanishes [cf. (2.4)] for the Born–Infeld field[11,28] and that the waves,

$$\dot{a}^{(\alpha\beta)}\varphi_\alpha\varphi_\beta = 0, \tag{2.74}$$

are to be found in Einstein–Schrödinger theory too.[6]

However, in nonlinear electrodynamics we do not find the other family of waves of the Einstein–Schrödinger field. If we put

$$g'_{\alpha\beta} = a_{(\alpha\beta)} = \zeta^{\frac{1}{2}}g_{\alpha\beta},$$

and hence

$$\dot{g}'^{\alpha\beta} = g^{\alpha\beta}/\zeta^{\frac{1}{2}}, \quad g' = \zeta^2 g,$$

this family is given by[29]

$$\left(\dot{a}^{(\alpha\beta)} - 2\frac{g'}{a}\dot{g}'^{\alpha\beta}\right)\varphi_\alpha\varphi_\beta = 0,$$

i.e.,

$$\{\tau^{\alpha\beta} + (Q - \zeta)g^{\alpha\beta}\}\varphi_\alpha\varphi_\beta = 0;$$

it would correspond to the change of ζ into $-\zeta$, which would not be acceptable since, by (2.23) and (2.35), ζ must be positive.

Another difference must be stressed: in Einstein–Schrödinger theory, the nonsymmetric tensor is assumed to be continuous in the first order[6] and to admit discontinuities of the second order, which is natural because the field equations involve second-order derivatives of this tensor. On the contrary we have here

$$\delta a_{\alpha\beta} \neq 0.$$

Thus if we want to try to write the field equations of nonlinear electrodynamics in terms of the components of a nonsymmetric tensor and its first- and second-order derivatives, we had better choose this tensor in such a way that

$$\delta a_{\alpha\beta} = 0.$$

For this purpose we might think of the vectors (2.66). (A set of four vectors is used by Møller in his tetrad theory.[30])

For instance, the tensor

$$a_{\alpha\beta} = g_{\alpha\beta} + f_{\alpha\beta},$$
$$f_{\alpha\beta} = V_{(0)\alpha}V_{(1)\beta} - V_{(1)\alpha}V_{(0)\beta} + \eta_{\alpha\beta\gamma\delta}V^{\gamma}_{(2)}V^{\delta}_{(0)},$$

is analogous to (2.72) when the ratio P/ζ is small,

$$g_{\alpha\beta} + f_{\alpha\beta} \simeq g_{\alpha\beta} + (F_{\alpha\beta}/\zeta^{\frac{1}{2}}), \quad P/\zeta \ll 1.$$

Furthermore,

$$f_{\alpha\beta}V^{\beta}_{(2)} = 0 \rightarrow |f_{\alpha\beta}| = 0$$

and

$$a = |a_{\alpha\beta}| = 2\,|a_{(\alpha\beta)}| = 2g.$$

This is a singular case in Einstein–Schrödinger theory, but it also allows for the choice of another connection.[28,31]

ACKNOWLEDGMENTS

We are much grateful to Professor L. Infeld and Professor A. Lichnerowicz for their encouragement and their kind interest in our work. It is a pleasure to thank Professor H. Wergeland for his warm hospitality and Professor A. Mercier for his encouraging comments.

[1] G. Boillat, J. Math. Phys. **10**, 452 (1969).
[2] J. Hadamard, *Leçons sur la propagation des ondes* (Hermann & Cie, Paris, 1903).
[3] G. Boillat, *La propagation des ondes* (Gauthier-Villars, Paris, 1965). For the case of nonhyperbolic fields see G. Boillat, Compt. Rend. **270A** (1970) (to be published).
[4] P. D. Lax, Ann. Math. Studies **33**, 211 (1954); Commun. Pure Appl. Math. **10**, 537 (1957).
[5] A. Jeffrey and T. Taniuti, *Non-linear Wave Propagation* (Academic Press Inc., New York, 1964).
[6] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson & Cie., Paris, 1955).
[7] H. C. Kranzer, Math. Rev. **33**, 6148 (1967); J. Bazer and W. B. Ericson, Astrophys. J. **129**, 758 (1959).
[8] A. H. Taub, Arch. Ratl. Mech. **3**, 312 (1959); A. Lichnerowicz, Commun. Math. Phys. **1**, 328 (1966).
[9] Y. Choquet-Bruhat, Astron. Acta **6**, 354 (1960); P. Reichel, Institute of Mathematical Sciences, New York University, Report NYO-7697, 1958; K. O. Friedrichs and H. Kranzer, Institute of Mathematical Sciences, New York University, Report NYO-6486, 1958.
[10] G. Boillat, Compt. Rend. **263A**, 646 (1966).
[11] M. Born and L. Infeld, Proc. Roy. Soc. (London) **144A**, 425 (1934).
[12] G. Boillat, Compt. Rend. **262A**, 1285 (1966).
[13] G. Boillat, Ann. Inst. Henri Poincaré **5A**, 217 (1966).
[14] G. Boillat (to be published). See Ref. 1.
[15] J. Lameau, Cahiers Phys. **19**, 229 (1965). Cf. R. Pellicer and R. J. Torrence, J. Math. Phys. **10**, 1718 (1969).
[16] G. Boillat, Compt. Rend. **264A**, 209 (1967).
[17] L. Mariot, Compt. Rend. **238**, 2055 (1954); **239**, 1189 (1954).
[18] J. L. Synge, *Relativity: The Special Theory* (North-Holland Publ. Co., Amsterdam, 1958).
[19] G. Boillat, Compt. Rend. **262A**, 1364 (1966).
[20] G. Boillat, Phys. Letters **27A**, 192 (1968).
[21] G. Valiron, *Equations fonctionnelles: Applications* (Masson & Cie., Paris, 1950), pp. 557–60.
[22] G. Boillat, Compt. Rend. **262A**, 884 (1966).
[23] G. Boillat, Compt. Rend. **264A**, 1113 (1967).
[24] D. Blokhintsev and V. Orlov, Zh. Eksp. Teor. Fiz. **25**, 513 (1953); T. Taniuti, Progr. Theoret. Phys. (Kyoto) Suppl. **9**, 69 (1958).
[25] M. Lutzky and J. S. Toll, Phys. Rev. **113**, 1649 (1959).
[26] W. Heisenberg and H. Euler, Z. Physik **98**, 714 (1936).
[27] This condition also derives from the variational principle if ζ is varied arbitrarily. Moreover, if we assume that both values of ζ can be used in building the Lagrangian, we are led to the results of Sec. 2H, i.e., to the exceptional Lagrangian.
[28] M.-A. Tonnelat, *Les théories unitaires de l'électromagnétisme et de la gravitation* (Gauthier-Villars, Paris, 1965).
[29] F. Maurer-Tison, Ann. Sci. École. Norm. Super. **76**, 185 (1959).
[30] C. Møller, Mat.-Fys. Skr. Danske Vid. Selsk. **1**, No. 10 (1961).
[31] V. Hlavatý, J. Ratl. Mech. Anal., **2** 2 (1953).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Korteweg–deVries Equation and Generalizations. V. Uniqueness and Nonexistence of Polynomial Conservation Laws

MARTIN D. KRUSKAL, ROBERT M. MIURA,* AND CLIFFORD S. GARDNER†
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
AND
NORMAN J. ZABUSKY
Bell Telephone Laboratories, Inc., Whippany, New Jersey

(Received 5 September 1969)

The conservation laws derived in an earlier paper for the Korteweg–deVries equation are proved to be the only ones of polynomial form. An algebraic operator formalism is developed to obtain explicit formulas for them.

1. INTRODUCTION

Research in recent years exhibits an increase of interest in nonlinear dispersive wave phenomena. One of the simplest nonlinear dispersive wave equations is the Korteweg–deVries (KdV) equation

$$u_t + uu_x + u_{xxx} = 0, \tag{1}$$

where subscripts denote partial differentiations. This equation is conservative and dispersive, in sharp contrast to the much studied Burgers' equation which would obtain if u_{xxx} were replaced by $-\nu u_{xx}$, $\nu > 0$. Burgers' equation can describe a constant pressure, incompressible, viscous fluid. In most applications (see[1] Papers I and III for references to physical applications of the KdV equation), a dispersion parameter appears as a coefficient of the u_{xxx} term in the KdV equation. However, there is no loss of generality (for present purposes) in confining ourselves to (1), because of the scaling properties discussed in Appendix A.

In Paper II we proved that the KdV equation possesses an infinite sequence of *polynomial conservation laws* in the form

$$T_t + X_x = 0,$$

where T, the *conserved density*, and $-X$, the *flux* of T, are polynomials (not explicitly dependent on x or t) in u and its derivatives. In this paper, we give a detailed discussion of these conservation laws (also called *conservation equations*). We frequently use the abbreviations c.d., p.c.d., and c.l. for conserved density, polynomial conserved density, and conservation law.

Two c.l. for the KdV equation are obtained immediately,

$$u_t + (\tfrac{1}{2}u^2 + u_{xx})_x = 0,$$
$$(\tfrac{1}{2}u^2)_t + (\tfrac{1}{3}u^3 + uu_{xx} - \tfrac{1}{2}u_x^2)_x = 0.$$

The first is simply the equation itself rewritten, and the second is obtained after multiplying by u and rewriting.

Whitham[2] needed a third c.l. to study nonlinear dispersive waves by a method of averaging, hence he sought (and found) one, with a term u^3 in its p.c.d.. A fourth c.l., with a term u^4 in its p.c.d., was found by Kruskal and Zabusky.[3] In an effort to understand the results of certain numerical computations, they developed an asymptotic theory for the KdV equation when a small parameter δ^2 multiplies the u_{xxx} term. If δ^2 is set equal to zero, then for a large class of (smooth) initial data, the solution becomes discontinuous after a finite time due to the nonlinearity. However, if δ^2 is small but not zero, then, as the solution tends to become discontinuous, the u_{xxx} term becomes important in spite of its small coefficient and keeps the solution smooth. Moreover, small-wavelength finite-amplitude oscillations develop, initially occupying a small region at the near-discontinuity but then spreading indefinitely. (For Burgers' equation with $\nu \to 0$ no such oscillations develop, but rather a conventional hydrodynamic shock structure.) The asymptotic theory of the KdV equation developed by Kruskal and Zabusky in this region of oscillations is a nonlinear generalization of the WKB method. They obtained a system of equations for the evolution of the solution on the "macroscopic" scale (together with a detailed description of the oscillations on the "microscopic" scale). Solutions of the limit system ($\delta^2 \to 0$) exhibit discontinuities of an unusual type which may be considered as "reversible (nondissipative) shock waves." A count of the appropriate number of jump conditions across such shock waves led to the search for and discovery of the fourth conservation equation and to the investigation presented here.

Definitions and operators are introduced in Sec. 2 along with a number of commutation formulas. In Sec. 3 we prove that the KdV equation has no p.c.d. other than those established in Paper II.

Further, we outline the method of undetermined coefficients used to establish them there. However, this method does not yield an explicit formula for all p.c.d.; in Sec. 4 we find such formulas by other means.

We had originally intended to include a section proving uniqueness and nonexistence theorems for some generalizations of the KdV equation. However, the material has accumulated to such an extent as to justify our collecting it into an eventual separate Paper VII in this series.

2. DEFINITIONS AND OPERATORS

In this paper we deal extensively with polynomials and eventually formal infinite power series in u and its x derivatives. In fact, the words *polynomial* and *series* are always to be understood in this sense. Here we introduce some definitions to be used throughout.

We denote the t and x differentiation operators by

$$\partial_t \equiv \frac{\partial}{\partial t} \quad \text{and} \quad \mathfrak{D} \equiv \frac{\partial}{\partial x},$$

and the ith-order x derivative of u by u_i. In much of the following we treat the u_i as independent variables.

Any term in a polynomial (or series) has the form $cu_0^{a_0}u_1^{a_1}\cdots u_l^{a_l}$ where $a_i \geq 0$ are integers with $a_l \geq 1$ if $l \neq 0$; l is the *order* of the term. (Throughout this paper c and c_i, $i = 0, 1, \cdots$, denote arbitrary constants.) Of any two distinct terms we call that one *dominant* which has larger l, or the same l but larger a_l, or the same l and a_l but larger a_{l-1}, etc. Obviously dominance is an ordering relation (it is transitive and irreflexive) and is complete (of any two distinct terms, one dominates the other).

The total number of factors in a term is its *degree*

$$m \equiv \sum_{i=0}^{l} a_i. \tag{2}$$

The total number of differentiations will be called the *derivative index* (or simply *index*)

$$n \equiv \sum_{i=1}^{l} i a_i. \tag{3}$$

The *rank* of a term (introduced in Paper II) is defined by

$$r \equiv m + \tfrac{1}{2}n = \sum_{i=0}^{l} (1 + \tfrac{1}{2}i)a_i; \tag{4}$$

a term scales as the $(-2r)$th power of c under the scale transformation (see Appendix A). For example, the last two terms in the KdV equation are of equal rank $\frac{5}{2}$. From the above definitions it is clear that acting on a term with the operator $\mathfrak{D}$ leaves the degree of that term unchanged but raises its index by 1 and its rank by $\frac{1}{2}$, whereas acting on a term with ∂_t raises the rank by $\frac{3}{2}$ but has a nonuniform effect on the degree and on the index. A *polynomial of degree m* has m as the l.u.b. of the degrees of all its terms, and similarly for the order, index, and rank.

Associated with the three "labelings" degree, index, and rank are the linear differential operators

$$\mathcal{M} \equiv \sum_i u_i \partial_i, \tag{5}$$

$$\mathcal{N} \equiv \sum_i i u_i \partial_i, \tag{6}$$

$$\mathcal{R} \equiv \mathcal{M} + \tfrac{1}{2}\mathcal{N} = \sum_i (1 + \tfrac{1}{2}i) u_i \partial_i, \tag{7}$$

where $\sum_i$ denotes $\sum_{i=0}^{\infty}$ and $\partial_i \equiv \partial/\partial u_i$. (For i negative, we shall interpret ∂_i as zero.) The effect of applying $\mathcal{M}$, $\mathcal{N}$, or $\mathcal{R}$ to any term is to multiply that term by its degree, index, or rank. Thus any term, indeed any polynomial of uniform degree (i.e., with all terms of the same degree), is an eigenpolynomial of $\mathcal{M}$ with eigenvalue m, and similarly for $\mathcal{N}$ and $\mathcal{R}$. For each of these labelings we define *slice operators*, i.e., projection operators that select terms of a certain degree, index, or rank. Thus, for example, the m-degree slice operator $\mathcal{M}_m$, acting on a polynomial, leaves unaffected all terms of degree m and annihilates all terms of degree different from m. (Because of the eigenproperty, such an m-degree slice operator could be defined formally as $\mathcal{M}_m \equiv \delta_{\mathcal{M},m}$, where the right side is the Kronecker delta function, here equal to unity if the eigenvalue of $\mathcal{M}$ operating on a term is m, and otherwise zero.)

Besides $\mathcal{R}$, we have occasion to consider other operators linear in $\mathcal{M}$ and $\mathcal{N}$. The operator $c_1\mathcal{M} + c_2\mathcal{N} + c_3$ has the effect of multiplying any term by the number $c_1m + c_2n + c_3$. We define $(c_1\mathcal{M} + c_2\mathcal{N} + c_3)^{-1}$ to be the operator whose effect is to *divide* any term by $c_1m + c_2n + c_3$, or annihilate any term for which this number vanishes. If there are no annihilated terms we have a true inverse, otherwise a pseudo-inverse; in any case it commutes with the original operator.

In Sec. 4 we use series (of nonnegative powers) extensively. Many of the properties (e.g., vanishing and equality) and operations (e.g., addition, multiplication, differentiation, integration, exponentiation, $\mathcal{M}$, $\mathcal{N}$, $\mathcal{R}$, and the slicings) generalize naturally and obviously from polynomials to series. A series is considered to have a property if arbitrarily high polynomial truncates of it do; in particular, it vanishes only if all its coefficients do. The result of an operation on a series consists of all terms common to the results of the same operation on arbitrarily high truncates;

for all the parenthetically mentioned operations, except multiplication and exponentiation, this amounts to operating term by term. The definitions of order, degree, index, and rank for polynomials have been so expressed as to apply also to series, now allowing infinity as a value. (A series of finite *rank* is necessarily a polynomial.)

Expressed as a differential operator in the space of functions of the u_i, the x differentiation operator is

$$\mathfrak{D} = \sum_i u_{i+1}\partial_i. \tag{8}$$

Lemma 1: A polynomial P of degree m and order l is either (a) a constant ($m = 0$), whereupon $\mathfrak{D}P = 0$, or (b) not a constant ($m > 0$), whereupon $\mathfrak{D}P(\neq 0)$ is of order $l + 1$ and is linear in u_{l+1}.

Proof: Case (a) is obvious. Case (b) follows immediately from (8), the dominant term in $\mathfrak{D}P$ coming from $i = l$. ∎

Using the KdV equation (1), the t differentiation operator can be written

$$\begin{aligned}\partial_t &= \sum_i (\partial_t u_i)\partial_i = -\sum_i [\mathfrak{D}^i(u_0u_1 + u_3)]\partial_i \\ &= -(\mathfrak{B} + \mathfrak{D}_3),\end{aligned} \tag{9}$$

where

$$\mathfrak{B} \equiv \tfrac{1}{2}\sum_i \sum_{j=0}^{i+1} \binom{i+1}{j} u_j u_{i-j+1}\partial_i, \tag{10}$$

$\binom{i+1}{j}$ being the binomial coefficient, and $\mathfrak{D}_3$ is a special case of

$$\mathfrak{D}_j \equiv \sum_i u_{i+j}\partial_i. \tag{11}$$

Other special cases of (11) are

$$\mathfrak{D}_0 = \mathcal{M}, \quad \mathfrak{D}_1 = \mathfrak{D}.$$

We present here a number of results not needed until Sec. 4. First, a number of commutation formulas all obvious or straightforwardly derived:

$$\partial_i\mathfrak{D}_j = \mathfrak{D}_j\partial_i + \partial_{i-j}, \tag{12}$$

$$\mathfrak{D}_i\mathfrak{D}_j = \mathfrak{D}_j\mathfrak{D}_i, \tag{13}$$

$$\partial_i\mathcal{N} = (\mathcal{N} + i)\partial_i, \tag{14}$$

$$\mathfrak{D}_i\mathcal{N} = (\mathcal{N} - i)\mathfrak{D}_i, \tag{15}$$

$$\partial_0\mathfrak{B} = \mathfrak{B}\partial_0 + \mathfrak{D}, \tag{16}$$

$$\partial_1\mathfrak{B} = \mathfrak{B}\partial_1 + \mathcal{M} + \mathcal{N}, \tag{17}$$

$$\mathfrak{D}\mathfrak{B} = \mathfrak{B}\mathfrak{D}, \tag{18}$$

$$\mathfrak{D}\partial_t = \partial_t\mathfrak{D}. \tag{19}$$

The commutativity of x and t differentiations expressed by (19) has already been used in (9).

We next introduce

$$\mathcal{Y} \equiv \sum_i (-\mathfrak{D})^i\partial_i, \tag{20}$$

the Euler operator of the calculus of variations. [For a function $F(u_0, \cdots, u_l)$, $\mathcal{Y}F$ is the functional derivative of $\int F\,dx$, aside from boundary terms.] Since we are dealing here with formal algebraic expressions, we give a purely algebraic proof of the familiar

Lemma 2: If P is a series [polynomial] with no constant term, then P is the derivative of some series [polynomial] if and only if $\mathcal{Y}P = 0$.

Proof: If $P = \mathfrak{D}Q$, then (12) for $j = 1$ gives

$$\mathcal{Y}\mathfrak{D}Q = \sum_i (-\mathfrak{D})^i\partial_i\mathfrak{D}Q = \sum_i (-\mathfrak{D})^i(\mathfrak{D}\partial_i + \partial_{i-1})Q$$

which is seen to vanish by shifting indices (since $\partial_{-1} = 0$).

Conversely, if $\mathcal{Y}P = 0$, we multiply by u_0 to restore the degree and integrate by parts repeatedly, obtaining

$$\begin{aligned}0 = u_0\mathcal{Y}P &= u_0\sum_i (-\mathfrak{D})^i\partial_i P \\ &= u_0\partial_0 P - \mathfrak{D}\left(u_0\sum_{i=1}^{\infty}(-\mathfrak{D})^{i-1}\partial_i P\right) \\ &\quad + u_1\sum_{i=1}^{\infty}(-\mathfrak{D})^{i-1}\partial_i P \\ &= \cdots = \mathcal{M}P - \mathfrak{D}\sum_j u_j\sum_{i=j+1}^{\infty}(-\mathfrak{D})^{i-j-1}\partial_i P.\end{aligned} \tag{21}$$

We apply $\mathcal{M}^{-1}$ and note that $\mathcal{M}^{-1}\mathcal{M}P = P$ because only constant terms are annihilated by $\mathcal{M}$, and P has none. Furthermore $\mathcal{M}^{-1}$ commutes with $\mathfrak{D}$: for, by (13), $\mathcal{M}(= \mathfrak{D}_0)$ does, so that

$$\mathcal{M}^{-1}(\mathcal{M}\mathfrak{D})\mathcal{M}^{-1} = \mathcal{M}^{-1}(\mathfrak{D}\mathcal{M})\mathcal{M}^{-1},$$

while $\mathcal{M}^{-1}\mathcal{M} = \mathcal{M}\mathcal{M}^{-1}$ is the identity operator except for constant terms, which, by (8), do not occur in derivatives and are annihilated by $\mathfrak{D}$. Thus we have

$$P = \mathfrak{D}\left(\mathcal{M}^{-1}\sum_j u_j\sum_{i=j+1}^{\infty}(-\mathfrak{D})^{i-j-1}\partial_i P\right). \tag{22}$$

Note that the infinite summations are (formally) well defined since we could have worked with derivative index slices, in each instance of which the summations would obviously be finite. For the same reason, if P is a polynomial, so is the expression in brackets. ∎

We will want operators which when applied to powers of $\mathfrak{D}$ produce something simpler. We seek

such operators in the form $\mathcal{O} \equiv \sum_i c_i(-\mathcal{D})^{i-r}\partial_i$ (which reduces to $\mathcal{Y}$ for $c_i = 1$). This gives

$$\mathcal{O}\mathcal{D} = \sum_i c_i(-\mathcal{D})^{i-r}(\mathcal{D}\partial_i + \partial_{i-1})$$
$$= \sum_i (c_{i+1} - c_i)(-\mathcal{D})^{i+1-r}\partial_i,$$

which suggests choosing $c_i = \binom{i}{r}$ and defining the operators

$$\mathcal{Y}_r \equiv \sum_{i=r}^{\infty} \binom{i}{r}(-\mathcal{D})^{i-r}\partial_i. \tag{23}$$

Then $\mathcal{Y}_r\mathcal{D} = \mathcal{Y}_{r-1}$ so that

$$\mathcal{Y}_r\mathcal{D}^r = \mathcal{Y}_0 \equiv \mathcal{Y}. \tag{24}$$

Incidentally, a basic property of the operators $\mathcal{Y}_r$ is stated in the following generalization of Lemma 2. However, we make no use of it.

Lemma 3: If P is a series [polynomial] with no constant term, then P is the $(r+1)$-order derivative of some series [polynomial] if and only if $\mathcal{Y}_s P = 0$ for $s = 0, 1, \cdots, r$.

Proof: Straightforward, inductive, and omitted. ▮

3. POLYNOMIAL CONSERVED DENSITIES

A. Preliminaries

Our general objective is to find conserved densities. The existence of an infinite sequence of p.c.d., one for each integral rank (here always meaning uniform rank), was proved in Paper II.

Since differentiations with respect to x and t commute, any x-derivative term is a *trivial* c.d. Given any c.d. we can immediately obtain any number of others simply by adding x-derivative terms. It is therefore natural to call two polynomials P_1 and P_2 *equivalent* if their difference is an x derivative; we denote this by

$$P_1 \doteqdot P_2. \tag{25}$$

(The terminology is justified by the obvious properties of reflexivity, symmetry, and transitivity.)

Among the many equivalents of a polynomial we seek a canonical representative. Let us integrate by parts to reduce order whenever possible. For instance $u_0u_2u_4 \doteqdot -u_1u_2u_3 - u_0u_3^2 \doteqdot \frac{1}{2}u_2^3 - u_0u_3^2$. A polynomial is *irreducible* if no term has its highest-derivative factor u_i occurring linearly, i.e., if each of its terms has order $l = 0$ or has $l \geq 1$ and $a_l \geq 2$, since then we cannot integrate by parts appropriately.

Our first two theorems state that the irreducible polynomials can serve as canonical representatives.

Theorem 1 (Existence): Every polynomial has an irreducible equivalent.

Proof: Appropriate integrations by parts can be applied successively only a finite number of times, since the order always decreases. ▮

Theorem 2 (Uniqueness): Any two equivalent irreducible polynomials are equal.

Proof: Their difference, an irreducible derivative, vanishes by Lemma 1. ▮

We also have occasion to speak of *equivalent operators*, meaning operators whose difference can be written in the form $\mathcal{D}\mathcal{A}$ for some operator $\mathcal{A}$. We keep the same notation, so that, e.g., $\mathcal{Y}_0 \doteqdot \partial_0$ by (20).

B. Uniqueness of Conserved Densities of Rank r

In this section we prove for each r the uniqueness theorem that, up to multiplication by a constant, there is at most one irreducible p.c.d. T_r of uniform rank r. We give two proofs, one of which shows further that the "extreme" term u_0^r necessarily appears (i.e., with nonzero coefficient), and the other that so does the opposite extreme term u_{r-2}^2. From either of these, we deduce the nonexistence theorem that for half-integral r there are no nontrivial p.c.d. at all.

We rely upon the structural

Lemma 4: The dominant term of the irreducible equivalent of the result of applying $\mathcal{B}$ to an irreducible term $u_0^{a_0} \cdots u_l^{a_l}$ may be obtained by multiplying that term by $[-1 + a_1 + 3a_2 + 4a_3 + \cdots + (l+1)a_l]u_1$, unless $l = 0$, in which case $\mathcal{B}u_0^{a_0} \doteqdot 0$.

Proof: The result is obtained by straightforward calculation followed by the appropriate integration by parts to reduce order. It is unnecessary to give the details; in any case a slightly more general lemma will be presented in Paper VII. ▮

A basic result of this section is

Theorem 3: If T_r is any nonzero irreducible p.c.d. of uniform rank r and $T_r^{(m)}$ its highest degree slice, then the dominant term Q of $T_r^{(m)}$ is cu_0^r $(c \neq 0)$.

Proof: The highest degree slice of $(\mathcal{B} + \mathcal{D}_3)T_r \doteqdot 0$ is $\mathcal{B}T_r^{(m)} \doteqdot 0$. By Lemma 4, the dominant term of the irreducible equivalent of $\mathcal{B}T_r^{(m)}$ would be just the dominant term of the irreducible equivalent of $\mathcal{B}Q$, if the latter did not vanish. Therefore $\mathcal{B}Q \doteqdot 0$. Then by Lemma 4 again, $Q = cu_0^r$. ▮

As immediate corollaries we have two of our main theorems.

Theorem 4 (Uniqueness): The KdV equation has only one linearly independent irreducible p.c.d. of rank r.

Proof: Given any two, we form a linear combination of them such that the u_0^r terms cancel. This is then a p.c.d. with no u_0^r term and hence vanishes, so the two are linearly dependent. ▮

Theorem 5 (Nonexistence): The only p.c.d. of half-integral rank (e.g., containing terms such as $u_0^{a_0}u_1^{a_1}$ with a_1 odd) for the KdV equation are trivial.

Proof: There is no polynomial term u_0^r for r a half-integer. ▮

An alternative proof of these results relies on the structural

Lemma 5: The dominant term of the irreducible equivalent of the result of applying $\mathfrak{D}_3$ to an irreducible term Q may be obtained by multiplying Q by a nonzero constant (depending on Q) and replacing each of the three highest factors u_l, u_l, and u_k $(k \leq l)$ by u_{l+1}, u_{l+1}, and u_{k+1}, unless Q has degree less than three, in which case $\mathfrak{D}_3 Q \doteq 0$.

Proof: Same comments apply as for Lemma 4. ▮

Analogous to Theorem 3 is

Theorem 6: The dominant term of any nonzero irreducible p.c.d. of uniform rank r is cu_{r-2}^2 if $r \geq 2$, cu_0 if $r = 1$, and c if $r = 0$ $(c \neq 0)$.

Proof: Analogous to that of Theorem 3, but, if anything, simpler because we need not take any kind of slice before selecting the dominant term. ▮

It is perhaps worth remarking, as a corollary of the preceding results, that the only irreducible p.c.d. of the "truncated" KdV equation $u_t + uu_x = 0$ are polynomials in u_0 alone. Similarly, the only irreducible p.c.d. of the alternatively truncated (i.e., linearized) KdV equation $u_t + u_{xxx} = 0$ are linear combinations of 1, u_0, and u_i^2, $i = 0, 1, 2, \cdots$.

C. Method of Undetermined Coefficients

In Paper II we exhibited ten irreducible p.c.d. for the KdV equation and seven of the associated fluxes. (An eleventh is given[5] in Appendix B.) Here we outline the method of undetermined coefficients (MUC) used to obtain those p.c.d. and fluxes.

We write T_r as a general linear combination of all possible irreducible terms of rank r. Then we seek to determine the coefficients such that $\partial_t T_r$ will be equivalent to zero, i.e., equal to an x derivative. By Theorem 3 T_r contains the term u_0^r, the coefficient of which we choose to be $1/r$ for computational convenience. (Later we will choose it to be $1/r!$ instead. Alternatively, we could have fixed the coefficient of u_{r-2}^2, by Theorem 6.)

We next work out $\partial_t T_r$, use the KdV equation to eliminate explicit t derivatives, and convert the result to its irreducible equivalent by the integration-by-parts algorithm described in Sec. 3A. We demand that the final expression vanish identically, and obtain thereby conditions on the original coefficients.

To illustrate MUC consider the case $r = 6$. All possible irreducible terms are represented in

$$T_6 = \tfrac{1}{6}u_0^6 + b_1u_0^3u_1^2 + b_2u_0^2u_2^2 + b_3u_1^4 + b_4u_0u_3^2 + b_5u_2^3 + b_6u_4^2.$$

The first term, for instance, gives

$$\partial_t(\tfrac{1}{6}u_0^6) = u_0^5(-u_0u_1 - u_3) = -\mathfrak{D}(\tfrac{1}{7}u_0^7 + u_0^5u_2 - \tfrac{5}{2}u_0^4u_1^2) - 10u_0^3u_1^3,$$

which contributes only $-10u_0^3u_1^3$ to the final irreducible expression which is to vanish. We eventually obtain an (overdetermined) inhomogeneous linear algebraic system of seven equations for the six b_i which turns out to have a unique solution. The result is

$$T_6 = \tfrac{1}{6}u_0^6 - 10u_0^3u_1^2 + 18u_0^2u_2^2 - 5u_1^4 - \tfrac{108}{7}u_0u_3^2 + \tfrac{120}{7}u_2^3 + \tfrac{36}{7}u_4^2.$$

It is evident that in obtaining the p.c.d., we simultaneously obtain the corresponding fluxes, so long as we keep all the $\mathfrak{D}$ terms.

An interesting aspect of MUC for obtaining irreducible p.c.d. is that for $r \geq 6$, the number of equations exceeds the number of unknowns b_i. (Nevertheless, in addition to the uniqueness assured by Theorem 4, the theory in Paper II guarantees the existence of a solution.) Table I lists,[5] for each rank r up to 20, the number of undetermined coefficients (which is one less than the number of irreducible terms of that rank) and the number of possible equations (i.e., the number of irreducible terms of rank $r + \frac{3}{2}$). Up to $r = 11$, every possible $r + \frac{3}{2}$ term actually occurs and so yields an equation, and presumably this is true in general.

TABLE I. Number of irreducible terms less one, and number of possible equations, up to rank 20.

Rank r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. of b_i	0	0	1	2	3	6	9	13	21	31	44	66	94	133	191	268	372	520	717	982
No. of $r + \frac{3}{2}$ terms	0	0	1	2	3	7	11	17	28	43	63	96	139	199	287	406	566	792	1092	1496

4. OPERATOR FORMULA FOR CONSERVED DENSITIES

In Paper II we proved that there exists an infinite sequence of p.c.d. for the KdV equation and obtained recursion formulas for generating them. MUC, as described in Sec. 3C, is straightforward for deriving individual p.c.d. but is not well suited for obtaining an explicit general formula. In this section our main objective is to derive an operator-product formula for T_r.

In Paper VII we will show that the generalized equation

$$u_{0_t} + u_0^p u_1 + u_q = 0, \quad p \geq 1, \quad q \geq 2, \tag{26}$$

for odd q has exactly three independent p.c.d. unless $p = 1$ or 2 and $q = 3$ (in which case it has infinitely many as proved in Paper II), and for even q has exactly one. The formalism to be developed in this section (for the KdV equation) depends essentially on Galilean invariance, requiring $p = 1$, but would go through for any q. However, the results would be meaningless (except for $q = 3$) because the supposed p.c.d. do not exist.

A. Exponential Series

A comparison of the p.c.d. in Paper II shows that consecutive T_r are related by[6]

$$\partial_0 T_r = (r-1)T_{r-1}, \quad r \geq 2. \tag{27}$$

As it stands, this relationship is of limited use even though it provides a complete prescription for going from T_r to T_{r-1}. We cannot go up in rank from T_{r-1} to T_r because nonzero "constants" of integration (functions of $u_1, u_2, \cdots$) appear in T_r.

Our remaining results depend heavily on

Lemma 6: If T and X are the c.d. and flux of a c.l. for the KdV equation, then so are

$$\tilde{T} \equiv \partial_0 T, \quad \tilde{X} \equiv \partial_0 X - T. \tag{28}$$

Proof: Making the Galilean transformation $x' = x + ct$, $t' = t$, and $u_0' = u_0 + c$ (see Appendix A), $\partial_t T + \mathfrak{D}X = 0$ can be written

$$\partial_{t'} T(u_0 + c, u_1, \cdots) + \mathfrak{D}' X(u_0 + c, u_1, \cdots) = 0,$$

where $\partial_{t'} = \partial_t - c\mathfrak{D}$, $\mathfrak{D}' = \mathfrak{D}$. Expanding with respect to c, the coefficient of the linear term is

$$\partial_t(\partial_0 T) + \mathfrak{D}(\partial_0 X - T) = 0. \quad \blacksquare \tag{29}$$

Alternatively, following Wiley,[6] we can derive (29) by operating on $\partial_t T + \mathfrak{D}X = 0$ with ∂_0 and using the commutation formulas

$$\partial_0 \partial_t = \partial_t \partial_0 - \mathfrak{D}, \quad \partial_0 \mathfrak{D} = \mathfrak{D}\partial_0,$$

obtained from (12) and (16). (*A posteriori*, one might have been led to this by applying $\mathcal{Y}$ in order to annihilate $\mathfrak{D}X$ and then have observed that $\mathcal{Y} \doteqdot \partial_0$.)

We can eliminate the factor $r - 1$ in (27) by rescaling each T_r so that u_0^r has coefficient $1/r!$ rather than $1/r$. Then instead of (27) we have

$$\partial_0 T_r = T_{r-1}, \tag{30}$$

valid for all $r \geq 0$ if we define $T_0 \equiv 1$ and $T_{-1} \equiv 0$.

We define the formal infinite sum

$$T \equiv \sum_r T_r, \tag{31}$$

a kind of generating function from which each T_r can be recovered by taking the r-rank slice. Then the formal sum of the infinite sequence of relations (30) can be written as the single relation

$$\partial_0 T = T, \tag{32}$$

which formally integrates to

$$T = e^{u_0} A(u_1, u_2, \cdots). \tag{33}$$

Series of this exponential form, namely e^{u_0} times a series in $u_1, u_2, \cdots$, will be called *e-series*. For the most part in the following we can confine ourselves to the class of *e-series* because it is obviously closed under many of our operators, e.g., $\mathfrak{D}$, ∂_i, and $\mathcal{N}$. However,

$$\begin{aligned}\partial_t[e^{u_0}B(u_1, u_2, \cdots)] &= -e^{u_0}[(u_0u_1 + u_3)B + (\mathfrak{B} + \mathfrak{D}_3)B] \\ &= -e^{u_0}u_0(u_1B + \mathfrak{D}B) - e^{u_0}C(u_1, u_2, \cdots)\end{aligned}$$

is not purely exponential. The offending terms have been collected and written first; note that only the $j = 0$ and $j = i + 1$ terms from the inner summation in $\mathfrak{B}$ have factors u_0. Nevertheless, since we only use the results of operating with ∂_t up to equivalence (i.e.,

modulo x derivatives), we can write

$$\begin{aligned}\partial_t(e^{u_0}B) &= -\{\mathfrak{D}[e^{u_0}(u_0-1)]\}B - e^{u_0}u_0\mathfrak{D}B - e^{u_0}C \\ &\doteqdot e^{u_0}(u_0-1)\mathfrak{D}B - e^{u_0}u_0\mathfrak{D}B - e^{u_0}C \\ &= -e^{u_0}(\mathfrak{D}B + C),\end{aligned}$$

which is an e-series.

B. Canonical e-Series

Factors u_1 can always be successively eliminated from an e-series by integrating by parts with e^{u_0}, e.g.,

$$\begin{aligned}e^{u_0}u_1^2u_3 &\doteqdot -e^{u_0}(u_2u_3 + u_1u_4) \\ &\doteqdot e^{u_0}(-u_2u_3 + u_5).\end{aligned}$$

This process terminates because the power of u_1 keeps decreasing. An e-series E will be called *canonical* if it is independent of u_1,

$$\partial_1 E = 0.$$

(We are abandoning irreducibility because it has no comparably algebraic expression.) The justification of the terminology is given by the above existence property and

Theorem 7 (Uniqueness): Any two equivalent canonical e-series are equal.

Proof: Their difference E is equivalent to zero, so $E = \mathfrak{D}F$. Thus

$$0 = \partial_0 E - E = \mathfrak{D}(\partial_0 F - F),$$

since ∂_0 commutes with $\mathfrak{D}$. From Lemma 1 $\partial_0 F - F = c$, so that (absorbing c into F) $F = e^{u_0}B(u_1, u_2, \cdots)$. But E is canonical, so that

$$\begin{aligned}0 = \partial_1 E &= (\mathfrak{D}\partial_1 + \partial_0)F = \mathfrak{D}\partial_1 F + F \\ &= e^{u_0}(u_1\partial_1 B + \mathfrak{D}\partial_1 B + B)\end{aligned}$$

using (12). Dropping the exponential and taking an arbitrary n-index slice gives

$$0 = u_1\partial_1 B^{(n)} + \mathfrak{D}\partial_1 B^{(n)} + B^{(n)}$$

where $B^{(n)} \equiv \mathcal{N}_n B$ is a polynomial. If $B^{(n)}$ did not vanish, its highest degree slice would satisfy the same equation without the middle term, which is impossible because the operator $u_1\partial_1 + 1$ has the effect of multiplying each term by a positive constant. Hence, all $B^{(n)} = 0$, $F = 0$, and $E = 0$. ∎

C. Operator Formula for Conserved Densities

From now on we use T to denote the canonical equivalent of (31) rather than (31) itself. Next we derive an equality (not just equivalence) for T which is the basis of the further analysis. To T corresponds a flux X such that

$$(\mathfrak{B} + \mathfrak{D}_3)T = \mathfrak{D}X. \tag{34}$$

We make use of the e-series property ($\partial_0 T = T$) by applying ∂_0 and subtracting (34) itself. In view of the commutation formulas (12) and (16) and Lemma 1, we obtain

$$T = \partial_0 X - X, \tag{35}$$

after absorbing into X an otherwise free constant as in the proof of Theorem 7; (35) characterizes the u_0 dependence of X as

$$X = u_0 T + Y \tag{36}$$

where Y is an e-series.

We make use of canonicity ($\partial_1 T = 0$) by applying ∂_1 to (34). In view of the commutation relations (12) and (17) we obtain

$$(\mathcal{M} + \mathcal{N})T = \mathfrak{D}\partial_1 X + \partial_0 X.$$

Eliminating $\partial_0 X$ by (35), this can be written

$$(\mathcal{K} - 1)T = \mathfrak{D}\partial_1 X + X, \tag{37}$$

where the effect of the operator

$$\mathcal{K} \equiv \mathcal{M} + \mathcal{N}$$

is to multiply each term by the number of its factors and differentiations combined.

Formally, solving (37) for the X term alone, we can substitute the resulting formula into itself recursively. Since $\partial_1(\mathcal{K} - 1)T = 0$, this gives $X = (\mathcal{K} - 1)T + 0 + 0 + \cdots$, and indeed

$$X = (\mathcal{K} - 1)T \tag{38}$$

is evidently a solution and is consistent with (35), and (36), by (12) and (14). To prove that it is *the* solution we need a uniqueness theorem. But we have already proved such a theorem for the e-series F in the proof of Theorem 7, and it applies for X of the form (36) since Y is an e-series.

Putting (38) back into (34) gives

$$(\mathfrak{B} + \mathfrak{D}_3)T = \mathfrak{D}(\mathcal{K} - 1)T \tag{39}$$

as an equation for T alone.

This can be rank-sliced to give the recursion formula

$$(\mathfrak{B} + \mathfrak{D}_3)T_r = \mathfrak{D}(\mathcal{K} - 1)T_{r+1}. \tag{40}$$

In using this to generate the T_r, it is most convenient to start iterating from $T_1 = u_0$.

The difficulty with solving this in general for T_{r+1} in terms of T_r is to invert $\mathfrak{D}$. This operator has no inverse because only polynomials of very special form are derivatives, and unessentially because it annihilates

constants. (The difficulty does not occur in carrying out the recursion explicitly step by step, since at each stage the left side is guaranteed to be a derivative by the existence results of Paper II.) However, $\mathfrak{D}$ has (many) left inverses (excluding constants), one such being

$$\mathfrak{D}^{-1} \equiv \mathcal{M}^{-1} \sum_j u_j \sum_{i=j+1}^{\infty} (-\mathfrak{D})^{i-j-1} \partial_i$$

obtained from (22). (Other expressions for $\mathfrak{D}^{-1}$ and a detailed discussion of the algebra of these and other operators will be given in a forthcoming paper.)

Applying $\mathfrak{D}^{-1}$ and then $(\mathcal{K} - 1)^{-1}$, we iterate to obtain

$$T_{r+1} = [(\mathcal{K} - 1)^{-1}\mathfrak{D}^{-1}(\mathcal{B} + \mathfrak{D}_3)]^r u_0. \quad (41)$$

This is an explicit representation of the p.c.d., as desired. Note that $(\mathcal{K} - 1)^{-1}$ annihilates only terms of the form cu_0, which are of rank 1 and hence never occur.

Alternatively, in order to avoid $\mathfrak{D}^{-1}$ with its arbitrariness and (moderate) complexity, we apply $\mathfrak{D}^{r-1}$ to (39) and commute operators to obtain for $\mathfrak{D}^r T_{r+1}$ the recursion formula

$$(\mathcal{K} - r - 1)\mathfrak{D}^r T_{r+1} = (\mathcal{B} + \mathfrak{D}_3)\mathfrak{D}^{r-1} T_r.$$

There is no difficulty in inverting $(\mathcal{K} - r - 1)$, since every term of T_{r+1} has degree $m \geq 2$ and $\mathfrak{D}^r$ raises derivative index n by r, so that $m + n - r - 1 > 0$; doing so and iterating gives for $\mathfrak{D}^r T_{r+1}$ the explicit formula

$$\mathfrak{D}^r T_{r+1} = (\mathcal{K} - r - 1)^{-1}(-\partial_t)(\mathcal{K} - r)^{-1} \times (-\partial_t) \cdots (\mathcal{K} - 2)^{-1}(-\partial_t) u_0. \quad (42)$$

Fortunately we can bypass the need to invert $\mathfrak{D}^r$ by applying ∂_r, because $\partial_r \mathfrak{D} \doteqdot \partial_{r-1}$ by (12) and so $\partial_r \mathfrak{D}^r \doteqdot \partial_0$. Since $\partial_0 T_{r+1} = T_r$ by (30), we thus have an equivalent representation of the p.c.d.

$$T_r \doteqdot (-1)^{r-1}\partial_r(\mathcal{K} - r - 1)^{-1} \times \partial_t(\mathcal{K} - r)^{-1}\partial_t \cdots (\mathcal{K} - 2)^{-1}\partial_t u_0. \quad (43)$$

Instead of ∂_r we could have applied $\mathcal{Y}_r$, by (24) obtaining $\mathcal{Y}_0 T_{r+1}$ on the left. This in itself is no improvement, being equivalent to $\partial_0 T_{r+1}$, but it can be used to remedy a slight disadvantage of formula (43): though ∂_t (and the associated inverse "numerical" operator involving $\mathcal{K}$) is applied r times on the right as in (41), we obtain a p.c.d. of rank merely r, rather than $r + 1$. To restore the rank lowered by $\mathcal{Y}_0$, we multiply by u_0 and observe that $u_0 \mathcal{Y}_0 \doteqdot \mathcal{M}$, as seen in (21). Since $\mathcal{M}^{-1}\mathfrak{D} = \mathfrak{D}\mathcal{M}^{-1} \doteqdot 0$, applying $\mathcal{M}^{-1}$ yields

$$T_{r+1} \doteqdot (-1)^{r-1}\mathcal{M}^{-1} u_0 \mathcal{Y}_r(\mathcal{K} - r - 1)^{-1} \times \partial_t(\mathcal{K} - r)^{-1}\partial_t \cdots (\mathcal{K} - 2)^{-1}\partial_t u_0. \quad (44)$$

ACKNOWLEDGMENTS

The work presented here was supported by the Air Force Office of Scientific Research under Contract AF 49(638)-1555. One author (R. M. M.) also acknowledges support from the U.S. Atomic Energy Commission, Contract AT(30-1)-1480.

APPENDIX A: TRANSFORMATION PROPERTIES OF THE KdV EQUATION

It is of interest to find transformations of the dependent and independent variables that leave the KdV equation invariant. It is obviously invariant to arbitrary translations of the independent variables (since they do not occur explicitly)

$$x \to x + c_1, \quad t \to t + c_2.$$

Similarly, it is invariant to the scale transformation

$$x \to cx, \quad t \to c^3 t, \quad u \to c^{-2} u.$$

In particular, for $c = -1$ the signs of x and t are reversed and we note a kind of reversibility of the solution. Further the KdV equation is Galilean-invariant, i.e., invariant under the transformation

$$x \to x + ct, \quad t \to t, \quad u \to u + c.$$

This property is essential for our results in Sec. 4.

Aside from the invariance transformations, there are (real-valued) transformations that change the signs of individual terms in the KdV equation. The sign of the t-derivative term (relative to the other two) is changed by either $t \to -t$ or $x \to -x$, that of the nonlinear term by $u \to -u$, and that of the dispersive term by the combination $u \to -u$, $t \to -t$ (or $u \to -u$, $x \to -x$).

Incidentally, the KdV equation possesses a similarity solution, i.e., a solution U invariant under the above scale transformation; U satisfies the ordinary differential equation

$$\frac{d^3U}{d\eta^3} + (U - \eta)\frac{dU}{d\eta} - 2U = 0,$$

where

$$u(x, t) \equiv (3t)^{-\frac{2}{3}} U(\eta), \quad \eta \equiv x(3t)^{-\frac{1}{3}}.$$

APPENDIX B: CONSERVED DENSITY OF RANK $r = 11$

$$\begin{aligned} T_{11} = {} & \tfrac{1}{11} u_0^{11} - 45u_0^8 u_1^2 + 216u_0^7 u_2^2 - 1260u_0^5 u_1^4 \\ & - 648u_0^6 u_3^2 + 4320u_0^5 u_2^3 + 22680u_0^4 u_1^2 u_2^2 \\ & - 7560u_0^2 u_1^6 + 1296u_0^5 u_4^2 - 32400u_0^4 u_2 u_3^2 \\ & - 38880u_0^3 u_1^2 u_3^2 + 45360u_0^3 u_2^4 + 238464u_0^2 u_1^2 u_3^2 \end{aligned}$$

$$+ 136080u_0u_1^4u_2^2 - \tfrac{19440}{11}u_0^4u_5^2 + \tfrac{544320}{11}u_0^3u_2u_4^2$$
$$+ 38880u_0^2u_1^2u_4^2 - \tfrac{15240960}{11}u_0^2u_1u_3^3$$
$$- \tfrac{4906656}{11}u_0^2u_2^2u_3^2 - 668736u_0u_1^2u_2u_3^2$$
$$- 58320u_1^4u_3^2 + \tfrac{1788480}{11}u_0u_2^5$$
$$+ 423792u_1^2u_2^4 + \tfrac{233280}{143}u_0^3u_6^2$$
$$- \tfrac{6531840}{143}u_0^2u_1u_5^2 - \tfrac{233280}{11}u_0u_1^2u_5^2$$
$$+ \tfrac{58786560}{143}u_0^2u_4^3 + \tfrac{44416512}{143}u_0u_1u_3u_4^2$$
$$+ \tfrac{53141184}{143}u_0u_2^2u_4^2 + \tfrac{26593920}{11}u_1^2u_2u_4^2$$
$$- \tfrac{22091616}{143}u_0u_3^4 - \tfrac{12908160}{13}u_1u_2u_3^3$$
$$- \tfrac{131803200}{143}u_2^3u_3^2 - \tfrac{139968}{143}u_0^2u_7^2$$
$$+ \tfrac{33592320}{143}u_0u_2u_6^2 + \tfrac{6998400}{143}u_1^2u_6^2$$
$$- \tfrac{11757312}{143}u_0u_4u_5^2 - \tfrac{134369280}{143}u_1u_3u_5^2$$
$$- \tfrac{177759360}{143}u_2^2u_5^2 + \tfrac{37511424}{143}u_2u_4^3$$
$$+ \tfrac{544477520}{143}u_3^2u_4^2 + \tfrac{8398080}{2431}u_0u_8^2$$
$$- \tfrac{125971200}{2431}u_2u_7^2 + \tfrac{50388480}{221}u_4u_6^2$$
$$- \tfrac{251942400}{46189}u_9^2.$$

* Present address: Courant Institute for Mathematical Sciences, New York University, New York, New York 10012.

† Present address: Department of Mathematics, The University of Texas, Austin, Texas 78712.

[1] The papers in this series will be referred to by Roman numerals. The first three, which have already appeared, are: (I) R. M. Miura, J. Math. Phys. **9**, 1202 (1968); (II) R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968); (III) C. H. Su and C. S. Gardner, J. Math. Phys. **10**, 536 (1969).

[2] G. B. Whitham, Proc. Roy. Soc. (London) **A283**, 238 (1965).

[3] N. J. Zabusky, in *Proceedings of the Symposium on Nonlinear Partial Differential Equations*, W. F. Ames, Ed. (Academic Press Inc., New York, 1967), p. 223.

[4] We indicate the end of a proof by the symbol ∎.

[5] We thank Dr. D. Stevens of the Courant Institute of Mathematical Sciences, New York University, for letting us publish some of his computer results (private communication, 1969). He computed p.c.d. for the KdV equation (in particular the new result for $r = 11$) on the AEC CDC-6600 computer using both MUC and our explicit formula (43); see Sec. 4. MUC was faster, requiring 2.2 seconds of actual computation time for ranks $r = 3$ to 11 all together. On the other hand, it took 6.4 seconds to compute the p.c.d. for $r = 7$ using (43), while for $r = 8$ integers were generated greater than the standard maximum size that can be handled correctly by the machine (48 binary digits). MUC encounters this same difficulty, but not until $r = 12$. Earlier Dr. H. R. Lewis of the Los Alamos Scientific Laboratory, New Mexico (private communication, 1966) computed p.c.d. from (43) using the FORMAC algebraic manipulating system on an IBM 7094 computer; however, for $r = 6$, the available storage space was exceeded.

[6] This formula was first proved and pointed out to us by Dr. D. S. Wiley, presently with the State of New Jersey, Department of Higher Education (private communication, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Nonphysical Region Singularities of the *S* Matrix

PETER GODDARD
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England

(Received 9 May 1969)

The work of a previous paper, which examined the relationship between the assumption of Cutkosky-type discontinuity formulas to the positive-α condition for the singularity of anomalous thresholds, is extended using more powerful methods. The method is first described in general terms. It is then justified using homological techniques, and specific expressions are calculated for distortion terms. Finally, some specific cases are considered and more elaboration versions of the cancellation processes previously found are shown to occur.

1. INTRODUCTION

In a previous paper[1] (hereafter referred to as I), we considered the way that the positive-α condition for anomalous thresholds and the assumption of Cutkosky-type discontinuity formulas with "vanishing cycle" contours both lead to criteria for the singularity of connected *S*-matrix elements. This permitted the testing of the consistency of these conjectures on the basis of the usual *S*-matrix-theory assumptions. We remarked on the technical difficulty of performing finite continuations of unitarity integrals outside the physical region. The method we used to perform these continuations can be described as a metrical one in that we took a specific form for the unitarity integral as a repeated integral in the space of (complex) internal momenta, the order of the integrations being chosen with reference to the possible distortions. The different orders of integration result in different but "homologous" contours of integration, whose equivalence follows from Cauchy's theorem. The lack of symmetry in the way the integration contour had to be chosen and the increasing complexity of the integrals, taken with the delicate cancellation processes found in any but the simplest examples, make it clear that

$$+ 136080u_0u_1^4u_2^2 - \tfrac{19440}{11}u_0^4u_5^2 + \tfrac{544320}{11}u_0^3u_2u_4^2$$
$$+ 38880u_0^2u_1^2u_4^2 - \tfrac{15240960}{11}u_0^2u_1u_3^3$$
$$- \tfrac{4906656}{11}u_0^2u_2^2u_3^2 - 668736u_0u_1^2u_2u_3^2$$
$$- 58320u_1^4u_3^2 + \tfrac{1788480}{11}u_0u_2^5$$
$$+ 423792u_1^2u_2^4 + \tfrac{233280}{143}u_0^3u_6^2$$
$$- \tfrac{6531840}{143}u_0^2u_1u_5^2 - \tfrac{233280}{11}u_0u_1^2u_5^2$$
$$+ \tfrac{58786560}{143}u_0^2u_4^3 + \tfrac{44416512}{143}u_0u_1u_3u_4^2$$
$$+ \tfrac{53141184}{143}u_0u_2^2u_4^2 + \tfrac{26593920}{11}u_1^2u_2u_4^2$$
$$- \tfrac{22091616}{143}u_0u_3^4 - \tfrac{12908160}{13}u_1u_2u_3^3$$
$$- \tfrac{131803200}{143}u_2^3u_3^2 - \tfrac{139968}{143}u_0^2u_7^2$$
$$+ \tfrac{33592320}{143}u_0u_2u_6^2 + \tfrac{6998400}{143}u_1^2u_6^2$$
$$- \tfrac{11757312}{143}u_0u_4u_5^2 - \tfrac{134369280}{143}u_1u_3u_5^2$$
$$- \tfrac{177759360}{143}u_2^2u_5^2 + \tfrac{37511424}{143}u_2u_4^3$$
$$+ \tfrac{544477520}{143}u_3^2u_4^2 + \tfrac{8398080}{2431}u_0u_8^2$$
$$- \tfrac{125971200}{2431}u_2u_7^2 + \tfrac{50388480}{221}u_4u_6^2$$
$$- \tfrac{251942400}{46189}u_9^2.$$

* Present address: Courant Institute for Mathematical Sciences, New York University, New York, New York 10012.

† Present address: Department of Mathematics, The University of Texas, Austin, Texas 78712.

[1] The papers in this series will be referred to by Roman numerals. The first three, which have already appeared, are: (I) R. M. Miura, J. Math. Phys. **9**, 1202 (1968); (II) R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968); (III) C. H. Su and C. S. Gardner, J. Math. Phys. **10**, 536 (1969).

[2] G. B. Whitham, Proc. Roy. Soc. (London) **A283**, 238 (1965).

[3] N. J. Zabusky, in *Proceedings of the Symposium on Nonlinear Partial Differential Equations*, W. F. Ames, Ed. (Academic Press Inc., New York, 1967), p. 223.

[4] We indicate the end of a proof by the symbol ∎.

[5] We thank Dr. D. Stevens of the Courant Institute of Mathematical Sciences, New York University, for letting us publish some of his computer results (private communication, 1969). He computed p.c.d. for the KdV equation (in particular the new result for $r = 11$) on the AEC CDC-6600 computer using both MUC and our explicit formula (43); see Sec. 4. MUC was faster, requiring 2.2 seconds of actual computation time for ranks $r = 3$ to 11 all together. On the other hand, it took 6.4 seconds to compute the p.c.d. for $r = 7$ using (43), while for $r = 8$ integers were generated greater than the standard maximum size that can be handled correctly by the machine (48 binary digits). MUC encounters this same difficulty, but not until $r = 12$. Earlier Dr. H. R. Lewis of the Los Alamos Scientific Laboratory, New Mexico (private communication, 1966) computed p.c.d. from (43) using the FORMAC algebraic manipulating system on an IBM 7094 computer; however, for $r = 6$, the available storage space was exceeded.

[6] This formula was first proved and pointed out to us by Dr. D. S. Wiley, presently with the State of New Jersey, Department of Higher Education (private communication, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Nonphysical Region Singularities of the *S* Matrix

PETER GODDARD
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England

(Received 9 May 1969)

The work of a previous paper, which examined the relationship between the assumption of Cutkosky-type discontinuity formulas to the positive-α condition for the singularity of anomalous thresholds, is extended using more powerful methods. The method is first described in general terms. It is then justified using homological techniques, and specific expressions are calculated for distortion terms. Finally, some specific cases are considered and more elaboration versions of the cancellation processes previously found are shown to occur.

1. INTRODUCTION

In a previous paper[1] (hereafter referred to as I), we considered the way that the positive-α condition for anomalous thresholds and the assumption of Cutkosky-type discontinuity formulas with "vanishing cycle" contours both lead to criteria for the singularity of connected *S*-matrix elements. This permitted the testing of the consistency of these conjectures on the basis of the usual *S*-matrix-theory assumptions. We remarked on the technical difficulty of performing finite continuations of unitarity integrals outside the physical region. The method we used to perform these continuations can be described as a metrical one in that we took a specific form for the unitarity integral as a repeated integral in the space of (complex) internal momenta, the order of the integrations being chosen with reference to the possible distortions. The different orders of integration result in different but "homologous" contours of integration, whose equivalence follows from Cauchy's theorem. The lack of symmetry in the way the integration contour had to be chosen and the increasing complexity of the integrals, taken with the delicate cancellation processes found in any but the simplest examples, make it clear that

the range of cases easily accessible to this approach would be small. In Sec. 2, we describe in general terms a more powerful approach to the problem which lacks these disadvantages. In Sec. 3, we justify in rigorous terms the discussion of Sec. 2. Section 4 deals with calculations necessary to fix signs and establish cancellations. In Sec. 5, we return from these homological considerations to applications of the method and show how the cancellation process associated with distortion by the triangle anomalous threshold arises in this approach. Sections 6 and 7 deal with further particular cases. Section 8 summarizes some further work and problems.

2. DESCRIPTION OF METHOD

We have to consider finite continuations of unitarity integrals, and of distortion terms arising from them, outside the physical region. These take the form of integrals over certain internal loop momenta k, restricted to various surfaces $S_i(k,p)$, $1 \leq i \leq N$, by mass-shell δ functions dependent on the external momenta p. The integrand will have certain singularity surfaces $S_i(k,p)$, $N+1 \leq i \leq N'$, specified by Landau equations, which depend on the masses of the particles in the theory. As p is varied, the contour of integration will have to vary, and we now outline when and how such continuations may be performed.

Consider the general problem of the integral, with respect to variables k, over a contour h, contained in certain surfaces $S_i(k,p)$, $1 \leq i \leq N$, and bounded by surfaces $S_i(k,p)$, $N'+1 \leq i \leq N''$, of a function $A(k,p)$, which is analytic except for singularity surfaces $S_i(k,p)$, $N+1 \leq i \leq N'$. We suppose the $S_i(k,p)$ are analytic manifolds. The sort of situation we encounter is that, for some range of values of p, the $S_i(k,p)$ are in general position (that is, that they do not pinch anywhere), and there is a real integration contour h, not intersecting the singularity surfaces. Cauchy's theorem then tells us that this contour may be deformed in various ways, leaving the value of the integral unchanged. The amount of freedom this gives the contours of multidimensional integrals is difficult to visualize. The only easily picturable case of one complex dimension is not usually sufficiently general, and intuitive geometrical methods need to be supplemented by more rigorous mathematical tools. At a fixed value of p, Cauchy's theorem permits us to perform any continuous deformation of the contour which leaves the $S_i(k,p)$ invariant. Such a deformation will be specified by maps $k \to f_t(k)$, $0 \leq t \leq 1$, where $f_t(k)$ depends continuously on t and k (and has a continuous inverse for each t) and leaves $S_i(k,p)$ invariant, $1 \leq i \leq N''$, and $f_0(k) = k$ (that is, the map starts as the identity). Similarly, when continuing in p, it is only necessary to "continuously," rather than analytically, continue the contour in order to analytically continue the integral. To continue along a path $p = \lambda(u)$, $0 \leq u \leq 1$, on which the $S_i(k,p)$ are in general position, one needs to construct continuous maps $k \to f_u(k)$, $0 \leq u \leq 1$, such that $(k, \lambda(0)) \to (f_u(k), \lambda(u))$ gives an isomorphism of $S_i(k, \lambda(0))$ onto $S_i(k, \lambda(u))$ and $f_0(k) = k$. Such a set of maps can be constructed given the requirement of general position (and also a compactness condition). (This is a rough explanation of the "ambiant isotopy" concept. For details of this and subsequent notions see Sec. 3 and the references there.)

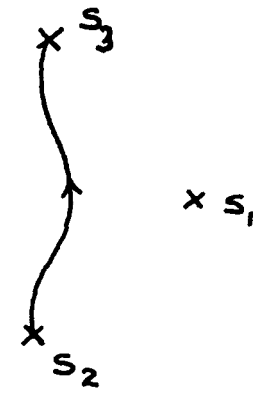


FIG. 1.

The possibility of constructing such a deformation is obvious in certain simple cases. Consider the 1-dimensional integration from $s_2(p)$ to $s_3(p)$ of a function with a pole at $k = s_1(p)$, illustrated in Fig. 1. Continuous deformation is always possible as long as s_1, s_2, s_3 remain finite and distinct.

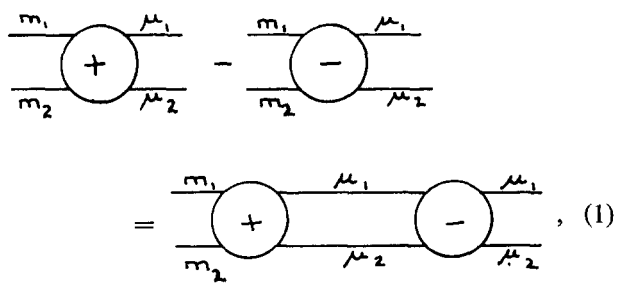
Returning to our general case, the integral $\int_h A$ thus continues into $\int_{f_u h} A$ along the path λ. f_u is not unique, but the corresponding possibilities for $f_u h$ are contours whose equivalence for A follows from Cauchy's theorem. The problem of continuing the integral all the way along λ is that of finding what f_1 maps h into. Now, the properties f is required to have are specified in terms of the $S_i(k,p)$ and, given these surfaces, the problem of continuing the contour h is independent of A. That is, in order to decide how and whether the contour may be continued along a given path, it is necessary only to know the singularity surfaces of A (and the boundary surfaces, etc., of the contour). The problem can be considered for any surfaces $S_i(k,p)$ independently of the existence of an analytic function A. This is crucial to our argument.

Now consider continuous variations of the path λ for fixed initial and final points of the continuation, $\lambda(0)$ and $\lambda(1)$. This continuous variation will be specified by a continuous function $\tau_v(u)$ of (u, v), with

$\tau_0(u) = \lambda(u)$, $0 \leq u \leq 1$ and $\tau_v(0) = \lambda(0)$, $\tau_v(1) = \lambda(1)$ so that the end points remain fixed. Provided that the $S_i(k, p)$ are in general position at each point of the surface of values of p, $\tau_v(u)$, $0 \leq u \leq 1$, $0 \leq v \leq 1$, traced out by the continuous variation of λ, then the "final" path will induce maps f_1' which define equivalent contours $f_1'h$ at $\lambda(1)$ to those defined by maps induced by $\tau = \lambda$. That is, paths that may be continuously deformed into one another result in equivalent continuations of contours, given fixed end points.

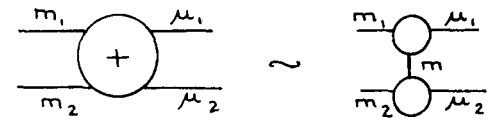
In the specific applications, keeping these general ideas in mind, we consider the continuation of unitarity integrals corresponding to normal thresholds occurring below the physical region for the relevant process, from just below this physical region to just above the normal threshold corresponding to the integral. Since the $S_i(k, p)$ depend on the masses of the theory only, we can consider paths of continuation for the contour on which the masses vary and even become complex. We may continuously deform paths along which the integral may be continued (paths at constant physical values of the masses, as specified by the theory) into paths along which the integrand is not defined, but where it is easier to decide what happens in the contour continuation. We will now show how this enables us to obtain information about continuation in a general case from the obvious solution in a particular case. The approach of trying to solve problems connected with distortions of unitarity integral contours, during finite continuations outside the physical region, by using the known behavior of other continuations has previously been employed by Polkinghorne,[2] who used Feynman integrals as a "mathematical probe."

Let us consider this, as in the introduction to I, in the context of the relation of the positive-α condition for a triangle anomalous threshold to the distortion of a 2-particle unitarity integral by a crossed channel pole. The extended unitarity equation considered there was



(1)

where $\mu_1 + \mu_2 \to m_1 + m_2$ has a t-channel pole at $t = m^2$, corresponding to a particle of mass m:



(at $t = m^2$). (2)

The continuation considered was from just below $(m_1 + m_2)^2$, where the contour of the unitarity integral is specified by the requirement that one obtains the appropriate Cutkosky-type discontinuity for the $(m_1 + m_2)^2$ normal threshold, to just above $(\mu_1 + \mu_2)^2$. A real (undistorted) initial contour in the loop momenta k will always exist, for stable values of the masses, just below threshold. If, more particularly, $m_1 = m_2$ and $\mu_1 = \mu_2$ with μ_1 just less than m_1 and m just less than $m_1 + \mu_1$, it is easy to see that this real k contour remains undistorted on our path continuation, which is illustrated in Fig. 2 (because the set of values t takes on the real contour during continuation varies continuously with the masses and so may be restricted to any neighborhood of $t = 0$, which is the sole value it takes when $\mu_1 = m_1$, by taking μ_1 near enough m_1). Denote this path by λ_0. Let λ_v, $0 \leq v \leq 1$, each be paths at fixed values of the masses, depending continuously on v, and free of pinches of the $S_i(k, p)$ (the variables p are now to be understood to include the masses), and suppose λ_1 is not the general value of the masses which is of interest. Now, we cannot immediately extract information about what happens along λ_1 from what is known to happen along λ_0, because they have different end points since they correspond to different values of the masses. In order to obtain a path with the same end points as λ_1, one has to add to λ_0 the paths traced out by the end points of λ_v, taken appropriate directions (see Fig. 3).

The path obtained by first following ν_0 and then λ_0 and then ν_1 has the same end points as λ_1 and may clearly be obtained from it by continuous deformation. If ν_0 is taken just below the $(m_1 + m_2)^2$ threshold and near enough to it [that is, the initial point of the path λ is kept near enough $(m_1 + m_2)^2$], the integration contour may be kept as the usual real undistorted contour specified by

$$k_1^2 = \mu_1^2, \quad k_2^2 = \mu_2^2, \quad k_1 + k_2 = P, \quad k_i^0 > 0. \tag{3}$$

Thus, continuation along ν_0 takes the contour required in (1) at the general value of the masses of interest,

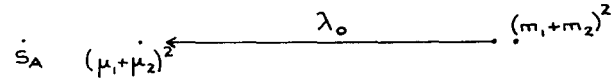
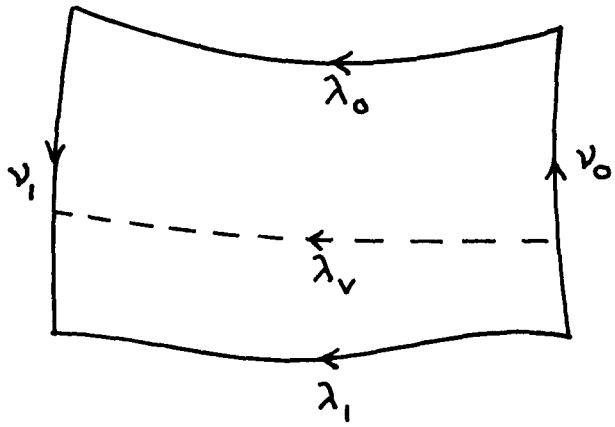


FIG. 2. Complex s plane; triangle anomalous threshold.



$\nu_0(v) = \lambda_{1-v}(0)$

$\nu_1(v) = \lambda_v(1)$

FIG. 3.

just below threshold, into that whose continuation along λ_0 is clearly undistorted; that is, it continues into the vanishing cycle at $(\mu_1 + \mu_2)^2$. We must now discuss what happens on ν_1 to this vanishing contour. The vanishing contour appropriate to a pinch of a given set of boundary, singularity, and constraint surfaces is uniquely determined (up to the usual equivalence resulting from Cauchy's theorem) apart from sign. Any contour appropriate to these surfaces will, on continuation along a small loop in p space about the pinch point, just have some multiple of the vanishing contour added to it. Consequently, we can always generate the vanishing contour in this way. Hence, we will have the vanishing contour at the end of λ_0 continuing into the vanishing contour at the end of λ_1 along ν_1, provided that one ensures that a small loop about $(\mu_1 + \mu_2)^2$, θ_0, with base point at $\nu_1(0) = \lambda_0(1)$, can be continuously deformed through loops θ_v with base point $\nu_1(v)$ to a loop θ_1, about $(\mu_1 + \mu_2)^2$ with base point $\nu_1(1) = \lambda_1(1)$ (see Fig. 4). Then, the path θ_1 may be continuously deformed into $\nu_1\theta_0\nu_1^{-1}$ and so, for a suitable contour h, being continued around θ_1 is equivalent to being continued along the following path: ν_1 taken in the reverse direction [this gives h' at $\nu_1(0)$, say], taken around θ_0 (which adds a multiple of e', the vanishing contour corresponding to θ_0), and taken back along ν_1 (this takes h' back into h and e' into e). Thus, any contour of the appropriate type has a multiple of e added to it on continuation around θ_1. Thus, e, which is the image of the vanishing contour e' corresponding to θ_0, is the vanishing contour corresponding to θ_1. So, provided that we can deform θ as well as λ, we can now infer that on λ_1 the contour appropriate in (1) just below $(m_1 + m_2)^2$ continues into the vanishing contour just above $(\mu_1 + \mu_2)^2$. Because λ_1 has to be obtained by this sort of continuous deformation, it is quite possible that it necessarily passes through nonreal values of $s = P^2$, between its end points at the $(\mu_1 + \mu_2)^2$ and $(m_1 + m_2)^2$ thresholds; indeed, if it does pass through real values only, there is no distortion. In general, the path may loop other pinch points before ending just above $(\mu_1 + \mu_2)^2$. In this case, as we shall now see, we can infer what the initial contour is in terms of contours which vanish at various thresholds when continued along the real s axis, avoiding intervening thresholds by negative imaginary contours.

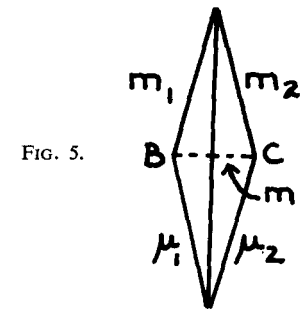


FIG. 5.

As we vary the masses from the extreme situation corresponding to Fig. 5, we must vary the paths λ, θ (see Fig. 6a), so that the pinch points s_A (which is the triangle anomalous threshold corresponding to a

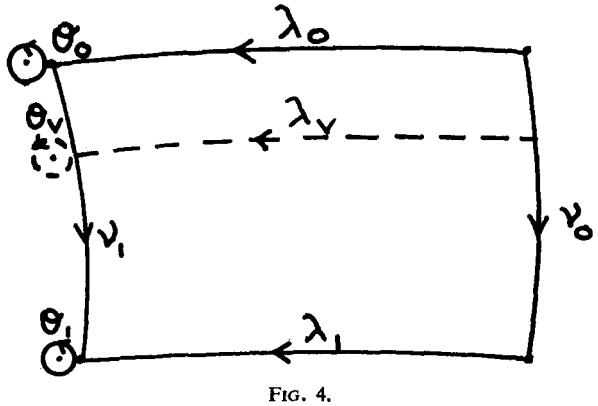


FIG. 4.

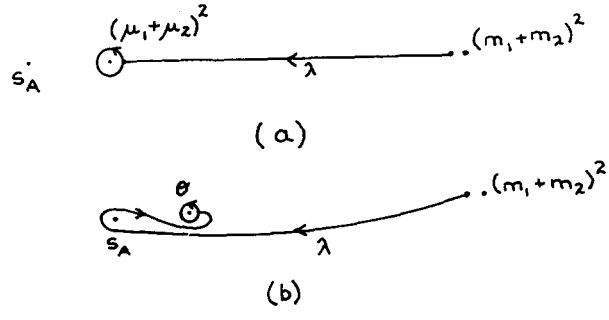
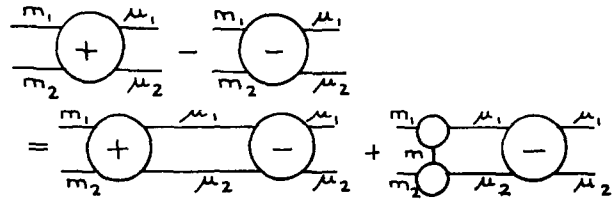


FIG. 6. Complex s planes.

pinch of the pole at $t = m^2$, with the two constraining mass-shell surfaces) and the normal threshold (corresponding to a pinch of the mass-shell surfaces) are kept away from them. (Pseudothresholds, etc., must also be avoided, but these do not come near in any case.) Clearly, this deformation of λ and θ will be trivial until s_A approaches $(\mu_1 + \mu_2)^2$. As long as $s_A < (\mu_1 + \mu_2)^2$ during the continuation λ may be kept as the usual path along the real s axis. As this is a Euclidean-region threshold, a coincidence between s_A and $(\mu_1 + \mu_2)^2$ is easily detected in the dual diagram. The position of s_A is given by the length of BC in Fig. 5. Figure 5 has mixed α's, and one can clearly vary the masses to reach any other mixed-α situation, keeping $s_A < (\mu_1 + \mu_2)^2$. (For details about the triangle Landau curve see Eden *et al.*, Ref. 3, pp. 57–65.) Thus, we conclude that there is no distortion in the mixed-α case. To reach a positive-α configuration (see Fig. 7) it would be necessary to move through a coincidence of s_A and $(\mu_1 + \mu_2)^2$ or to avoid this by taking a detour through complex values of the masses, during which s_A would loop $(\mu_1 + \mu_2)^2$. Notice, again, that this crucial coincidence which is the dividing point between the two cases is easily detected in the dual diagram. s_A can be made to loop $(\mu_1 + \mu_2)^2$ either way, and we choose the more convenient one for our subsequent argument. As s_A moves round, the path λ must be deformed around before it, for if s_A were to cross λ (or θ), we would have the no pinch condition on the paths violated. Figure 6(a) shows the effect of avoiding coincidence by a clockwise detour, that is, a detour in a negative sense. Once this detour has been made and we are in the situation of Fig. 6(b), any other positive-α situation can be reached without a coincidence of s_A and $(\mu_1 + \mu_2)^2$ and, thus, the corresponding path λ will be as in Fig. 6(b). Thus, in the positive-α situation one has that the contour required just below $(m_1 + m_2)^2$ continues into the vanishing cycle at $(\mu_1 + \mu_2)^2$ if first continued around s_A in the manner described. Consequently, we may obtain the contour one should start with just below $(m_1 + m_2)^2$ by reversing this contribution and beginning with the vanishing cycle just above $(\mu_1 + \mu_2)^2$. If this is continued around s_A, one adds a multiple of the vanishing contour (corresponding to the pinch at s_A) to this contour. The details of this and, in particular, the signs are discussed (for a general case) in Sec. 4. However, leaving aside the details of signs of terms for the moment, above $(\mu_1 + \mu_2)^2$ the extended unitarity relation may be written as in (4), where the contours for the integrals on the right side will continue into vanishing contours, at the



(4)

appropriate thresholds, along a path which goes through decreasing real values of s, except for small detours in the lower-half complex s plane to avoid intervening thresholds. Thus, separating these two contributions on the right-hand side of (4), we replace Fig. 6(b) by Fig. 8 and obtain the appropriate contour just below $(m_1 + m_2)^2$ by continuing up vanishing contours at s_A and $(\mu_1 + \mu_2)^2$ along paths indicated in this diagram. Further distortions, in more complicated cases, where there are other singularities present, can be discussed in more terms of Fig. 8. In this way, the most general case considered is reduced to that of continuing a Cutkosky-type integral from just below the real s axis to just above the corresponding anomalous or normal threshold.

For definiteness we have discussed the method in the context of the simplest particular case; the application to a general case should, however, be clear from this. In such a case, we shall be dealing with a general unitarity integral corresponding to a threshold below the physical region and start with a diagram like Fig. 6a. Points at which the character of λ changes will correspond to flattenings of the dual diagrams of certain anomalous thresholds of the amplitude. This is exemplified in Sec. 5 and subsequent sections.

Before moving to a rigorous justification of the analysis of this section, we note that up to now we

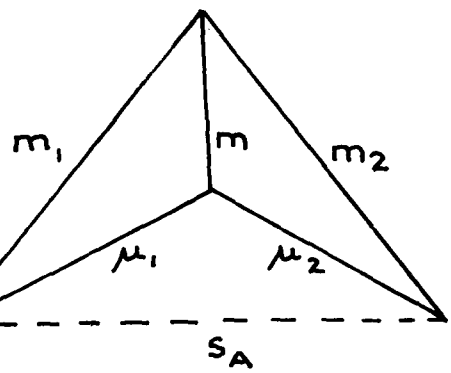


FIG. 7.

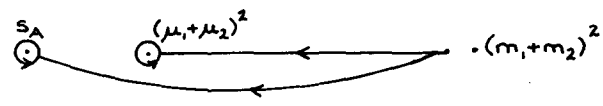


FIG. 8.

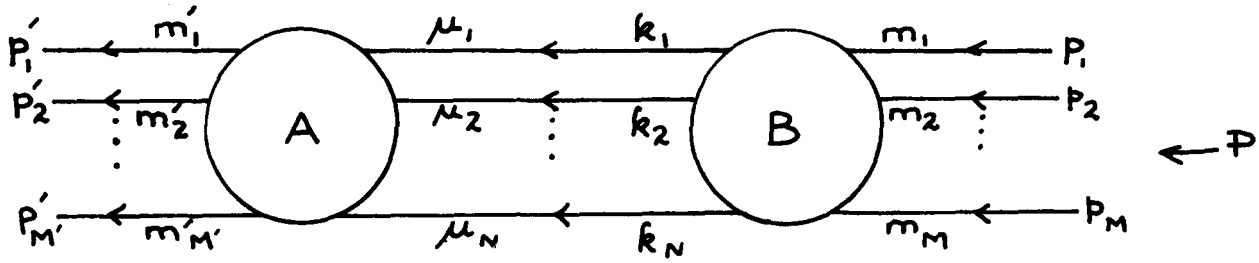


FIG. 9.

have omitted mention of technical problems with the method arising from noncompactness and singularities of the individual $S_i(k, p)$. These are dealt with in the more detailed discussion of the next section.

3. JUSTIFICATION OF THE METHOD HOMOLOGICAL TECHNIQUES

In this section and the next, we shall apply ideas developed by various authors on the application of homological techniques to the sort of integrals of functions of several complex variables occurring in particle physics. Accounts of this work can be found in the books of Pham[4] and Hwa and Teplitz.[5] Elsewhere, Pham[6] had used these techniques to discuss aspects of the consistency of the physical-region singularity structure, suggested by perturbation theory, within S-matrix theory. Pham did not use unitarity; subsequently, Bloxham, Olive, and Polkinghorne[7] have proved this singularity structure for simple Landau curves on the basis of the usual S-matrix-theory postulates. We wish to apply these homological techniques to study extended unitarity equations outside the physical region and to develop a method for performing finite continuations.

First, let us consider expressions for unitarity integrals. Consider the unitarity integral (Fig. 9) for the process $m_1 + \cdots + m_M \to m_1' + \cdots + m_{M'}'$ corresponding to the $(\sum_{i=1}^N \mu_i)^2$ threshold, which occurs below the physical region. We write

$$P = \sum_{i=1}^{M} p_i' = \sum_{i=1}^{M'} p_i'.$$

For $S = p^2 > (\sum \mu_i)^2$, this integral is given by (5). Further for real P in this region, the integral is restricted by the δ functions to a real integration contour in k space

$$(-2\pi i)^N \frac{i^{N-1}}{(2\pi)^{4(N-1)}} \times \int AB\left[\prod_{i=1}^{N} \delta^{(+)}(k_i^2 - \mu_i^2)\, d^4k_i\right]\delta_4\left(P - \sum_{i=1}^{N} k_i\right), \quad (5)$$

specified by

$$k_i^2 = \mu_i^2, \quad 1 \le i \le N, \quad \sum_{i=1}^{N} k_i = P, \quad k_i^0 > 0. \quad (6)$$

Because singularities of the integrand may cross the real contour defined by (6) when we are outside the physical region, the values of the expression (5) for various real P with $P^2 > (\sum_{i=1}^N \mu_i)^2$ will not necessarily be related by analytic continuation along the real P^2 axis, even when there is no intervening singularity. In particular, this applies to its values at points just below the physical region and points just above the $(\sum_{i=1}^N \mu_i)^2$ threshold. In the latter region, the integration is over a vanishing contour. We shall write (5) in a form capable of being continued easily [cf. Ref. 8, Eq. (2.2)f.].

Let $s_i(k) = k_i^2 - \mu_i^2$ and S_i be the surfaces $s_i = 0$ in $\mathbb{C}^{4l}$, where $l = N - 1$, the space of the complex k_i, $1 \le i \le N - 1$, and where we have set $k_N = P - \sum_{i=1}^{N-1} k_i$. Then, the s_i are in general position in $\mathbb{C}^{4l}$ for $P^2 \ne (\sum \mu_i)^2$ (or any pseudothreshold), and (5) can be written, just below the physical region, as

$$(2\pi i)^N \frac{i^l}{(2\pi)^{4l}} \int_{h_1} f \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N}, \quad (7)$$

where $f = AB$, $d^{4l}k = \bigwedge_{i=1}^{l} (\bigwedge_{\mu=0}^{3} dk_i^\mu)$ (the repeated exterior product of the dk_i^μ) and h_1 is a suitable element of the compact homology group

$$H_{4l-N}^c\left[\bigcap_{i=1}^{N} S_i \sim \bigcup_{i=N+1}^{N'} S_i\right]^{\sim}.$$

Here the $S_i(k, p)$, $N < i \le N'$, are singularity surfaces of f and

$$\tilde{\Sigma}_p \equiv \left[\bigcap_{i=1}^{N} S_i \sim \bigcup_{i=N+1}^{N'} S_i\right]^{\sim}$$

is a covering space of

$$\Sigma_p \equiv \bigcup_{i=1}^{N} S_i \sim \bigcup_{i=N+1}^{N'} S_i$$

on which f is defined (that is, single valued).

Just above $s = (\sum \mu_i)^2$, (5) will equal (7) with h_1 replaced by $\pm e$, where e is the "vanishing class"[4,9] of

$H^c_{4l-N}(\tilde{\Sigma}_p)$ at this threshold (on the appropriate sheet) and the sign depends on the orientation chosen for e. We now make this choice: Let $k_i = (\kappa_i^0, i\kappa_i)$ and $\kappa_i = (\kappa_i^0, \boldsymbol{\kappa}_i)$. Then, there is a real "vanishing cell"[4,9] in the κ_i coordinates just below the threshold at $(\sum_{i=1}^N \mu_i)^2$. The order of the κ_i thus induces an orientation of the vanishing cell $\mathbf{e}$ (cf. the definition of $d^{4l}k$) and, hence, of e. Now, in (7), h_1 must be replaced by an element of $H_{4l-N}(\Sigma_p)$, agreeing with the homology class of the current (compare Boyling, Ref. 8)

$$\prod_{i=1}^{N} \delta^{(+)}(s_i)\, ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N . \tag{8}$$

Noticing that the normals to the S_i are all outwards to the vanishing cell e for s just below $(\sum_{i=1}^N \mu_i)^2$, one obtains the following expression for the value of (5) just above threshold:

$$(-1)^{l+1}(-2\pi i)^N \frac{i^l}{(2\pi)^{4l}} \int_e f \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N} . \tag{9}$$

Now, let G be a Landau diagram for the process

$$m_1 + m_2 + \cdots + m_M \to m_1' + m_2' + \cdots + m_{M'}'$$

corresponding to a threshold in the Euclidean region. To this graph we associate a Cutkosky integral (9), where e is the vanishing cycle at the threshold, corresponding to G, at which s takes its largest value (here we will consider only fixed-s thresholds), e is oriented as described above, and k_i are internal momenta for G. f is a product of amplitudes and coupling constants corresponding to the vertices.

Now consider the following situation. Let Z be a compact n-dimensional analytic manifold and P an analytic manifold. Let $S_i(p)$, $1 \le i \le N'$, be analytic submanifolds of Z in general position for $p \in P$. Write $\mathcal{S}_p = \bigcup_{i=1}^{N'} S_i(p)$. Then, any path λ: $[0, 1] \to P$ induces a Z-ambiant isotopy class from $\mathcal{S}_{\lambda(0)}$ to $\mathcal{S}_{\lambda(1)}$.[4,10] This in turn will induce an isomorphism of the homology groups $H_{n-N}(\tilde{\Sigma}_{\lambda(0)}) \leftrightarrow H_{n-N}(\tilde{\Sigma}_{\lambda(1)})$, where $\tilde{\Sigma}$ is a covering space of

$$\Sigma = \bigcup_{p \in P} \left\{ \bigcap_{i=1}^{N} S_i(p) \sim \bigcup_{i=N+1}^{N'} S_i(p) \right\} \subseteq Z \times P$$

and $\tilde{\Sigma}_p = \tilde{\pi}^{-1}(p)$, $\tilde{\pi}$ being the natural projection $\tilde{\Sigma} \to P$. This ambiant isotopy class and, hence, the isomorphism is dependent only on the homotopy class of λ in P (relative to fixed end points).

Further, let $(u, v) \to \lambda(u)$ be a continuous map of the unit square $[0, 1] \times [0, 1] \to P$. Write $\nu_0(v) = \lambda_{1-v}(0)$ and $\nu_1(V) = \lambda_v(1)$, as in Sec. 2. Then, λ_1 is homotopic to ν_1, λ_0, ν_0 and, consequently, these paths induce the same isomorphisms: $H_{n-N}(\tilde{\Sigma}_{\lambda_1(0)}) \leftrightarrow H_{n-N}(\tilde{\Sigma}_{\lambda_1(1)})$. Suppose, now, that $P = P' \sim L$, where P' and L are analytic manifolds and L is a submanifold of P' of (complex) codimension one, and that some subset of the S_i has a simple (quadratic zero) pinch[4,10] on L. Let $(u, v) \to \theta_v(u)$ be a continuous map of the unit square $[0, 1] \times [0, 1] \to P$, with $\theta_v(0) = \theta_v(1) = \lambda_v(1)$ $[= \nu_1(v)]$ and such that θ_v is a simple loop[4] about L for all $v \in [0, 1]$. Using λ_*, etc., to denote isomorphisms of the group $H_{n-N}(\tilde{\Sigma}_p)$, and e_v to denote the corresponding vanishing class (on the appropriate sheet) associated with the loop θ, it is easy to see that $\nu_{1*}e_0 = e_1$. So, since $\lambda_{1*} = \nu_{1*} \cdot \lambda_{0*} \cdot \nu_{0*}$, if $h \in H_{n-N}(\tilde{\Sigma}_{\lambda(0)})$ and $\lambda_{0*}h = e_0$, then $\lambda_{1*}(\nu_{0*}^{-1}h) = e_1$.

The difficulties with applying these results about $Z \times P$ to our situation are of the type usually met in the application of homological methods to S-matrix theory or Feynman integrals. While they have, as yet, prevented the completion of attempts to completely analyze situations of interest,[5,11,12] they are not important for the limited applications we have in mind. In applying this discussion, we would wish to take $Z = \mathbb{C}^{4l}$, P to be some subset of the space of external momenta and internal and external masses, with the S_i defined as at the beginning of this section. The difficulties are that Z would not then be compact and that the S_i, $N + 1 \le i \le N'$, would not always be manifolds (because we are dealing with points outside the physical region and here they may have cusps). The former difficulty is associated with second-type singularities. To overcome it, it is necessary to compactify $\mathbb{C}^{4l}$ in some way. The more obvious ways of compactifying this space[8,11,12] will have the difficulty that, when the S_i are replaced by the closure of their images in this compactification, they are, for any but the simplest graphs, no longer in general position[8,11,13] at infinity. In consequence, the compactification procedure has to be supplemented by either "stratification"[5,12] or a "resolution of singularities."[11,12] The fact that the "permanent pinches" at infinity can be resolved means that the above argument is applicable to our case, provided that P does not contain any "second-type singularities." Similarly, the singular points of the S_i may be removed, and they will not bother us provided that the contour is kept away from them.

By considering P to include paths along which the masses vary, we are considering a bigger space than that on which f is defined, and paths along which the continuation of (5) has no meaning. However, for arbitrary values of the masses satisfying the stability conditions, the relations (6), or equivalently the current (8), specify a real integration contour for the

unitarity integral just below the physical region. Our approach now, as described in Sec. 2, is to take λ_0 to be a path along the real s axis down to just above the $(\sum_{i=1}^N \mu_i)^2$ threshold, at sufficiently extreme values of the masses to make it clear that $\lambda_{0*}h_0 = e_0$, where h_0 is the element of $H_{4l-N}(\tilde{\Sigma}_p)$ defined by (6) just below the "physical region." For Euclidean region singularities, stability ensures that, if we choose λ_v so that ν_0 is kept near enough threshold (see Fig. 4), the real contour (6) will be left "undistorted," that is, $\nu_{0*}h_1 = h_0$. L is taken (firstly) to be the $(\sum_{i=1}^N \mu_i)^2$ threshold. So using $\nu_{1*}e_0 = e_1$, we have $\lambda_{1*}h_0 = e_1$. When the pinches occur in the Euclidean region or below, the masses may be kept real as they are varied from their extreme values (corresponding to λ_0) until some anomalous threshold is about to coincide with $(\sum_{i=1}^N \mu_i)^2$. This coincidence would make deformation of θ impossible. If it is avoided by a complex detour, the threshold will loop $(\sum_{i=1}^N \mu_i)^2$ and we will reach a situation corresponding to Fig. 6(b), which we can then replace by a diagram like Fig. 8. In general, we get this encircling because the coincidence of thresholds corresponds to the effective tangency of the Landau curves, which the thresholds trace out as the masses are varied. After the contact we replace λ by the two paths indicated in Fig. 8, and h_v is then the sum of the images of the respective vanishing contours under the induced isomorphisms of the homology groups. We may then consider continuous deformations of these paths. In this way, the problem is reduced to the investigation for the Cutkosky-integration contour defined by (5); the effective contacts of the Landau curve correspond to a pinch of all the S_i, $1 \leq i \leq N$, with those Landau surfaces corresponding to the pinch of a subset of these S_i, or a subset of the S_i, $1 \leq i \leq N'$, containing the first N. In fact, in the cases we consider here, it turns out that, as regards situations of the latter type, it is only necessary to consider effective contacts of the surface corresponding to a pinch of the S_i, $1 \leq i \leq N$, with the surfaces corresponding to pinches of these with one other S_i, $i > N$. In the next section, we discuss how to attach the correct sign, etc., to these additional terms.

4. EFFECTIVE CONTACTS

We consider again the Cutkosky integral (9)

$$(-2\pi i)^N(-1)^{l+1}\frac{i^l}{(2\pi)^{4l}}\int_{e_{B_1}} f\frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N}, \tag{10}$$

where e_{B_1} is the appropriate vanishing class for a pinch of all the S_i, $1 \leq i \leq N$. We study the effective contact between the Landau curve corresponding to a simple

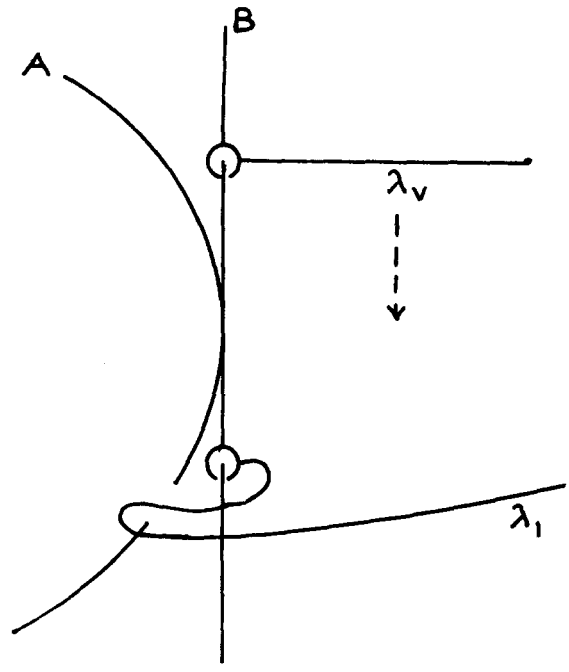


FIG. 10.

pinch of the S_i, $1 \leq i \leq N$, (labeled B in Fig. 10) and that corresponding to a simple pinch of these together with S_{N+1} (labeled A in Fig. 10). In the cases we discuss, the effective contact is of the type discussed by Pham.[9] The effect on the path λ of continuation past the contact of A and B is illustrated in Fig. 10 [compare with Fig. 6(a, b)]. Figure 11 defines simple loops α_1, β_1, α_2, β_2 (no two of which are homotopic) about various parts of A and B. (Compare Pham, Ref. 4, pp. 93ff.) We orientate the vanishing spheres $\mathbf{e}_{A_1}$, $\mathbf{e}_{A_2}$, $\mathbf{e}_{B_1}$, $\mathbf{e}_{B_2}$, corresponding to α_1, β_1, α_2, β_2 as indicated in Sec. 3. Then, following Pham,[4] using $\tilde{e}$ for "vanishing cycles" and e for vanishing spheres, we have

$$KI[\mathbf{e}_{A_1}, \tilde{e}_{B_1}] = (-1)^{\frac{1}{2}n(n-1)}, \quad n = 4l. \tag{11}$$

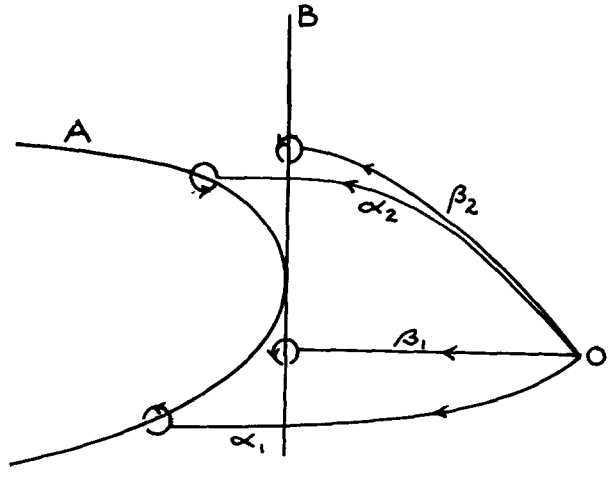


FIG. 11.

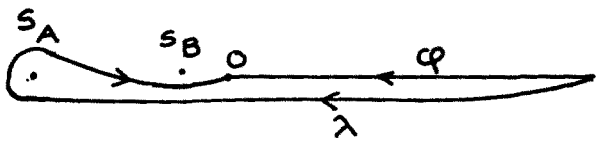


FIG. 12.

Now suppose f has a pole on S_{N+1}: $f \sim f_1/S_{N+1}$. So with the notation of the last section, $h_1 = \lambda_{*1}^{-1} e_{B_1} = \varphi_*^{-1}\alpha_{1*} e_{B_1}$ where φ is the path directly through real values of P indicated in Fig. 12 [cf. Fig. 6(b)]. Thus, using the Picard–Lefschetz theorem[4] (following Ref. 4, we use Leray's original sign convention), as

$$e_{B_1} = \partial^{(N)} \cdot \partial^{(N-1)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{B_1} \in H^c_{n-N}\left(\bigcap_{i=1}^{N} S_i \sim S_{N+1}\right),$$

$$\begin{aligned} \alpha_{1*} e_{B_1} &= e_{B_1} + (-1)^{\frac{1}{2}(n-N+1)(n-N+2)} \\ &\quad \times KI[\partial^{(N)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{A_1}, e_{B_1}] \delta^{(N+1)} e_{A_1} \\ &= e_{B_1} + (-1)^{\frac{1}{2}(n+1)(n+2)} KI[\mathbf{e}_{A_1}, \check{e}_{B_1}] \delta^{(N+1)} e_{A_1} \\ &= e_{B_1} - \delta^{(N+1)} e_{A_1}. \end{aligned} \tag{12}$$

Thus, in the situation of Fig. 6(b), (7) is equal to (10) together with the additional term

$$-(-2\pi i)^N (-1)^{l+1} \frac{i^l}{(2\pi)^{4l}} \times \int_{\delta^{(N+1)} e_{A_1}} f \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N}. \tag{13}$$

Using the Cauchy theorem of Leray's residue calculus, this becomes

$$(-2\pi i)^{N+1} (-1)^{l+1} \frac{i^l}{(2\pi)^{4l}} \times \int_{e_{A_1}} \operatorname{Res}\left[f \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N} \right]. \tag{14}$$

$$s_{N+1} f|_{S_{N+1}} = f_1|_{S_{N+1}}$$

and

$$\begin{aligned} \operatorname{Res}\left[f \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N} \right] &= \left[\frac{s_{N+1} f \, d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N} \right] \Big/ ds_{N+1} \\ &= f_1 \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_{N+1}}. \end{aligned} \tag{15}$$

Thus (14) becomes

$$(-2\pi i)^{N+1} (-1)^{l+1} \frac{i^l}{(2\pi)^{4l}} \times \int_{e_{A_1}} f_1 \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_{N+1}}, \tag{16}$$

which is of the required form (10).

Now, suppose S_{N+1} is a branch point of f; we will assume (on an inductive basis) that the discontinuity of f around this branch point will be given by

$$\operatorname{disc}_{N+1} f = (-2\pi i)^{N'} (-1)^{l'+1} \frac{i^{l'}}{(2\pi)^{4l'}} \times \int_{e_{A'}} g \frac{d^{4l'}k'}{ds'_1 \wedge ds'_2 \wedge \cdots \wedge ds'_{N'}}, \tag{17}$$

where l' is the number of loops, N' the number of lines, g the product of amplitudes, and $e_{A'}$ the vanishing sphere corresponding to the graph G' of which S_{N+1} will be a positive-α arc. e_{B_1} is now the vanishing sphere on some sheet of the covering space of $\bigcap_{i=1}^{N} S_i \sim S_{N+1}$. In Pham's notation his generalization of the Picard–Lefschetz formula takes the form[5,9]

$$\begin{aligned} \alpha_{1*} {}_f e_{B_1} &= {}_f e_{B_1} + \sum_{f'} (-1)^{\frac{1}{2}(n-N+1)(n-N+2)} \\ &\quad \times KI[\partial^{(N)} \cdot \cdots \cdot \partial^{(1)}{}_{f'} e_{A_1}, {}_f e_{B_1}] \\ &\quad \times (1 - \omega_{N+1})_* \partial^{(N)} \cdot \cdots \cdot \partial^{(1)}{}_{f'} \mathbf{e}_{A_1}, \end{aligned} \tag{18}$$

$$\alpha_{1*} e_{B_1} = e_{B_1} - (1 - \omega_{N+1})_* \partial^{(N)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{A_1}. \tag{19}$$

Thus, in this case, in the situation of Fig. 6(b), (7) is equal to (10) together with

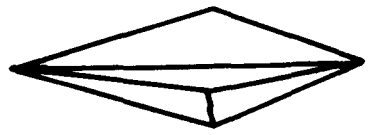
$$-(-2\pi i)^N (-1)^{l+1} \frac{i^l}{(2\pi)^{4l}} \times \int_{\partial^{(N)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{A_1}} \operatorname{disc}_{N+1} f \frac{d^{4l}k}{ds_1 \wedge ds_2 \wedge \cdots \wedge ds_N}. \tag{20}$$

So, if G'' is the graph obtained by inserting G' into G (so that, in Pham's notation,[6] the sequence $G' \to G'' \to G$ is exact), combining (17) and (20), we find that the additional term into (10) will be

$$(-2\pi i)^{N''} (-1)^{l''+1} \frac{i^{l''}}{(2\pi)^{4l''}} \times \int_{\partial^{(N)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{A_1} \otimes e_{A'}} g \frac{d^{4l''}k''}{ds''_1 \wedge ds''_2 \wedge \cdots \wedge ds''_{N''}}, \tag{21}$$

where $s''_i = s_i$, $1 \le i \le N$, $s''_{i+N} = s_i$, $1 \le i \le N'$, $N'' = N' + N$, $l'' = l' + l$. Further by comparison with Ref. 6 (p. 150) one sees that the fibration of the ball $\partial^{(N)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{A_1}$ by the sphere $e_{A'}$ is the sphere $e_{A''}$ and, thus, (21) is of the form (10). This concludes our discussion of the second of the two effective contact situations described at the end of Sec. 3.

FIG. 13.



To discuss what happens when a subset of the S_i, $1 \leq i \leq N$, pinch, change notation so that (10) becomes

$$(-2\pi i)^{N''}(-1)^{l''+1}\frac{i^{l''}}{(2\pi)^{4l''}} \times \int_{e_{A''}} g \frac{d^{4l''}k''}{ds_1'' \wedge ds_2'' \wedge \cdots \wedge ds_{N''}''}, \quad (22)$$

and the pinching subsets are S_i, $1 \leq i \leq N < N''$. Then, (22) can be written in the form (21). Referring to Fig. 11, but moving in the opposite direction to that indicated there, after the effective contact the appropriate vanishing cell will be $\mathbf{e}_{A_2}$ rather than $\mathbf{e}_{A_1}$. Using $\mathbf{e}_{A_1} + \mathbf{e}_{A_2} = \mathbf{e}_B$ as elements of $H(C^{4l}, \bigcup^{N+1} S_i)$, we have $\partial^{(N)} \cdot \cdots \cdot \partial^{(1)} e_{A_1} = -\partial^{(N)} \cdot \cdots \cdot \partial^{(1)} \mathbf{e}_{A_2}$; (22) will be changed by having $e_{A''}$ replaced by $e_{A_2''}$, where the subscript 2 indicates that this is the sphere appropriate to the other side of the contact and that there is an over-all change of sign. This minus sign is responsible for the cancellation mechanism discussed in I and to be rederived in the next section.

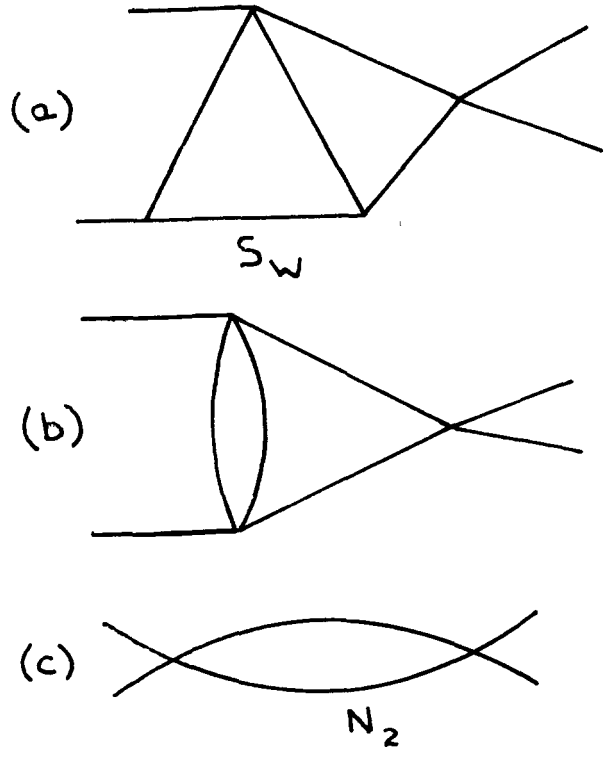
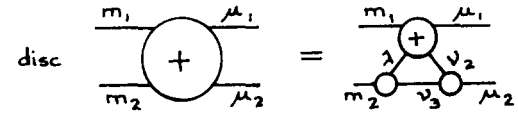


FIG. 14

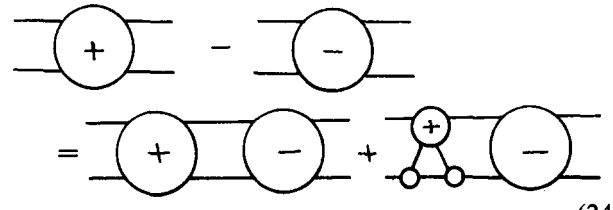
5. APPLICATION TO DISTORTION BY THE TRIANGLE SINGULARITY: THE CANCELLATION PROCESS

We consider the situation discussed in Sec. 5 of I. In Eq. (1) the amplitude now has, instead of a pole, a t-channel cut corresponding to a triangle anomalous threshold with the discontinuity



(23)

Figure 13 shows the extreme situation (analogous to Fig. 5) in which we can infer that there is no distortion. We shall assume that the threshold of (23) is singular, satisfies the positive-α condition, and has the discontinuity (23) in line with the consistency arguments of Sec. 2. Assuming the positive-α condition for the distorting triangle singularity means that we do not have to consider coincidences between the Landau surfaces of Figs. 14(a, b). We can pass from the situation of Fig. 13 to any positive-α situation for Fig. 14(a) (see Fig. 15), with only a coincidence of the thresholds of Figs. 14(a, c) occurring. Thus, from the discussion of the last section, the appropriate form for (1) just above the $(\mu_1 + \mu_2)^2$ threshold in the positive-α situation is



(24)

which is the desired form for consistency. But we must also discuss the appropriate form for (1) in the situation illustrated in Fig. 16. Consideration of this diagram indicates that the extended unitarity equation (1) should sometimes contain a 3-particle unitarity

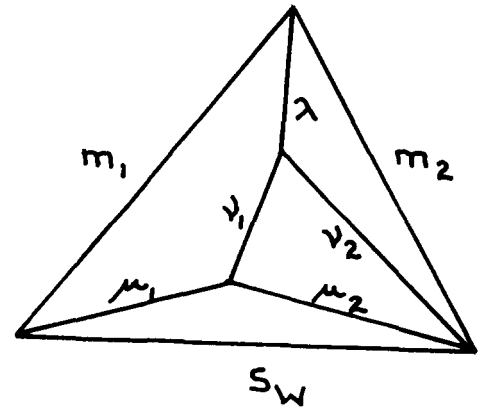


FIG. 15

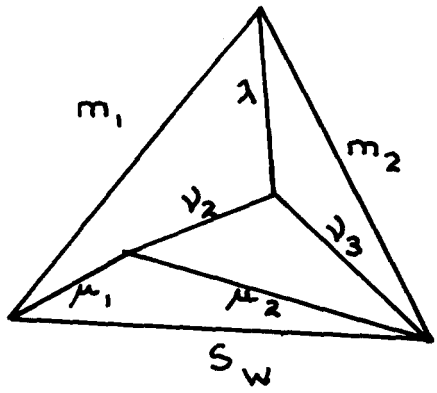
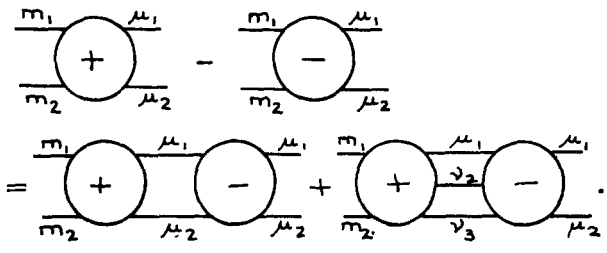


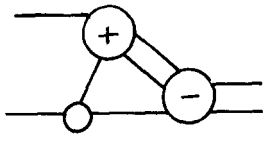
FIG. 16.

integral corresponding to the $(\mu_1 + \nu_2 + \nu_3)^2$ threshold



(25)

We must consider the continuation of this unitarity integral as well and coincidences of the thresholds of Fig. 14 with those of Fig. 17. To get from a situation in which the 3-particle integral is undistorted to the situation of Fig. 16 we must first pass through (or rather avoid by a complex detour) a coincidence of the thresholds of Figs. 17(a, b). If the masses are in the proportions indicated by Fig. 15, Fig. 17(b) would have positive α's and there would be an additional term (26) in the extended unitarity equation (25):



(26)

To go from Fig. 15 to Fig. 16 we must avoid a coincidence of Fig. 14(a) with Fig. 17(b). If we use expressions such as (26) to denote the precisely defined Cutkosky integrals of Secs. 3 and 4 and if we apply the last part of the latter section, we see that, in the situation of Fig. 16, the distortion of the 2-particle integral by the triangle threshold will give rise to a term which is minus the Cutkosky integral, whereas there is a double distortion term from the 3-particle integral which is just the Cutkosky integral. These terms cancel, leaving in the extended unitarity relation just one integral for each positive α anomalous threshold, over a contour which vanishes at this threshold. Labeling the thresholds as indicated in the Figs. 14 and 17, we draw a table of diagrams in Fig. 18 to indicate the continuation between the situations with the various signs for the α's. The 2-particle continuation to s_W in Fig. 18(d) illustrates the path that introduces an extra minus sign relative to the 3-particle distortion in this situation.

Note that in a full analysis one still has to discuss stable values of the masses which do not give real roots for s_W (as in I, Sec. 6, for example). However, this part is completed when one notes that one can continue to such values from mixed-α arcs without coincidences of thresholds occurring. Thus, the approach outlined in Secs. 2–4 has reduced the problem to one of plane Euclidean geometry for these thresholds.

6. APPLICATION TO THE THRESHOLDS OF FIG. 19 AND FIG. 23

We now apply the above method to a case not considered in I (Fig. 19). While this threshold would be very difficult to discuss from the point of view of the previous paper, it is a simple application of the new method. Figure 20 indicates an extreme position in which we can infer the absence of distortion in the 3-particle integral. Figure 21 shows thresholds whose coincidences with that of Fig. 19 must be avoided. We indicate the transition from the situation of Fig. 19 to any positive-α situation in a series of diagrams like those of Fig. 18 (see Fig. 22). Thus, there are no cancellation processes associated with Fig. 19. We can now consider distortion by such a threshold in the crossed channel (Fig. 23). This will be discussed in a little detail; the new features it possesses are also found in the next section. Assuming the positive-α condition for the distortion threshold (Fig. 19, crossed), establishing the required correspondence

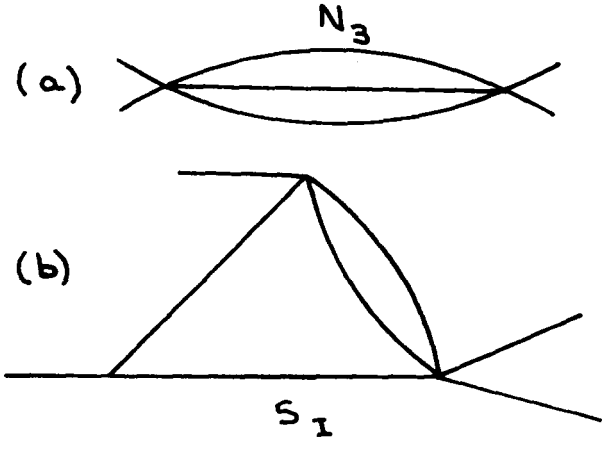


FIG. 17.

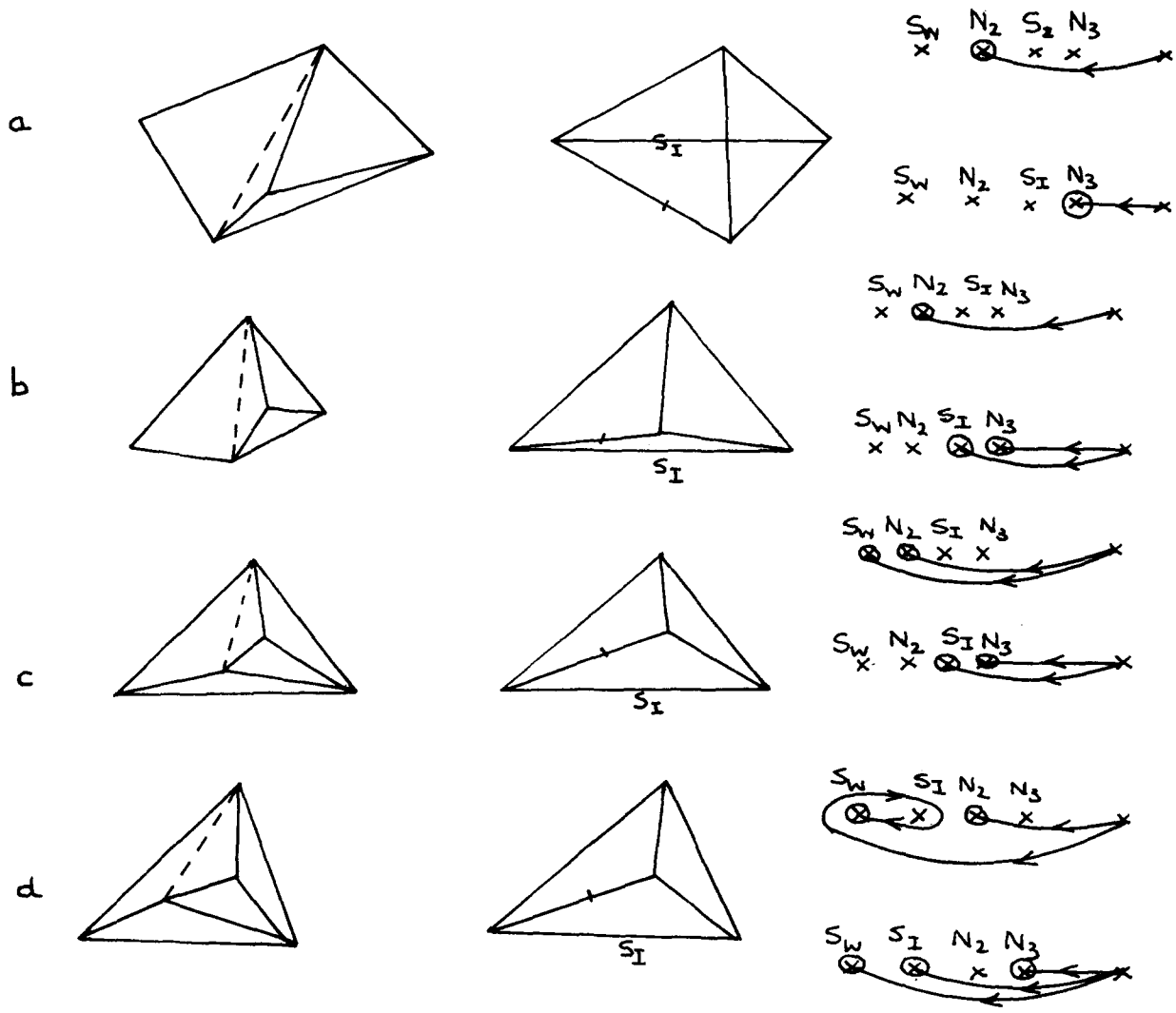


FIG. 18. Appropriate versions of Figs. 15 and 16, appropriate versions of dual diagram of Fig. 17b, and continuation paths (2- and 3-particle unitary integrals in the physical region), from left to right, respectively; detours taken to avoid a coincidence of: (b) S_I and N_3; (c) S_W and N_2; (d) S_W and S_I.

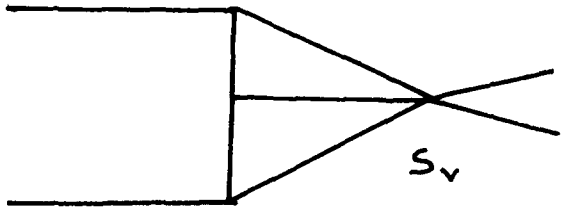


FIG. 19.

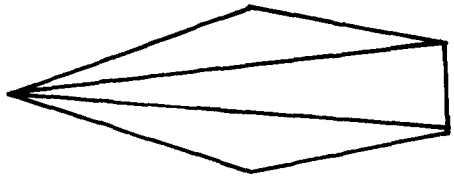
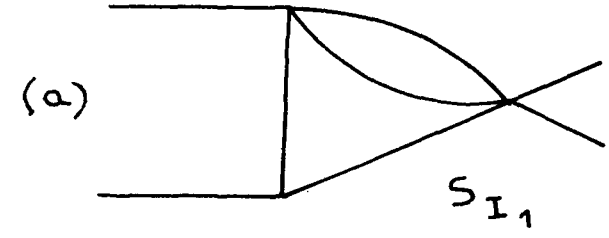


FIG. 20.



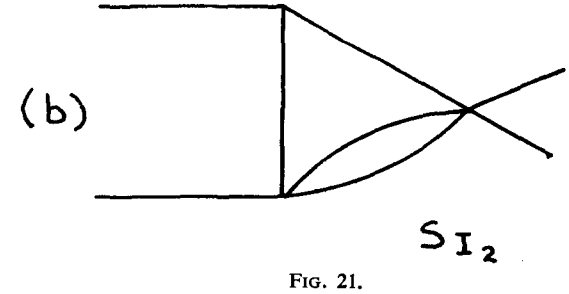


FIG. 21.

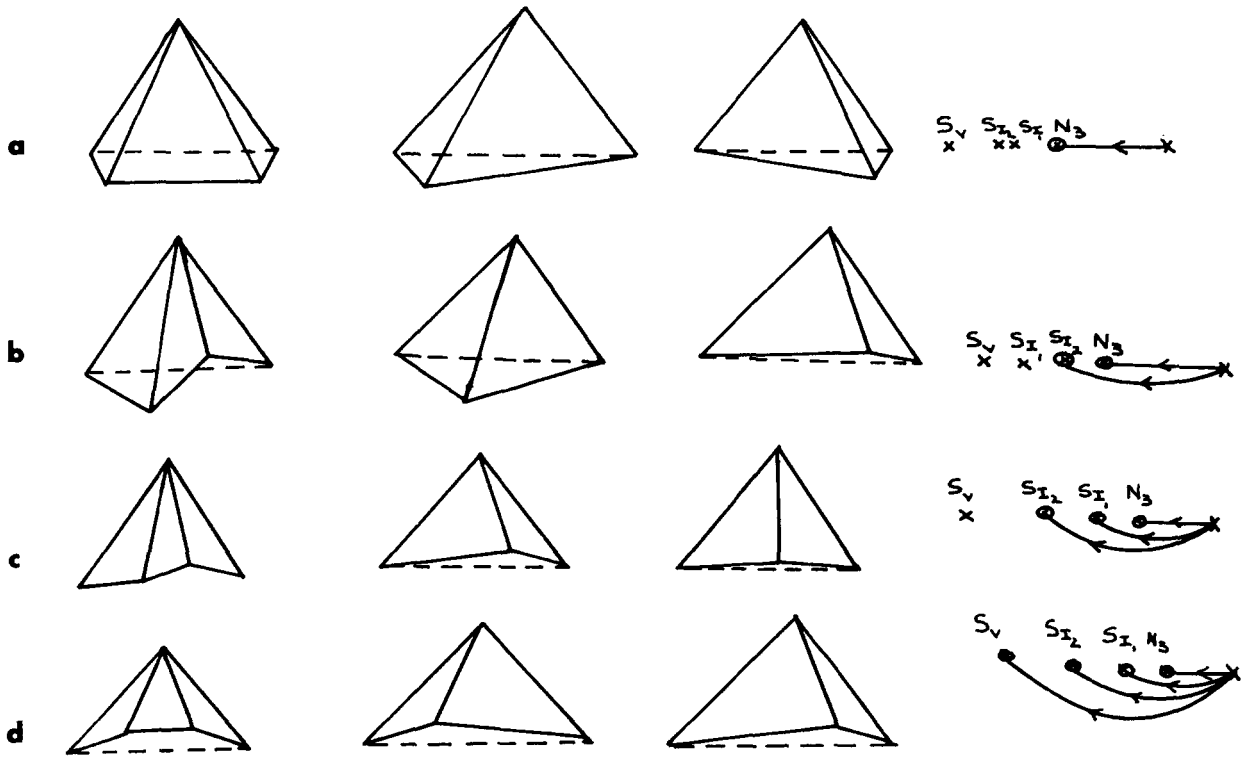


FIG. 22. Appropriate versions of dual diagrams of Figs. 19, 21a, and 21b, and continuation paths (physical region), from left to right, respectively; detours taken to avoid a coincidence of: (b) S_{I_2} and N_3; (c) S_{I_1} and N_3; (d) S_{I_2} and S_V.

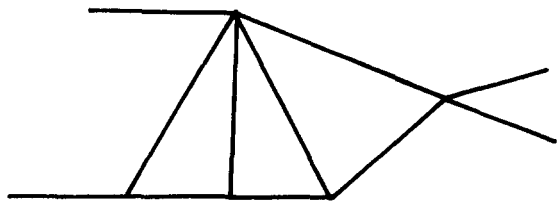


FIG. 23.

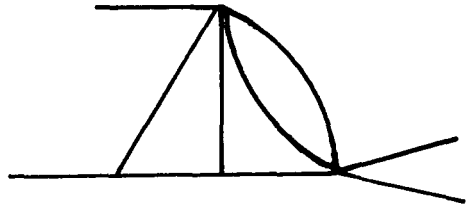


FIG. 24.

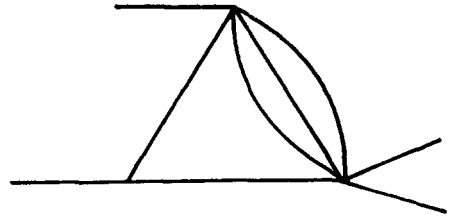


FIG. 25.

between the positive-α condition for Fig. 23 and the extended unitarity relations involves considering cancellations with distortions of the distortion term associated with the threshold of Fig. 24. This term itself is involved in cancellation effects in the thresholds of Fig. 25. Figure 26 illustrates values of the masses for which both cancellations are taking place. We can get to this situation from a nondistortion situation by detouring the following coincidences of thresholds in this order: first between Fig. 25 and the 4-particle normal threshold, then between Fig. 24 and the 3-particle normal threshold, and finally between Fig. 23 and the 2-particle normal threshold. This gets one into a situation in which all α's are positive. To get to Fig. 26, it remains to detour contacts between Figs. 23 and 24 and between Figs. 24 and 25. Thus, in terms of distortions, the 4-, 3-, and 2-particle unitarity integrals produce distortions corresponding to Figs. 25, 24, and 23, respectively, the distortion of the 2-particle integral is then cancelled by a distortion of the distortion of the 3-particle one, and then the distortion of the 3-particle integral is itself cancelled by a distortion of the 4-particle integral leaving only the term corresponding to Fig. 25, which is the only threshold satisfying the positive-α condition. So the cancellation effects are quite subtle, and a general

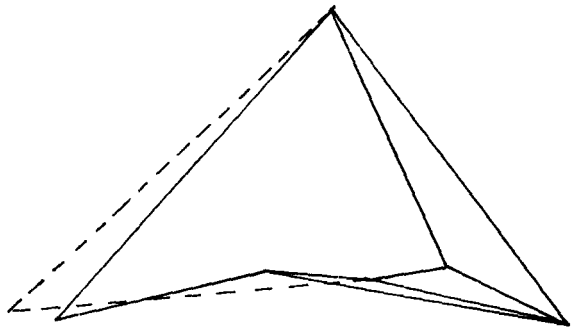


FIG. 26.

proof of the consistency of the Cutkosky discontinuity formula and the positive-α condition for, say, fixed s-anomalous thresholds on some sort of "inductive" basis would have to encompass all such possibilities. A further obstacle to this project is provided by the case considered in the next section.

7. DISTORTION BY THE ACNODE CURVE

The "inductive" idea referred to at the end of the last section is that the "consistency" should have already been proved for the distorting singularity. However, the case we now consider shows that this sort of induction is not possible if we only discuss fixed s-singularities (and use crossing). The acnode diagram (see Eden *et al.*, Ref. 3, Sec. 2.8) is an example of a curve depending on both s and t which enters the Euclidean region. Figure 27 illustrates an anomalous threshold corresponding to distortion of a 2-particle integral by the acnode curve and Fig. 28 a dual diagram of Fig. 27, which is Euclidean and satisfies the positive-α condition (note that the distorting singularity has a Landau curve which is not a manifold because it has cusps). Figure 27 has to be considered in conjunction with Fig. 29 because of cancellations. This, in turn, requires the consideration of Fig. 30. Figures 27, 29, and 30 are in a similar relation to Figs. 23, 24, and 25, and one gets the same complicated structure of cancellation effects in certain cases; the analysis goes through as in the last section.

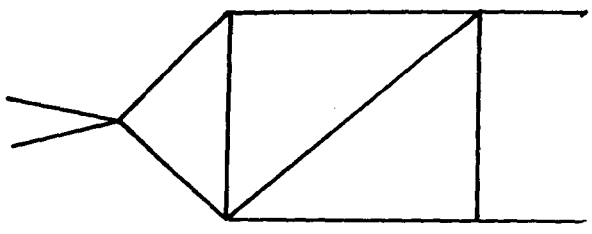


FIG. 27.

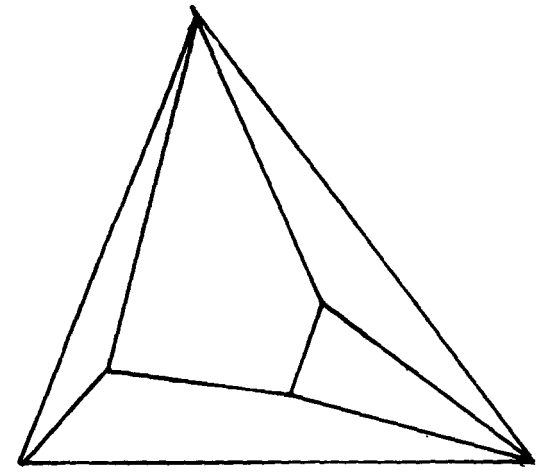


FIG. 28.

8. COMMENTS

In Sec. 7, we assumed that the distorting singularity belongs to the positive-α arc of the acnode curve and has the appropriate discontinuity. The verification of the consistency of these assumptions, so far discussed only for fixed s threshold, would require a generalization in the application of the methods of Secs. 2–4. This will be the subject of a further paper.

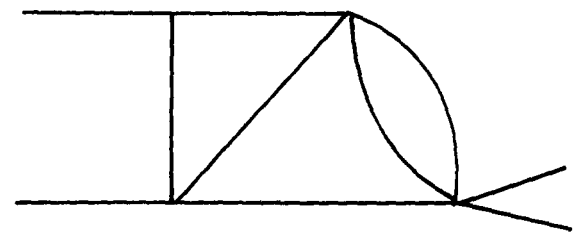


FIG. 29.

The methods developed in this part make a much wider range of cases accessible than those of I. They appear capable of effecting the test of consistency of the Cutkosky discontinuity formulas (with vanishing contours) and the positive-α criterion, for, at least, fixed-s Euclidean region thresholds. They reduce this problem of distortion of multidimensional complex

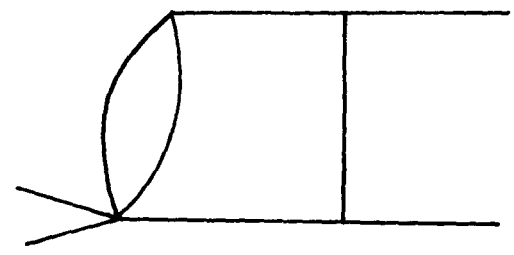


FIG. 30.

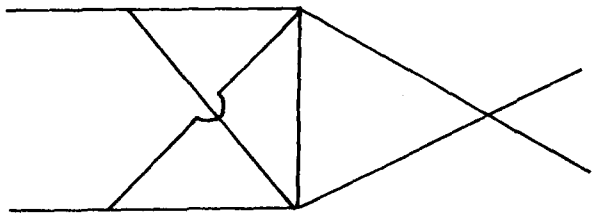


FIG. 31.

integrals to one in plane Euclidean geometry; but the geometry can get very complicated and so far neither a counter example nor a general proof has been found. In addition to the examples considered above, a nonplanar diagram (Fig. 31) has been considered. No new problems arise here apart from the great increase complexity of the dual diagram geometry.

As in I, all the examples we have considered are consistent with the conjecture that the condition that physical sheet singularities have discontinuities, given by a Cutkosky-type expression integrated over a "flat" or vanishing contour, is equivalent to the condition that the physical sheet singularity structure is identical with that given by finite-order perturbation theory.

ACKNOWLEDGMENTS

I should like to thank Professor J. C. Polkinghorne for a lot of advice and encouragement with this work. I should also like to thank the Science Research Council for a grant.

[1] P. Goddard, Nuovo Cimento **59A**, 335 (1969).
[2] J. C. Polkinghorne, J. Math. Phys. **7**, 2230 (1966).
[3] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, Cambridge, England, 1966).
[4] F. Pham, *Introduction à l'étude topologique des singularities de Landau* (Memorial des Sciences Mathematiques, Fasc. 164) (Gauthier-Villars, Paris, 1967).
[5] R. C. Hwa and V. L. Teplitz, *Homology and Feynman Integrals* (W. A. Benjamin, New York, 1966).
[6] F. Pham, Ann. Inst. Henri Poincaré **6**, 89 (1967).
[7] M. J. W. Bloxham, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **10**, 494, 545, 553 (1969).
[8] J. B. Boyling, Nuovo Cimento **44**, 379 (1966).
[9] F. Pham, Bull. Soc. Math. France **93**, 333 (1965) (reprinted in Hwa and Teplitz, Ref. 5).
[10] D. Fotiadi, M. Froissart, J. Lascoux, and F. Pham, Topology **4**, 159 (1965) (reprinted in Hwa and Teplitz, Ref. 5).
[11] M. J. Westwater, thesis, Cambridge University, Cambridge, England, 1966. Some of this is published in Westwater, Helv. Phys. Acta **40**, 596 (1967).
[12] J. B. Boyling, Nuovo Cimento **53A** 351 (1968).
[13] D. B. Fairlie, P. V. Landshoff, J. Nuttall, and J. C. Polkinghorne, J. Math. Phys. **3**, 594 (1962) (reprinted in Hwa and Teplitz, Ref. 5).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Boson Expansions for an Exactly Soluble Model of Interacting Fermions with $SU(3)$ Symmetry*

SHOU YUNG LI, ABRAHAM KLEIN, AND R. M. DREIZLER
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 7 July 1969)

An exactly soluble three-level model for a system of fermions with Hamiltonian formed from generators of the group $SU(3)$ is studied. A basis for the representation of which the ground state is a member is constructed. It is demonstrated that in this representation, the generators can each be expanded in a series as functions of a pair of "kinematical" boson operators; the series are uniquely determined to satisfy the operator algebra and the invariants of the representation by a method of Marumori. It is seen that the lowest anharmonic approximation to the Hamiltonian and other operators yields excellent numerical agreement with exact results for all regimes of interaction strength considered. An alternative description of the system in terms of a dynamically more meaningful boson, called the "physical" boson, is shown to be appropriate for relatively weak coupling where one has a near harmonic spectrum.

I. INTRODUCTION

This paper represents a contribution to the problem of the description of the low-lying excitation spectrum of a system of fermions by means of boson variables. Within the context of nuclear physics, which provided the stimulus for the present study, a method for the expansion of fermion-pair operators as series of boson operators was first suggested by Beliaev and Zelevinsky[1] (BZ). This method has been generalized by Sørensen[2] and applied by him to the physically most important case of quadrupole vibrations of heavy nuclei. A modified method, developed by Marumori *et al.*,[3,4] takes more accurate account of the restrictions of the Pauli principle by transforming the fermion space into a truncated boson space through a nonunitary transformation.

Insight into the connection between the two methods was provided in a study by Pang, Klein, and Dreizler[5] of an exactly soluble model[6] of Meshkov, Glick, and Lipkin (MGL). For this model, it was shown that the two expansion methods, when completely formulated, gave exactly the same series expansions and yielded quite satisfactory approximations to the exact solution. From this model, from the present work, and from additional studies in course, we can give the general formulation of the problem that follows.

The commutators of fermion-pair operators define a Lie algebra; the first task is to obtain a representation of the generators of this algebra in terms of boson operators, the latter *always* fewer in number than the generators themselves. The solution determined solely by the condition that all the commutators be satisfied must contain arbitrary constants, at least sufficient in number to specify any representation of the Lie algebra accessible to pairs of fermions. The values of the constants should, in this case, be uniquely determined by the independent Casimir invariants of the algebra. (This last statement is missing from the work of BZ and Sørensen, but is necessary and sufficient to establish its equivalence to that of Marumori *et al.*) The problem of boson representation can be completely settled by the purely "geometrical" considerations described above; we refer to the resulting boson operators as *kinematical.*

Having determined one such representation without any reference to dynamics, one may subject it to unitary transformation. Thus, the well-known boson of the random phase approximation is linearly related to the kinematical boson. The problem of diagonalizing the Hamiltonian can, more generally, be interpreted as the introduction of a nonlinear transformation in the boson space. These alternative boson representations always contain more than the minimum number of constants, the additional parameters to be determined by dynamical requirements.

In the present work, we use Marumori's method for an exactly soluble model with $SU(3)$ symmetry which contains the model of MGL as a limiting case.[7] Some of the properties of the model are described in Secs. II and III. In Secs. IV and V, Marumori's method is used to obtain kinematical-boson expansion series for the Hamiltonian and for various transition operators. By contrast, we introduce in Sec. VI the concept of "physical" boson, which provides a straightforward way of measuring the anharmonicity of a vibrational Hamiltonian.

In Sec. VII, we discuss and compare our various numerical results. These include: (i) exact diagonalization of the Hamiltonian for the $SU(3)$ representation containing the ground state; (ii) diagonalization of a

quartic approximation in the kinematical bosons; (iii) discussion of limits of validity of a vibrational picture.

II. THE $SU(3)$ MODEL

Assume three N-fold degenerate single-particle shells with energies ϵ_1, ϵ_2, and ϵ_3, respectively. A single-particle state is, therefore, specified by two quantum numbers m and σ, where $m = 1, 2, \cdots, N$ numbers the substates within a shell and $\sigma = 1, 2, 3$ is the quantum number designating the different shells. It is also assumed that the two-body interactions V_{12}, V_{13}, and V_{23}, existing between shells, are of monopole–monopole type which scatter pairs of particles, change the quantum number σ, but keep m constant. This assumption about the interaction, as will be seen in the following, allows one to write the Hamiltonian in terms of $SU(3)$ generators and is, therefore, an essential assumption in the construction of the model.

With the above assumptions, the Hamiltonian is taken as

$$H = \sum_{m\sigma} \epsilon_\sigma a^\dagger_{m\sigma} a_{m\sigma} + \left(V_{12} \sum_{mm'} a^\dagger_{m2} a^\dagger_{m'2} a_{m'1} a_{m1} + \text{h.c.} \right) + \left(V_{13} \sum_{mm'} a^\dagger_{m3} a^\dagger_{m'3} a_{m'1} a_{m1} + \text{h.c.} \right) + \left(V_{23} \sum_{mm'} a^\dagger_{m3} a^\dagger_{m'3} a_{m'2} a_{m2} + \text{h.c.} \right), \tag{1}$$

where $a_{m\sigma}$ and its Hermitian conjugate $a^\dagger_{m\sigma}$ are second-quantized fermion operators which satisfy the anti-commutation relations

$$\{a_{m\sigma}, a^\dagger_{m'\sigma'}\} = \delta_{mm'}\delta_{\sigma\sigma'} \tag{2}$$

and

$$\{a_{m\sigma}, a_{m'\sigma'}\} = 0. \tag{3}$$

Next, if one defines the operators

$$G_{\sigma\sigma'} = \sum_m a^\dagger_{m\sigma} a_{m\sigma'}, \tag{4}$$

then, from (2) and (3), one obtains the following commutation relations:

$$[G_{\alpha\beta}, G_{\gamma\rho}] = \delta_{\beta\gamma} G_{\alpha\rho} - \delta_{\alpha\rho} G_{\gamma\beta}. \tag{5}$$

The nine operators $G_{\sigma\sigma'}$, $\sigma, \sigma' = 1, 2, 3$, are consequently the generators of a $U(3)$ algebra which becomes an $SU(3)$ algebra if we remember the condition of particle-number conservation,

$$\sum_\sigma G_{\sigma\sigma} = N. \tag{6}$$

With (6), we definitely restrict ourselves to particle number equal to the degeneracy of one of the shells. Finally, if the V_{ij} are assumed to be real as well as Hermitian, i.e.,

$$V_{ij} = V^*_{ji} = V_{ji},$$

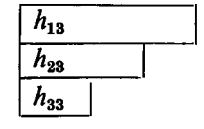
then the Hamiltonian (1) will have the form

$$H = \sum_\sigma \epsilon_\sigma G_{\sigma\sigma} + V_{12}(G^2_{21} + G^2_{12}) + V_{13}(G^2_{31} + G^2_{13}) + V_{23}(G^2_{32} + G^2_{23}). \tag{7}$$

It is thus an $SU(3)$-model Hamiltonian which has no matrix element between states belonging to two different irreducible representations of $SU(3)$.

III. THE GROUND-STATE BAND

In this section, the irreducible representation of $SU(3)$ to which the ground state belongs is discussed, and we find all the states belonging to that representation. For this purpose, the $SU(3) \supset SU(2) \supset SU(1)$ classification is used for the states of a given irreducible representation. Also the standard notation used by Moshinsky[8] is adopted here. In this notation an irreducible representation of $SU(3)$ corresponding to the Young tableaux,



will be denoted as $[h_{13}, h_{23}, h_{33}]$. A state belonging to this representation and classified according to $SU(3) \supset SU(2) \supset SU(1)$ will be written as

$$\left| \begin{matrix} h_{13} & h_{23} & h_{33} \\ & h_{12} & h_{22} \\ & & h_{11} \end{matrix} \right\rangle .$$

The weight of the above state is defined to be the set of numbers $(h_{11}, h_{12} + h_{22} - h_{11}, h_{13} + h_{23} + h_{33} - h_{12} - h_{22})$ which are actually the eigenvalues of G_{11}, G_{22}, and G_{33} operating on the state. The highest weight state or the generating state in the representation $[h_{13}, h_{23}, h_{33}]$ is the state

$$|h_{13} \quad h_{23} \quad h_{33}\rangle \equiv \left| \begin{matrix} h_{13} & h_{23} & h_{33} \\ & h_{13} & h_{23} \\ & & h_{13} \end{matrix} \right\rangle, \tag{8}$$

which has the property

$$G_{12}|h_{13} \quad h_{23} \quad h_{33}\rangle = G_{13}|h_{13} \quad h_{23} \quad h_{33}\rangle = G_{23}|h_{13} \quad h_{23} \quad h_{33}\rangle = 0. \tag{9}$$

A. The Ground State

According to the model described in Sec. II, in the absence of interaction an N-particle system will have

the ground state

$$|N\rangle = a_{11}^{\dagger}a_{21}^{\dagger}\cdots a_{N1}^{\dagger}|0\rangle, \tag{10}$$

where $|0\rangle$ is the vacuum state. By using the anti-commutation relations (2) and (3), one has

$$\begin{aligned} G_{11}|N\rangle &= \left(\sum_{m=1}^{N} a_{m1}^{\dagger}a_{m1}\right)a_{11}^{\dagger}a_{21}^{\dagger}\cdots a_{N1}^{\dagger}|0\rangle \\ &= \sum_{m=1}^{N} a_{11}^{\dagger}a_{21}^{\dagger}\cdots a_{m1}^{\dagger}(1 - a_{m1}^{\dagger}a_{m1})a_{m+1,1}^{\dagger}\cdots a_{N1}^{\dagger}|0\rangle \\ &= N|N\rangle. \end{aligned} \tag{11}$$

Similarly, one can show that

$$G_{22}|N\rangle = G_{33}|N\rangle = 0 \tag{12}$$

and

$$G_{12}|N\rangle = G_{13}|N\rangle = G_{23}|N\rangle = 0. \tag{13}$$

From (11) and (12), one thus concludes that the ground state has the weight $(N\ 0\ 0)$ and, therefore, can only belong to the representation $[N\ 0\ 0]$. From (13) one also sees that $|N\rangle$ is the generating state in that representation.

B. The Ground-State Band

Using the $SU(3) \supset SU(2) \supset SU(1)$ chain, one may represent a general state in $[N\ 0\ 0]$ as

$$\left|\begin{matrix} N & 0 & 0 \\ h_{12} & 0 & \\ h_{11} & & \end{matrix}\right\rangle,$$

which can be derived from the generating state

$$|N\rangle \equiv \left|\begin{matrix} N & 0 & 0 \\ N & 0 & \\ N & & \end{matrix}\right\rangle$$

by the lowering operators[8]

$$L_{21} = G_{21} \tag{14}$$

and

$$L_{31} = G_{31}(G_{11} - G_{22} + 1) + G_{21}G_{32}. \tag{15}$$

The commutation relation

$$[L_{31}, G_{12}] = 0 \tag{16}$$

is easily proved from Eq. (5). This means that L_{31}, when applied to a state of highest weight in the subgroup $SU(2)$, gives a new state also of highest weight in $SU(2)$. Thus, we have

$$\begin{aligned} \left|\begin{matrix} N & 0 & 0 \\ h_{12} & 0 & \end{matrix}\right\rangle &\equiv \left|\begin{matrix} N & 0 & 0 \\ h_{12} & 0 & \\ h_{12} & & \end{matrix}\right\rangle \\ &= \mathcal{N}_{h_{12}}(L_{31})^{N-h_{12}}|N\rangle, \end{aligned} \tag{17}$$

where $\mathcal{N}_{h_{12}}$ is a normalization factor

$$\mathcal{N}_{h_{12}} = \frac{(h_{12}+1)!}{(N+1)!}\left(\frac{h_{12}!}{N!\,(N-h_{12})!}\right)^{\frac{1}{2}}. \tag{18}$$

In this way, a general state is, therefore, given by the formula

$$\begin{aligned} \left|\begin{matrix} N & 0 & 0 \\ h_{12} & 0 & \\ h_{11} & & \end{matrix}\right\rangle &= \mathcal{N}_{h_{11}}^{h_{12}}(L_{21})^{h_{12}-h_{11}}\left|\begin{matrix} N & 0 & 0 \\ h_{12} & 0 & \end{matrix}\right\rangle \\ &= \mathcal{N}_{h_{11}}^{h_{12}}\mathcal{N}_{h_{12}}(L_{21})^{h_{12}-h_{11}}(L_{31})^{N-h_{12}}|N\rangle, \end{aligned} \tag{19}$$

with

$$\mathcal{N}_{h_{11}}^{h_{12}} = \left(\frac{h_{11}!}{(h_{12}-h_{11})!\,h_{12}!}\right)^{\frac{1}{2}}. \tag{20}$$

Evaluation of the commutator $[L_{21}, L_{31}]$ shows that it is not zero. Consequently, in Eq. (19) there are two noncommuting operators. However, by using the commutation relation (5) and the property of the ground state, one can put (19) into a form quite suitable for the start of a boson expansion.

From the equations starting with

$$\begin{aligned} L_{31}|N\rangle &= [G_{31}(G_{11} - G_{22} + 1) + G_{21}G_{32}]|N\rangle \\ &= G_{31}(N+1)|N\rangle, \end{aligned}$$

$$\begin{aligned} L_{31}^{2}|N\rangle &= [G_{31}(G_{11} - G_{22} + 1) + G_{21}G_{32}]L_{31}|N\rangle \\ &= [G_{31}(G_{11} - G_{22} + 1) + G_{21}G_{32}]G_{31}(N+1)|N\rangle \\ &= G_{31}^{2}(N+1)N|N\rangle, \end{aligned}$$

and continuing in a now obvious way, we have finally

$$\begin{aligned} L_{31}^{N-h_{12}}|N\rangle &= G_{31}^{N-h_{12}}(N+1)N\cdots(h_{12}+2)|N\rangle \\ &= \frac{(N+1)!}{(h_{12}+1)!}G_{31}^{N-h_{12}}|N\rangle. \end{aligned} \tag{21}$$

Thus,

$$\left|\begin{matrix} N & 0 & 0 \\ h_{12} & 0 & \\ h_{11} & & \end{matrix}\right\rangle = C(h_{11}, h_{12})(G_{21})^{h_{12}-h_{11}}(G_{31})^{N-h_{12}}|N\rangle, \tag{22}$$

with

$$\begin{aligned} C(h_{11}, h_{12}) &= \frac{(N+1)!}{(h_{12}+1)!}\mathcal{N}_{h_{11}}^{h_{12}}\mathcal{N}_{h_{12}} \\ &= \left(\frac{h_{11}!}{(h_{12}-h_{11})!\,(N-h_{12})!\,N!}\right)^{\frac{1}{2}}. \end{aligned} \tag{23}$$

We have thus reached an expression of the ground-state band in terms of the generating state and two commuting operators. Equation (22) will be the basis of a boson expansion.

IV. THE BOSON EXPANSION

A. Association with a Harmonic Spectrum

If we write a state in $[N \ 0 \ 0]$ as

$$|pq\rangle \equiv \left| \begin{matrix} N & 0 & 0 \\ & h_{12} & 0 \\ & & h_{11} \end{matrix} \right\rangle$$

with $p = h_{12} - h_{11}$ and $q = N - h_{12}$, then Eq. (21) becomes

$$|pq\rangle = c(p, q)(G_{12}^{\dagger})^p (G_{13}^{\dagger})^q |N\rangle, \tag{24}$$

where

$$c(p, q) = \left(\frac{(N - p - q)!}{N!\, p!\, q!}\right)^{\frac{1}{2}}. \tag{25}$$

Accordingly, the states belonging to $[N \ 0 \ 0]$ will have a one-to-one correspondence to a subset of 2-dimensional harmonic oscillator eigenstates, i.e., the set defined by

$$|pq)_B = \left(\frac{1}{p!\, q!}\right)^{\frac{1}{2}} A^{\dagger p} B^{\dagger q} |N)_B, \tag{26}$$

where A and B are two commuting boson operators

$$[A, A^{\dagger}] = [B, B^{\dagger}] = 1,$$
$$[A, B] = [A, B^{\dagger}] = 0,$$

and the sum of p and q in (26) is restricted by the inequality

$$0 \leq p + q \leq N.$$

B. The Purpose of Boson Expansions

The Hamiltonian in Eq. (7) is diagonal when we set $V_{12} = V_{13} = V_{23} = 0$. In other words, when all the interactions are turned off, the Hamiltonian has a pure harmonic spectrum, with eigenstates (24) or its image (26). Therefore, with this correspondence in mind, if we assume, for example, the expansion

$$G_{12}^{\dagger} = A^{\dagger} \sum_{i,j=0}^{\infty} d_{ij} A^{\dagger i} A^i B^{\dagger j} B^j,$$

along with similar expansions for the other generators $G_{\sigma\sigma'}$, and consequently for the Hamiltonian itself, then it can be expected that, at least for small interactions, these series will converge rapidly. Hence, one can cut off at low-order terms and, thus, greatly simplify the problem and at the same time have an accurate solution.

In the following a method due to Marumori[3,4] is used for getting the boson expansions of the various operators in the $SU(3)$ model.

C. Marumori's Method of Boson Expansion

We first introduce an image transformation U between states in the $SU(3)$ fermion space (24) and states in the boson space (26) such that

$$|pq\rangle_F = U^{\dagger} |pq\rangle_B, \quad p + q \leq N, \tag{27}$$

$$0 = U^{\dagger} |pq)_B, \quad p + q > N. \tag{28}$$

An operator in the $SU(3)$ space will then have an image in the boson space

$$\Theta_B = U \Theta_F U^{\dagger}. \tag{29}$$

We have to notice, however, that the operator U has no inverse and that the matrix element

$${}_B(pq| \, \Theta_B \, |p'q')_B$$

is not zero only for $p + q$ and $p' + q' \leq N$. Therefore, in a sense, we are dealing with the whole fermion space but only a subspace of the entire boson space.

From (27) and (28), one can construct U as

$$\begin{aligned} U &= \sum_{q=0}^{N} \sum_{p=0}^{N-q} |pq)_{B\ F}\langle pq| \\ &\equiv \sum_{(p,q)} |pq)_{B\ F}\langle pq| \\ &= \sum_{(p,q)} \frac{A^{\dagger p} B^{\dagger q}}{(p!\, q!)^{\frac{1}{2}}} |N)_{B\ F}\langle N| \, c(p, q) G_{13}^q G_{12}^p, \end{aligned} \tag{30}$$

and (29) becomes

$$\begin{aligned} \Theta_B &= U \Theta_F U^{\dagger} \\ &= \sum_{(p,q)(r,s)} \Theta_{pq,rs} \frac{A^{\dagger p} B^{\dagger q}}{(p!\, q!)^{\frac{1}{2}}} |N)_{B\ B}(N| \frac{B^s A^r}{(r!\, s!)^{\frac{1}{2}}}, \end{aligned} \tag{31}$$

with

$$\Theta_{pq,rs} \equiv \langle pq| \, \Theta_F \, |rs\rangle_F.$$

Next, we assume that the operator $|N)_{B\ B}(N|$ can be expressed in the form

$$|N)_{B\ B}(N| = \sum_{m,n=0}^{\infty} c_{mn} A^{\dagger m} A^m B^{\dagger n} B^n. \tag{32}$$

Substituting into the completeness relation

$$\sum_{p,q=0}^{\infty} \frac{A^{\dagger p} B^{\dagger q}}{(p!\, q!)^{\frac{1}{2}}} |N)_{B\ B}(N| \frac{B^q A^p}{(p!\, q!)^{\frac{1}{2}}} = 1$$

or, equivalently,

$$\begin{aligned} |N)_{B\ B}(N| = 1 &- \sum_{p=1}^{\infty} \frac{A^{\dagger p}}{(p!)^{\frac{1}{2}}} |N)_{B\ B}(N| \frac{A^p}{(p!)^{\frac{1}{2}}} \\ &- \sum_{q=1}^{\infty} \frac{B^{\dagger q}}{(q!)^{\frac{1}{2}}} |N)_{B\ B}(N| \frac{B^q}{(q!)^{\frac{1}{2}}} \\ &- \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \frac{A^{\dagger p} B^{\dagger q}}{(p!\, q!)^{\frac{1}{2}}} |N)_{B\ B}(N| \frac{B^q A^p}{(p!\, q!)^{\frac{1}{2}}} \end{aligned}$$

and equating the corresponding terms, we get

$$\text{(i)} \quad c_{00} = 1, \tag{33}$$

$$\text{(ii)} \quad \sum_{m=1}^{\infty} c_{m0} A^{\dagger m} A^m = -\sum_{p=1}^{\infty} \sum_{m=0}^{\infty} \frac{c_{m0}}{p!} (A^{\dagger})^{p+m} A^{p+m}.$$

Applying the transformation

$$u = p + m, \quad v = m \tag{34}$$

to the right-hand side, we get

$$-\sum_{u=1}^{\infty}\left[\sum_{v=0}^{u-1}\frac{c_{v0}}{(u-v)!}\right]A^{\dagger u}A^{u}.$$

A recurrence relation,

$$c_{m0} = -\sum_{v=0}^{m-1}\frac{c_{v0}}{(m-v)!}, \quad m \geq 1, \tag{35}$$

is thus obtained. We have

$$\text{(iii)}\ \sum_{n=1}^{\infty}c_{0n}B^{\dagger n}B^{n} = -\sum_{q=1}^{\infty}\sum_{n=0}^{\infty}\frac{c_{0n}}{q!}(B^{\dagger})^{q+n}B^{q+n}.$$

By a transformation similar to (34), we have

$$c_{0n} = -\sum_{v=0}^{n-1}\frac{c_{0v}}{(n-v)!}, \quad n \geq 1. \tag{36}$$

With

$$\begin{aligned}\text{(iv)}\ \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}c_{mn}A^{\dagger m}A^{m}B^{\dagger n}B^{n} &= -\sum_{p=1}^{\infty}\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}\frac{c_{mn}}{p!}(A^{\dagger})^{p+m}A^{p+m}B^{\dagger n}B^{n}\\ &\quad -\sum_{q=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=0}^{\infty}\frac{c_{mn}}{q!}A^{\dagger m}A^{m}(B^{\dagger})^{q+n}B^{q+n}\\ &\quad -\sum_{p=1}^{\infty}\sum_{q=1}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{c_{mn}}{p!\,q!}\\ &\quad \times (A^{\dagger})^{p+m}A^{p+m}(B^{\dagger})^{q+n}B^{q+n},\end{aligned}$$

and using the same transformation (34) in each term, we finally get

$$\begin{aligned}c_{mn} &= -\sum_{v=0}^{m-1}\frac{c_{vn}}{(m-v)!}\\ &\quad -\sum_{v=0}^{n-1}\frac{c_{mv}}{(n-v)!}\\ &\quad -\sum_{v=0}^{m-1}\sum_{v'=0}^{n-1}\frac{c_{vv'}}{(m-v)!\,(n-v')!}, \quad m,n \geq 1.\end{aligned} \tag{37}$$

Substituting Eq. (32) into (31) yields the basic equation for boson expansions,

$$\begin{aligned}\Theta_B = \sum_{(p,q)(r,s)}\sum_{m,n=0}^{\infty}&\frac{c_{mn}\Theta_{pq,rs}}{(p!\,q!\,r!\,s!)^{\frac{1}{2}}}\\ &\times (A^{\dagger})^{p+m}A^{r+m}(B^{\dagger})^{q+n}B^{s+n},\end{aligned} \tag{38}$$

in which the coefficients c_{mn} are given by Eqs. (33), (35), (36), and (37). Note that $c_{mn} = c_{nm}$.

Some of the values for c_{mn} are listed below:

$$\begin{aligned}&c_{00} = 1, \quad c_{01} = -1, \quad c_{02} = \tfrac{1}{2},\\ &\qquad\quad c_{10} = -1, \quad c_{20} = \tfrac{1}{2},\\ &c_{11} = 1, \quad c_{12} = c_{21} = -\tfrac{1}{2}.\end{aligned}$$

V. EXPANSIONS OF THE HAMILTONIAN

A. Matrix Elements of Operators in the Hamiltonian

Matrix elements of $SU(3)$ generators between states belonging to the $SU(3) \supset SU(2) \supset SU(1)$ chain have been calculated by Moshinsky.[8] They can be expressed in terms of the quantum numbers p and q as follows:

$${}_F\langle pq|\, G_{11}\,|rs\rangle_F = \delta_{pr}\delta_{qs}(N-p-q), \tag{39}$$

$${}_F\langle pq|\, G_{22}\,|rs\rangle_F = \delta_{pr}\delta_{qs}p, \tag{40}$$

$${}_F\langle pq|\, G_{33}\,|rs\rangle_F = \delta_{pr}\delta_{qs}q, \tag{41}$$

$${}_F\langle pq|\, G_{12}\,|rs\rangle_F = \delta_{p+1,r}\delta_{qs}[(p+1)(N-p-q)]^{\frac{1}{2}}, \tag{42}$$

$${}_F\langle pq|\, G_{21}\,|rs\rangle_F = \delta_{p-1,r}\delta_{qs}[p(N-p-q+1)]^{\frac{1}{2}}, \tag{43}$$

$${}_F\langle pq|\, G_{13}\,|rs\rangle_F = \delta_{pr}\delta_{q+1,s}[(q+1)(N-p-q)]^{\frac{1}{2}}, \tag{44}$$

$${}_F\langle pq|\, G_{31}\,|rs\rangle_F = \delta_{pr}\delta_{q-1,s}[q(N-p-q+1)]^{\frac{1}{2}}, \tag{45}$$

$${}_F\langle pq|\, G_{23}\,|rs\rangle_F = \delta_{p-1,r}\delta_{q+1,s}[p(q+1)]^{\frac{1}{2}}, \tag{46}$$

$${}_F\langle pq|\, G_{32}\,|rs\rangle_F = \delta_{p+1,r}\delta_{q-1,s}[(p+1)q]^{\frac{1}{2}}. \tag{47}$$

Using (39) in Eq. (38), one can then expand G_{11} as a series in the boson operators A and B:

$$\begin{aligned}(G_{11})_B = \sum_{p=0}^{N}\sum_{q=0}^{N-p}\sum_{m,n=0}^{\infty}&\frac{c_{mn}(N-p-q)}{p!\,q!}\\ &\times (A^{\dagger})^{p+m}A^{p+m}(B^{\dagger})^{q+n}B^{q+n}.\end{aligned} \tag{48}$$

Since ${}_B\langle pq|\, G_{11}\,|p'q'\rangle_B$ vanishes if either $p+q$ or $p'+q'$ exceeds N, all the terms in (48) with $p+q+m+n > N$ actually have no effect at all. For convenience of calculation, we can, therefore, extend all the summations in (48) to infinity, but at the same time keep a record of the above-mentioned restriction by using a prime on the summation symbol, as in the following equation:

$$\begin{aligned}(G_{11})_B = \sum_{p,q,m,n=0}^{\infty}{}'&\frac{c_{mn}(N-p-q)}{p!\,q!}\\ &\times (A^{\dagger})^{p+m}A^{p+m}(B^{\dagger})^{q+n}B^{q+n}.\end{aligned} \tag{49}$$

After applying Eq. (32) together with the completeness relation in boson space, the above equation reduces to a simple form

$$(G_{11})_B = N - A^{\dagger}A - B^{\dagger}B. \tag{50}$$

By the same method, we also find

$$(G_{22})_B = A^{\dagger}A, \tag{51}$$

$$(G_{33})_B = B^{\dagger}B. \tag{52}$$

Next, the matrix element for G_{12}^2 can be derived from (42):

$$
\begin{aligned}
&{}_F\langle pq|\, G_{12}^2\, |rs\rangle_F \\
&= [(p+2)(p+1)(N-p-q)(N-p-q-1)]^{\frac{1}{2}} \\
&\quad \times \delta_{p+2,r}\delta_{qs}.
\end{aligned}
$$

Again substituting into (38), we get

$$
\begin{aligned}
(G_{12}^2)_B &= \sum_{p,q,m,n=0}^{\infty}{}' \frac{c_{mn}[(N-p-q)(N-p-q-1)]^{\frac{1}{2}}}{p!\,q!} \\
&\quad \times (A^\dagger)^{p+m} A^{p+m+2} (B^\dagger)^{q+n} B^{q+n} \\
&= \left(\sum_{i,j=0}^{\infty}{}' f_{ij} A^{\dagger i} A^i B^{\dagger j} B^j \right) A^2, \qquad (53)
\end{aligned}
$$

where

$$
f_{ij} = \sum_{m=0}^{i} \sum_{n=0}^{j} \left(\frac{c_{mn}(N+m+n-i-j)(N+m+n-i-j-1)}{(i-m)!\,(j-n)!} \right)^{\frac{1}{2}}
$$

and the first few values of f_{ij} are

$$
\begin{aligned}
f_{00} &= [N(N-1)]^{\frac{1}{2}}, \\
f_{01} &= [(N-1)(N-2)]^{\frac{1}{2}} - [N(N-1)]^{\frac{1}{2}} = f_{10}, \\
f_{02} &= \tfrac{1}{2}[(N-2)(N-3)]^{\frac{1}{2}} - [(N-1)(N-2)]^{\frac{1}{2}} \\
&\quad + \tfrac{1}{2}[N(N-1)]^{\frac{1}{2}} = f_{20}, \\
f_{11} &= [(N-2)(N-3)]^{\frac{1}{2}} - 2[(N-2)(N-2)]^{\frac{1}{2}} \\
&\quad + [N(N-1)]^{\frac{1}{2}}.
\end{aligned}
$$

Similarly, we obtain the expansion for G_{13}^2,

$$
(G_{13}^2)_B = \left(\sum_{i,j=0}^{\infty}{}' f_{ij} A^{\dagger i} A^i B^{\dagger j} B^j \right) B^2, \qquad (54)
$$

and G_{23}^2, in which case the series can be summed up to give a closed form

$$
(G_{23}^2)_B = A^{\dagger 2} B^2. \qquad (55)
$$

Also expansions for the conjugate operators are found to be

$$
(G_{21}^2)_B = (G_{12}^{\dagger 2})_B = (G_{12}^2)_B^\dagger. \qquad (56)
$$

B. The Hamiltonian as a Boson Expansion

From Eq. (7) we have, after utilizing Eqs. (50)–(56), the transformed Hamiltonian in the boson space

$$
\begin{aligned}
H_B &= \epsilon_1(N - A^\dagger A - B^\dagger B) + \epsilon_2 A^\dagger A + \epsilon_3 B^\dagger B \\
&\quad + (V_{12}A^{\dagger 2} + V_{13}B^{\dagger 2}) \left(\sum_{i,j}{}' f_{ij} A^{\dagger i} A^i B^{\dagger j} B^j \right) \\
&\quad + \left(\sum_{i,j}{}' f_{ij} A^{\dagger i} A^i B^{\dagger j} B^j \right) (V_{12}A^2 + V_{13}B^2) \\
&\quad + V_{23}(A^{\dagger 2}B^2 + A^2 B^{\dagger 2}). \qquad (57)
\end{aligned}
$$

In practice, it is only possible for one to calculate a few leading terms in the above expansion and compare the result with the exact solution of Eq. (7). We shall examine the extent to which the simplest approximations agree with the exact solution as a function of the relevant coupling parameters.

C. The Harmonic Approximation

If we include terms in (57) up to the second order of the operators A and B, then we have the harmonic approximation

$$
\begin{aligned}
H_B^{(2)} &= \epsilon_1 N + (\epsilon_2 - \epsilon_1) A^\dagger A + (\epsilon_3 - \epsilon_1) B^\dagger B \\
&\quad + [N(N-1)]^{\frac{1}{2}} V_{12}(A^{\dagger 2} + A^2) + V_{13}(B^{\dagger 2} + B^2). \qquad (58)
\end{aligned}
$$

It is always possible to set $\epsilon_1 = 0$ and $\epsilon_2 - \epsilon_1 = 1$, the latter corresponding to measuring the energy in units of $\epsilon_2 - \epsilon_1$. Using the following parameters,

$$
e \equiv (\epsilon_3 - \epsilon_1)/(\epsilon_2 - \epsilon_1) = \epsilon_3, \quad e > 1, \qquad (59)
$$

$$
x \equiv NV_{12}/(\epsilon_2 - \epsilon_1) = NV_{12}, \qquad (60)
$$

$$
y \equiv NV_{13}/(\epsilon_3 - \epsilon_1) = NV_{13}/e, \qquad (61)
$$

$$
M_2 = x[1 - (1/N)]^{\frac{1}{2}}, \qquad (62)
$$

$$
M_3 = y[1 - (1/N)]^{\frac{1}{2}}, \qquad (63)
$$

in Eq. (58) gives

$$
\begin{aligned}
H_B^{(2)} &= A^\dagger A + M_2(A^{\dagger 2} + A^2) \\
&\quad + e[B^\dagger B + M_3(B^{\dagger 2} + B^2)]. \qquad (64)
\end{aligned}
$$

The two operators A and B do not couple together in the harmonic approximation. Eigenvalues of (64) can be obtained either by two separate canonical transformations or by a straightforward diagonalization process.

D. The Quartic Approximation

Including the fourth-order terms in A and B gives us the quartic approximation

$$
\begin{aligned}
H_B^{(4)} &= A^\dagger A + M_2(A^{\dagger 2} + A^2) + L_2(A^{\dagger 3}A + A^\dagger A^3) \\
&\quad + e\{B^\dagger B + M_3(B^{\dagger 2} + B^2) + L_3(B^{\dagger 3}B + B^\dagger B^3)\} \\
&\quad + L_2(A^{\dagger 2}B^\dagger B + B^\dagger B A^2) \\
&\quad + L_3 e(B^{\dagger 2}A^\dagger A + A^\dagger A B^2) \\
&\quad + \tfrac{1}{2}(K_2 + eK_3)(A^{\dagger 2}B^2 + B^{\dagger 2}A^2), \qquad (65)
\end{aligned}
$$

in which the parameters z, L_2, etc.,are defined by

$$z = NV_{23}/(\epsilon_2 - \epsilon_1) = NV_{23},$$
$$L_2 \equiv M_2[1 - (2/N) - 1]^{\frac{1}{2}},$$
$$L_3 \equiv M_3[1 - (2/N) - 1]^{\frac{1}{2}},$$
$$K_2 \equiv (z/N) = V_{23},$$
$$K_3 \equiv (z/Ne) = (V_{23}/e),$$

along with the definitions (59)–(63).

In (65), A and B are coupled together in several terms. Therefore, the best way to get eigenvalues is to diagonalize (65) directly.

E. The Expanded Quartic Approximation

It may be interesting to calculate the coefficients in (65) up to the order of $1/N$ and then compute the eigenvalues. This is because in some realistic cases it may not be possible to calculate these coefficients exactly. One can probably only expand them in powers of $1/N$ and take a few leading terms. Thus, one will have the expanded quartic approximation which is essentially the same as Eq. (65) except that all the coefficients are computed to the order $1/N$:

$$\begin{aligned} H_{BE}^{(4)} = {} & A^\dagger A + x(1 - 1/2N)(A^{\dagger 2} + A^2) \\ & - (2/N)x(A^{\dagger 3}A + A^\dagger A^3) \\ & + e\{B^\dagger B + y(1 - 1/2N)(B^{\dagger 2} + B^2) \\ & - (2/N)y(B^{\dagger 3}B + B^\dagger B^3)\} \\ & - (2/N)(A^{\dagger 2}B^\dagger B + B^\dagger BA^2) \\ & - (2e/N)(B^{\dagger 2}A^\dagger A + A^\dagger AB^2) \\ & + (z/N)(A^{\dagger 2}B^2 + B^{\dagger 2}A^2). \end{aligned} \tag{66}$$

VI. THE PHYSICAL BOSON

It will be seen later that the boson expansion series in Sec. V actually turn out to be very good expansions. We shall call the boson operators A and B in Eq. (26) the "kinematical" bosons, since their definitions have nothing to do with the dynamics of the Hamiltonian. In fact, with these bosons the quartic approximation (65) already gives almost exact energy eigenvalues over an extremely wide range of interactions.

In this section, we want to introduce by contrast the concept of "physical" boson and describe some of its properties. It should be emphasized that the reason for the name is that, if our model corresponded to an actual physical system, then this is the boson which we would introduce phenomenologically if the experimental low-lying spectrum is nearly harmonic. The conditions for the validity of such a description will emerge from our results.

A. The Definition of Physical Boson

We notice that when all interactions in the Hamiltonian (7) are turned off, i.e., when $V_{12} = V_{13} = V_{23} = 0$, the exact energy eigenstates are just the states $|pq\rangle$ in (24), and they have a one-to-one correspondence to the harmonic basis $|pq)_B$ in (26). As interactions are gradually turned on, this kind of association will still be possible. Therefore, one can label the exact eigenstates as $|pq\rangle_p$ with the two quantum numbers p and q. The physical-boson operators α and β are now defined by the following equations:

$$|pq\rangle_p = (p!\,q!)^{-\frac{1}{2}}\alpha^{\dagger p}\beta^{\dagger q}|00\rangle_p, \tag{67}$$
$$[\alpha, \alpha^\dagger] = [\beta, \beta^\dagger] = 1, \tag{68}$$
$$[\alpha, \beta] = [\alpha^\dagger, \beta] = 0. \tag{69}$$

As we have mentioned, the physical-boson operators are identical with those kinematical-boson operators in the limit when $V_{12} = V_{13} = V_{23} = 0$. Consequently, the deviation between the physical and the kinematical bosons, as interactions are turned on, is a measure of the anharmonicity of the spectrum.

B. The Hamiltonian

Since the states $|pq\rangle_p$ are exact energy eigenstates, the Hamiltonian is diagonal in that basis. The most general form of the Hamiltonian, in terms of α and β, must then be

$$\begin{aligned} H = {} & h_{00} + (h_{10}\alpha^\dagger\alpha + h_{01}\beta^\dagger\beta) \\ & + (h_{20}\alpha^{\dagger 2}\alpha^2 + h_{11}\alpha^\dagger\alpha\beta^\dagger\beta + h_{02}\beta^{\dagger 2}\beta^2) \\ & + \cdots + (h_{N0}\alpha^{\dagger N}\alpha^N + h_{N-1,1} \\ & \times \alpha^{\dagger N-1}\alpha^{N-1}\beta\beta + \cdots + h_{0N}\beta^{\dagger N}\beta^N) \\ = {} & \sum_{u=0}^{N}\sum_{v=0}^{N-u} h_{uv}\alpha^{\dagger u}\alpha^u\beta^{\dagger v}\beta^v. \end{aligned} \tag{70}$$

There are exactly $\frac{1}{2}(N+1)(N+2)$ constants h_{uv} to fit the $\frac{1}{2}(N+1)(N+2)$ energy levels. An exactly harmonic spectrum would give the result $h_{uv} = 0$ for all $u + v > 1$; hence, for a nearly harmonic spectrum we will expect that

$$h_{10}, h_{01} \gg h_{20}, h_{11}, h_{02} \gg \cdots \gg h_{N0}, h_{N-1,1} \cdots h_{0N}. \tag{71}$$

C. The Transition Operators

Because of the form of the model Hamiltonian (7), the energy eigenstates are always separated into four independent subsets such that the Hamiltonian itself has no matrix element between two eigenstates belonging to two different subsets. These four subsets are connected to each other by the three operators G_{12}, G_{13}, and G_{23}, together with their Hermitian conjugates. If we define transition operators T_1, T_2, and

T_3 as

$$T_1 \equiv \tfrac{1}{2}(G_{12} + G_{21}), \tag{72}$$

$$T_2 \equiv \tfrac{1}{2}(G_{13} + G_{31}), \tag{73}$$

$$T_3 \equiv \tfrac{1}{2}(G_{23} + G_{32}), \tag{74}$$

then their kinematical boson expansions are readily obtained from Eqs. (38), (42), and (44). In terms of physical bosons, T_1, for example, has the form

$$\begin{aligned} T_1 = {} & \alpha^\dagger[\xi^{(1)}_{00} + (\xi^{(1)}_{10}\alpha^\dagger\alpha + \xi^{(1)}_{01}\beta^\dagger\beta) \\ & + (\xi^{(1)}_{20}\alpha^{\dagger 2}\alpha^2 + \xi^{(1)}_{11}\alpha^\dagger\alpha\beta^\dagger\beta + \xi^{(1)}_{02}\beta^{\dagger 2}\beta^2) \\ & + \cdots + (\xi^{(1)}_{N-1,0}\alpha^{\dagger N-1}\alpha^{N-1} + \xi^{(1)}_{N-2,1}\alpha^{\dagger N-2}\alpha^{N-2}\beta^\dagger\beta \\ & + \cdots + \xi^{(1)}_{0,N-1}\beta^{\dagger N-1}\beta^{N-1})] \\ & + \alpha^{\dagger 3}[\xi^{(3)}_{00} + (\xi^{(3)}_{10}\alpha^\dagger\alpha + \xi^{(3)}_{01}\beta^\dagger\beta) \\ & + \cdots + (\xi^{(3)}_{N-3,0}\alpha^{\dagger N-3}\alpha^{N-3} \\ & + \cdots + \xi^{(3)}_{0,N-3}\beta^{\dagger N-3}\beta^{N-3})] \\ & + \cdots + (\alpha^\dagger)^{2[\frac{1}{2}N]+1}\xi^{(2[\frac{1}{2}N]+1)}_{00} + \text{h.c.} \\ \equiv {} & \sum_{j=0}^{[\frac{1}{2}N]}[(\alpha^\dagger)^{2j+1}X^{(j)} + X^{(j)*}\alpha^{2j+1}], \end{aligned} \tag{75}$$

with

$$X^{(j)} \equiv \sum_{u=0}^{N-(2j+1)} \sum_{v=0}^{N-(2j+1)-u} \xi^{(j)}_{uv}\alpha^{\dagger u}\alpha^u B^{\dagger v}B^v, \tag{76}$$

where $\xi^{(j)}_{uv}$ are constants and $[\frac{1}{2}N]$ is the largest integer *less than* $\frac{1}{2}N$.

Two energy eigenstates differing by an even number of α-quanta will belong to the same subset mentioned before and are not connected by T_1. This explains why there are no terms like $(\alpha^\dagger)^{2j}$ in the expansion (75).

The constants $\xi^{(j)}_{uv}$ are determined by fitting the physical transition amplitudes, e.g.,

$$\begin{aligned} \xi^{(1)}_{00} &= {}_p\langle 10|\, T_1\, |00\rangle_p, \\ \xi^{(1)}_{10} &= {}_p\langle 20|\, T_1\, |10\rangle_p - \xi^{(1)}_{00}, \\ \xi^{(3)}_{00} &= {}_p\langle 30|\, T_1\, |00\rangle_p/\sqrt{6}, \\ \xi^{(3)}_{10} &= ({}_p\langle 40|\, T_1\, |10\rangle_p/\sqrt{6}) - \xi^{(3)}_{00}, \end{aligned} \tag{77}$$

etc. The number of constants is just the same as the number of physical transition amplitudes. In this way we have, as in Eq. (70), a theoretically exact fit.

For the sake of completeness, we will write down the physical boson expansions for T_2 and T_3 as well. The meaning of these equations should be self-evident:

$$T_2 = \sum_{j=0}^{[\frac{1}{2}N]}[(\beta^\dagger)^{2j+1}Y^{(j)} + Y^{(j)*}\beta^{2j+1}], \tag{78}$$

$$Y^{(j)} \equiv \sum_{u=0}^{N-(2j+1)} \sum_{v=0}^{N-(2j+1)-u} \eta^{(j)}_{uv}\beta^{\dagger u}\beta^u\alpha^{\dagger v}\alpha^v, \tag{79}$$

$$\begin{aligned} T_3 = {} & \sum_{i=0}^{[\frac{1}{2}N]} \sum_{j=0}^{[\frac{1}{2}N]-i} [(\alpha^\dagger)^{2i+1}Z^{(ij)}\beta^{2j+1} \\ & + (\beta^\dagger)^{2j+1}Z^{(ij)*}\alpha^{2i+1}], \end{aligned} \tag{80}$$

$$Z^{(ij)} = \sum_{u=0}^{N-(2i+1)-(2j+1)} \sum_{v=0}^{N-(2i+1)-(2j+1)-u} \xi^{(ij)}_{uv}\alpha^{\dagger u}\alpha^u\beta^{\dagger v}\beta^v. \tag{81}$$

VII. DISCUSSION OF RESULTS

We have developed both kinematical- and physical-boson representations for the $SU(3)$ model. We shall here describe some of the numerical results obtained from the application of these forms to the Hamiltonian (7).

In Sec. IV.C we have noted that the exact energy eigenstates always separate into four independent subsets, namely,

$$\begin{aligned} &\text{I:} && \{|pq\rangle_p;\ p \text{ even},\ q \text{ even}\}, \\ &\text{II:} && \{|pq\rangle_p;\ p \text{ odd},\ q \text{ even}\}, \\ &\text{III:} && \{|pq\rangle_p;\ p \text{ even},\ q \text{ odd}\}, \\ &\text{IV:} && \{|pq\rangle_p;\ p \text{ odd},\ q \text{ odd}\}. \end{aligned}$$

As an example, Fig. 1 shows the $SU(3)$ states $|pq\rangle$ divided into such four subsets, which are eigenstates of the unperturbed Hamiltonian

$$H = \sum_\sigma \epsilon_\sigma G_{\sigma\sigma},$$

with $e = (\epsilon_3 - \epsilon_1)/(\epsilon_2 - \epsilon_1) = 2.5$ and $N = 6$. Here N reminds us of the fact that these states belong to the $SU(3)$ representation $[N\ \ 0\ \ 0]$. Otherwise, one would not be able to describe these states by using only two quantum numbers p and q.

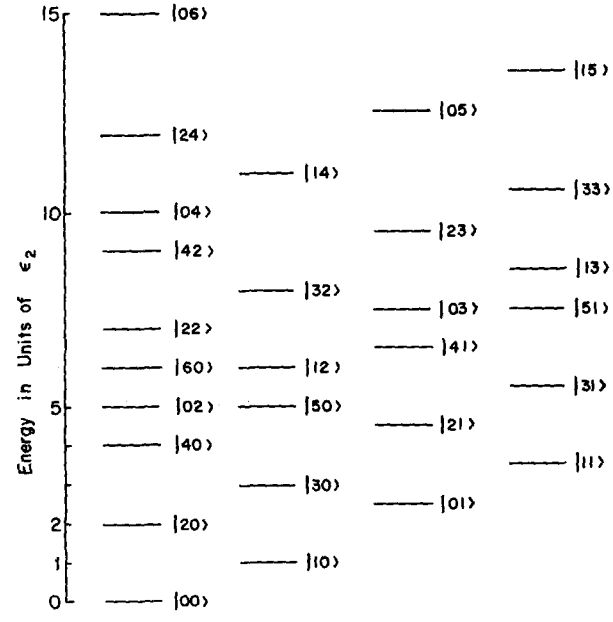


FIG. 1. Level scheme of the zero-order Hamiltonian for $N = 6$, $e = 2.5$.

TABLE I. Exact and approximate energy values for a subset of eigenstates $|p0\rangle$ with $p = 0, 1, \cdots, N$, which correspond to MGL eigenstates.

(a) $N = 6, y = z = 0$				(b) $N = 14, y = z = 0$			
$x = 0.05$				$x = 0.05$			
Exact solution	Harmonic approx.	Expanded quartic approx.	Quartic approx.	Exact solution	Harmonic approx.	Expanded quartic approx.	Quartic approx.
6.00208	5.98956	6.00234	6.00221	14.00232	13.96975	14.00261	14.00255
5.00416	4.99373	5.00433	5.00420	13.00597	12.98371	13.00621	13.00614
4.00290	3.99791	4.00275	4.00279	12.00778	11.99768	12.00769	12.00766
3.00000	2.99373	2.99987	2.99995	11.00804	10.99301	11.00798	11.00796
1.99709	1.98956	1.99700	1.99708	10.00710	9.99768	10.00703	10.00704
0.99583	0.99373	0.99578	0.99583	9.00524	8.99301	9.00518	9.00519
0.00208	−0.00208	−0.00210	−0.00208	8.00277	7.99767	8.00272	8.00274
$x = 0.25$				7.00000	6.99301	6.99995	6.99998
				5.99722	5.98836	5.99718	5.99721
6.05233	5.72439	6.05872	6.05545	4.99476	4.98371	4.99472	4.99475
5.10158	4.83463	5.10564	5.10257	3.99289	3.97905	3.99287	3.99289
4.06845	3.94487	4.06462	4.06561	2.99195	2.98371	2.99193	2.99195
3.00000	2.75498	2.99620	2.99867	1.99222	1.98836	1.99220	1.99222
1.93155	1.72439	1.92944	1.93127	0.99402	0.99301	0.99401	0.99402
0.89841	0.83463	0.89710	0.89838	−0.00232	−0.00232	−0.00232	−0.00232
−0.05233	−0.05512	−0.05279	−0.05233				

A. Reduction to $SU(2)$ Model

The $SU(3)$ model reduces to the $SU(2)$ model of MGL[6] if we set $V_{13} = V_{23} = 0$ in Eq. (7). Among the energy eigenstates $|pq\rangle_p$, a subset $|p0\rangle_p$ with $p = 0, 1, \cdots, N$ will correspond to the energy eigenstates of this model. Table I(a) gives the exact energy eigenvalues together with the harmonic approximation [Eq. (64)], the expanded quartic approximation [Eq. (66)], and the quartic-approximation [Eq. (65)] eigenvalues of these states for $N = 6$, $x = 0.05$ and 0.25 (corresponding to $\delta = 0.1$ and 0.5, respectively, in Ref. 5), and $y = z = 0$. Table I(b) gives the same results for $N = 14$, $x = 0.05$, and $y = z = 0$. They are in agreement with the results in Ref. 5, except that here we find more accurate quartic and expanded quartic results than those in Ref. 5. This is because we have obtained these values from Eqs. (65) and (66) through direct diagonalization rather than by first making a canonical transformation. The latter procedure introduces numerical errors associated with the incorrect use of the same vector space for the transformed boson as for the kinematical boson. In our procedure, the vector space of the kinematical boson is a simple image of the space of the $SU(3)$ representation. (This is the expression of the Pauli principle.) Further canonical transformation in the boson space "distorts the shape" of this vector space. If we continue to use the shape appropriate to the kinematical boson in subsequent diagonalization procedures, we introduce errors which become appreciable, particularly for the high-lying energy levels.

TABLE II. Energy spectra for $e = 2.5$, $N = 6$, and $x = y = z = \frac{1}{8}N^3 = 108$.

Exact solution	Expanded quartic approx.	Quartic approx.
776.55	791.17	781.72
753.91	763.51	755.77
603.55	619.18	610.20
601.70	609.46	603.33
457.82	470.07	463.11
353.61	359.51	355.03
297.45	301.71	299.31
247.51	250.85	248.68
245.32	250.73	246.65
116.36	123.34	119.56
103.13	105.81	104.43
98.306	97.016	97.837
82.371	83.140	82.838
49.031	48.541	48.847
−100.91	−109.13	−105.09
−118.98	−119.48	−119.11
−150.48	−150.87	−150.74
−150.70	−155.17	−152.80
−194.50	−198.31	−196.44
−276.44	−280.67	−278.01
−281.65	−285.51	−283.31
−286.06	−290.52	−287.05
−304.54	−309.74	−305.77
−383.90	−393.05	−387.70
−472.86	−481.81	−476.03
−474.57	−486.15	−478.20
−690.02	−699.34	−691.78
−704.81	−718.27	−709.31

TABLE III. A comparison between the eigenvalues of the exact Hamiltonian and the physical-boson expansion for $e = 2.5$, $N = 6$.

(a) $x = y = z = 0.1$; the values of the parameters of Eq. (82) are $h_{00} = -0.02885$, $h_{10} = 0.9990$, $h_{01} = 2.4824$, $h_{20} = 0.00545$, $h_{11} = 0.01363$, $h_{02} = 0.01644$.		(b) $x = y = z = 0.5$; the values of the parameters of Eq. (82) are $h_{00} = -0.6710$, $h_{10} = 1.0422$, $h_{01} = 2.2285$, $h_{20} = 0.04015$, $h_{11} = 0.21403$, $h_{02} = 0.32315$.	
Exact solution	Physical-boson expansion	Exact solution	Physical-boson expansion
15.027	15.359	15.670	22.394
13.525	13.779	14.122	19.047
12.545	12.712	13.555	16.934
12.017	12.216	12.403	15.998
11.032	11.152	11.772	14.019
10.505	10.669	10.658	13.247
10.031	10.098	10.630	12.121
9.5170	9.6077	9.9129	11.402
8.9962	9.1393	8.8924	10.795
8.5136	8.5569	8.8526	9.6376
8.0051	8.0803	8.0902	9.0839
7.5000	7.6257	7.5450	8.6416
7.4936	7.5171	7.2784	7.9533
6.9994	7.0323	7.0271	7.4531
6.5016	6.5694	6.4064	7.0642
6.0023	6.1286	6.0045	6.7867
5.9862	5.9951	5.8020	5.9025
5.4958	5.5241	5.3028	5.5670
5.0119	5.0751	5.1589	5.3430
4.9689	4.9689	4.4322	4.4322
4.4821	4.4897	4.0872	4.1502
4.0071	4.0325	4.0270	3.9796
3.4662	3.4662	2.8137	2.8137
2.9944	3.0008	2.8057	2.6965
2.4536	2.4536	1.5575	1.5575
1.9800	1.9800	1.4937	1.4937
0.97014	0.97014	0.37117	0.37117
−0.028845	−0.028845	−0.67103	−0.67103

B. The Kinematical-Boson Expansions

For the $SU(3)$ model, the kinematical-boson representations again prove to be remarkably good expansions. The harmonic approximation, equivalent to the random phase approximation, breaks down for $\max(|x|, |y|, |z|) \gtrsim 0.5$. But, because the Hamiltonians (65) and (66) are diagonalized directly, the expanded quartic approximation turns out to be a good approximation even for larger interactions, contrary to the statements in Ref. 5. The result shows that within the range $\max(|x|, |y|, |z|) \leqslant \frac{1}{2}N^3$, for $N = 6$, the discrepancy between quartic approximation and exact solution is less than 0.7%, and that between expanded quartic approximation and exact solution is less than

TABLE IV. A comparison between some exact values and the corresponding physical-boson approximations of the transition amplitudes for $e = 2.5$, $N = 14$.

	(a) $x = y = z = 0.1$; the parameters of Eq. (84) are $\xi^{(1)}_{00} = 3.4079$, $\xi^{(0)}_{11} = -0.07711$, $\xi^{(1)}_{10} = -6.7199$, $\xi^{(3)}_{00} = 0.00283$, $\xi^{(3)}_{10} = 0.0019468$, $\xi^{(3)}_{01} = -0.0006029$.		(b) $x = y = z = 0.5$; the parameters of Eq. (84) are $\xi^{(1)}_{00} = 2.3220$, $\xi^{(1)}_{10} = -2.3230$, $\xi^{(1)}_{01} = -6.7480$, $\xi^{(3)}_{00} = 0.000119$, $\xi^{(3)}_{10} = -0.000176$, $\xi^{(3)}_{01} = -0.27783$.	
	Exact value	Physical-boson approx.	Exact value	Physical-boson approx.
$\lvert_p\langle 30 \rvert T_1 \lvert 20\rangle_p\rvert$	5.620	5.635	2.553	4.025
$\lvert_p\langle 40 \rvert T_1 \lvert 30\rangle_p\rvert$	6.301	6.353	3.608	9.294
$\lvert_p\langle 21 \rvert T_1 \lvert 11\rangle_p\rvert$	4.562	4.793	2.736	9.544
$\lvert_p\langle 41 \rvert T_1 \lvert 11\rangle_p\rvert$	0.00909	0.0204	0.697	1.361

3%. For larger N the agreement is even better. Table II shows the result for $N = 6$, $e = 2.5$, and $x = y = z = \frac{1}{2}N^3 = 108$. The rapid convergence of the quartic approximation even for exceedingly large coupling constant was initially a complete surprise, since it can be shown that,in contrast to the weak coupling limit, all terms in the series for generators and Hamiltonian are of the same order in N. The reason for the convergence is that we encounter *alternating* series in the essential places. A detailed discussion of this point will be given in connection with a separate study of the strong coupling limit.

C. The Physical-Boson Expansion

To see whether the exact energy spectrum is a nearly harmonic one, we take the first six terms in Eq. (70):

$$H_p = h_{00} + (h_{10}\alpha^\dagger\alpha + h_{01}\beta^\dagger\beta) + (h_{20}\alpha^{\dagger 2}\alpha^2 + h_{11}\alpha^\dagger\alpha\beta^\dagger\beta + h_{02}\beta^{\dagger 2}\beta^2). \quad (82)$$

The lowest levels in the exact energy spectrum are used to get the six constants in (82). For a nearly harmonic spectrum we expect

$$|h_{20}|, |h_{11}|, |h_{02}| \lesssim h_{10}/N, h_{01}/N. \quad (83)$$

Next, the entire spectrum of H_p is calculated using these values in (82). It is found that, for max $(|x|, |y|, |z|) \lesssim 0.5$, Eq. (82) gives fairly good results, but somewhere in the domain $0.5 \lesssim \max(|x|, |y|, |z|) \lesssim 1$, i.e., past the point where RPA breaks down, the condition (83) breaks down and Eq. (82) is no longer meaningful. Results for $N = 6$, $e = 2.5$, with $x = y = z = 0.1$ and 0.5, are given in Tables III(a) and III(b), respectively.

A similar situation is found for the transition amplitude when we use the approximation

$$T_1 \simeq \alpha^\dagger[\xi_{00}^{(1)} + (\xi_{10}^{(1)}\alpha^\dagger\alpha + \xi_{01}^{(1)}\beta^\dagger\beta)] + \alpha^{\dagger 3}[\xi_{00}^{(3)} + (\xi_{10}^{(3)}\alpha^\dagger\alpha + \xi_{01}^{(3)}\beta^\dagger\beta)] + \text{h.c.} \quad (84)$$

with a criterion similar to (83). Tables IV(a) and IV(b) list the values of the six constants in (84) as well as some of the transition amplitudes under the conditions $N = 14$, $e = 2.5$, with $x = y = z = 0.1$ and 0.5. Also, the exact values of those transition amplitudes are given for comparison. They are calculated by using the formula

$$\begin{aligned} {}_p\langle pq|\, T_1\, |p'q'\rangle_p &= \sum_{u=0}^{N}\sum_{v=0}^{N-u}\sum_{u'=0}^{N}\sum_{v'=0}^{N-u'} \langle uv|\, \tfrac{1}{2}(G_{12}+G_{21})\, |u'v'\rangle\, c_{uv}^{(pq)*} c_{u'v'}^{(p'q')} \\ &= \tfrac{1}{2}\sum_{(u,v)} c_{uv}^{(pq)*}\{c_{u+1,v}^{(p'q')}[(u+1)(N-u-v)]^{\frac{1}{2}} \\ &\quad + c_{u-1,v}^{(p'q')}[u(N-u-v+1)]^{\frac{1}{2}}\} \\ &= \tfrac{1}{2}\sum_{(u,v)} (c_{uv}^{(pq)*} c_{u-1,v}^{(p'q')} + c_{u-1,v}^{(pq)*} c_{uv}^{(p'q')}) \\ &\quad \times [u(N-u-v+1)]^{\frac{1}{2}}, \end{aligned} \quad (85)$$

where $|pq\rangle_p$ are exact eigenstates, $|uv\rangle$ are $SU(3)$ states, and

$$|pq\rangle_p = \sum_{u=0}^{N}\sum_{v=0}^{N-u} c_{uv}^{(pq)}\, |uv\rangle, \quad (86)$$

with $c_{uv}^{(pq)}$ obtained through diagonalization of the exact Hamiltonian (7) in the $SU(3)$ basis.

In summary, we have studied two boson representations of an exactly soluble model. The kinematical boson, most closely tied to the group-theoretical structure, has no dynamical significance except in the limit of vanishing coupling; it provides, nonetheless, excellent dynamical approximations when used beyond the leading order. On the other hand, the physical-boson representation has been seen to represent a possible physical interpretation of the model when the coupling constants are not too large, approximately below the limiting values for validity of the random phase approximation. This is the region of small anharmonicity in which the physical boson clearly represents a vibration degree of freedom. Detailed study and physical interpretation of the strong coupling limit will be considered in a separate paper.

ACKNOWLEDGMENT

We are grateful to Dr. S. C. Pang for enlightening discussions.

* Supported in part by the U.S. Atomic Energy Commission.

[1] S. Beliaev and V. Zelevinsky, Nucl. Phys. **39**, 582 (1962).

[2] B. Sørensen, Nucl. Phys. **A97**, 1 (1967); Progr. Theoret. Phys. (Kyoto) **39**, 1468 (1968); Nucl. Phys. **A119**, 65 (1968).

[3] T. Marumori *et al.*, Progr. Theoret. Phys. (Kyoto) **31**, 1009 (1964).

[4] T. Marumori *et al.*, Progr. Theoret. Phys (Kyoto) **32**, 726 (1964).

[5] S. C. Pang, A. Klein, and R. M. Dreizler, Ann. Phys. **49**, 477 (1968).

[6] H. Lipkin, N. Meshkov, and A. Glick, Nucl. Phys. **62**, 118 (1965).

[7] The general class of models was suggested to us by Dr. N. Meshkov, who has studied different approximation methods with a different choice of Hamiltonian. We are indebted to Dr. Meshkov for informing us of her work.

[8] M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, Science Publ., New York, 1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Trajectory of Charged Particles in Weakly Inhomogeneous Magnetic Fields

DAVID M. COOK
Department of Physics, Lawrence University, Appleton, Wisconsin 54911*

(Received 18 July 1969)

The detailed time dependence of the trajectory, and particularly of the acceleration, of a charged particle moving nonrelativistically in a weakly inhomogeneous, slowly varying electromagnetic field is calculated by a method that involves superimposing a conventional perturbation expansion in the weak field on the expansion of Berkowitz and Gardner and of Kruskal, the latter expansion being asymptotically valid in the limit of small Larmor radius and slowly varying fields. Fourier transforms are introduced to facilitate isolation of rapidly varying terms from the most general solutions for the coefficients in the expansion. The calculations are carried to first order in the weak inhomogeneity and (at least formally) to all orders in Larmor radius. The resulting velocity reproduces the familiar transverse drift velocities of first-order orbit theory, emphasizes that in the presence of fields varying slowly in time the polarization and curvature drifts are to be calculated from the time derivative of the fields as seen from a frame of reference fixed with respect to the guiding center of the particle's orbit, and permits calculation of higher-order contributions to the drift velocities, contributions of second order in Larmor radius being explicitly evaluated; the resulting acceleration prepares the way for a study of the effect of weak inhomogeneities on the radiative output of single particles.

I. INTRODUCTION

The radiation emitted by a plasma has been calculated by a variety of methods and with a variety of different approximations.[1,2] Under suitable conditions,[3–5] some of this radiation can be interpreted as the incoherent superposition of radiation from individual charged particles accelerated by their interaction with whatever electromagnetic fields are present within the plasma. The radiation emitted by a single particle, however, is calculated most directly from knowledge of its trajectory as an explicit function of time.[6] This paper presents a calculation of the trajectory of a charged particle moving nonrelativistically in a uniform magnetic induction field on which is superimposed a weak and slowly varying electromagnetic inhomogeneity. The resulting velocity reproduces the well-known drift velocities; the resulting acceleration prepares the way for a study of the effect of weak inhomogeneities on the radiative output of single particles.[7]

The condition of slow variation imposed on the fields assures the validity of an approach based on the concept of a guiding center and enters the calculation in the following way: An expansion of the trajectory in a perturbation series in the weak inhomogeneity is sought. To avoid the appearance of terms that are valid only for times that are short compared to the gyroperiod, however, a perturbation expansion more involved than the most straightforward expansion must be adopted. The starting point of the present approach is an expansion that has been described by Berkowitz and Gardner[8] and by Kruskal,[9] an expansion that is valid only for fields varying slowly both in space and in time. Consequently, the fields here considered must be so restricted.

In this paper, trajectories are calculated by superimposing a conventional perturbation expansion in the weak inhomogeneity on the expansion of Berkowitz and Gardner and of Kruskal. In Sec. II, the theoretical background is summarized and equations, correct to *all* orders in the Larmor radius, are obtained for the coefficients in a general expansion. Appropriate boundary conditions are enumerated and, in particular, the need to eliminate rapidly varying terms from the solutions is inferred. For the *constant* magnetic induction field treated in Sec. III, the general equations decouple and their solution provides both a didactic example and the zeroth-order approximation for the weakly inhomogeneous field. The trajectory in a weakly inhomogeneous field is considered in Sec. IV. Terms varying rapidly with time are isolated by introducing Fourier transforms of the fields.

II. BASIC THEORY

Fundamental Equation for the Trajectory

The position vector $\mathbf{r}$ of a particle (charge e, mass m) moving nonrelativistically in a specified electromagnetic field $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{B}(\mathbf{r}, t)$ satisfies[10]

$$\epsilon\ddot{\mathbf{r}}(t) = \dot{\mathbf{r}}(t) \times \mathbf{B}(\mathbf{r}(t), t) + c\mathbf{E}(\mathbf{r}(t), t), \tag{1}$$

where c is the speed of light, t is time, and

$$\epsilon = mc/e. \tag{2}$$

Equation (1) is to be solved subject to the general initial conditions

$$\mathbf{r}(0) = \mathbf{r}_0, \quad \dot{\mathbf{r}}(0) = \mathbf{v}_0. \tag{3}$$

Because the total energy radiated per cycle by a particle gyrating nonrelativistically, even in magnetic fields as large as 10^6 Gauss, is a very small fraction of the kinetic energy of the particle,[11] the emitted radiation can be calculated using a trajectory determined by assuming no radiation. Forces arising from the radiation reaction may therefore be omitted from Eq. (1).

A dimensionless presentation of Eq. (1) can be accomplished by introducing a length L, characterizing the distance over which the fields change appreciably, and by defining B_0 to be a typical magnitude of the magnetic induction field. If the dimensionless quantities $\mathbf{r}/L$, v_0t/L, $\mathbf{B}/B_0$, $\mathbf{E}/B_0$, and c/v_0 are introduced into Eq. (1) and each dimensionless quantity is represented by the same symbol as is used for its dimensional counterpart, the resulting equation is formally identical with Eq. (1) except that the symbol ϵ now represents the dimensionless parameter mcv_0/eB_0L. Thus, ϵ measures the ratio of the Larmor radius ($\approx mcv_0/eB_0$) to the characteristic distance L. Apart from making clear the physical significance of the parameter ϵ, the dimensionless presentation has no particular advantage and Eq. (1) will be considered directly in dimensional form.[12]

Expansion in Larmor Radius

According to Berkowitz and Gardner[8] and to Kruskal,[9] the series[13]

$$\mathbf{r}(t) = \sum \epsilon^{|n|} f^n \mathbf{R}_n(t, \epsilon), \tag{4}$$

in which the $\mathbf{R}_n$ are coefficients to be determined,

$$f = \exp(i\omega_0\tau), \tag{5}$$

$$\omega_0 = B_0/\epsilon, \tag{6}$$

$$\tau(t) = \int_0^t [B(\mathbf{R}_0(t', \epsilon), t')/B_0]\, dt', \tag{7}$$

$$\text{and} \quad i = \sqrt{-1},$$

expresses a solution to Eq. (1) that is asymptotically valid when $\epsilon \to 0$, or, in view of the discussion of the previous subsection, when

$$\epsilon = mcv_0/eB_0L \ll 1, \tag{8}$$

or when the Larmor radius is small and the fields vary slowly in space.

In this expansion, $B(\mathbf{R}_0, t) = |\mathbf{B}(\mathbf{R}_0, t)|$ and each $\mathbf{R}_n(t, \epsilon)$ can itself be expanded in a power series in ϵ beginning with a term of zeroth order. Reality of $\mathbf{r}(t)$ implies that

$$\mathbf{R}_n^*(t, \epsilon) = \mathbf{R}_{-n}(t, \epsilon) \tag{9}$$

and in particular that $\mathbf{R}_0(t, \epsilon)$—identifiable as the position of the guiding center[14,15]—is real. Correctness of the expansion presupposes further that the fields satisfy[16]

$$\mathbf{E}(\mathbf{R}_0, t) \cdot \mathbf{B}(\mathbf{R}_0, t) = O(\epsilon) \tag{10}$$

and that the fields vary slowly in time. [See Eq. (20).] In Eq. (10), the symbol $O(\epsilon)$ stands for terms of order ϵ and higher that are omitted.

Equations for the Coefficients

Following the procedure used by Kruskal, one finds equations determining the functions $\mathbf{R}_n$ by substituting Eq. (4) into Eq. (1), expanding the fields in a Taylor series about $\mathbf{R}_0$, and equating separately the coefficients of each power of f. Appropriate expressions for $\dot{\mathbf{r}}$ and $\ddot{\mathbf{r}}$, obtained by differentiating Eq. (4), are

$$\dot{\mathbf{r}} = B_0 \sum \epsilon^{|n|-1} f^n \mathfrak{D}_n \mathbf{R}_n \tag{11}$$

and

$$\ddot{\mathbf{r}} = B_0^2 \sum \epsilon^{|n|-2} f^n (\mathfrak{D}_n)^2 \mathbf{R}_n, \tag{12}$$

where the operator[17]

$$\mathfrak{D}_n = \frac{\epsilon}{B_0}\frac{d}{dt} + in\frac{B}{B_0} \tag{13}$$

has been introduced. A suitable expansion of $\mathbf{B}(\mathbf{r}, t)$ about $\mathbf{R}_0$ can be obtained from the Taylor series[18]

$$\begin{aligned}\mathbf{B}(\mathbf{r}, t) &= \mathbf{B}(\mathbf{R}_0 + \textstyle\sum' \epsilon^{|n|} f^n \mathbf{R}_n, t) \\ &= \exp(\textstyle\sum' \epsilon^{|n|} f^n \xi_n)\mathbf{B}(\mathbf{y}, t),\end{aligned} \tag{14}$$

where

$$\xi_n = \mathbf{R}_n \cdot \nabla_y \tag{15}$$

and ∇_y involves differentiations with respect to the components of $\mathbf{y}$; $\mathbf{y}$ is to be set equal to $\mathbf{R}_0$ *after* the derivatives have been evaluated.[19] As is explored in greater detail in the Appendix, the terms of the familiar expansion of the exponential in Eq. (14) may be rearranged to display explicitly the coefficient of each power of f; the result is

$$\mathbf{B}(\mathbf{r}, t) = \sum \epsilon^{|n|} f^n \mathcal{M}_n\{\mathbf{B}\}, \tag{16}$$

where $\mathcal{M}_n$ is a differential operator that does not contain f but that does depend on $\mathbf{R}_n$, ∇_y, and ϵ. The operators $\mathcal{M}_n$ are linear, mutually commuting operators displaying in particular the property

$$\mathcal{M}_n^* = \mathcal{M}_{-n}. \tag{17}$$

In Eq. (16) and subsequently, the symbol $\mathcal{M}_n\{\mathbf{B}\}$ represents a sequence of steps in which $\mathbf{B}$ is first evaluated at an auxiliary argument $\mathbf{y}$ and then, after differentiations with respect to the components of $\mathbf{y}$ have been performed, $\mathbf{y}$ is replaced with $\mathbf{R}_0$. The

expansion

$$\mathbf{E}(\mathbf{r}, t) = \sum \epsilon^{|n|} f^n \mathcal{M}_n\{\mathbf{E}\} \tag{18}$$

follows directly from Eq. (16).

Equations (11), (12), (16), and (18) can now be substituted into Eq. (1). After a rearrangement of the summation indices in the term $\dot{\mathbf{r}} \times \mathbf{B}$, the coefficient of f^n may be extracted to provide the equation[20]

$$(\mathfrak{D}_n)^2 \mathbf{R}_n = (\epsilon c/B_0)\mathcal{M}_n\{\mathbf{E}/B_0\} + \sum \epsilon^{|m|-|n|+|n-m|}(\mathfrak{D}_m \mathbf{R}_m) \times \mathcal{M}_{n-m}\{\mathbf{B}/B_0\} \tag{19}$$

satisfied by $\mathbf{R}_n$. Separation of Eq. (19) from the general equation, however, can be justified only if there are no terms in Eq. (19) that vary with a frequency comparable to that of f, for the presence of such terms in Eq. (19) for any n would compel rejection of the method by which the coefficients of f^n were identified. If B in Eq. (7) varies slowly in time, the integral may be evaluated approximately for short times by setting $B = B_0$; the variation of f is thus characterized approximately by the frequency ω_0 and correctness of Eq. (19) therefore requires (1) that the fields be restricted by the condition

$$\omega_1/\omega_0 \ll 1, \tag{20}$$

where ω_1 characterizes the frequency of the field variation, and (2) that the solutions for $\mathbf{R}_n$ contain no terms varying with frequencies comparable to or larger than ω_0. As will be later confirmed, elimination of rapidly varying terms in $\mathbf{R}_n$ provides the proper number of constraints to determine the additional constants that have arisen unavoidably in the replacement of a single second-order differential equation [Eq. (1)] by several such equations [Eq. (19)].

Summary of Procedure

In summary, the method adopted involves solving Eq. (19), subject to the initial conditions of Eq. (3) and to the constraint that all terms in $\mathbf{R}_n$ vary slowly compared to the gyroperiod. The initial conditions cannot be imposed until after the $\mathbf{R}_n$ have been substituted into Eq. (4), and the entire procedure is meaningful only when the fields vary slowly over distances comparable to the Larmor radius and over times comparable to the gyroperiod.

III. THE CONSTANT FIELD

Although the solution for the trajectory of a charged particle moving in a uniform magnetic induction field can be obtained more easily by much less powerful methods, a full solution for this almost trivial case provides a prototype calculation in which the method is not obscured by the details of algebraic manipulation. The appearance and elimination of rapidly varying terms can be illustrated in a simple context. Furthermore, the solution in a uniform field is the zeroth-order approximation for the solution in a weakly inhomogeneous field.

Consider then the field

$$\mathbf{B}(\mathbf{r}, t) = B_0\hat{\mathbf{k}}, \quad \mathbf{E}(\mathbf{r}, t) = 0, \tag{21}$$

where $\hat{\mathbf{k}}$ is a unit vector in the direction of the magnetic field. The conditions (8) and (20) are clearly satisfied, for L is infinite and ω_1 is zero. Given Eq. (A6) in the Appendix, one finds that Eq. (19) reduces to[21]

$$(\mathcal{E}_n)^2 \mathbf{Z}_n(t) - \mathcal{E}_n \mathbf{Z}_n(t) \times \hat{\mathbf{k}} = 0, \tag{22}$$

where

$$\mathcal{E}_n = \frac{\epsilon}{B_0}\frac{d}{dt} + in. \tag{23}$$

The general solution of Eq. (22) is most conveniently written in the form

$$\mathbf{Z}_n = a_n \mathbf{e}_+ e^{-i(n+1)\omega_0 t} + b_n \mathbf{e}_- e^{-i(n-1)\omega_0 t} + [c_n\hat{\mathbf{i}} + d_n\hat{\mathbf{j}} + (g_n + h_n t)\hat{\mathbf{k}}]e^{-in\omega_0 t}, \tag{24}$$

where

$$\mathbf{e}_\pm = \hat{\mathbf{i}} \mp i\hat{\mathbf{j}} \tag{25}$$

and $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are unit vectors that together with $\hat{\mathbf{k}}$ define a right-handed Cartesian coordinate system. Suppression of terms varying as or more rapidly than $\exp(i\omega_0 t)$ requires all of the constants except c_0, d_0, g_0, h_0, b_1, and a_{-1} to be zero. Furthermore, Eq. (9) then implies that $b_1^* = a_{-1}$. Imposition of the initial conditions leads ultimately to values for the nonzero constants, and the solution assumes the final form

$$\begin{aligned} \mathbf{Z}_0 &= \mathbf{r}_0 + (\epsilon/B_0)\mathbf{v}_0 \times \hat{\mathbf{k}} + \dot{z}_0 t\hat{\mathbf{k}}, \\ \mathbf{Z}_1 &= (\mathbf{Z}_{-1})^* = -i(v_{0-}/2B_0)\mathbf{e}_-, \\ \mathbf{Z}_n &= 0, \quad |n| \geq 2, \end{aligned} \tag{26}$$

where $v_{0-} = \dot{x}_0 - i\dot{y}_0$ and $\dot{x}_0$, $\dot{y}_0$, $\dot{z}_0$ are the components of $\mathbf{v}_0$. Equation (26) leads to the familiar helix when substituted into Eq. (4).

IV. THE WEAKLY INHOMOGENEOUS FIELD

Equations for the First-Order Correction

Let the electromagnetic field now consist of a strong uniform magnetic induction field on which is superimposed a weak and slowly varying, but otherwise arbitrary inhomogeneity, i.e., let

$$\begin{aligned} \mathbf{B}(\mathbf{r}, t) &= B_0\hat{\mathbf{k}} + \delta\mathbf{B}_1(\mathbf{r}, t), \\ \mathbf{E}(\mathbf{r}, t) &= \delta\mathbf{E}_1(\mathbf{r}, t), \end{aligned} \tag{27}$$

where $\hat{\mathbf{k}}$ is a unit vector in the direction of the strong

component of the field, $|\mathbf{B}_1(\mathbf{r}, t)|$ is of order B_0, and δ is a (dimensionless) small parameter.[22] Furthermore, let[23]

$$\mathbf{R}_n(t, \epsilon) = \mathbf{Z}_n(t, \epsilon) + \delta\mathbf{F}_n(t, \epsilon) + O(\delta^2). \quad (28)$$

Then, the expansions

$$\mathbf{B}(\mathbf{R}_0, t) = B_0\hat{\mathbf{k}} + \delta\mathbf{B}_1(\mathbf{Z}_0, t) + O(\delta^2), \quad (29)$$

$$\mathbf{E}(\mathbf{R}_0, t) = \delta\mathbf{E}_1(\mathbf{Z}_0, t) + O(\delta^2), \quad (30)$$

$$B(\mathbf{R}_0, t) = B_0 + \delta B_{13}(\mathbf{Z}_0, t) + O(\delta^2), \quad (31)$$

$$\tau(t) = t + \delta\int_0^t [B_{13}(t')/B_0]\, dt' + O(\delta^2), \quad (32)$$

where

$$B_{13}(t) = B_{13}(\mathbf{Z}_0(t), t), \quad (33)$$

can be derived.[24,25] Substitution of Eqs. (13), (28), and (31) into Eqs. (4) and (11) and evaluation of the results at $t = 0$ lead to the conclusion that $\mathbf{Z}_n$ must satisfy the same initial conditions as were imposed on the solution to Eq. (22); initial conditions on $\mathbf{F}_n$ will be presented in Eqs. (44) and (45).

Equations determining $\mathbf{Z}_n$ and $\mathbf{F}_n$ are obtained by substituting Eqs. (28)–(31) into Eq. (19) and extracting separately the coefficient of each power of δ; the results are

$$(\mathcal{E}_n)^2\mathbf{Z}_n - \mathcal{E}_n\mathbf{Z}_n \times \hat{\mathbf{k}} = 0, \quad (34)$$

$$\mathcal{E}_n\mathcal{E}_{n\pm1}F_{n\pm} = \Lambda_{n\pm}, \quad (35)$$

$$(\mathcal{E}_n)^2F_{n3} = \Lambda_{n3}, \quad (36)$$

where the + and − components of any vector $\mathbf{A}$ are defined by

$$A_\pm = A_1 \pm iA_2 \quad (37)$$

and the vector itself has the expression

$$\mathbf{A} = \tfrac{1}{2}A_+\mathbf{e}_+ + \tfrac{1}{2}A_-\mathbf{e}_- + A_3\hat{\mathbf{k}}. \quad (38)$$

Since Eqs. (34) and (22) are identical and are both solved subject to the same initial conditions, the solution to Eq. (34) is given by Eq. (26). Thus, $\mathbf{\Lambda}_n$ may ultimately be written as follows:

$$\mathbf{\Lambda}_0(t) = (\epsilon/B_0)\mathcal{N}_0\{\mathbf{L}/B_0\} - i\epsilon^2[\mathbf{Z}_{-1} \times \mathcal{N}_1\{\mathbf{B}_1/B_0\} - \mathbf{Z}_1 \times \mathcal{N}_{-1}\{\mathbf{B}_1/B_0\}], \quad (39)$$

$$\mathbf{\Lambda}_1(t) = -i\mathbf{Z}_1\mathcal{E}_1\{B_{13}/B_0\} + (\epsilon/B_0)\mathcal{N}_1\{\mathbf{L}/B_0\} - i\epsilon^2\mathbf{Z}_{-1} \times \mathcal{N}_2\{\mathbf{B}_1/B_0\} + i\mathbf{Z}_1 \times \mathcal{N}_0\{\mathbf{B}_1/B_0\}, \quad (40)$$

$$\mathbf{\Lambda}_n(t) = (\epsilon/B_0)\mathcal{N}_n\{\mathbf{L}/B_0\} - i\epsilon^2\mathbf{Z}_{-1} \times \mathcal{N}_{n+1}\{\mathbf{B}_1/B_0\} + i\mathbf{Z}_1 \times \mathcal{N}_{n-1}\{\mathbf{B}_1/B_0\}, \quad n \geq 2, \quad (41)$$

where

$$\mathbf{L} = c\mathbf{E}_1 + \dot{z}_0\hat{\mathbf{k}} \times \mathbf{B}_1 \quad (42)$$

and $\mathbf{\Lambda}_n$ for $n < 0$ can be obtained from

$$\mathbf{\Lambda}_{-n} = \mathbf{\Lambda}_n^*. \quad (43)$$

In Eqs. (39)–(42), $\mathcal{N}_n$ is the operator obtained from $\mathcal{M}_n$ by replacing $\mathbf{R}_n$ with $\mathbf{Z}_n$; the notation $\mathcal{N}_n\{\mathbf{B}_1\}$ implies the following sequence of steps: (1) evaluate $\mathbf{B}_1$ at $\mathbf{r} = \mathbf{y}$; (2) perform the differentiations implicit in $\mathcal{N}_n$; (3) set $\mathbf{y} = \mathbf{Z}_0$. Explicit evaluations of $\mathcal{N}_n$ are presented in the Appendix. Finally, initial conditions to be imposed on $\mathbf{F}_n$, determined by substituting Eqs. (13), (28), and (31) into Eqs. (4) and (11) and evaluating the results at $t = 0$, may be expressed in the form

$$\sum \epsilon^{|n|}\mathbf{F}_n(0) = 0, \quad (44)$$

$$\sum \epsilon^{|n|-1}\mathcal{E}_n\mathbf{F}_n(0) = -\mathbf{v}_{0\perp}B_{13}(0)/B_0^2, \quad (45)$$

where $\mathbf{v}_{0\perp} = \dot{x}_0\hat{\mathbf{i}} + \dot{y}_0\hat{\mathbf{j}}$ is the (vector) component of $\mathbf{v}_0$ perpendicular to the uniform field. In Eq. (45) and in similar occurrences hereafter, $\mathcal{E}_n$ acts on $\mathbf{F}_n$ *before* the indicated evaluation at $t = 0$.

Solution for the First-Order Correction

In this subsection, the solution of Eqs. (35) and (36) subject to the two requirements, (1) that no rapidly varying terms appear and (2) that the initial conditions of Eqs. (44) and (45) be satisfied, is obtained. Particular solutions to Eqs. (35) and (36) are conveniently expressed in terms of an operator $(\mathcal{E}_n)^{-1}$ whose action on a function $\varphi(t)$ is defined by[26]

$$(\mathcal{E}_n)^{-1}\varphi(t) = \omega_0\int\frac{d\omega}{2\pi}\frac{e^{-i(\omega-i\mu)t}}{(-i)(\omega - n\omega_0 - i\mu)}\bar{\varphi}(\omega), \quad (46)$$

where $\bar{\varphi}(\omega)$ is the Fourier transform[27] of $\varphi(t)$

$$\varphi(t) = (2\pi)^{-1}\int d\omega e^{-i(\omega-i\mu)t}\bar{\varphi}(\omega). \quad (47)$$

(A limit $\mu \to 0^+$ is implied but not written; the function of μ is to assure convergence of the integral for $t > 0$.) The definition of Eq. (47) is consistent with the requirement that $\mathcal{E}_n(\mathcal{E}_n)^{-1}\varphi(t) = \varphi(t)$. Furthermore, $(\mathcal{E}_n)^{-1}$ so defined is a linear operator and commutes with $\mathcal{E}_m$ and $(\mathcal{E}_m)^{-1}$ for all m. Particular solutions to Eqs. (35) and (36) now assume the forms

$$p_{n\pm} = (\mathcal{E}_n)^{-1}(\mathcal{E}_{n\pm1})^{-1}\Lambda_{n\pm}, \quad (48)$$

$$p_{n3} = (\mathcal{E}_n)^{-2}\Lambda_{n3}, \quad (49)$$

respectively, which may be combined after the manner of Eq. (38) to produce

$$\mathbf{p}_n = (\mathcal{E}_n)^{-2}\mathbf{\Lambda}_{n\parallel} + (\mathcal{E}_n)^{-1}(\mathcal{E}_{n-1})^{-1}(\mathcal{E}_{n+1})^{-1} \times [\mathbf{\Lambda}_{n\perp} \times \hat{\mathbf{k}} + \mathcal{E}_n\mathbf{\Lambda}_{n\perp}]. \quad (50)$$

In Eq. (50), $\mathbf{\Lambda}_{n\parallel}$ is the portion of $\mathbf{\Lambda}_n$ parallel to the uniform field and $\mathbf{\Lambda}_{n\perp} = \mathbf{\Lambda}_n - \mathbf{\Lambda}_{n\parallel}$.

The general solution to Eqs. (35) and (36) is found by adding to each particular solution an arbitrary solution to the corresponding homogeneous equation. These complimentary solutions are readily determined by standard means, and one finds finally that

$$F_{n\pm} = p_{n\pm} + A_{n\pm}e^{-in\omega_0 t} + C_{n\pm}e^{-i(n\pm1)\omega_0 t}, \quad (51)$$

$$F_{n3} = p_{n3} + (A_{n3} + K_n\omega_0 t)e^{-in\omega_0 t}, \quad (52)$$

where $A_{n\pm}$, $C_{n\pm}$, A_{n3}, and K_n are integration constants.

It will now be argued that $(\mathcal{E}_n)^{-1}\varphi(t)$ is slowly varying (in time) if $\varphi(t)$ is itself slowly varying. If Eq. (46) defining $(\mathcal{E}_n)^{-1}\varphi(t)$ is evaluated as a contour integral for $t > 0$ (the only region of interest), the contour must be closed in the lower half ω plane. The obvious singularity of the integrand, however, lies in the upper half ω plane. Thus, only the singularities in $\bar{\varphi}(\omega)$ that occur in the lower half ω plane will contribute. If $\varphi(t)$ is slowly varying, the Fourier transform $\bar{\varphi}(\omega)$ must be small, except when $\mathrm{Re}\,\omega$ is itself small. The only significant values of ω in the integral of Eq. (46) are small and $(\mathcal{E}_n)^{-1}\varphi(t)$ is therefore slowly varying (Q E D).

Since the fields and $\mathbf{Z}_n(t)$ are slowly varying, $\mathbf{\Lambda}_n$, which depends on the fields and on $\mathbf{Z}_n$, must be slowly varying. Applying the arguments of the previous paragraph to Eqs. (48) and (49), one concludes that $\mathbf{p}_n$ must be slowly varying and finally that Eqs. (51) and (52) display explicitly the rapidly varying terms in the general solution for $\mathbf{F}_n$. Suppression of these terms is accomplished if all constants except $A_{0\pm}$, C_{1-}, $C_{-1,+}$, A_{03}, and K_0 (hereafter, $A_\pm$, C_-, C_+, A_3, and K) are zero. The most general solution for $\mathbf{F}_n$ consistent with the absence of rapidly varying terms is therefore given by

$$\mathbf{F}_n = \mathbf{p}_n + [\mathbf{A} + K\omega_0 t\hat{\mathbf{k}}]\delta_{n,0} + \tfrac{1}{2}C_-\mathbf{e}_-\delta_{n,1} + \tfrac{1}{2}C_+\mathbf{e}_+\delta_{n,-1}, \quad (53)$$

where $\delta_{n,m}$ is zero if $n \neq m$ and one if $n = m$. Substitution of Eq. (53) into Eqs. (44) and (45) leads to the expressions

$$\mathbf{C} = [B_{13}(0)B_0^{-2}\mathbf{v}_0 + \sum \epsilon^{|n|-1}\mathcal{E}_n\mathbf{p}_n(0)] \times \hat{\mathbf{k}}, \quad (54)$$

$$\mathbf{A} = -\epsilon\mathbf{C} - \sum \epsilon^{|n|}\mathbf{p}_n(0), \quad (55)$$

$$K = -\sum \epsilon^{|n|}\mathcal{E}_n p_{n3}(0), \quad (56)$$

for the constants $\mathbf{C}$, $\mathbf{A}$, and K.

Taken together, Eqs. (39)–(41), (50), and (54)–(56) permit the solution of Eq. (53) to be expressed directly in terms of the field inhomogeneities; the result is the goal of this subsection.

Position, Velocity, Acceleration

In this subsection, some of the results for the position, velocity, and acceleration of the charged particle will be compiled and a few of them will be expanded to obtain the first few terms in a power series in ϵ. When Eqs. (53) and (26) are substituted into Eq. (28), one finds

$$\mathbf{R}_0 = \mathbf{Z}_0 + \delta[\mathbf{p}_0 + \mathbf{A} + K\omega_0 t\hat{\mathbf{k}}], \quad (57)$$

$$\mathbf{R}_1 = (\mathbf{R}_{-1})^* = \delta\mathbf{p}_1 + D\mathbf{e}_-, \quad (58)$$

$$\mathbf{R}_n = \delta\mathbf{p}_n, \quad |n| \geq 2, \quad (59)$$

where

$$2D = -(iv_{0-}/B_0)[1 - \delta(B_{13}(0)/B_0)] + i\delta\sum \epsilon^{|n|-1}\mathcal{E}_n p_{n-}(0), \quad (60)$$

for the coefficients in Eq. (4). The coefficients in Eqs. (11) and (12) then have the expressions

$$\mathcal{D}_0\mathbf{R}_0 = \delta\mathcal{E}_0\mathbf{p}_0 + \hat{\mathbf{k}}[(\epsilon\dot{z}_0/B_0) + \delta K], \quad (61)$$

$$\mathcal{D}_1\mathbf{R}_1 = (\mathcal{D}_{-1}\mathbf{R}_{-1})^* = \delta\mathcal{E}_1\mathbf{p}_1 + iD[1 + \delta(B_{13}/B_0)]\mathbf{e}_-, \quad (62)$$

$$\mathcal{D}_n\mathbf{R}_n = \delta\mathcal{E}_n\mathbf{p}_n, \quad |n| \geq 2, \quad (63)$$

$$(\mathcal{D}_n)^2\mathbf{R}_n = \delta(\mathcal{E}_n)^2\mathbf{p}_n, \quad n \neq \pm1, \quad (64)$$

$$(\mathcal{D}_1)^2\mathbf{R}_1 = [(\mathcal{D}_{-1})^2\mathbf{R}_{-1}]^* = \delta(\mathcal{E}_1)^2\mathbf{p}_1 - D[1 - i\delta\mathcal{E}_2(B_{13}/B_0)]\mathbf{e}_-. \quad (65)$$

Substituted into Eqs. (4), (11), and (12), Eqs. (61)–(65) lead to expressions for $\mathbf{r}$, $\dot{\mathbf{r}}$, and $\ddot{\mathbf{r}}$ as explicit functions of $\mathbf{E}_1$ and $\mathbf{B}_1$. In this context, f is obtained from Eqs. (5) and (32). The results at this point include terms of all orders in ϵ but only of zeroth- and first-order in δ.

Certain of these results are more useful when expanded in powers of ϵ. The resulting double expansion in the two parameters ϵ and δ can be justified only if ϵ and δ are independent. Now, ϵ as given by Eq. (8) depends on B_0 and L (which might depend on δ) but also depends on v_0. Since v_0 is certainly independent of the field, ϵ is independent of δ (which is determined exclusively by the field) as required.

A few preliminary expansions facilitate the later expansion of Eqs. (64) and (65). With reference to Eqs. (10), (39)–(41), and (A10), one finds first that $E_{13} = O(\epsilon)$ and then that

$$(B_0/\epsilon)\mathbf{\Lambda}_0 = (1 + \epsilon^2\zeta\zeta^*)(\mathbf{L}/B_0) - \tfrac{1}{2}(\epsilon\,|\mathbf{v}_{0\perp}|^2/B_0^2)\nabla B_{13} + O(\epsilon^3), \quad (66)$$

$$\mathbf{\Lambda}_1 = -\frac{iv_{0-}}{2B_0^2}\left(B_{1+}\hat{\mathbf{k}} - \frac{i\epsilon}{B_0}\frac{dB_{13}}{dt}\mathbf{e}_-\right) + \frac{\epsilon}{B_0^2}\zeta(\mathbf{L}) + O(\epsilon^2), \quad (67)$$

$$\mathbf{\Lambda}_n = (n!)^{-1}\zeta^{n-1}[\tfrac{1}{2}(iv_{0-}n/B_0^2)(B_{13}\mathbf{e}_- - B_{1+}\hat{\mathbf{k}}) + (\epsilon/B_0^2)\zeta(\mathbf{L})] + O(\epsilon^2), \quad n \geq 2. \quad (68)$$

Equation (43) permits determination of $\mathbf{\Lambda}_n$ for $n < 0$. In Eqs. (66)–(68), all fields are evaluated at $(\mathbf{Z}_0, t)$, ∇ differentiates with respect to the components of $\mathbf{r}_0$, and

$$\zeta = -\frac{iv_{0-}}{2B_0}\left(\frac{\partial}{\partial x_0} + i\frac{\partial}{\partial y_0}\right). \tag{69}$$

The Maxwell equation $\nabla \cdot \mathbf{B} = 0$ was used in obtaining Eq. (67). Furthermore, Eq. (23) supports the expansion

$$(\mathcal{E}_n)^{-1} = -\frac{i}{n}\sum_{m=0}^{\infty}\left(\frac{i\epsilon}{nB_0}\right)^m\left(\frac{d}{dt}\right)^m \tag{70}$$

when $n \neq 0$, and Eq. (46) leads to

$$(\mathcal{E}_0)^{-1}\varphi(t) = (B_0/\epsilon)\int_0^t \varphi(t')\,dt' + (\mathcal{E}_0)^{-1}\varphi(0), \tag{71}$$

where the second term on the right is simply a constant whose value is related to the behavior of $\varphi(t)$ for $t < 0$; the disappearance of constants of this sort from all expressions having physical significance supports an earlier conjecture (Footnote 27) that the results are insensitive to the manner in which the electromagnetic field is established. If the quantities

$$\alpha(t) = \frac{cE_{13}}{\epsilon} - \frac{|\mathbf{v}_{0\perp}|^2}{2B_0}\frac{\partial B_{13}}{\partial z_0}, \tag{72}$$

$$\beta(t) = c\nabla \cdot \mathbf{E}_1 - \dot{z}_0(\nabla \times \mathbf{B}_1)_3, \tag{73}$$

$$\boldsymbol{\gamma}_n(t) = [n/(n-1)]B_{13}\mathbf{e}_- - B_{1+}\hat{\mathbf{k}} \tag{74}$$

are introduced, then the additional expansions

$$\begin{aligned}\frac{B_0}{\epsilon}\mathcal{E}_0\mathbf{p}_{0\perp} &= \frac{\mathbf{L}}{B_0} \times \hat{\mathbf{k}} + \frac{\epsilon}{B_0^2}\frac{d\mathbf{L}_\perp}{dt} \\ &+ \frac{\epsilon|\mathbf{v}_{0\perp}|^2}{2B_0^2}\hat{\mathbf{k}} \times \nabla B_{13} \\ &- \frac{\epsilon^2|\mathbf{v}_{0\perp}|^2}{2B_0^2}\frac{d}{dt}\nabla_\perp\frac{B_{13}}{B_0} \\ &+ \epsilon^2\left(\zeta\zeta^* - \frac{1}{B_0^2}\frac{d^2}{dt^2}\right)\left(\frac{\mathbf{L}}{B_0} \times \hat{\mathbf{k}}\right) + O(\epsilon^3),\end{aligned} \tag{75}$$

$$(B_0/\epsilon)\mathcal{E}_0\mathbf{p}_{0\parallel} = (\mathcal{E}_0)^{-1}\mathbf{\Lambda}_0\,(0) + \hat{\mathbf{k}}\int_0^t \alpha(t')\,dt' + O(\epsilon^2), \tag{76}$$

$$\begin{aligned}\mathcal{E}_1\mathbf{p}_{1\perp} = -(iv_{0-}\mathbf{e}_-/4B_0^2)[-iB_{13} &+ \int_0^t \beta(t')\,dt' \\ &+ (\epsilon/B_0)(\mathcal{E}_0)^{-1}\beta(0) + O(\epsilon)],\end{aligned} \tag{77}$$

$$\mathcal{E}_1\mathbf{p}_{1\parallel} = -\frac{v_{0-}\hat{\mathbf{k}}}{2B_0^2}\left(1 + \frac{i\epsilon}{B_0}\frac{d}{dt}\right)B_{1+} + O(\epsilon^2), \tag{78}$$

$$\mathcal{E}_n\mathbf{p} = (v_{0-n}/2n!\,B_0^2)\zeta^{n-1}\boldsymbol{\gamma}_n + O(\epsilon), \quad n \geq 2, \tag{79}$$

can be obtained from Eqs. (50) and (66)–(68); the Maxwell equations $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times c\mathbf{E} = -\partial\mathbf{B}/\partial t$ have been freely used. Finally, Eqs. (75)–(79) lead to the expansion

$$\begin{aligned}D = -\frac{iv_{0-}}{2B_0}\left(1 - \delta\frac{B_{13}(0)}{2B_0} + i\delta\frac{\epsilon}{2B_0^2}(\mathcal{E}_0)^{-1}\beta(0)\right) \\ - \delta\frac{L_-(0)}{2B_0^2} + O(\epsilon\delta)\end{aligned} \tag{80}$$

when substituted into Eq. (60).

The expansions of the previous paragraph now can be combined with Eqs. (64) and (65) to yield ultimately the expressions

$$(\mathfrak{D}_0)^2\mathbf{R}_0 = \delta\frac{\epsilon^2}{B_0^2}\left(\frac{d}{dt}\frac{\mathbf{L} \times \hat{\mathbf{k}}}{B_0} + \alpha\hat{\mathbf{k}} + O(\epsilon)\right) + O(\delta^2), \tag{81}$$

$$\begin{aligned}(\mathfrak{D}_1)^2\mathbf{R}_1 = X\Big[-\delta\frac{B_{1+}}{B_0}\hat{\mathbf{k}} + \mathbf{e}_-\Big(1 + \tfrac{3}{2}\delta\frac{B_{13} - B_{13}(0)}{B_0} \\ - \frac{i\delta}{2B_0}\int_0^t \beta(t')\,dt'\Big) + O(\epsilon\delta)\Big] + O(\delta^2),\end{aligned} \tag{82}$$

$$(\mathfrak{D}_n)^2\mathbf{R}_n = [i\delta v_{0-}/2(n-1)!\,B_0]\zeta^{n-1}\boldsymbol{\gamma}_n + O(\epsilon\delta), \tag{83}$$

where Eq. (83) applies for $n \geq 2$ and

$$X = \frac{iv_{0-}}{2B_0}\left(1 + \delta\frac{B_{13}(0)}{B_0}\right) + \delta\frac{L_-(0)}{2B_0^2}. \tag{84}$$

Values for $n < 0$ may be calculated from the relation

$$(\mathfrak{D}_{-n})^2\mathbf{R}_{-n} = [(\mathfrak{D}_n)^2\mathbf{R}_n]^*. \tag{85}$$

Substitution of Eqs. (81)–(83) into Eq. (12) gives finally the acceleration of the charged particle as an explicit function of time.

The Transverse Drift Velocities

The familiar drift velocity of a charged particle can be extracted from the above results. The vector $\mathbf{R}_0$ has already been identified as the position vector of the guiding center.[28] The velocity of the guiding center is thus $\dot{\mathbf{R}}_0$ and its component normal to the magnetic induction field—the drift velocity $\mathbf{v}_d$—is given by

$$\mathbf{v}_d = \dot{\mathbf{R}}_0 - (\dot{\mathbf{R}}_0 \cdot \mathbf{B})\mathbf{B}/B^2. \tag{86}$$

Expanded to first order in δ, this expression becomes

$$\mathbf{v}_d = \delta\left(\frac{B_0}{\epsilon}\mathcal{E}_0\mathbf{p}_{0\perp} - \dot{z}_0\frac{\mathbf{B}_{1\perp}}{B_0}\right) + O(\delta^2). \tag{87}$$

Substituting from Eq. (75), one then finds that

$$\mathbf{v}_d = -\frac{\delta c}{B_0}\hat{\mathbf{k}} \times \mathbf{E}_{1\perp} + \frac{\delta\epsilon c}{B_0^2}\frac{d\mathbf{E}_{1\perp}}{dt} + \frac{\delta\epsilon \dot{z}_0}{B_0^2}\hat{\mathbf{k}} \times \frac{d\mathbf{B}_{1\perp}}{dt} + \frac{\delta\epsilon |\mathbf{v}_{0\perp}|^2}{2B_0^2}\hat{\mathbf{k}} \times \nabla B_{13} - \frac{\delta\epsilon^2 |\mathbf{v}_{0\perp}|^2}{2B_0^2}\frac{d}{dt}\nabla_\perp \frac{B_{13}}{B_0} + \delta\epsilon^2\left(\zeta\zeta^* - \frac{1}{B_0^2}\frac{d^2}{dt^2}\right)\left(\frac{\mathbf{L}}{B_0} \times \hat{\mathbf{k}}\right) + O(\delta\epsilon^3, \delta^2). \tag{88}$$

The first four terms on the right-hand side of Eq. (88) are to be compared with the usual first-order drifts. When expanded through terms of order δ in the weakly inhomogeneous field, Chandrasekhar's results[29] for the electric and gradient drifts lead to expressions identical to the first and fourth terms in Eq. (88); his results for the polarization (p) and curvature (c) drifts become[30]

$$\mathbf{v}_p = \delta\left(\frac{\epsilon c}{B_0^2}\frac{d\mathbf{E}_1}{dt} + O(\epsilon^2)\right) + O(\delta^2), \tag{89}$$

$$\mathbf{v}_c = \delta\left(\frac{\epsilon \dot{z}_0^2}{B_0^2}\hat{\mathbf{k}} \times \frac{\partial \mathbf{B}_{1\perp}}{\partial z_0} + O(\epsilon^2)\right) + O(\delta^2). \tag{90}$$

These results agree with the second and third terms in Eq. (90) if it is noted (1) that

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \dot{z}_0\frac{\partial}{\partial z_0} \tag{91}$$

when the time derivative acts on a field evaluated at space–time argument $(\mathbf{Z}_0, t)$, (2) that Chandrasekhar's results are restricted to a static magnetic field ($\partial\mathbf{B}/\partial t = 0$), and (3) that E_{13} is $O(\epsilon)$, thereby permitting replacement of $\mathbf{E}_1$ with $\mathbf{E}_{1\perp}$ in Eq. (89). On the basis of the present calculation one concludes that, when the fields vary slowly in time as well as space, both the polarization and the curvature-drift velocities are determined by the time derivative of the fields as seen from a frame of reference fixed with respect to the guiding center.

The remaining terms in Eq. (88) express contributions to $\mathbf{v}_d$ that are second order in Larmor radius; the present result thus goes beyond the usual first-order orbit theory. In contrast to the first-order drifts, which involve first derivatives of the fields, these second-order terms involve second derivatives of the fields.

A Constant Inhomogeneity

An inhomogeneity in which $\mathbf{B}_1$ is constant and $\mathbf{E}_1$ is zero provides a second point of contact between the above results and exactly known solutions. No generality is lost by selecting a coordinate system in which the field $\mathbf{B}_1$ lies in the (y,z) plane. Thus, let

$$\mathbf{B} = (B_0 + \delta B_{13})\hat{\mathbf{k}} + \delta B_{12}\hat{\mathbf{j}}, \qquad \mathbf{E} = 0. \tag{92}$$

With reference to Eqs. (39)–(41) and to the Appendix, one then finds that only $\mathbf{\Lambda}_{-1}$, $\mathbf{\Lambda}_0$, and $\mathbf{\Lambda}_1$ are nonzero. Upon substitution into Eq. (50), the values obtained for $\mathbf{\Lambda}_n$ lead to

$$\mathbf{p}_0 = [\dot{z}_0 B_{12}(t - T)/B_0]\hat{\mathbf{j}}, \tag{93}$$

$$\mathbf{p}_1 = \mathbf{p}_{-1}^* = -\tfrac{1}{2}v_{0-}B_{12}\hat{\mathbf{k}}/B_0^2, \tag{94}$$

$$\mathbf{p}_n = 0, \quad |n| \geq 2. \tag{95}$$

Here the constant T arises from a term like the second term on the right in Eq. (71); its value is related to the manner in which the constant field is turned on and does not influence the final result for the trajectory. Substitution of Eqs. (93)–(95) into Eqs. (54)–(56) and (60) leads to an evaluation of the constants $\mathbf{C}$, $\mathbf{A}$, K, and D. Finally, all of these results can be substituted into Eqs. (57)–(59) to obtain, for the nonzero values of $\mathbf{R}_n$, the expressions

$$\mathbf{R}_0 = \mathbf{r}_0 + (\epsilon/B_0^2)[1 - \delta(B_{13}/B_0)][(\dot{y}_0 B_0 - \delta\dot{z}_0 B_{12})\hat{\mathbf{i}} - \dot{x}_0 B_0\hat{\mathbf{j}} + \delta\dot{x}_0 B_{12}\hat{\mathbf{k}}] + \dot{z}_0 t[\hat{\mathbf{k}} + \delta(B_{12}/B_0)\hat{\mathbf{j}}] + \delta\dot{y}_0 t(B_{12}/B_0)\hat{\mathbf{k}} + O(\delta^2), \tag{96}$$

$$\mathbf{R}_1 = \mathbf{R}_{-1}^* = -\tfrac{1}{2}(\delta v_{0-}B_{12}/B_0^2)\hat{\mathbf{k}} + \tfrac{1}{2}\mathbf{e}_-[-(iv_{0-}/B_0)(1 - \delta(B_{13}/B_0)) + (\delta\dot{z}_0 B_{12}/B_0^2)] + O(\delta^2). \tag{97}$$

When evaluated via Eq. (32), the function τ is

$$\tau = t[1 + \delta(B_{13}/B_0)] + O(\delta^2). \tag{98}$$

These results are correct to *all* orders in ϵ.

These same results can be obtained by applying Eq. (26) to the present field. The symbols in Eq. (26), however, must be interpreted in terms of a coordinate system defined by the *total* field rather than by the strong component of the field alone. Furthermore, the symbol B_0 in Eq. (26) must be interpreted as the magnitude of the *total* field. With these recognitions, one can expand Eq. (26) in a power series in δ to obtain Eqs. (96) and (97). With similar treatment, Eq. (7) leads to the same function f as does Eq. (98).

V. SUMMARY

The object of the present paper has been to derive analytic expressions for the acceleration of a particle moving nonrelativistically in a (prescribed) weakly inhomogeneous electromagnetic field. The final solution is obtained by substituting Eqs. (81)–(83) into

Eq. (12). The function f is determined by substituting Eq. (32) into Eq. (5). Incidental to the main objective, it has been pointed out that, in the presence of fields varying slowly in time, the polarization and curvature drifts involve total time derivatives of the fields as seen from the guiding center of the particle trajectory. Furthermore, expressions for the drift velocities to second order in Larmor radius have been presented [Eq. (88)].

VI. ACKNOWLEDGMENT

Several helpful conversations with Professor M. Krook during the author's graduate study are gratefully acknowledged.

APPENDIX: EVALUATION OF THE OPERATORS $\mathcal{M}_n$ AND $\mathcal{N}_n$

Equations (14) and (16) together define $\mathcal{M}_n$; as an operator relation, this definition is

$$\exp\left(\sum{}' \epsilon^{|n|} f^n \xi_n\right) = \sum \epsilon^{|n|} f^n \mathcal{M}_n . \tag{A1}$$

One basic property of $\mathcal{M}_n$ reflects Eq. (9). With reference to Eqs. (5), (7), and (9), one finds first that $f^* = f^{-1}$ and then that the left-hand side of Eq. (A1) is invariant to complex conjugation. Similar invariance of the right-hand side requires that

$$\mathcal{M}_{-n} = \mathcal{M}_n^* . \tag{A2}$$

Here, as elsewhere in this paper, the separate equating of coefficients of different powers of f is justified by the same conditions as are presented in the paragraph containing Eq. (19).

Additional properties of these operators can be demonstrated more easily with reference to the equation

$$\sum \epsilon^{|n|} f^n \mathcal{M}_n = \sum_{j=0}^{\infty} [(j!)^{-1} (\sum{}' \epsilon^{|p|} f^p \xi_p)^j], \tag{A3}$$

obtained by expanding the exponential in Eq. (A1). The operator $\mathcal{M}_n$ is to be obtained by reordering the terms in the right-hand side of Eq. (A3) so as to sort out the coefficient of the quantity $\epsilon^{|n|} f_n$.[31] From this procedure and from the linearity of ξ_n, one concludes that the $\mathcal{M}_n$ will themselves be linear operators. Furthermore, because none of the coefficients in Eq. (A3) depends on the auxiliary variable $\mathbf{y}$, the $\mathcal{M}_n$ will commute among themselves.

Still further features of the operators $\mathcal{M}_n$ can be inferred if the jth power of the primed sum in Eq. (A3) is written as a j-fold sum

$$(\cdots)^j = \sum_{p_1}{}' \cdots \sum_{p_j}{}' (\epsilon^{\Sigma_i |p_i|} f^{\Sigma_i p_i} \boldsymbol{\pi}_i \xi_{p_i}), \tag{A4}$$

where the index i assumes values $1 \leq i \leq j$. The contribution of this term to the coefficient of f^n in Eq. (A3) is composed of all terms in the multiple sum for which $\sum_i p_i = n$. Since $\sum_i |p_i|$ and $\sum_i p_i$ are both even or both odd, all terms in the coefficient of f^n will contain only even or only odd powers of ϵ according as n is even or odd. Furthermore, since $\sum |p_i| \geq |\sum p_i|$, the power of ϵ appearing in each term that makes a contribution to the coefficient of f^n is at least as great as $|n|$. These conclusions apply for all j. Consequently, the explicit appearance of a factor $\epsilon^{|n|}$ in the left-hand side of Eq. (A3) means that each $\mathcal{M}_n$ will be an *even* power series in ϵ starting with a term of order ϵ^0.

The explicit expansion of the right-hand side of Eq. (A3) in powers of ϵ is facilitated by noting that all terms in the jth power of the primed sum involve ϵ to a power at least as great as $j(|p_i| \geq 1)$ in Eq. (A4). Hence, the jth term in the right-hand side of Eq. (A3) will contribute only if the expansion is carried to order ϵ^j or higher. A more detailed evaluation of the first few operators leads to

$$\begin{aligned}
\mathcal{M}_0 &= 1 + \epsilon^2 \xi_1 \xi_{-1} + \epsilon^4 (\xi_2 \xi_{-2} + \tfrac{1}{4} \xi_1^2 \xi_{-1}^2 \\
&\quad + \tfrac{1}{2} \xi_1^2 \xi_{-2} + \tfrac{1}{2} \xi_2 \xi_{-1}^2) + O(\epsilon^6), \\
\mathcal{M}_1 &= \xi_1 + \epsilon^2 (\xi_2 \xi_{-1} + \tfrac{1}{2} \xi_1^2 \xi_{-1}) + O(\epsilon^4), \\
\mathcal{M}_2 &= \xi_2 + \tfrac{1}{2} \xi_1^2 + \epsilon^2 (\tfrac{1}{6} \xi_1^3 \xi_{-1} + \xi_3 \xi_{-1} + \xi_2 \xi_1 \xi_{-1}) + O(\epsilon^4), \\
\mathcal{M}_3 &= \xi_3 + \xi_2 \xi_1 + \tfrac{1}{6} \xi_1^3 + O(\epsilon^2), \\
\mathcal{M}_4 &= \xi_4 + \xi_3 \xi_1 + \tfrac{1}{2} \xi_2^2 + \tfrac{1}{2} \xi_2 \xi_1^2 + (\tfrac{1}{24}) \xi_1^4 + O(\epsilon^2).
\end{aligned} \tag{A5}$$

$\mathcal{M}_n$ for $n < 0$ may be found by using Eq. (A2). Since every term in $\mathcal{M}_n$ for $n \neq 0$ and all but the first term in $\mathcal{M}_0$ contain at least one derivative (i.e., one power of ξ),

$$\mathcal{M}_n\{\mathbf{C}\} = \mathbf{C}\delta_{n,0}, \tag{A6}$$

where $\mathbf{C}$ is a constant vector.

The expansion of Eq. (A3) becomes particularly tractable when the $\mathbf{R}_n$ have the form given in Eq. (26), for the primed sum can then be raised easily to any power by the binomial theorem. Thus,

$$\sum \epsilon^{|n|} f^n \mathcal{N}_n = \sum_{s=0}^{\infty} \sum_{m=0}^{s} \frac{\epsilon^s}{(s-m)!\, m!} f^{s-2m} \zeta^{s-m} (\zeta^*)^m, \tag{A7}$$

where $\mathcal{N}_n$ is the operator obtained from $\mathcal{M}_n$ when $\mathbf{R}_n$ is replaced with $\mathbf{Z}_n$ and $\zeta = \mathbf{Z}_1 \cdot \nabla_y$. Rearrangement of the summation indices leads ultimately to the expressions

$$\mathcal{N}_{2r} = \sum_{s=|r|}^{\infty} \epsilon^{2s-|2r|} \frac{\zeta^{s+r} (\zeta^*)^{s-r}}{(s+r)!\,(s-r)!}, \tag{A8}$$

$$\mathcal{N}_{2r+1} = \sum_{2s+1=|2r+1|}^{\infty} \epsilon^{2s+1-|2r+1|} \frac{\zeta^{s+r+1} (\zeta^*)^{s-r}}{(s+r+1)!\,(s-r)!}. \tag{A9}$$

More explicitly, the first few of these operators are

$$\begin{aligned}\mathcal{N}_0 &= 1 + \epsilon^2\zeta\zeta^* + \tfrac{1}{4}\epsilon^4(\zeta\zeta^*)^2 + O(\epsilon^6),\\ \mathcal{N}_1 &= \zeta + \tfrac{1}{2}\epsilon^2\zeta^2\zeta^* + O(\epsilon^4),\\ \mathcal{N}_2 &= \tfrac{1}{2}\zeta^2 + \tfrac{1}{6}\epsilon^2\zeta^3\zeta^* + O(\epsilon^4),\\ \mathcal{N}_n &= (n!)^{-1}\zeta^n + O(\epsilon^2), \quad n \geq 0.\end{aligned} \tag{A10}$$

The operators $\mathcal{N}_{-n}$ are obtained from those given through the use of Eq. (A2).

* The work here reported was begun while the author was a predoctoral student at Harvard University and, in substantially different notation, was a part of his doctoral dissertation. Fellowship support from the National Science Foundation and from Harvard University during the author's graduate study is acknowledged.

[1] G. Bekefi, *Radiation Processes in Plasmas* (John Wiley and Sons, New York, 1966). This book concludes with an extensive bibliography.

[2] J. M. Dawson, Plasma Physics Laboratory, Princeton University, Report MATT-428, 1965.

[3] D. J. Rose and M. Clark, Jr., *Plasmas and Controlled Fusion* (John Wiley and Sons, New York, 1961), Sec. 11.4.

[4] J. L. Hirshfield, D. E. Baldwin, and S. C. Brown, Phys. Fluids **4**, 198 (1961).

[5] W. E. Drummond and M. N. Rosenbluth, Phys. Fluids **3**, 45 (1960).

[6] J. D. Jackson, *Classical Electrodynamics* (John Wiley and Sons, New York, 1962), p. 480.

[7] It is believed that treatment of a prescribed inhomogeneity is a first step to a statistical treatment of a randomly fluctuating inhomogeneity.

[8] J. Berkowitz and C. Gardner, Commun. Pure Appl. Math. **12**, 501 (1959).

[9] M. Kruskal, Project Matterhorn, Princeton University, Report PM-S-33 (NYO-7903), 1958.

[10] Gaussian units are employed throughout.

[11] H. Rosner, Republic Aviation Corporation, Farmingdale, New York, Report AFSWC-TR-58-47, 1958.

[12] T. G. Northrop, *The Adiabatic Motion of Charged Particles* (Interscience Publishers, New York, 1963), pp. 2–3.

[13] Unless limits are more explicitly specified, the symbol $\sum$ represents a sum over the range from $n = -\infty$ to $n = +\infty$. If the summation index is not obvious, it will be indicated as in $\sum_n$.

[14] M. Kruskal, Ref. 9, p. 3.

[15] T. G. Northrop, Ref. 12, p. 37.

[16] M. Kruskal, Ref. 9, pp. 3, 9.

[17] Until modified after Eq. (33), all fields with unspecified arguments are understood to be evaluated at argument $\mathbf{R}_0, t$.

[18] The symbol $\sum'$ represents a sum over the limits from $n = -\infty$ to $n = +\infty$, *omitting* the value $n = 0$.

[19] The auxiliary variable $\mathbf{y}$ assures that $\nabla_{\mathbf{y}}$ will commute with the rest of the exponent and facilitates the formal expansion of the exponential.

[20] The triangle inequality, $|m| + |n - m| \geq |n|$, assures that the power to which ϵ is raised under the sum in Eq. (19) is positive.

[21] The symbol $\mathbf{Z}_n$ will be used for the solution for the field of Eq. (21); the symbol $\mathbf{R}_n$ will be reserved for the solution in the more general case.

[22] The functions $\mathbf{E}_1(\mathbf{r}, t)$ and $\mathbf{B}_1(\mathbf{r}, t)$ must, of course, satisfy Maxwell's equations.

[23] The symbols $\mathbf{Z}_n$ and $\mathbf{F}_n$ represent the *Z*eroth- and *F*irst-order contributions to $\mathbf{R}_n$ in an expansion in powers of δ.

[24] Components of a vector $\mathbf{Q}$ will be indicated by the symbol Q_i, with i assuming the possible values 1, 2, and 3. (Later + and − will be added.) The subscript 3 refers to the component in the direction of the strong field, while subscripts 1 and 2 refer to components in two other directions which together with the third direction form a right-handed orthogonal Cartesian coordinate system. Components of a vector already having a subscript (e.g., $\mathbf{B}_1$) will be denoted by a second subscript (e.g., B_{1i}).

[25] For the rest of this paper, fields with unspecified arguments will be understood to be evaluated at argument $\mathbf{Z}_0, t$.

[26] Unless otherwise specified, integrals extend from $-\infty$ to $+\infty$.

[27] In order that the Fourier transform of $\mathbf{\Lambda}_n$ be well defined, it may be necessary to imagine some artificial means of turning the inhomogeneity on and off at times outside the time interval of immediate interest to the problem, so that at very remote times in the past and future the inhomogeneity is zero. One might expect that the solution for the trajectory in the time interval of interest would be insensitive to the manner in which the fields are turned on and off.

[28] Different definitions of the position of the guiding center are possible. These definitions agree up to and including terms of first order in ϵ. Some ambiguity arises in higher-order terms. (See T. G. Northrop, Ref. 12, p. 37.) In the present context, the natural choice for the guiding center position to all orders in ϵ is $\mathbf{R}_0$.

[29] S. Chandrasekhar, *Plasma Physics* (University of Chicago Press, Chicago, 1960), p. 65.

[30] From his derivation of the polarization drift (Ref. 29, p. 20), it appears that Chandrasekhar means d/dt even though he writes $\partial/\partial t$.

[31] It is assumed that the series and the function on which it acts are such as to permit the necessary rearrangement of terms. In application, calculations will be carried to a given order in ϵ and all sums in Eq. (A3) will, therefore, be truncated. If the sums are viewed in this light, rearrangement of terms is certainly justified.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Approximation of Linear Transport Processes

H. J. HEJTMANEK
Department of Mathematics, University of Colorado, Boulder 80302

(Received 25 April 1969; Revised Manuscript Received 11 August 1969)

The semigroup generated by the linear transport operator is approximated by the corresponding discrete linear transport process. It can be shown that, in the case of a 2-dimensional μ space (x, μ), the generating operators converge and the discrete linear-transport process is uniformly quasibounded. A stability condition can be formulated.

1. INTRODUCTION

The transport of neutrons, electrons, ions, and photons in matter has the following common features: (1) These particles move on a straight line with constant velocity; (2) They can collide with particles of the supporting medium resulting in absorption, scattering, or multiplication. We are interested in the time development of the particle density-function $n(\mathbf{x}, \mathbf{v}, t)$. This is a function in the 6-dimensional μ space of statistical mechanics: $(\mathbf{x}, \mathbf{v}) \in R_6$, $\mathbf{x} = (x_1, x_2, x_3)$, and $\mathbf{v} = (v_1, v_2, v_3)$. Then $n(\mathbf{x}, \mathbf{v}, t)$ is a solution of an equation of integro-differential type[1]:

$$\begin{aligned} &\frac{\partial n}{\partial t} = -(T + A)n, \\ &Tn = \mathbf{v} \operatorname{grad}_x n, \\ &An = v \sum (\mathbf{v}) - \int_S v' \sum_s (\mathbf{v}', \mathbf{v}) n(\mathbf{x}, \mathbf{v}', t)\, d\mathbf{v}'. \end{aligned} \tag{1.1}$$

The operator T corresponds to the motion on a straight line with constant velocity; the operator A corresponds to the collision process. $\sum (v)$ is the total cross section for any collision process and $\sum_s (\mathbf{v}', \mathbf{v}')$ is the differential scattering cross section, if no multiplication occurs. If multiplication by fission occurs, then we have to add the corresponding term to $\sum_s$. We shall assume that the μ space is not all R_6, but a compact subset of it: $R \times S$, R being a bounded, closed, and convex subset of R_3, and S the 3-dimensional sphere with radius $v_{\max}$.

We shall choose as space of functions $n(\mathbf{x}, \mathbf{v}, t)$ the Banach space $X = L^p(R \times S)$, $p \geq 1$. Then $n(t)$ is a vector-valued function from $t \geq 0$ to X. If we replace the partial derivative by the strong derivative in X, we get the following abstract Cauchy problem[2,3]

$$\begin{aligned} &\dot{n} = -(T + A)n, \\ &n(0) = n_0 \in D(T + A). \end{aligned} \tag{1.2}$$

The domain of $T + A$ consists of all $n \in X$ such that $(T + A)n$ makes sense and $(T + A)n \in X$. We restrict $D(T + A)$ further to all $n(\mathbf{x}, \mathbf{v}) \in X$ such that $n(\mathbf{x}_B, \mathbf{v}_I) \equiv 0$, $\mathbf{x}_B$ being a point on the boundary of R and $\mathbf{v}_I$ pointing to the inside of R. This restriction corresponds to the situation that R is imbedded into a collision-free medium having no particles at $t = 0$. The operator $T + A$ with this domain is not normal. Because of this property it is very interesting to ask for the spectrum of $T + A$ and for the semigroup $V(t) = \exp[-(T + A)t]$ generated by $-(T + A)$. Another problem would be to find an approximating discrete semigroup $V_n(k\tau_n)$, because then we are able to decide whether a numerical calculation made on a computer makes $V_n(k\tau_n)$ converge to $V(t)$, if $n \to \infty$ in some sense. Problems of similar kind are treated in Refs. 3 and 4. $V_n(k\tau_n)$ is defined and described in Sec. 4.

The aim of this paper is to construct an approximating sequence of discrete semigroups $V_n(k\tau_n)$ that converges to $V(t)$ for $n \to \infty$ and to formulate the stability condition. We find it more convenient not to treat the problem in full generality, but to restrict to the monoenergetic problem of plane geometry with isotropic scattering:

$$\begin{aligned} &R = (x: |x| \leq a), \quad S = (\mu: |\mu| \leq 1), \\ &\qquad n = n(x, \mu, t), \\ &\dot{n} = -(T + A)n, \\ &n(0) = n_0 \in D(T + A), \\ &Tn = \mu \frac{\partial n}{\partial x}, \\ &An = n - \tfrac{1}{2}c \int_{-1}^{+1} n(x, \mu', t)\, d\mu', \quad c > 0. \end{aligned} \tag{1.3}$$

The operator $T + A$ has as domain all functions that are absolutely continuous in x for all μ, such that $(T + A)n \in X$, $n(-a, \mu > 0) \equiv 0$, and

$$n(+a, \mu < 0) \equiv 0.$$

The spectrum of the operator has been studied.[5]

Remark: If we substitute for n,

$$n(x, \mu, t) = e^{-t}\varphi(x, \mu, t), \tag{1.4}$$

then we get the problem

$$\begin{aligned} \dot{\varphi} &= -(T + A_1)\varphi, \\ \varphi(0) &= n_0 \in D(T + A_1), \\ T\varphi &= \mu \frac{\partial \varphi}{\partial x}, \\ A_1\varphi &= -\tfrac{1}{2}c\int_{-1}^{+1} \varphi(x, \mu', t)\, d\mu'. \end{aligned} \tag{1.5}$$

So the semigroup generated by $-(T + A_1)$ is written $V_1(t)$, and the corresponding approximating semigroup $V_{1n}(k\tau_n)$. The result of this paper is the contruction of an approximating linear-transport process V_{1n} or V_n depending on the discrete time interval τ_n. With this matrix V_{1n} or V_n we can find the particle distribution at a certain time $k\tau_n$ $(k = 1, 2, \cdots)$, if the initial distribution g at $k = 0$ was given. If the process tends to an equilibrium, we should calculate

$$V_n(k\tau_n)g = [V(\tau_n)]^k g$$

for large k, like the neutron distribution in a critical reactor. The stability condition (5.24) is a sufficient condition for the convergence of the approximation procedure for $\tau_n \to 0$ $(n \to \infty)$.

2. THE COLLISION-FREE TRANSPORT-OPERATOR T

If we choose $p = 1$ in $X = L^p(R \times S)$, then the norm of this Banach space equals the total number of particles in $R \times S$, if $n(x, \mu) \geq 0$:

$$\|n\| = \int_{-1}^{+1} d\mu \int_{-a}^{+a} dx\, |n(x, \mu)|. \tag{2.1}$$

Theorem 1: The operator $-T$ decomposes the spectral plane in the spectrum $\sum(-T) = (\lambda\colon \mathrm{Re}\,(\lambda) \leq 0)$ and the resolvent set $P(-T) = (\lambda\colon \mathrm{Re}(\lambda) > 0)$. For every $\lambda \in P(-T)$ we have

$$\|(T + \lambda)^{-1}\| \leq [\mathrm{Re}\,(\lambda)]^{-1}. \tag{2.2}$$

Proof: Take any λ with $\beta = \mathrm{Re}\,(\lambda) > 0$. We shall show it belongs to $P(-T)$. The equation

$$(T + \lambda)n = g\,,\ g \in X, \tag{2.3}$$

has the formal solution

$$\begin{aligned} n(x, \mu) &= \frac{1}{\mu}\int_{-a}^{x} \exp\left(-\frac{\lambda}{\mu}(x - \xi)\right) g(\xi, \mu)\, d\xi, \\ &\qquad \text{for} \quad \mu > 0, \\ &= -\frac{1}{\mu}\int_{x}^{+a} \exp\left(-\frac{\lambda}{\mu}(x - \xi)\right) g(\xi, \mu)\, d\xi, \\ &\qquad \text{for} \quad \mu < 0. \end{aligned} \tag{2.4}$$

We have to show that, if $g \in X$, then

$$n \in X = L'(R \times S). \tag{2.5}$$

Substituting $x - \xi = t$, we get

$$|n(x, \mu)| \leq \int_0^{a+x} \exp\left(-\frac{\beta}{\mu}t\right) |g(x - t, \mu)|\, dt, \qquad \text{for} \quad \mu > 0. \tag{2.6}$$

We define $g(x, \mu) = 0$, if $|x| > a$ and so we can write

$$|n(x, \mu)| \leq \frac{1}{\mu}\int_0^{\infty} \exp\left(-\frac{\beta}{\mu}t\right) |g(x - t, \mu)|\, dt, \qquad \text{for} \quad \mu > 0, \tag{2.7}$$

$$\begin{aligned} &\int_0^1 d\mu \int_{-a}^{+a} dx\, |n(x, \mu)| \\ &\qquad \leq \int_0^1 d\mu \frac{1}{\mu}\int_0^{\infty} \exp\left(-\frac{\beta}{\mu}t\right) dt \int_{-a}^{+a} |g(x, \mu)|\, dx \\ &\qquad = \frac{1}{\beta}\int_0^1 d\mu \int_{-a}^{+a} dx\, |g(x, \mu)|. \end{aligned} \tag{2.8}$$

A similar result holds for $\mu < 0$ and so we get, for all the μ range,

$$\begin{aligned} \|n\| &\leq \beta^{-1}\|g\|, \\ \|(T + \lambda)^{-1}\| &\leq [\mathrm{Re}\,(\lambda)]^{-1}. \end{aligned} \tag{2.9}$$

Next we shall prove that every λ with $\beta = \mathrm{Re}\,(\lambda) < 0$ belongs to the spectrum. We construct a set of functions depending on the parameter $\delta > 0$ with the properties

$$\begin{aligned} &n_\delta \in X, \quad \|n_\delta\| \geq C > 0\,, \text{ for } \delta < \delta_0\,, \\ &\lim_{\delta \to 0} \|(T + \lambda)n_\delta\| = 0. \end{aligned} \tag{2.10}$$

We define[5]

$$\begin{aligned} n_\delta(x, \mu) &= \frac{1}{\delta\mu} \exp\left[-(\lambda/\mu)(x - a)\right](x + a)/a, \\ &\qquad \text{for} \quad \delta^2 < \mu < \delta, \\ &= 0, \qquad \text{otherwise.} \end{aligned} \tag{2.11}$$

Then we get

$$\begin{aligned} &\|n_\delta(x, \mu)\| \\ &= \frac{1}{\delta}\int_{\delta^2}^{\delta} d\mu \frac{1}{\mu}\int_{-a}^{+a} \exp\left(-\frac{\beta}{\mu}(x - a)\right)\frac{x + a}{a}\, dx \\ &= \frac{1}{\delta}\int_{\delta^2}^{\delta} \left\{\frac{2}{-\beta} + \frac{\mu}{\alpha\beta^2}\left[\exp\left(\frac{2a}{\mu}\beta\right) - 1\right]\right\} d\mu \\ &\geq \frac{1}{\delta}\int_{\delta^2}^{\delta} \left(\frac{2}{-\beta} - \frac{\mu}{a\beta^2}\right) d\mu \\ &= \frac{2}{-\beta}(1 - \delta) - \frac{1}{2a\beta^2}(\delta - \delta^3) \to \frac{2}{-\beta} > 0, \\ &\qquad \text{for} \quad \delta \to 0. \end{aligned} \tag{2.12}$$

Then we have to calculate

$$(T+\lambda)n_\delta = \frac{1}{\delta a}\exp\left(-\frac{\lambda}{\mu}(x-a)\right), \quad (2.13)$$

and

$$\begin{aligned}\|(T+\lambda)n_\delta\| &= \frac{1}{\delta a}\int_{\delta^2}^{\delta} d\mu \int_{-a}^{+a} dx \exp\left(-\frac{\beta}{\mu}(x-a)\right)\\ &= \frac{1}{-\beta\delta a}\int_{\delta^2}^{\delta}\mu\left(1-\exp\left(\frac{2a}{\mu}\beta\right)\right)d\mu\\ &\le \frac{1}{-\beta\delta a}\int_{\delta^2}^{\delta}\mu\, d\mu = \frac{1}{-2\beta a}(\delta-\delta^3)\to 0, \quad \text{for} \quad \delta\to 0.\end{aligned} \quad (2.14)$$

Since $\sum(-T)$ is a closed set, it equals the set $(\lambda\colon \mathrm{Re}\,(\lambda)\le 0)$. Q E D

So $-T$ satisfies the condition of the Hille–Yosida theorem (Appendix A), and $-T$ is the infinitesimal generator of a semigroup

$$U(t)n_0(x,\mu) = e^{-Tt}n_0(x,\mu) = n_0(x-\mu t,\mu). \quad (2.15)$$

Of course, this semigroup describes the collision-free motion of particles in μ space. For the general T from (1.1) we would get the semigroup

$$U(t)n_0(\mathbf{x},\mathbf{v}) = n_0(\mathbf{x}-\mathbf{v}t,\mathbf{v}). \quad (2.16)$$

This has been known since the work of Case *et al.* (Ref. 6).

3. THE TRANSPORT OPERATOR $T + A_1$

We are now interested in the solution of the abstract Cauchy problem with the infinitesimal generator $-(T+A_1)$. We can try to regard A_1 as a perturbation to the operator T.

Theorem 2: A_1 is a bounded operator equal to cP, where P is a projection in X

Proof: If $u = A_1 v$, $v \in X$, then

$$\begin{aligned}u(x,\mu) &= cPv(x,\mu),\\ Pv(x,\mu) &= \tfrac{1}{2}\int_{-1}^{+1} v(x,\mu')\,d\mu',\\ \|Pv\| &\le \|v\|.\end{aligned} \quad (3.1)$$

Taking $v(x,\mu)\equiv 1$, we get $\|P\| = 1$. For A_1 we have

$$\|A_1\| = c; \quad (3.2)$$

$P^2 = P$ follows from an easy calculation. Q E D

Due to the theorem of Appendix B, $-(T+A_1)$ is the infinitesimal generator of a semigroup

$$\begin{aligned}V_1(t) &= e^{-(T+A_1)t},\\ \|V_1(t)\| &\le e^{-ct}.\end{aligned} \quad (3.3)$$

Remark: For the operator A we get

$$\begin{aligned}V(t) &= e^{-(T+A)t} = e^{t}e^{-(T+A_1)t},\\ \|V(t)\| &\le e^{(1-c)t}.\end{aligned} \quad (3.4)$$

This is the best estimate for $\|V(t)\|$: Just choose as the initial distribution a cloud of particles having a finite distance from the boundary of R. Then the particles behave for a short time as if they are in an infinite medium, that is, $\|V(t)\| = e^{(1-c)t}$.

4. THE DISCRETE TRANSPORT SEMIGROUP

We are going to construct the discrete transport semigroup $U_n(k\tau_n)$ approximating $U(t)$, and then the corresponding one $V_n(k\tau_n)$ for $V(t)$. The μ space $R\times S$ is divided into a finite number of cells by chopping the x interval $[-a,+a]$ into m_n equal parts and the μ interval $[-1,+1]$ into $2m_n'+1$ equal parts. h_n and h_n' are the length of these parts, that is,

$$h_n = \frac{2a}{m_n}, \quad h_n' = \frac{2}{2m_n'+1}. \quad (4.1)$$

Then each cell can be labeled by a pair of integers (i,j): $i = 1,\cdots,m_n$ and $j = -m_{n-1}',\cdots,-1,0,+1,\cdots,+m_n'$. The number of particles in cell (i,j) is written $\xi_{i,j}$. We define the set of all vectors $(\xi_{i,j})$, $\xi_{i,j}$ real, the Banach space X_n, by introducing the norm

$$\begin{aligned}v &= (\xi_{i,j}),\\ \|v\| &= \sum_{i,j}|\xi_{i,j}|.\end{aligned} \quad (4.2)$$

To the operator $T\in C(X)$ [where $C(X)$ is the set of all closed operators from X into X], there corresponds a matrix operator $T_n\in B(X_n)$ [where $B(X)$ is the set of all bounded operators from X into X]:

$$\begin{aligned}&v = T_n u, \quad u = (\xi_{i,j}), \quad v = (\eta_{i,j}),\\ &\eta_{i,j} = jh_n'\frac{\xi_{i,j}-\xi_{i-j,j}}{jh_n} = \frac{h_n'}{h_n}(\xi_{i,j}-\xi_{i-j\,j}),\end{aligned} \quad (4.3)$$

with the convention $\xi_{i,j} = 0$, whenever $i<1$ or $i>m_n$. This takes care of the boundary condition that no particles enter R through the boundary of R. $-T_n$ is the generator of a semigroup $U_n(\tau_n)$ defined by[3]

$$U_n = U_n(\tau_n) = 1 - \tau_n T_n, \quad (4.4)$$

$$\begin{aligned}&v = U_n u, \quad u = (\xi_{i,j}), \quad v = (\eta_{i,j}),\\ &\eta_{i,j} = \xi_{i,j} - \tau_n(h_n'/h_n)(\xi_{i,j}-\xi_{i-j,j}).\end{aligned} \quad (4.5)$$

Let us choose τ_n such that $\tau_n h_n' h_n^{-1} = 1$ for each value of $n = 1, 2, \cdots$. So we get a connection between the discrete time interval τ_n and the cell widths h_n and h_n'. We get

$$\eta_{i,j} = \xi_{i-j,j}. \tag{4.6}$$

We should have expected (4.6), because for $-T$ the semigroup $U(t)$ has the property

$$U(t)n(x, \mu) = n(x - \mu t, \mu) \tag{4.7}$$

that corresponds exactly to Eq. (4.6).

We turn now to the operator $T + A$ and construct the corresponding matrix operators $T_n + A_n$ as

$$v = (T_n + A_n)u, \quad u = (\xi_{i,j}), \quad v = (\eta_{i,j}),$$

$$\eta_{i,j} = \frac{h_n'}{h_n}(\xi_{i,j} - \xi_{i-j,j}) - \tfrac{1}{2}c \sum_{l=-m_n'}^{+m_n'} \xi_{i,l} h_n' + \xi_{i,j}. \tag{4.8}$$

Then $-(T_n + A_n)$ is the generator of a semigroup $V_n(\tau_n)$, that is,

$$V_n = V_n(\tau_n) = 1 - \tau_n(T_n + A_n),$$
$$v = V_n u, \quad u = (\xi_{i,j}), \quad v = (\eta_{i,j}),$$

and

$$\eta_{i,j} = \xi_{i-j,j} + \tfrac{1}{2}c \sum_{l=-m_n'}^{+m_n'} \xi_{i,l} \tau_n h_n' - \tau_n \xi_{i,j}. \tag{4.9}$$

We define A_{1n} by

$$v = A_{1n}u, \quad n = (\xi_{i,j}), \quad v = (\eta_{i,j}),$$

$$\eta_{i,j} = -\tfrac{1}{2}c \sum_{l=-m_n'}^{+m_n'} \xi_{i,l} h_n'. \tag{4.10}$$

$-(T_n + A_{1n})$ is the generator of a semigroup $V_{1n}(\tau_n)$, where

$$V_{1n} = V_{1n}(\tau_n) = 1 - \tau_n(T_n + A_{1n}). \tag{4.11}$$

Finally we remark that we have

$$V_n(k\tau_n) = [V_n(\tau_n)]^k, \quad V_{1n}(k\tau_n) = [V_{1n}(\tau_n)]^k.$$

5. CONVERGENCE AND STABILITY

Let $X = L^1(R \times S)$ and $X_n = C^{m_n \times m_n'}$ be the set of all $(m_n \times m_n')$-dimensional vectors $v = (\xi_{i,j})$, $i = 1, \cdots, m_n$ and $j = -m_n', \cdots, -1, 0, +1, \cdots, +m_n'$, with the corresponding norm,

$$\|v\| = \sum_{i,j} |\xi_{i,j}|. \tag{5.1}$$

Let P_n be the operator from X into X_n:

$$v = P_n u, \quad u \in X, \quad v = (\xi_{i,j}) \in X_n,$$

and

$$\xi_{i,j} = \int_i dx \int_j d\mu \, u(x, \mu), \tag{5.2}$$

where the integration is over the cell (i, j).

Lemma 1: The operator P_n has the following properties:

(a) $\|P_n\| = 1$,
(b) $\|P_n u\| \to \|u\|$ for $n \to \infty$,
(c) there exists for every $v \in X_n$ a $u \in X$ such that $v = P_n u$. (5.3)

Proof:

(a) For every $u \in X$ we have

$$\xi_{i,j} = \int_i dx \int_j d\mu u(x, \mu)$$

and

$$\|P_n u\| = \sum_{i,j} |\xi_{i,j}| = \sum_{i,j} \left| \int_i dx \int_j d\mu u(x, \mu) \right|$$
$$\leq \sum_{i,j} \int_i dx \int_j d\mu \, |u(x, \mu)| = \|u\|. \tag{5.4}$$

If we choose $u(x, \mu) = 1$, we see that $\|P_n\| = 1$.

(b) First, for every $u(x, \mu) \geq 0$, we get

$$\|P_n u\| = \|u\|. \tag{5.5}$$

Every $u \in X$ has the form $u = u_+ - u_-$, $u_+ \geq 0$, and $u_- \geq 0$. So we get

$$\|P_n u\| = \|P_n(u_+ - u_-)\| = \|P_n u_+\| + \|P_n u_-\|$$
$$= \|u_+\| + \|u_-\| = \|u\|.$$

(c) For every $v \in X_n$, we take for $u(x, \mu)$,

$$u(x, \mu) = \sum_{i,j} \xi_{i,j} X_{i,j}(x, \mu), \tag{5.6}$$

where $X_{i,j}(x, \mu)$ is the characteristic function of all (i, j).

Definition: A sequence $u_n \in X_n$ is said to converge to $u \in X$, iff

$$\|u_n - P_n u\| \to 0 \quad \text{for} \quad n \to \infty. \tag{5.7}$$

Lemma 2: Every $P_n u$ converges to u.

This follows from the definition and requires no calculation.

Definition: The sequence $A_n \in B(X_n)$ is said to converge strongly to $A \in C(X)$, iff

$$\|A_n P_n u - P_n A u\| \to 0, \quad \text{for} \quad n \to \infty$$
and every $u \in X$. (5.8)

Lemma 3: T_n converges to T for all u of a core of $D(T)$. (A core of a closed operator T is a dense subset of the graph of T.)

Proof: If we set

$$T_nP_nu = (\eta_{i,j}),$$
$$\eta_{i,j} = \frac{h_n'}{h_n}(\xi_{i,j} - \xi_{i-j,j}), \tag{5.9}$$

and

$$P_nTu = (\xi_{i,j}),$$
$$\xi_{i,j} = \int_i dx \int_j d\mu\mu \frac{\partial u(x,\mu)}{\partial x}, \tag{5.10}$$

then we form

$$\begin{aligned}\eta_{i,j} - \xi_{i,j} &= \frac{h_n'}{h_n}\left[\int_i dx \int_j d\mu u(x,\mu) - \int_{i-j} dx \int_j d\mu u(x,\mu)\right] \\ &\quad - \int_i dx \int_j d\mu\mu \frac{\partial u(x,\mu)}{\partial x} \\ &= \int_i dx \int_j d\mu \left[\frac{h_n'}{h_n}(u(x,\mu) - u(x - jh_n',\mu)) - \mu\frac{\partial n(x,\mu)}{\partial x}\right] \\ &= \int_i dx \int_j d\mu \left[jh_n' \frac{1}{jh_n}\int_{x-jh_n'}^{x} \frac{\partial u(\sigma,\mu)}{\partial x} d\sigma - \mu\frac{\partial n}{\partial x}(x,\mu)\right].\end{aligned} \tag{5.11}$$

The polynomials form a core of T. If we substitute for $u(x, \mu)$ a polynomial, then the argument of the integral is uniformly convergent, so that we get

$$\sum_{i,j} |\eta_{i,j} - \xi_{i,j}| < \epsilon \quad \text{if} \quad n > n_0(\epsilon). \tag{5.12}$$

Theorem 3: $T_n + A_n$ converges to $T + A$ for all u of a core of $D(T)$.

Proof: $T_n \underset{s}{\to} T$ (where $\underset{s}{\to}$ means convergence in the strong sense) follows from Lemma 3, so we have only to show that

$$A_n \underset{s}{\to} A,$$
$$A_nP_nu = (\eta_{i,j}),$$
$$\eta_{i,j} = -\tfrac{1}{2}c \sum_{l=-m_n'}^{+m_n'} \xi_{i,l}h_n' + \xi_{i,j}, \tag{5.13}$$

$$P_nAu = (\xi_{i,j}),$$
$$\begin{aligned}\xi_{i,j} &= \int_i dx \int_j d\mu \left[-\tfrac{1}{2}c\int_{-1}^{+1} u(x,\mu')\,d\mu' + u(x,\mu)\right] \\ &= -\tfrac{1}{2}ch_n' \sum_{l=-m_n'}^{+m_n'} \xi_{i,l} + \xi_{i,j},\end{aligned} \tag{5.14}$$

so we get

$$\eta_{i,j} - \xi_{i,j} = 0. \tag{5.15}$$

Theorem 4: $V_{1n}(\tau_n) = 1 - \tau_n(T_n + A_{1n})$ is uniformly quasibounded, i.e.,

$$\|V_{1n}(k\tau_n)\| \le e^{\beta k\tau_n}, \tag{5.16}$$

where β does not depend on k and n.

Proof:

$$u = (\xi_{i,j}), \quad v = (\eta_{i,j}),$$
$$v = [1 - \tau_n(T_n + A_{1n})]u, \tag{5.17}$$

$$\begin{aligned}\|v\| &= \sum_{i,j} |\eta_{i,j}| \le \sum_{i,j} |\xi_{i-j,j}| + \tfrac{1}{2}c\tau_n \sum_{i,j}\sum_{l} |\xi_{i,l}|\, h_n' \\ &\le \|u\| + \tfrac{1}{2}c\tau_n(2m_n' + 1)h_n' \|u\| \qquad (5.18) \\ &= (1 + c\tau_n)\|u\|. \qquad (5.19)\end{aligned}$$

So we finally get

$$\|1 - \tau_n(T_n + A_{1n})\| \le 1 + c\tau_n$$

and

$$\|V_{1n}(k\tau_n)\| \le (1 + c\tau_n)^{\tau_n^{-1}k\tau_n}, \tag{5.20}$$

and if we put

$$\sup_{n\to\infty}(1 + c\tau_n)^{(\tau_n)^{-1}} = e^c,$$
$$\|V_{1n}(k\tau_n)\| \le e^{ck\tau_n}. \tag{5.21}$$

For $V_n(\tau_n k)$, we have

$$\|V_n(k\tau_n)\| \le e^{(c-1)k\tau_n}, \tag{5.22}$$

with $\beta = c - 1$, and independent of k and n.

We can now apply the theorem of Appendix C, and we get that

$$V_n(k\tau_n) \to V(t), \tag{5.23}$$

where the limit has to be taken in the sense of Appendix C. The stability condition (which corresponds to the uniform quasiboundedness) is

$$\tau_n = h_n'/h_n \to 0, \quad \text{for} \quad n \to \infty, \tag{5.24}$$

meaning that we have to make the division in the μ direction finer than that in the x direction. The stability condition is a sufficient one for the convergence in the above sense of the approximation procedure. In other words, if we violate the condition (5.24) by an inappropriate choice of τ_n, h_n', and h_n, with $\tau_n h_n \neq h_n'$, we cannot expect, in general, that the procedure converges for τ_n—the time interval—tending to zero. If in a certain approximation of the linear transport process τ_n and h_n—the time interval and the x interval—are given, then we have to choose h_n'—the μ interval—such that (5.24) is satisfied to be sure that a numerical calculation will give better and better results for $n \to \infty$.

For more details, see Ref. 4.

APPENDIX A

Theorem[3]: If $T \in C(X)$ with a dense domain $D(T)$, if the negative real axis belongs to the resolvent set P

(T), and if the resolvent satisfies the inequality

$$\|(T+\xi)^{-1}\| \leq \xi^{-1} \quad \xi > 0, \tag{A1}$$

then $-T$ is the infinitesimal generator of a semigroup

$$U(t) = e^{(-Tt)}, \quad \text{for} \quad t \geq 0. \tag{A2}$$

This semigroup has the following properties:

(1) $U(t)$ is strongly continuous in t for $t \geq 0$ with $U(0) = 1$ and $\|U(t)\| \leq 1$;

(2) $n(t) = U(t)n_0$, $n_0 \in D(T)$ is a unique solution of the abstract Cauchy problem

$$\begin{aligned} \dot{n} &= -T_n, \\ n(0) &= n_0 \in D(T). \end{aligned} \tag{A3}$$

This implies $n(t) \in D(T)$ for all $t \geq 0$.

APPENDIX B

Theorem[3]*:* If T satisfies the assumptions of Appendix A, and if A is a bounded operator, then $T + A$ is a closed operator with domain $D(T + A) = D(T)$ dense in X, the part $\lambda \leq -\|A\|$ of the real axis belongs to the resolvent set of $T + A$, and the resolvent of $T + A$ satisfies the inequality

$$\|(T + A + \xi)^{-1}\| \leq (\xi - \|A\|)^{-1}, \quad \text{for} \quad \xi > \|A\|. \tag{B1}$$

$-(T + A)$ is then infinitesimal generator of a semigroup

$$V(t) = e^{-(T+A)t}, \quad \text{for} \quad t \geq 0. \tag{B2}$$

This semigroup has properties (1) and (2) of Appendix A, except that we have to replace $\|U(t)\| \leq 1$ by $\|V(t)\| \leq e^{-\|A\|t}$.

$V(t)$ can be constructed from $U(t)$ by the relations

$$\begin{aligned} V(t) &= \sum_{n=0}^{\infty} U_n(t), \\ U_0(t) &= U(t), \\ U_{n+1}(t) &= -\int_0^t U(t-s)AU_n(s)\,ds, \\ &\qquad \text{for} \quad n = 0, 1, 2, \cdots. \end{aligned} \tag{B3}$$

APPENDIX C

Theorem[3]*:* If $U_n(k\tau_n)$ is uniformly quasibounded, i.e.,

$$\|U_n(k\tau_n)\| \leq M e^{\beta k\tau_n} \tag{C1}$$

for all values of τ_n and k, M, and β independent of τ_n and k, and if $T_n \xrightarrow{s} T$ in the generalized sense [that is, T_n converges to T in the generalized sense, if $\hat{\delta}(T_n, T) \to 0$ for $n \to 0$; for the gap topology and the distance $\hat{\delta}$ between two operators, see Ref. 3], then $U_n(k\tau_n) \to U(t)$.

If $T_n \to T$ for a core of $D(T)$, then $T_n \to T$ in the generalized sense.

[1] B. Davidson, *Neutron Transport Theory* (Oxford University Press, London, 1957).

[2] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups*, American Mathematical Society Colloquium Publications 31 (American Mathematical Society, Providence, R.I., 1957).

[3] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1966).

[4] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems* (Interscience Publishers, Inc., New York, 1957).

[5] C. M. Wing, *An Introduction to Transport Theory* (John Wiley & Sons, Inc., New York, 1962).

[6] K. M. Case, F. de Hoffmann, and G. P. Placzek, *Introduction to the Theory of Neutron Diffusion* (U.S. Government Printing Office, Washington, D.C., 1953).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Application of Iteration Technique to the Integral Equations of Scattering Theory

P. DOLESCHALL AND J. RÉVAI
Central Research Institute for Physics, Budapest, Hungary

(Received 17 June 1969)

An iterative method based on the generalization of the Fredholm reduction is developed for the solution of the integral equations of the scattering theory.

In nearly all the fields of scattering theory in practical applications one faces the problem of solving integral equations.[1–5] The analytical treatment of these equations is possible only in a very few cases of special potentials. Therefore, one has to seek the ways of numerical solution.

There are two basic methods:

(a) the "discretization" of the integrals: that is, the substitution of the integrals with a finite sum using some quadrature formulas leading to a system of linear equations which has to be solved;

(b) iterative methods.

The first method is always applicable, though the singularities of the equations cause an extra problem, for which there are several methods of solution.[6–9] The accuracy of this method is finally determined by the number of terms in the sum standing for the integral. In the case of iteration, the accuracy can be increased, in principle, to any extent, but its application is allowed only under certain conditions which generally are not fulfilled. The advantages of this method would make it desirable to establish a general method with the aid of which the originally non-iterable integral equation could be substituted by an equivalent iterable one. In this paper, we try to solve this problem using a generalization of the Fredholm reduction[6] of singular integral equations.

In the quantum theory of scattering, the most frequently[3–5] appearing integral equations are those of the Lippman–Schwinger type, which can be written in the following form (omitting the energy parameter):

$$f(x) = c \cdot k(x, p_1) + \int k(x, y)t(y)f(y)\,dy, \quad (1)$$

where p_1 and c are energy-dependent constants.

The function $k(x, y)$ can generally be deduced from the potential and is assumed to be a square-integrable function. In the simplest case, $t(y)$ corresponds to the energy denominator $(p_1^2 - y^2 + i\epsilon)^{-1}$, but in a general case it can be more complicated, e.g., it can have poles at $p_1, \cdots, p_n$.

Let us write (1), for $x = p_i$, $i = 1, 2, \cdots, n$,

$$f(p_i) = c \cdot k(p_i, p_1) + \int k(p_i, y)t(y)f(y)\,dy. \quad (2)$$

If we now multiply (2) by a still unknown function $d_i(x)$, sum over i, and subtract the resulting equation from (1), we shall obtain

$$f(x) - \sum_{i=1}^{n} d_i(x) \cdot f(p_i) = c \cdot \left[k(x, p_1) - \sum_{i=1}^{n} d_i(x)k(p_i, p_1)\right] + \int \left(k(x, y) - \sum_{i=1}^{n} d_i(x) \cdot k(p_i, y)\right) t(y)f(y)\,dy. \quad (3)$$

Let us now choose the $d_i(x)$ in such a way that the expression in the round brackets under the integral would give zero at the poles p_j of $t(y)$:

$$d_i(x) = \sum_{j=1}^{n} k(x, p_j)M_{ji}^{-1}, \quad (4)$$

where

$$M_{ij} = k(p_i, p_j).$$

Here we must assume that $\det(M) \neq 0$. Substituting (4) into (3), we get the following equation:

$$f(x) = \sum_{i,j=1}^{n} k(x, p_i)M_{ij}^{-1}f(p_j) + \int L_n(x, y)t(y)f(y)\,dy, \quad (5)$$

where

$$L_n(x, y) = k(x, y) - \sum_{i,j=1}^{n} k(x, p_i)M_{ij}^{-1}k(p_j, y) = k(x, y) - k_n(x, y). \quad (6)$$

Thus, we have achieved the following: the kernel $L_n(x, y)t(y)$ of Eq. (5) is not singular. This procedure is called Fredholm reduction.

Let us now write $f(x)$ as

$$f(x) = \sum_{i,j=1}^{n} g_i(x)M_{ij}^{-1}f(p_j). \quad (7)$$

Thus, for the new unknown functions $g_i(x)$, we get the equations

$$g_i(x) = k(x, p_i) + \int L_n(x, y)t(y)g_i(y)\,dy. \qquad (8)$$

If we have the solutions $g_i(x)$ of (8), using (7) and (2), the original $f(x)$ can be determined.

It has to be noted that the existence and uniqueness problems arising in connection with these new equations have not been satisfactorily solved yet.[6]

The described procedure can be performed for an arbitrary sequence of points $x_1, x_2, \cdots, x_m$, containing the points $p_1, \cdots, p_n$, too. The function $L_m(x, y)$ obtained in this way will be zero along the $2m$ straight lines $x = x_i$, $y = x_i$, since at the same places $k(x, y) = k_m(x, y)$. Hence, it can be seen that the function $k_m(x, y)$, introduced in (6), coincides with the m-point Bateman approximation[10–12] of the function $k(x, y)$.

It can be shown[11] that a function of the form

$$K(x, y) = \sum_{i=1}^{m} A_i(x)B_i(y) \qquad (9)$$

and its m-point Bateman approximation are identical. Further, it is a well-known fact that any square-integrable kernel $k(x, y)$ can be approximated with so-called degenerate kernels of the form (9), if m is large enough. These arguments make it plausible that

$$\int |k(x, y) - k_m(x, y)|^2\,dx\,dy$$

can be made arbitrarily small if the mesh points x_i, $i = 1, 2, \cdots, m$, are suitably chosen and m is large enough. Then, it can be directly shown that

$$\int |L_m(x, y)t(y)|^2\,dx\,dy < 1. \qquad (10)$$

The relation (10) is a sufficient condition of iterability of Eq. (5).

Now, it has to be pointed out that, using the described method based on the combination of the generalized Fredholm reduction and the Bateman approximation, the question of existence and uniqueness is solved automatically since iterable Fredholm equations of the second kind always have one and only one solution.

The described method can be easily generalized for the case of multichannel processes, i.e., for systems of integral equations.

The practical applicability of this method is naturally very sensitive to the choice of the number and the positions of the points x_i. The numerical investigation of this method for the case of some practically important kernels is in progress.

ACKNOWLEDGMENT

The authors are indebted to Dr. A. Békéssy, Hungarian Academy of Sciences, Research Institute of Mathematics, for valuable discussions.

[1] B. A. Lippman and J. Schwinger, Phys. Rev. **79**, 469 (1950).
[2] L. D. Faddeev, *Publications of the Steklov Mathematical Institute, No. 69: Mathematical Problems of the Quantum Theory of Scattering for a Three-Particle System* (Publications of the Steklov Mathematical Institute, Leningrad, 1963).
[3] C. Bloch, *Lecture at International School of Physics "Enrico Fermi," Varenna, 1965, Course* XXXVI: *An Introduction to the Many Body Theory of Nuclear Reactions, 1965.*
[4] C. Lovelace, Phys. Rev. **135**, B1225 (1964).
[5] E. O. Alt, P. Grassberg, and W. Sandhas, Nucl. Phys. **B2**, 167 (1967).
[6] K. L. Kowalski and D. Feldman, J. Math. Phys. **4**, 507 (1963).
[7] J. H. Hetherington and L. H. Schick, Phys. Rev. **137**, B935 (1965).
[8] R. Aaron and R. D. Amado, Phys. Rev. **150**, 857 (1966).
[9] J. Raynal, M. A. Melkanoff, and T. Sawada, Nucl. Phys. **A101**, 369 (1967).
[10] H. Bateman, Proc. Roy. Soc. (London), Ser. A.,**100**, 441 (1922).
[11] L. V. Kantorovich and V. I. Krylov, *Approximation Methods of Higher Mathematical Analysis* (State Publisher of the Technical-Theoretical Literature, Moscow-Leningrad, 1950).
[12] V. B. Beljajev and E. Vsecionko, Preprint P4-4144 (Joint Institute for Nuclear Research, Dubna, 1969) (in Russian).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Irreducible Cartesian Tensors. II. General Formulation*

J. A. R. COOPE AND R. F. SNIDER
Department of Chemistry, University of British Columbia, Vancouver, Canada

(Received 28 October 1966)

A general formulation is given of a method of reduction of Cartesian tensors, by Cartesian tensor operations, to tensors irreducible under the three-dimensional rotation group. The criterion of irreducibility is that a tensor be representable as a traceless symmetric tensor, its reduced or natural form, invariantly embedded in the space of appropriate order. The general formulation exploits the properties of invariant linear mappings between tensor spaces. Considered abstractly, such mappings bring out the structure of the theory and illuminate the relation to spherical tensor theory. On the other hand, any linear invariant mapping between tensor spaces is equivalent to a combination of operations with the elementary invariant tensors U and ϵ. The general abstract formation therefore has a direct operational representation in terms of the ordinary tensor operations of contraction and permutation of indices. An analogous formulation is given for spinors, and the relations between spinors, Cartesian tensors, and spherical tensors is discussed in the language of the present formalism. Lastly, several examples are given as to how the general formalism may be applied to groups other than the rotation group.

I. INTRODUCTION

This paper is concerned with the decomposition of Cartesian tensors by purely Cartesian methods into tensors irreducible under the three-dimensional rotation group. The theory has a generality for integral j similar to spherical tensor theory[1] and, indeed, it can be transformed into the latter, However, it is independent of spherical tensor theory and is appropriately adapted to solve problems in Cartesian tensors.

Cartesian tensors remain popular in many problems involving spherical symmetry[2] in spite of the power and generality of spherical tensor methods. They arise naturally in the course of ordinary vector algebra, and it is often desirable to have the decomposition into irreducible tensors adapted to this vector algebra. Their relation to the properties of three-dimensional space is more apparent than with spherical tensors, where one axis is arbitrarily distinguished, and, furthermore, they can often be more easily handled, i.e., without the necessity of tables of numerical coefficients. Cartesian tensors are, however, generally used without an appropriate irreducible tensor analysis being available. Even the appropriate orthogonality relations for irreducible Cartesian tensors do not seem to be generally appreciated, although recently, several papers have appeared tending toward a general formulation.[3-5]

In a paper henceforth referred to as Part I,[6] it was shown how the classes of traceless symmetric tensors, there called natural tensors, provide a criterion of irreducibility which permits problems of Cartesian tensor reduction to be handled by purely Cartesian methods. A form of irreducible tensor analysis can in fact be developed for natural tensors fully as general as spherical tensor analysis (except limited to integral j) with an analogous apparatus of n-j symbols.[7] The use of natural tensors in the reduction of general Cartesian tensors depends on the fact that any irreducible Cartesian tensor can be regarded as a natural tensor embedded in the tensor space of appropriate order.

Part I gave a concrete construction of the reduction of tensors by tensorial operations. The key to an abstract formulation through which the general theory emerges, and also its relation to spherical tensor theory, is the representation of the relevant tensor operations by invariant linear mappings. In spherical tensor theory, the mappings produced by the 3-j symbols are examples of just such invariant mappings. However, whereas in spherical tensor theory it is natural to represent invariant tensors by the numerical values of their components, in Cartesian theory it is natural to represent them in terms of the elementary invariant tensors U and ϵ. As these are merely abstract representations of the vector operations "dot" and "cross" and of the permutation of indices, the mappings have a direct operational significance. This scheme leads to a considerable gain in clarity and compactness.

Section II establishes in a very general way the abstract properties of invariant mappings of tensor spaces $\mathfrak{T}^n$, where invariance is with respect to the behavior under a general group G of transformations acting on the vector space $\mathfrak{T}$. Much of this general formulation simplifies considerably when applied to the rotation group (Sec. III). The only property of the proper rotation group required is that there are only two elementary invariant tensors U and ϵ. Section IV establishes the properties of the natural projections $\mathsf{E}^{(j)}$ for traceless symmetric tensors, while Sec. V gives the general method for reducing tensors of arbitrary order. The general formulation of Sec. II is required in

Sec. VI, where the reduction of spinors is considered. Irreducible spinors are thus analyzed as embedded symmetric spinors, the latter being the natural form for spinors. Section VII discusses the relation between spinors, Cartesian tensors, and spherical tensors in the language of the present formalism, while Sec. VIII considers how the formalism can be applied to certain other vector spaces and other groups.

The Cartesian Clebsch–Gordan problem is essentially a special case of the reduction considered here and is treated in Part III.[7] With a proper normalization of the Cartesian 3-j tensors, the Cartesian 6-j, 9-j, etc., symbols, being scalars, coincide with those of spherical tensor theory. This alone implies that, when the final results are to be scalar quantities, it is unnecessary to transform Cartesian tensors into spherical tensors.

Practical problems arising in Cartesian tensor applications are frequently of low tensorial order. Convenient basis sets for tensors of order $n \leq 6$ together with their metric properties have been obtained and will be published subsequently.[8]

II. INVARIANT TENSORS AND THEIR MAPPING PROPERTIES

Given a ν-dimensional vector space $\mathfrak{X}$ over the real or complex field, there exists[9] another ν-dimensional vector space $\overline{\mathfrak{X}}$, the dual vector space, which is effectively the space of linear functionals on $\mathfrak{X}$. As ν is assumed to be finite, $\mathfrak{X}$ is also isomorphic to the space of linear functionals on $\overline{\mathfrak{X}}$. When the distinction is important, the elements of $\mathfrak{X}$ are called contravariant vectors, the components relative to a given basis being designated by superscript[10] indices, the dual vectors being covariant with subscript indices. For every pair of integers n, m there is a tensor space $\mathfrak{X}_m^n$ of order[11] (n, m) which can be thought of as the $(n + m)$-fold tensor product of $n\mathfrak{X}$ factors and $m\overline{\mathfrak{X}}$ factors.[9] In particular, any element $\mathsf{T}(n \mid m)$ of $\mathfrak{X}_m^n$ is equivalent to a linear mapping from $\mathfrak{X}^m$ into $\mathfrak{X}^n$. Thus, if $\mathbf{x}(m) \in \mathfrak{X}^m$, then[12] $\mathbf{y}(n) = \mathsf{T}(n \mid m) \odot^m \mathbf{x}(m)$ is an element of $\mathfrak{X}^n$. None of these properties depend on the existence of a metric for $\mathfrak{X}$. In fact, all of the tensor properties discussed in this section are independent of any metric.

It is assumed that a group G of linear transformations (automorphisms) of $\mathfrak{X}$ is given. Then a representation of G is given by a set of elements of $\mathfrak{X}_1^1$, the identity 1, in component form being δ_j^i with respect to some (arbitrary) basis $\{\mathbf{e}_i\}$ of $\mathfrak{X}$. G induces the group G^n of automorphisms of $\mathfrak{X}^n$ with representation in $\mathfrak{X}_n^n$. An element $\mathsf{R}^n = \mathsf{R} \times \mathsf{R} \times \cdots \times \mathsf{R}$ of G^n is the n-fold Kronecker product[13] of the corresponding group element $\mathsf{R} \in \mathfrak{X}_1^1$ with itself. A tensor $\mathsf{T}(n \mid 0) \in \mathfrak{X}^n$ is defined as invariant if $\mathsf{R}^n\mathsf{T}(n \mid 0) = \mathsf{T}(n \mid 0)$ for all $R^n \in G^n$. More generally, a tensor $\mathsf{T}(n \mid m) \in \mathfrak{X}_m^n$ is invariant if $\mathsf{R}^n\mathsf{T}(n \mid m) = \mathsf{T}(n \mid m)\mathsf{R}^m$. The properties of these invariant tensors as maps between tensor spaces, and between subspaces irreducible under G in particular, are considered in the following.

The complete group-theoretical reduction of $\mathfrak{X}^n$ consists in finding a set of linearly independent projections $\mathbf{\Pi}_{j\alpha}(n \mid n) \in \mathfrak{X}_n^n$, where $\mathbf{\Pi}_{j\alpha}(n \mid n)$ projects out the αth irreducible subspace $H_{j\alpha}^n \subset \mathfrak{X}^n$ of symmetry type[14] j. Let D_j denote the *dimension* if the representation and $N_n^{(j)}$ the *multiplicity*, i.e., the number of independent, equivalent representations of the same symmetry j into which $\mathfrak{X}^n$ decomposes[15] (α is a label for these $N_n^{(j)}$ independent, equivalent representations which, of course, are not at all unique). In many practical problems, it is sufficient to decompose $\mathfrak{X}^n$ only into the total $D_jN_n^{(j)}$-dimensional, invariant, but in general reducible, subspaces $H_j^n = \sum_\alpha H_{j\alpha}^n$ of given symmetry. However, it is the completely irreducible subspaces which behave simply under invariant mappings.

An irreducible tensor[16] $\mathsf{A}^{(j)}(r_1 \cdots r_n) \equiv \mathsf{A}^{(j)}(n)$ of order $(n, 0)$ and symmetry j has ν^n components but lies in an irreducible subspace $H_{j\alpha}^n$ such that only D_j of its components are independent. If $\mathbf{e}_\sigma(r_1 \cdots r_n) \equiv \mathbf{e}_\sigma(n)$ are a set of basis tensors spanning $H_{j\alpha}^n$, then $\mathsf{A}^{(j)}(n) = \sum_\sigma A^{(j)\sigma}\mathbf{e}_\sigma(n)$. The D_j independent components $A^{(j)\sigma}$ can easily be obtained by introducing the basis $\{\mathbf{e}^\sigma(n)\} \subset \mathfrak{X}_n$ dual[9] to $\{\mathbf{e}_\sigma(n)\}$, defined such that[12]

$$\mathbf{e}^\tau(n) \odot^n \mathbf{e}_\sigma(n) = \delta_\sigma^\tau. \tag{1}$$

With this set of dual basis tensors, the component $A^{(j)\sigma}$ of $\mathsf{A}^{(j)}(n)$ is simply $\mathbf{e}^\sigma(n) \odot^n \mathsf{A}^{(j)}(n)$. An analysis of the three-dimensional rotation group by spherical tensors considers exactly these components $A^{(j)m}$ with m labeling the weights[13] of the irreducible representation.

Any tensor $\mathsf{A}^{(j)}(n)$ in the total subspace H_j^n of symmetry j can be considered as a tensor of symmetry j, but in general it is reducible. As there can be at most[17] $\min(D_j, N_n^{(j)})$ linearly independent tensors each with D_j components, any such tensor of symmetry j can be written as a sum of at most $\min(D_j, N_n^{(j)})$ (not, in general, $N_n^{(j)}$) linearly independent irreducible tensors.

An invariant mapping is defined as a (linear) transformation of the type

$$\mathsf{B}(m) = \mathsf{T}^{(0)}(m \mid n) \odot^n \mathsf{A}(n), \tag{2}$$

where $\mathsf{T}^{(0)}(m \mid n)$ is an invariant tensor (of symmetry type 0) and of order (m, n). These mappings are used

for representing isomorphisms between tensorial subspaces of the same symmetry j and are basic both for the general reduction of Cartesian tensors and in the Clebsch–Gordan problem. It is mainly the general reduction which is considered in this paper, while the properties of the Cartesian 3-j tensors are discussed in the succeeding article.[7]

Although formally the invariant mappings are accomplished by tensors with $m + n$ indices, the effect of the mapping can often be very simply interpreted. Thus, for example, all invariant spinors can be represented as combinations of the spinors ϵ_{rs}, δ^r_s, and ϵ^{rs} which, respectively, contract the contravariant order, permute the indices, or expand the contravariant order. A particular mapping of order $(m, m + 2)$ might in fact be just a contraction of the first two indices. Again, for some groups such as $GL(\nu)$, there are no invariant tensors with $m \neq n$, and consequently there are no invariant mappings between tensors of different order.[18] Certain general properties of invariant mappings are stated as the following three theorems.

Theorem 1: A projection from $\mathfrak{T}^n$ onto an invariant subspace is an invariant mapping.

Proof: Let $\mathbf{\Pi} = \mathbf{\Pi}(n \mid n)$ be a projection onto an invariant, but not necessarily irreducible, subspace H having the basis $\{\mathbf{e}_\sigma(n)\}$. Then $\mathbf{\Pi}\mathbf{e}_\sigma = \mathbf{e}_\sigma$ and, since $\mathsf{R}^n\mathbf{e}_\sigma$ is also in H, it follows that $\mathbf{\Pi}\mathsf{R}^n\mathbf{e}_\sigma = \mathsf{R}^n\mathbf{e}_\sigma = \mathsf{R}^n\mathbf{\Pi}\mathbf{e}_\sigma$. Similarly if $\{\mathbf{e}'_\tau(n)\}$ is a basis for the invariant subspace $\bar{H}$ complementary to H, then $\mathbf{\Pi}\mathsf{R}^n\mathbf{e}'_\tau = 0 = \mathsf{R}^n\mathbf{\Pi}\mathbf{e}'_\tau$. Thus $\mathbf{\Pi}\mathsf{R}^n = \mathsf{R}^n\mathbf{\Pi}$ for all $\mathsf{R}^n \in G^n$, and this implies that $\mathbf{\Pi}$ is an invariant tensor.

For any tensor $\mathsf{A}(n)$ in H, the decomposition of $\mathsf{A}(n)$ into components A^σ follows the formula

$$\mathbf{\Pi}\mathsf{A}(n) = \mathsf{A}(n) = \sum_\sigma A^\sigma\mathbf{e}_\sigma(n) = \sum_\sigma \mathbf{e}_\sigma(n)\mathbf{e}^\sigma(n) \odot^n \mathsf{A}(n). \tag{3}$$

Thus $\mathbf{\Pi}$ has the spectral representation,

$$\mathbf{\Pi}(n \mid n) = \sum_\sigma \mathbf{e}_\sigma(n)\mathbf{e}^\sigma(n). \tag{4}$$

Theorem 2: The image H^m of an irreducible subspace $H^n_{j\alpha}$ under any nonnull invariant map $\mathsf{T}^{(0)}(m \mid n)$ from $H^n_{j\alpha}$ into $\mathfrak{T}^m$ is an irreducible subspace of the same symmetry j. Thus $\mathsf{T}^{(0)}(m \mid n)$ has an inverse invariant mapping $\mathsf{T}^{(0)}(n \mid m)$.[19] (The latter part of this theorem is identical to Wigner's[20] Theorem 3 while Boerner[13] calls it Schur's Lemma.)

Proof: Let $\{\mathbf{e}_\sigma(n)\}$ be a basis of $H^n_{j\alpha}$. Then

$$\mathbf{a}_\sigma(m) \equiv \mathsf{T}^{(0)}(m \mid n) \odot^n \mathbf{e}_\sigma(n) \tag{5}$$

forms a basis for H^m. If $U^{(j)\tau}_\sigma(R)$ is the irreducible representation of G to which $H^n_{j\alpha}$ belongs, then since $\mathsf{T}^{(0)}(m \mid n)$ is invariant,

$$\begin{aligned}\mathsf{R}^m\mathbf{a}_\sigma(m) &= \mathsf{T}^{(0)}(m \mid n) \odot^n \mathsf{R}^n\mathbf{e}_\sigma(n) \\ &= \sum_\tau [\mathsf{T}^{(0)}(m \mid n) \odot^n \mathbf{e}_\tau(n)]U^{(j)\tau}_\sigma(R) \\ &= \sum_\tau \mathbf{a}_\tau(m)U^{(j)\tau}_\sigma(R).\end{aligned} \tag{6}$$

That is, $\mathbf{a}_\sigma$ forms a basis for the jth irreducible representation of G. This implies first that, since $\mathsf{T}^{(0)}(m \mid n)$ is nonnull, then all of $\mathbf{a}_\sigma$ are nonnull and, secondly, that the dimension of H^m is exactly D_j, the same as $H^n_{j\alpha}$, and hence that H^m is an irreducible subspace $H^n_{j\beta}$ of symmetry j. This in turn requires that the mapping be one-to-one, so that an inverse map $\mathsf{T}^{(0)}(n \mid m)$ exists such that[19]

$$\mathsf{T}^{(0)}(n \mid m) \odot^m \mathsf{T}^{(0)}(m \mid n) = \mathbf{\Pi}_{j\alpha}(n \mid n) \tag{7a}$$

and

$$\mathsf{T}^{(0)}(m \mid n) \odot^n \mathsf{T}^{(0)}(n \mid m) = \mathbf{\Pi}_{j\beta}(m \mid m), \tag{7b}$$

since the projection $\mathbf{\Pi}_{j\alpha}(n \mid n)$ is the identity in $H^n_{j\alpha}$, etc. Obviously $\mathsf{T}^{(0)}(n \mid m)$ is invariant since $\mathsf{T}^{(0)}(m \mid n)$ and $\mathbf{\Pi}_j(n \mid n)$ are.

Note that if the domain subspace $H^n_{j\alpha}$ were invariant but not irreducible (e.g., H^n_j) then, in general, the $\mathbf{a}_\sigma(m)$ could be linearly dependent since some of the irreducible subspaces contained in H^n_j could be mapped onto the same irreducible subspace in $\mathfrak{T}^m$.

Theorem 3: An invariant mapping $\mathsf{T}^{(0)}(n \mid n)$ of an irreducible subspace $H^n_{j\alpha}$ into itself leaves all tensors in $H^n_{j\alpha}$ unchanged except for a constant factor. (This is a form of Schur's Lemma.[20,13])

Proof: $\mathsf{T}^{(0)}(n \mid n)$ has the matrix representation

$$T^{(0)\sigma}_\tau \equiv \mathbf{e}^\sigma(n) \odot^n \mathsf{T}^{(0)}(n \mid n) \odot^n \mathbf{e}_\tau(n), \tag{8}$$

which commutes with all elements of the irreducible representation $U^{(j)\sigma}_\tau(R)$ because $\mathsf{T}^{(0)}(n \mid n)$ is invariant. By Schur's Lemma, $T^{(0)\sigma}_\tau = \lambda\delta^\sigma_\tau$ and consequently, for any tensor $\mathsf{A}(n)$ in $H^n_{j\alpha}$,

$$\begin{aligned}\mathsf{T}^{(0)} \odot^n \mathsf{A}(n) &= \sum_\sigma \mathsf{T}^{(0)} \odot^n \mathbf{e}_\sigma(n)A^\sigma \\ &= \lambda \sum_\sigma \mathbf{e}_\sigma(n)A^\sigma = \lambda\mathsf{A}(n).\end{aligned} \tag{9}$$

Another way of expressing this result is that $\mathsf{T}^{(0)}(n \mid n)$ is a multiple of the projection operator $\mathbf{\Pi}_{j\alpha}(n \mid n)$. In this statement it is again assumed that the domain and range of $\mathsf{T}^{(0)}(n \mid n)$ are precisely $H^n_{j\alpha}$. A more general result is that, for any invariant tensor $\mathsf{T}^{(0)}(n \mid n)$

mapping any subspace of $\mathfrak{X}^n$ into $\mathfrak{X}^n$,

$$\mathbf{\Pi}_{j\alpha}(n \mid n) \odot^n \mathsf{T}^{(0)}(n \mid n) \odot^n \mathbf{\Pi}_{j\alpha}(n \mid n) = \lambda \mathbf{\Pi}_{j\alpha}(n \mid n). \tag{10}$$

Theorem 2 is the basis of our method of Cartesian tensor analysis. The tensors $\mathsf{T}^{(0)}(m \mid n)$ and $\mathsf{T}^{(0)}(n \mid m)$ describe the isomorphism between $H^n_{j\alpha}$ and $H^m_{j\beta}$. Thus, if $\mathsf{A}^{(j)}(n) \in H^n_{j\alpha}$, then

$$\mathbf{a}^{(j)}(m) \equiv \mathsf{T}^{(0)}(m \mid n) \odot^n \mathsf{A}^{(j)}(n) \tag{11a}$$

describes an equivalent tensor in $H^m_{j\beta}$, while the inverse mapping is

$$\mathsf{A}^{(j)}(n) = \mathsf{T}^{(0)}(n \mid m) \odot^m \mathbf{a}^{(j)}(m). \tag{11b}$$

In general, there will be a least contravariant order for which contravariant tensors of a given symmetry can occur. Equations (11) imply that any irreducible tensor $\mathsf{A}^{(j)}$ can be represented as a tensor $\mathbf{a}^{(j)}$ of minimal order and thus simpler structure. The tensor $\mathsf{A}^{(j)}$ itself can be considered as consisting of the tensor $\mathbf{a}^{(j)}$ suitably embedded [by $\mathsf{T}^{(0)}(n \mid m)$] in the higher-order n space. Thus $\mathbf{a}^{(j)}$ will be called the reduced or natural form for the tensor. The reduced matrix element occurring in the Wigner–Eckart theorem is a special case of a scalar which is embedded to form an invariant tensor. More generally, any tensor $\mathsf{A}^{(j)}(n)$ of symmetry j will have an expansion

$$\mathsf{A}^{(j)}(n) = \sum_p \mathsf{T}^{(0)}_p(n \mid m) \odot^m \mathbf{a}^{(j)}_p(m), \tag{12}$$

where the sum is over the possible number

$$\min\,(D_j, N^{(j)}_n)$$

of independent irreducible tensors of symmetry j. In Sec. V, it is shown that there is a distinguished representation of $\mathsf{A}^{(j)}(n)$ in this form.

III. IRREDUCIBLE CARTESIAN TENSORS FOR THE PROPER ROTATION GROUP

We now specialize to the three-dimensional proper rotation group $SO(3)$ with $\mathfrak{X}$ three dimensional. The only information about the rotation group that is required is that the only elementary invariant tensors are the metric tensor U with components δ_{rs} whose invariance defines the rotation group and the anti-symmetric third-order tensor $\boldsymbol{\epsilon}$, with components ϵ_{rst} equal to 1, -1, or 0 according to whether rst is an even or odd permutation of 123 or not a permutation of 123, whose invariance implies that the rotations be proper.[21] The distinction between contravariant and covariant tensors is irrelevant in this case since the tensorial components are unchanged by lowering or raising indices with δ_{rs}.

All other invariant tensors can be constructed as linear combinations of tensor products of U and $\boldsymbol{\epsilon}$. An essentially equivalent statement is that the only "vector operations" are the dot and cross products. This fact is presumably "well known" but it can be proved inductively after showing that the Cartesian or outer product of an irreducible tensor with a single vector can be reduced by means of U and $\boldsymbol{\epsilon}$ alone.[6] By the usual parentage scheme, it is then possible to construct explicitly from U and $\boldsymbol{\epsilon}$ alone a set of invariant tensors which completely span the invariant subspace H^n_0 of $\mathfrak{X}^n$ for any n. Furthermore, the relation[22]

$$\begin{aligned}\epsilon_{rst}\epsilon_{uvw} = {} & \delta_{ru}\delta_{sv}\delta_{tw} + \delta_{rv}\delta_{sw}\delta_{tu} + \delta_{rw}\delta_{su}\delta_{tv} \\ & - \delta_{ru}\delta_{sw}\delta_{tv} - \delta_{rv}\delta_{su}\delta_{tw} - \delta_{rw}\delta_{sv}\delta_{tu}\end{aligned} \tag{13}$$

implies that all even-order invariant tensors, and particularly all projections into irreducible subspaces, can be constructed from U alone. Furthermore, in any invariant tensor, $\boldsymbol{\epsilon}$ need only appear linearly.

Corresponding to the discussion in Sec. II, it is now shown that the natural form for a tensor of symmetry j is a symmetric traceless tensor and that any irreducible tensor $\mathsf{A}^{(j)}(n) \in H^n_{j\alpha}$ can be represented as a symmetric traceless tensor embedded in $\mathfrak{X}^n$: If $\mathsf{A}^{(j)}(n)$ is not traceless symmetric, its order may be reduced by applying the contraction $\mathsf{U}\odot^2$ or the "cross-product relation" $\boldsymbol{\epsilon}\odot^2$ to at least one pair of indices. Each contraction lowers the order by two, whereas each cross-product lowers the order by one. By Theorem 2, the group symmetry type j is unchanged by this procedure, which ends only when $\mathsf{A}^{(j)}(n)$ has been reduced (by an invariant mapping) to a completely symmetric, traceless tensor $\mathbf{a}^{(j)}(m)$, i.e., symmetric and traceless in each pair of indices. Since the mapping of $\mathsf{A}^{(j)}(n)$ into traceless symmetric form is an invariant mapping from $H^n_{j\alpha}$ onto the image space $H^m \subset \mathfrak{X}^m$, it has, by Theorem 2, an inverse. Thus $\mathsf{A}^{(j)}(n)$ is represented [Eq. (11b)] as $\mathbf{a}^{(j)}(m)$, a symmetric traceless tensor, embedded in $\mathfrak{X}^n$.

The subspace of traceless symmetric tensors of any order is thus necessarily irreducible and of multiplicity $N^{(j)}_n = 1$. This last comment can be proved in the following interesting way. Since any invariant mapping from $\mathfrak{X}^n$ to $\mathfrak{X}^n$ is accomplished by an even-order tensor, it is a combination of U's alone. Since U can act either as a contraction, which causes a traceless symmetric tensor to vanish, or a permutation of indices, any nonnull invariant mapping $\mathsf{T}^{(0)}(n \mid n)$ acting on a symmetric traceless tensor is merely a permutation of indices and, hence, equivalent to a multiple of the identity map for symmetric traceless tensors. Hence there is one and only one irreducible

subspace of symmetry j whose tensors are symmetric traceless. This uniqueness implies that the symmetry number j can be assigned to an irreducible representation as precisely the order of the natural tensor. Since symmetric traceless tensors of order j have exactly $2j + 1$ linearly independent components,[6] the designation coincides with the usual one for the rotation group with all integral j representations being realized.[23] This discussion can be summarized by saying that any irreducible tensor $\mathsf{A}^{(j)}(n)$ can be uniquely represented in the form

$$\mathsf{A}^{(j)}(n) = \mathsf{T}^{(0)}(n \mid j) \odot^j \mathbf{a}^{(j)}(j), \tag{14}$$

where $\mathbf{a}^{(j)}(j)$ is symmetric traceless.

The Casimir operator $\mathcal{J}^2(n \mid n)$ is a special invariant mapping from $\mathfrak{X}^n$ onto $\mathfrak{X}^n$. It is formed as the sum of the squares of the Hermitian infinitesimal generators[24] $\mathcal{J}_x(n \mid n)$, $\mathcal{J}_y(n \mid n)$, $\mathcal{J}_z(n \mid n)$ in $\mathfrak{X}^n$ for rotations about the x, y, and z directions. For example, $\mathcal{J}_z(n \mid n)$ is a sum of n terms, one for each tensorial index, of the form[25] $\mathsf{U} \otimes \mathsf{U} \otimes \cdots \otimes \mathcal{J}_z(1 \mid 1) \otimes \cdots \otimes \mathsf{U}$, where $\mathcal{J}_z(1 \mid 1)$ acting on $\mathfrak{X}$ has the well-known matrix representation $\mathcal{J}_z(r \mid s) = -i\epsilon_{zrs}$. With the relations $\epsilon_{rst}\epsilon_{ruv} = \delta_{su}\delta_{tv} - \delta_{sv}\delta_{ut}$, the explicit form for $\mathcal{J}^2(n \mid n)$ becomes

$$\mathcal{J}^2(n \mid n) = 2n + 2\sum_{i<j}(P_{ij} - \delta_{ij}\,\mathrm{Tr}_{ij}), \tag{15}$$

where P_{ij} is a permutation of the ith and jth indices of $\mathfrak{X}^n$, while $\delta_{ij}\,\mathrm{Tr}_{ij}$ means,[26] "Take the contraction (trace) between the ith and jth indices, then introduce a Krönecker δ between these same indices." It is possible to reduce $\mathfrak{X}^n$ into its total invariant subspaces $H_j^n = \sum_\alpha H_{j\alpha}^n$ of given symmetry j by looking for the eigentensors of[27] $\mathcal{J}^2(n \mid n)$. Orthogonal projections into the H_j^n can also be constructed by Löwdin's procedure.[28] However, the use of $\mathcal{J}^2(n \mid n)$ is surprisingly complicated, indeed more complicated than the methods developed in this article. One result that is easily confirmed is that any symmetric traceless tensor $\mathsf{A}(n)$ of order n satisfies

$$\mathcal{J}^2(n \mid n)\mathsf{A}(n) = n(n+1)\mathsf{A}(n), \tag{16}$$

since all traces are zero, and each of the $\frac{1}{2}n(n-1)$ permutations reproduces $\mathsf{A}(n)$.

The Casimir operator for tensors may be contrasted to that for spinors,[29]

$$\mathcal{J}^2(n \mid n) = \tfrac{1}{4}n(4-n) + \sum_{i<j}P_{ij}. \tag{17}$$

The absence of the trace term in the spinor operator corresponds to the fact that irreducible spinors under the rotation group can be completely classified by the symmetric group, whereas the reduction of Cartesian tensors by means of the symmetric group requires a preliminary separation into traceless, double traceless, etc., parts according to the classical procedure.[30]

IV. NATURAL PROJECTIONS

The projection of $\mathfrak{X}^j$ onto the irreducible subspace H_j^j of traceless symmetric tensors of order j is denoted by $\mathsf{E}^{(j)} = \mathsf{E}(j \mid j)$. This *natural projection* is the essential "unit tensor" of "natural tensor analysis," equivalent to $\delta_{m'}^{m}$ of spherical tensor theory. Theorem 1 implies that $\mathsf{E}^{(j)}$ is invariant, of order $2j$ and, by the spectral resolution [Eq. (4)], separately symmetric traceless in both sets of j indices, since by definition it is symmetric traceless in one set. The converse is important.

Theorem 4: An invariant tensor $\mathsf{T}^{(0)}(j_1 \mid j_2)$, separately symmetric traceless in each set of indices, is a multiple of $\delta_{j_1,j_2}\mathsf{E}^{(j_1)}$.

Proof: Since the subspace H_j^j of traceless symmetric tensors is irreducible of symmetry j, it follows that $\mathsf{T}^{(0)}(j_1 \mid j_2)$ is an invariant mapping from $H_{j_2}^{j_2}$ onto $H_{j_1}^{j_1}$. By Theorem 2, this is nonnull only if $j_1 = j_2$, in which case, by Theorem 3, $\mathsf{T}^{(0)}(j_1 \mid j_1)$ is a multiple of the identity in $H_{j_1}^{j_1}$, namely $\mathsf{E}^{(j_1)}$.

The explicit construction of $\mathsf{E}^{(j)}$ is aided by introducing a notation suitable for handling sets of symmetrized indices. Since each member of a set of symmetrized indices is equivalent to every other, it is redundant to label them differently. Thus each index of a set of p symmetrized indices, say $r_1 \cdots r_p$, may be replaced by the generic symbol r with the understanding that when working with the tensor, the p r-indices must be symmetrized, e.g.,

$$\begin{aligned}
(\delta_{r,s})^2 &= \tfrac{1}{2}(\delta_{r_1s_1}\delta_{r_2s_2} + \delta_{r_1s_2}\delta_{r_2s_1}),\\
(\delta_{r,r})^2 &= \tfrac{1}{3}(\delta_{r_1r_2}\delta_{r_3r_4} + \delta_{r_1r_3}\delta_{r_2r_4} + \delta_{r_1r_4}\delta_{r_2r_3}),\\
\delta_{r,r}\delta_{r,s} &= \tfrac{1}{3}(\delta_{r_1r_2}\delta_{r_3s_1} + \delta_{r_1r_3}\delta_{r_2s_1} + \delta_{r_2r_3}\delta_{r_1s_1}).
\end{aligned} \tag{18}$$

In carrying out contractions on a pair of indices, one simply considers the different kinds of terms in the final result and then calculates each numerical coefficient as the fractional number of ways in which the particular term may arise.

$\mathsf{E}^{(j)}(r_1 \cdots r_j \mid s_1 \cdots s_j)$ is symmetric in both the r and s indices as well as to the interchange $r \leftrightarrow s$. Thus it can be expanded in the form

$$\mathsf{E}^{(j)} = \sum_{t=0}^{[\frac{1}{2}j]} c_t(\delta_{r,s})^{j-2t}(\delta_{r,r})^t(\delta_{s,s})^t, \tag{19}$$

where $[\frac{1}{2}j] = \frac{1}{2}j$ or $\frac{1}{2}(j-1)$ according to whether j is even or odd. The leading term is simply a symmetrizer, the next takes single traces, the third double

traces, and so on. The uniqueness implied by Theorem 4 implies that the successive terms are linearly independent. The coefficients c_t are thus determined by the condition that $\mathsf{E}^{(j)}$ is traceless in the r (and hence also in the s) indices. This condition is obtained by setting the trace over a pair of r indices in Eq. (19) equal to zero, namely,

$$\sum_{t=0}^{[\frac{1}{2}j]} c_t\{(j-2t)(j-2t-1)\delta_{r,s}^{j-2t-2}\delta_{r,r}^{t}\delta_{s\,s}^{t+1} + 2t(2j-2t+1)\delta_{r,s}^{j-2t}\delta_{r,r}^{t-1}\delta_{s,s}^{t}\} = 0, \quad (20)$$

which, by the linear independence of the different trace terms, leads to the recursion formula,

$$c_t = -\frac{(j-2t+2)(j-2t+1)}{2t(2j-2t+1)}c_{t-1}. \quad (21)$$

This is identical to the recursion formula for the coefficients of the Legendre polynomials. With the choice $c_0 = 1$ to make $\mathsf{E}^{(j)}$ idempotent,

$$c_t = (-1)^t\frac{(j!)^2}{(2j!)}\binom{j}{t}\binom{2j-2t}{j}. \quad (22)$$

Table I lists some of the $\mathsf{E}^{(j)}$ explicitly. The coefficients for the Legendre polynomials differ only in normalization and, in fact, the Legendre polynomial is given by

$$P_n(\hat{u}\cdot\hat{v}) = \frac{(2n)!}{2^n(n!)^2}(\hat{u})^n \odot^n \mathsf{E}^{(n)} \odot^n (\hat{v})^n, \quad (23)$$

where $\hat{u}$, $\hat{v}$ are unit vectors and $(\hat{u})^n$ is the decomposable tensor[9] $\hat{u}\otimes\hat{u}\otimes\cdots\otimes\hat{u}$. In particular, if $\hat{v}$ is in the $\hat{z}$ direction, Eq. (23) identifies the $\hat{z}\cdots\hat{z}$ component of $\mathsf{E}^{(n)}\odot^n(\hat{v})^n$ (except for normalization) with the spherical harmonic $Y_n^0(\hat{v})$. The $\mathsf{E}^{(j)}$ satisfy the recursion relation

$$\mathsf{E}^{(j+1)} - \delta_{r,s}\mathsf{E}^{(j)} + j^2(4j^2-1)^{-1}\mathsf{E}^{(j-1)}\delta_{r,r}\delta_{s,s} = 0, \quad (24)$$

differing essentially only in normalization from the Legendre polynomial recursion relation.

A j-fold contraction between r and s indices of $\mathsf{E}^{(j)}$ (j-fold trace) gives the rank[11] of the projection $\mathsf{E}^{(j)}$, i.e., the dimension of the representation

$$\mathrm{Tr}_j\,\mathsf{E}^{(j)} = 2j+1, \quad (25)$$

while a single trace (between r and s indices) gives a a multiple of $\mathsf{E}^{(j-1)}$,

$$\mathrm{Tr}_1\,\mathsf{E}^{(j)} = (2j+1)(2j-1)^{-1}\mathsf{E}^{(j-1)}. \quad (26)$$

Theorem 4 implies that the invariant part of the tensor product of two natural, i.e., traceless symmetric, tensors $\mathsf{A}^{(j)}(j)$, $\mathsf{B}^{(j')}(j')$ vanishes if $j\neq j'$ and is a multiple of $\mathsf{E}^{(j)}$ if $j'=j$; i.e.,

$$(\mathsf{A}^{(j)}(j)\otimes\mathsf{B}^{(j')}(j'))^{(0)} = \lambda\delta_{jj'}\mathsf{E}^{(j)}. \quad (27)$$

Using Eq. (25), one has

$$\mathsf{E}^{(j)}\odot^{2j}(\mathsf{A}^{(j)}(j)\otimes\mathsf{B}^{(j)}(j)) = \mathsf{A}^{(j)}(j)\odot^j\mathsf{B}^{(j)}(j) = (2j+1)\lambda, \quad (28)$$

from which λ is obtained and, finally,

$$(\mathsf{A}^{(j)}(j)\otimes\mathsf{B}^{(j')}(j'))^{(0)} = \frac{\mathsf{A}^{(j)}(j)\odot^j\mathsf{B}^{(j)}(j)}{2j+1}\delta_{jj'}\mathsf{E}^{(j)}. \quad (29)$$

This equation contains the orthogonality relations for natural (irreducible) tensors. Since the integral of a tensor over angles is just the isotropic (invariant) part times the volume integral, Eq. (29) implies that

$$\int\mathsf{A}^{(j)}(j)\otimes\mathsf{B}^{(j')}(j')\,d\Omega = 4\pi(2j+1)^{-1}\mathsf{A}^{(j)}(j)\odot^j\mathsf{B}^{(j)}(j)\delta_{jj'}\mathsf{E}^{(j)}. \quad (30)$$

In the quantum-mechanical treatment of angular momentum, the quantum trace Tr_Q takes the place of the volume integral so that if $\mathsf{A}^{(j)}(j)$ and $\mathsf{B}^{(j')}(j')$ are quantum-mechanical operators,

$$\mathrm{Tr}_Q\,\mathsf{A}^{(j)}(j)\otimes\mathsf{B}^{(j')}(j') = (2J+1)(2j+1)^{-1}\mathsf{A}^{(j)}(j)\odot^j\mathsf{B}^{(j)}(j)\delta_{jj'}\mathsf{E}^{(j)}, \quad (31)$$

where $[J(J+1)]^{\frac{1}{2}}\hbar$ is the magnitude of the angular momentum (note that[23] J has no relation to j). The $\mathsf{A}^{(j)}(j)\odot^j\mathsf{B}^{(j)}(j)$ is, since it is scalar, essentially the same as the corresponding quantity in spherical-tensor theory.

Natural Cartesian tensors constructed from a decomposable tensor and, in particular, those from a single vector $\mathbf{a}$,

$$\bar{\mathsf{Y}}^{(j)}(\mathbf{a}) \equiv \mathsf{E}^{(j)}\odot^j(\mathbf{a})^j, \quad (32)$$

are used quite widely, and have recently been discussed by several authors.[3,4] The bar ($\bar{Y}^{(j)}$) is used to emphasize the normalization $c_0 = 1$, while $\bar{P}_j(x)$ is the Legendre polynomial with the same normalization.

TABLE I. Natural projections for traceless symmetric tensors.

Order j	$\mathsf{E}(j\,\vert\,j)$
0	1
1	δ_{rs}
2	$(\delta_{rs})^2 - \frac{1}{3}\delta_{rr}\delta_{ss}$
3	$(\delta_{rs})^3 - \frac{3}{5}\delta_{rs}\delta_{rr}\delta_{ss}$
4	$(\delta_{rs})^4 - \frac{6}{7}(\delta_{rs})^2\delta_{rr}\delta_{ss} + \frac{3}{35}(\delta_{rr})^2(\delta_{ss})^2$
j	$(\delta_{rs})^j - \frac{j(j-1)}{2(2j-1)}(\delta_{rs})^{j-2}\delta_{rr}\delta_{ss} + \cdots$

One has

$$\bar{Y}^{(j)}(\mathbf{a}) \odot^{(j)} \bar{Y}^{(j)}(\mathbf{b}) = (\mathbf{a})^j \odot^j \bar{Y}^{(j)}(\mathbf{b}) = \bar{P}_j(\mathbf{a}\cdot\mathbf{b}) \quad (33)$$

as the spherical harmonic addition theorem, almost a trivial result in Cartesian form. Hence the Cartesian spherical harmonics $\bar{Y}^{(j)}(\hat{u})$ as such,[31] for a unit vector $\hat{u}$, satisfy the orthogonality relations,

$$\int \bar{Y}^{(j)}(\hat{u})\bar{Y}^{(j')}(\hat{u})\, d\hat{u} = 2^j(j!)^2[(2j)!]^{-1}4\pi(2j+1)^{-1}\delta_{jj'}\mathsf{E}^{(j)}, \quad (34)$$

with the numerical factor equal to $\bar{P}_j(1)$. The $\bar{Y}^{(j)}(\hat{u})$ can be used in the usual way as a complete orthogonal (but not normalized) set of angular functions for use in expanding an arbitrary function.[32] When $\mathbf{a}$ and $\mathbf{b}$ are quantum operators, commutation relations complicate the scalar $\bar{Y}^j(\mathbf{a})\cdot\bar{Y}^j(\mathbf{b})$ and lead, e.g., to the "quantum Legendre polynomials" of Kramers'.[4,33]

V. REDUCTION OF nth-ORDER TENSORS

An invariant mapping $\mathsf{T}^{(0)}_p(n\,|\,j)$, from the space H^j_j of symmetric traceless tensors of order j into $H^n_j \subset \mathfrak{X}^n$, defines a $(2j+1)$-dimensional subspace of H^n_j. Since H^n_j is $(2j+1)N^{(j)}_n$ dimensional and of symmetry j, there exist exactly $N^{(j)}_n$ independent invariant mappings from H^j_j into H^n_j. Each of these invariant mappings $\mathsf{T}^{(0)}_p(n\,|\,j)$ can be made symmetric traceless in the j indices, and it is assumed in the following that this has already been done; thus $\mathsf{T}^{(0)}_p(n\,|\,j)\odot^j \mathsf{E}^{(j)} = \mathsf{T}^{(0)}_p(n\,|\,j)$. Acting on $\mathfrak{X}^j$, such a tensor first selects out the subspace H^j_j and maps this into the subspace H^n_j. Consequently, if $\mathbf{\Pi}_j(n\,|\,n)$ is the projection operator into H^n_j, then

$$\mathbf{\Pi}_j(n\,|\,n)\odot^n \mathsf{T}^{(0)}_p(n\,|\,j) = \mathsf{T}^{(0)}_p(n\,|\,j).$$

The advantage of having such maps is that one can very easily construct tensors in H^n_j without having to assure, by methods in $\mathfrak{X}^n$, whether or not such tensors are, in fact, in H^n_j.

If $\mathsf{F}(n)$ is any nth-order tensor and $\{\mathsf{T}^{(0)}_p(n\,|\,j)\}$, $p = 1, 2, \cdots, N^{(j)}_n$, is a given linearly independent set of invariant maps from H^j_j into H^n_j, then the symmetry-j part of $\mathsf{F}(n)$, namely $\mathsf{F}^{(j)}(n)$, can be represented as

$$\mathsf{F}^{(j)}(n) = \sum_{p=1}^{N^{(j)}_n} \mathsf{T}^{(0)}_p(n\,|\,j)\odot^j \mathbf{f}^{(j)p}(j), \quad (35)$$

in terms of $N^{(j)}_n$ tensors $\mathbf{f}^{(j)p}(j)$ of order j. Now $\{\mathsf{T}^{(0)}_p(n\,|\,j)\}$ spans that subspace of $\mathfrak{X}^n_j$ consisting of invariant tensors symmetric and traceless in the j indices. If this set is used to generate a basis for $\mathfrak{X}^n_j$, then there is[9] a dual basis in $\mathfrak{X}^j_n$ whose invariant elements $\{\mathsf{T}^{(0)q}(j\,|\,n)\}$ satisfy[34]

$$\mathsf{T}^{(0)q}(j\,|\,n)\odot^{n+j}\mathsf{T}^{(0)}_p(n\,|\,j) = \delta^q_p(2j+1). \quad (36)$$

Since $\mathsf{T}^{(0)q}(j\,|\,n)\odot^n \mathsf{T}^{(0)}_p(n\,|\,j)$ is symmetric traceless in its covariant set of indices, it maps natural tensors of symmetry j into tensors of order j, which must then be symmetric traceless in their contravariant set of j indices. Consequently, by Theorem 4, this quantity is a multiple of $\mathsf{E}^{(j)}$ whose multiplicative factor can be deduced from a combination of Eqs. (25) and (36), so that

$$\mathsf{T}^{(0)q}(j\,|\,n)\odot^n \mathsf{T}^{(0)}_p(n\,|\,j) = \delta^q_p\mathsf{E}^{(j)}. \quad (37)$$

This identity allows an evaluation of the $\mathbf{f}^{(j)p}(j)$ tensors as

$$\mathbf{f}^{(j)p}(j) = \mathsf{T}^{(0)p}(j\,|\,n)\odot^n \mathsf{F}(n). \quad (38)$$

Equation (38) gives the extraction of the pth natural form tensor of symmetry j from $\mathsf{F}(n)$, while Eq. (35) gives their re-embedding in $\mathfrak{X}^n$. The invariant mappings $\mathsf{T}^{(0)}_p(n\,|\,j)$ and their duals define projections

$$\mathbf{\Pi}_{jp}(n\,|\,n) = \mathsf{T}^{(0)}_p(n\,|\,j)\odot^j \mathsf{T}^{(0)p}(j\,|\,n), \quad (39)$$

into the N^j_n irreducible subspaces H^n_{jp} whose direct sum is the total subspace H^n_j. They satisfy the idempotency relation

$$\mathbf{\Pi}_{jp}(n\,|\,n)\odot^n \mathbf{\Pi}_{j'q}(n\,|\,n) = \delta_{jj'}\delta_{pq}\mathbf{\Pi}_{jp}(n\,|\,n), \quad (40)$$

but are not necessarily orthogonal projections.

As no distinction exists between covariant and contravariant indices in the case of the rotation group, an n-fold dot product between $\mathsf{T}^{(0)}_p(n\,|\,j)$ and $\mathsf{T}^{(0)}_q(n\,|\,j)$ is defined and gives an invariant tensor, symmetric in both pairs of j indices, and hence a multiple of $\mathsf{E}^{(j)}$. Thus a metric g_{pq} is defined by

$$\mathsf{T}^{(0)}_p(n\,|\,j)\odot^n \mathsf{T}^{(0)}_q(n\,|\,j) = g_{pq}\mathsf{E}^{(j)}. \quad (41)$$

If g^{pq} are the components of the inverse metric satisfying $g^{pq}g_{qr} = \delta^p_r$, then

$$\mathsf{T}^{(0)p}(j\,|\,n) = \sum_q g^{pq}\mathsf{T}^{(0)}_q(n\,|\,j). \quad (42)$$

The resolution of the projection $\mathbf{\Pi}_j(n\,|\,n)$ into the total invariant subspace H^n_j is therefore

$$\mathbf{\Pi}_j(n\,|\,n) = \sum_{p,q} g^{pq}\mathsf{T}^{(0)}_p(n\,|\,j)\odot^j \mathsf{T}^{(0)}_q(n\,|\,j). \quad (43)$$

The metric g can also be used to express the scalar product of the symmetry-j part of two tensors $\mathsf{F}^{(j)}_1(n)$ and $\mathsf{F}^{(j)}_2(n)$,

$$\mathsf{F}^{(j)}_1(n)\odot^n \mathsf{F}^{(j)}_2(n) = \sum_{pq} g_{pq}\mathbf{f}^{(j)p}_1(j)\odot^j \mathbf{f}^{(j)q}_2(j), \quad (44)$$

in terms of their corresponding natural forms. The reduction of second- and third-order tensors using this formalism is given in Tables II and III, with some discussion of these cases in the Appendix.

TABLE II. Reduction of 2nd-order tensors by mapping from natural forms. Indices r refer to order n, s to order j.

j	$\mathsf{T}_p(n \mid j)$	g_{pq}	g^{pq}	$\mathsf{T}^p(j \mid n)$	$\Pi_j(n \mid n)$
2	$\mathsf{E}(2 \mid 2) = (\delta_{rs})^2 - \frac{1}{3}\delta_{rr}\delta_{ss}$	1	1	$\mathsf{E}(2 \mid 2)$	$\mathsf{E}(2 \mid 2)$
1	$\epsilon_{r_1r_2s}$	2	$\frac{1}{2}$	$\frac{1}{2}\epsilon_{sr_1'r_2'}$	$\frac{1}{2}(\delta_{r_1r_1'}\delta_{r_2r_2'} - \delta_{r_1r_2'}\delta_{r_1'r_2})$
0	$\delta_{r_1r_2}$	3	$\frac{1}{3}$	$\frac{1}{3}\delta_{r_1'r_2'}$	$\frac{1}{3}\delta_{r_1r_2}\delta_{r_1'r_2'}$

If the tensors $\{\mathsf{T}_p^{(0)}(n \mid j)\}$ have been chosen so that they are orthonormal in the sense $g_{pq} = \delta_{pq}$, then the extraction and re-embedding is particularly simple. This is always possible since g is, by Eq. (41), a symmetric, positive definite matrix. Orthogonal $\mathsf{T}_p(n \mid j)$'s are still arbitrary to the extent of a real orthogonal transformation, since such a transformation leaves the metric $g_{pq} = \delta_{pq}$ unchanged. Therefore, given any particular $\mathsf{F}(n)$, it is possible to make its natural forms orthogonal by diagonalizing the matrix $K^{pq} \equiv \mathbf{f}^{(j)p}(j) \odot^j \mathbf{f}^{(j)q}(j)$ by a real orthogonal transformation. Since the dimension of H_j^j is $(2j + 1)$, there can be at most $2j + 1$ orthogonal $\mathbf{f}^{(j)p}(j)$'s. Coupled with the fact that H_j^n is $(2j + 1)N_n^{(j)}$ dimensional, this means that at most $\min(2j + 1, N_n^{(j)})$ eigenvalues k_p of K^{pq} are nonzero (hence positive). Thus $\mathsf{F}^{(j)}(n)$ has a distinguished representation as a sum of $\min(2j + 1, N_n^{(j)})$ orthogonal natural tensors, separately embedded in different orthogonal subspaces H_{jp}^n, namely,

$$\mathsf{F}^{(j)}(n) = \sum_{p=1}^{\min(2j+1, N_n^{(j)})} \mathsf{T}_p^{(0)}(n \mid j) \odot^j \mathbf{f}^{(j)p}(j). \quad (45)$$

This representation is unique if the eigenvalues of K^{pq} are distinct. If, in particular, $\mathsf{F}^{(j)}(n)$ is irreducible, then there exists only one $\mathbf{f}^{(j)}$—that is, every invariant mapping $\mathsf{F}(n)$ into H_j^j gives a multiple of $\mathbf{f}^{(j)}$— and a unique embedding operator $\mathsf{T}^{(0)}(n \mid j)$, which reproduces $\mathsf{F}^{(j)}(n)$ from $\mathbf{f}^{(j)}$, can be found.

VI. ANALOGOUS REDUCTION OF SPINORS

The reduction of Cartesian tensors proceeded by mapping into the natural form tensors which are symmetric traceless tensors of appropriate order. Spinors can be similarly reduced by mapping into their natural forms which are symmetric spinors of appropriate order. The basic formalism is the same, as discussed in Sec. II, but whereas the distinction between the vector space and its dual is irrelevant for Cartesian tensors, the distinction is now very important. The emphasis on invariant spinors is closely related to Kramer's exploitation of "Spinor Invariants."[35]

The vector space $\mathfrak{X}$ is now two-dimensional and over the complex field. In agreement with standard terminology, we denote the two base vectors $\mathbf{e}_1(1)$ and $\mathbf{e}_2(1)$ by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. The dual basis is given by $\mathbf{e}^1(1) = \boldsymbol{\alpha}^*$ and $\mathbf{e}^2(1) = \boldsymbol{\beta}^*$; these elements are defined to satisfy the relations $\boldsymbol{\alpha}^* \odot \boldsymbol{\alpha} = \boldsymbol{\beta}^* \odot \boldsymbol{\beta} = 1$ and $\boldsymbol{\alpha}^* \odot \boldsymbol{\beta} = \boldsymbol{\beta}^* \odot \boldsymbol{\alpha} = 0$. In order to avoid confusion with the Cartesian tensor case, we will not write spinors in tensorially invariant form but only in component form relative to the basis $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$. Thus the spinor η^μ represents the abstract spinor $\boldsymbol{\eta} = \eta^\mu \mathbf{e}_\mu(1) = \eta^1\boldsymbol{\alpha} + \eta^2\boldsymbol{\beta}$.

The group of transformations is the two-dimensional special unitary group $SU(2)$, and it thus follows that there exists a unique (to a multiplicative factor) invariant metric,[36]

$$(\epsilon_{\mu\nu}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (46)$$

Besides this invariant covariant spinor and its inverse $\epsilon^{\mu\nu}$ having the same matrix form, the only other elementary invariant spinor is the identity transformation δ_ν^μ. By this it is meant that any invariant spinor can be written as a sum of tensor products of $\epsilon_{\mu\nu}$, $\epsilon^{\mu\nu}$, and δ_ν^μ, a property which can be proved by parentage.

TABLE III. Reduction of 3rd-order tensors by elementary contractions. Indices r refer to order j, s to order three.

Symmetry j	Elementary Contraction	$\mathsf{T}^p(j \mid n)$	g^{pq}	g_{pq}
3	none	$\mathsf{E}(3 \mid 3)$	1	1
2	$\epsilon_{rs_2s_3}$	$\mathsf{T}^1 = \Sigma_t \mathsf{E}^{(2)}(r_1r_2 \mid ts_1)\epsilon_{ts_2s_3}$	$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$	$\frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$
	$\epsilon_{rs_1s_3}$	$\mathsf{T}^2 = \Sigma_t \mathsf{E}^{(2)}(r_1r_2 \mid ts_2)\epsilon_{ts_1s_3}$		
1	$\delta_{s_2s_3}$	$\mathsf{T}^1 = \delta_{rs_1}\delta_{s_2s_3}$	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	$\frac{1}{10}\begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{pmatrix}$
	$\delta_{s_1s_3}$	$\mathsf{T}^2 = \delta_{rs_2}\delta_{s_1s_3}$		
	$\delta_{s_1s_2}$	$\mathsf{T}^3 = \delta_{rs_3}\delta_{s_1s_2}$		
0	$\epsilon_{s_1s_2s_3}$	$\mathsf{T} = \epsilon_{s_1s_2s_3}$	6	$\frac{1}{6}$

It is also equivalent to the theorem[37] that all spinor invariants can be constructed from combinations of the basic spinor invariants $\epsilon_{\mu\nu}u^\mu v^\nu = u^1v^2 - u^2v^1$. Finally, in view of the relation

$$\epsilon^{\mu\nu}\epsilon_{\sigma\tau} = \delta^\mu_\sigma\delta^\nu_\tau - \delta^\nu_\sigma\delta^\mu_\tau, \tag{47}$$

$\epsilon^{\mu\nu}$ and $\epsilon_{\mu\nu}$ need never both appear in the same invariant spinor.

The invariant metric $\epsilon_{\mu\nu}$ may be used to lower the order of a (contravariant) spinor. In particular, if a spinor $A^{(j)}(\mu_1\cdots\mu_n)$ is irreducible of symmetry j, it follows by Theorem 2 that $\epsilon_{\mu_l\mu_m}A^{(j)}(\mu_1\cdots\mu_n)$ either vanishes or is again of symmetry j. This process ends only if $A^{(j)}(\mu_1\cdots\mu_n)$ is a completely symmetric spinor, which is thus the natural or reduced form for spinors of symmetry j. By dimensionality or by applying the Casimir operator [Eq. (17)], it is seen that a natural spinor of symmetry j has order $2j$. Thus there exists an invariant spinor (basically a contraction) $T^{(0)}(2j \mid n)$ and its inverse $T^{(0)}(n \mid 2j)$ such that

$$a^{(j)}(2j) = T^{(0)}(2j \mid n) \odot^n A^{(j)}(n) \tag{48}$$

is in natural form, and

$$A^{(j)}(n) = T^{(0)}(n \mid 2j) \odot^{2j} a^{(j)}(2j) \tag{49}$$

describes the re-embedding of $a^{(j)}(2j)$ as an nth-order spinor.

The natural projection, i.e., for symmetric spinors, is

$$E^{(j)} \equiv E^{(j)}(2j \mid 2j) = (\delta^\mu_\nu)^{2j}, \tag{50}$$

where the previous symmetrization convention is used [see Eq. (18)]. As in the tensor case, a complete contraction between the μ and ν indices gives the dimension of the representation,

$$\mathrm{Tr}_{2j}\, E^{(j)} = 2j + 1, \tag{51}$$

whereas a single contraction gives

$$\mathrm{Tr}_1\, E^{(j)} = [(2j+1)/2j]E^{(j-\frac{1}{2})}. \tag{52}$$

The analog of Eq. (29), and indirectly of Theorem 4, is that the invariant part of the product of two symmetric contravariant spinors is

$$(A^{(j)}(\mu_1\cdots\mu_{2j})B^{(j')}(\mu_1\cdots\mu_{2j'}))^{(0)} = \frac{\boldsymbol{\epsilon}^{2j}\odot^{2j}A^{(j)}\otimes B^{(j)}}{2j+1}\,\delta_{jj'}(\epsilon^{\mu\nu})^{2j}. \tag{53}$$

As in Eqs. (30) and (31), the invariant part of $A^{(j)}\otimes B^{(j')}$ can be selected out in certain cases by integrating over angles, and in other cases by taking the quantum trace. Analogous results are valid for covariant spinors with $(\epsilon_{\mu\nu})^j$ replacing $(\epsilon^{\mu\nu})^j$, whereas with one covariant, one contravariant spinor $(\delta^\mu_\nu)^j = E^{(j)}$ appears. It is the latter that is strictly the analog of Eq. (29), all three results being simply contravariant-covariant forms of one another.

The reduction of nth-order spinors depends on the selection of a convenient set of $\sum_{j=0}^{\frac{1}{2}n} N^{(j)}_n$ linearly independent invariant spinors $T^{(0)}_p(n \mid 2j)$ symmetric in the $2j$ indices. For a given symmetry j, these describe the $N^{(j)}_n$ linearly independent ways of embedding a natural spinor of order $2j$ in the nth-order spinor space, whereas their duals $T^{(0)q}(2j \mid n)$, satisfying[34]

$$T^{(0)q}(2j \mid n) \odot^n T^{(0)}_p(n \mid 2j) = \delta^q_p E^{(j)}, \tag{54}$$

describe the extraction of these same natural spinors. A metric g_{pq} for the $T^{(0)}_p(n \mid 2j)$'s is defined by[38]

$$(\epsilon_{\mu\nu})^n_{\mathrm{NS}}T^{(0)}_p(\mu^n \mid \sigma^{2j})T^{(0)}_q(\nu^n \mid \tau^{2j}) = g_{pq}(\epsilon_{\sigma\tau})^{2j}, \tag{55}$$

and from its inverse g^{pq}, the dual basis

$$T^{(0)p}(\sigma^{2j} \mid \mu^n) = \sum_q g^{pq}(\epsilon_{\mu\nu})^n_{\mathrm{NS}}T^{(0)}_q(\nu^n \mid \tau^{2j})(\epsilon^{\sigma\tau})^{2j} \tag{56}$$

follows. The projection onto the total invariant nth-order spinor subspace of symmetry j is given by

$$\Pi_j(n \mid n) = \sum_p T^{(0)}_p(n \mid 2j) \odot^{2j} T^{(0)p}(2j \mid n). \tag{57}$$

Table IV gives the reduction of 4th-order spinors, wherein the embedding tensors are taken as simple tensor products of n-2 $j\epsilon^{\mu\nu}$'s. As an example, the

TABLE IV. Reduction of 4th-order spinors by elementary embeddings. Indices ν refer to $2j$th-order, μ to nth-order.

j	Embedding	$\mathrm{T}_p(4 \mid 2j)$	g_{pq}	g^{pq}
2	none	$(\delta^\mu_\nu)^4$	1	1
1	$\epsilon^{\mu_1\mu_4}$	$T_1 = \epsilon^{\mu_1\mu_4}\frac{1}{2}(\delta^{\mu_2}_{\nu_1}\delta^{\mu_3}_{\nu_2} + \delta^{\mu_3}_{\nu_1}\delta^{\mu_2}_{\nu_2})$	$\begin{pmatrix}2&1&1\\1&2&1\\1&1&2\end{pmatrix}$	$\frac{1}{4}\begin{pmatrix}3&-1&-1\\-1&3&-1\\-1&-1&3\end{pmatrix}$
	$\epsilon^{\mu_2\mu_4}$	$T_2 = \epsilon^{\mu_2\mu_4}\frac{1}{2}(\delta^{\mu_1}_{\nu_1}\delta^{\mu_3}_{\nu_2} + \delta^{\mu_2}_{\nu_1}\delta^{\mu_1}_{\nu_2})$		
	$\epsilon^{\mu_3\mu_4}$	$T_3 = \epsilon^{\mu_3\mu_4}\frac{1}{2}(\delta^{\mu_1}_{\nu_1}\delta^{\mu_2}_{\nu_2} + \delta^{\mu_2}_{\nu_1}\delta^{\mu_1}_{\nu_2})$		
0	$\epsilon^{\mu_1\mu_2}\epsilon^{\mu_3\mu_4}$	$T_1 = \epsilon^{\mu_1\mu_2}\epsilon^{\mu_3\mu_4}$	$\begin{pmatrix}4&2\\2&4\end{pmatrix}$	$\frac{1}{8}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}$
	$\epsilon^{\mu_2\mu_3}\epsilon^{\mu_4\mu_1}$	$T_2 = \epsilon^{\mu_2\mu_3}\epsilon^{\mu_4\mu_1}$		

projection onto the invariant spinors is

$$\Pi_0(\mu_1 \cdots \mu_4 \mid \nu_1 \cdots \nu_4) = \tfrac{1}{3}(\epsilon^{\mu_1\mu_2}\epsilon^{\mu_3\mu_4}\epsilon_{\nu_1\nu_2}\epsilon_{\nu_3\nu_4} + \epsilon^{\mu_2\mu_3}\epsilon^{\mu_4\mu_1}\epsilon_{\nu_2\nu_3}\epsilon_{\nu_4\nu_1} - \tfrac{1}{6}(\epsilon^{\mu_1\mu_2}\epsilon^{\mu_3\mu_4}\epsilon_{\nu_2\nu_3}\epsilon_{\nu_4\nu_1} + \epsilon^{\mu_2\mu_3}\epsilon^{\mu_4\mu_1}\epsilon_{\nu_1\nu_2}\epsilon_{\nu_3\nu_1}). \quad (58)$$

It is interesting to observe that the valence-bond method[39] of molecular quantum mechanics coincides in this formalism with the use of elementary embeddings to define the set of $T_p^{(0)}(n \mid 2j)$'s. Thus a valence-bond structure of spin s for an n-electron system is obtained by assigning $(n - 2s)$ electrons to $\frac{1}{2}(n - 2s)$ "bonds." Each bond is considered as a pair of spins, associated with atoms i and j coupled to spin zero in the manner $\alpha(i)\beta(j) - \beta(i)\alpha(j)$, while the remaining $2s$ electrons are coupled to maximum spin (s). In the present formalism the "bonds" are invariant spinors $\epsilon_{\mu_i\mu_j}$, and the state is a symmetric spinor of order $2s$ with an elementary embedding, i.e., through a tensor product with n-$2s$ $\epsilon^{\mu\nu}$'s. For example, a triplet state for four electrons is given by the embedding spinor

$$T_3^{(0)}(4 \mid 2) = \tfrac{1}{2}(\delta_{\nu_1}^{\mu_1}\delta_{\nu_2}^{\mu_2} + \delta_{\nu_1}^{\mu_2}\delta_{\nu_2}^{\mu_1})\epsilon^{\mu_3\mu_4}$$

which has components

$$\begin{aligned} T_3^{(0)}(4 \mid 2) &= \alpha\alpha \quad (\alpha\beta - \beta\alpha), \quad \nu_1 = \nu_2 = 1, \\ &= \tfrac{1}{2}(\alpha\beta + \beta\alpha)(\alpha\beta - \beta\alpha), \quad \nu_1 = -\nu_2 = 1, \\ &= \beta\beta \quad (\alpha\beta - \beta\alpha), \quad \nu_1 = \nu_2 = -1. \end{aligned} \quad (59)$$

If $\Psi(n)$ is the general spin state, the reduced spinors

$$\Psi_p(2j) = T^{(0)p}(2j \mid n) \odot^n \Psi(n)$$

are the expansion coefficients of Ψ with respect to the (unnormalized) valence-bond basis.

VII. RELATION TO SPHERICAL TENSORS

Spherical tensors, symmetric spinors, and, for integral j, traceless symmetric tensors are equivalent ways of describing tensorial quantities irreducible under the rotation group. The theories of all three can be developed independently, and we have attempted to emphasize the self-contained development of the Cartesian tensor theory. However, there are direct correspondences between them, and these are now given.

The correspondence between symmetric spinors and spherical tensors is particularly simple because, with the natural choice of spinor components, the first-order spinor is already a spherical tensor. Thus if[40]

$$e_m(\nu^{2j}) = \binom{2j}{j-m}^{-\frac{1}{2}} \delta_m^{3j-\Sigma_i\nu_i} \quad (60)$$

is a basis set for symmetric spinors of order $2j$, then the spherical tensor components of the symmetric spinor $A^{(j)}(\nu^{2j})$ are

$$A^{(j)m} = e^m(\nu^{2j}) \odot^{2j} A^{(j)}(\nu^{2j}), \quad (61)$$

where $\{e^m(\nu^{2j})\}$ is the basis dual to $\{e_m(\nu^{2j})\}$, and each $e^m(\nu^{2j})$ has the same numerical value as $e_m(\nu^{2j})$. In particular, if $A^{(j)}(\nu^{2j})$ is the decomposable spinor $(u\boldsymbol{\alpha} + v\boldsymbol{\beta})^{2j}$, then $A^{(j)m}$ is the spherical tensor

$$A^{(j)m} = \binom{2j}{j-m}^{\frac{1}{2}} u^{j+m}v^{j-m}. \quad (62)$$

This is the well-known relation,[35,41] which has been used to derive the rotation matrices[42] $D^{(j)}_{mm'}$ and the 3-j coefficients[43] for spherical tensors.[44] The spherical tensor metric $g_{mm'}$ for this basis set is also the standard one,

$$g_{m,m'} = \boldsymbol{\epsilon}^{2j} \odot^{2j} e_m(\nu^{2j}) \otimes e_n(\nu^{2j}) = (-1)^{j-m}\delta_{m,-m'}. \quad (63)$$

It is simple to relate a spherical tensor to a symmetric traceless Cartesian tensor if the latter is expressed in terms of the spherical vector basis set (in natural phase[45]) $\mathbf{e}_1(1) = -2^{-\frac{1}{2}}(\hat{x} - i\hat{y})$, $\mathbf{e}_0(1) = \hat{z}$, $\mathbf{e}_{-1}(1) = 2^{-\frac{1}{2}}(\hat{x} + i\hat{y})$. The relevant basis tensors are, in fact,

$$\mathbf{e}_m(m_1 \cdots m_j) = \binom{2j}{j-m}^{-\frac{1}{2}} 2^{\frac{1}{2}(j-\mu)}\delta_m^{\mu'}, \quad (64)$$

where $\mu = \sum_i |m_i|$, $\mu' = \sum_i m_i$. These agree with the spherical vector-basis set for $j = 1$ and have the metric $g_{m,m'} = (-1)^m\delta_{m,-m'}$. This metric differs from the spinor metric (63) by a factor $(-1)^j$ which could easily be obtained by changing the phase in the definition of $\mathbf{e}_m(1)$. However the above used form is also standard—it should only be remembered that there is a factor $(i)^j$ between the spherical tensor components as derived from a Cartesian tensor as opposed to those derived from a spinor.

The analog of Eq. (62) for a decomposable tensor $(a^m\mathbf{e}_m(1))^j = (x\hat{x} + y\hat{y} + z\hat{z})^j$ is a formula for the spherical harmonics,

$$\bar{Y}_j^m(\mathbf{a}) = \binom{2j}{j-m}^{-\frac{1}{2}} \Sigma_s 2^{\frac{1}{2}(j-m-2s)} \binom{j}{m+2s}\binom{m+2s}{s} \times (a^+)^{m+s}(a^0)^{j-m-2s}(a^-)^s, \quad (65)$$

where $\mu = m + 2s$. These have the projection-type normalization of Sec. IV,[46]

$$\int \bar{Y}_j^m(\hat{a})\,\bar{Y}_{j'}^{m'}(\hat{a})\,d\hat{a} = 4\pi(2j+1)^{-1}2^j(j!)^2[(2j)!]^{-1}\delta_{jj'}(-1)^m\delta_{m,-m'}. \quad (66)$$

Equation (65) is therefore equivalent to

$$Y_j^m(\theta, \phi) = \left[\frac{2j+1}{4\pi}(j-m)!\,(j+m)!\right]^{\frac{1}{2}} \times \Sigma_s \frac{(-1)^{m+s}\sin\theta^{m+2s}\cos\theta^{j-m-2s}e^{im\phi}}{2^{m+2s}(m+s)!\,s!\,(j-m-2s)!} \tag{67}$$

as the usual spherical harmonics normalized to unity.

The transformation from Cartesian components involves, in addition, the unitary transformation $a^r \to a^m$, namely

$$a^m = U_r^m a^r, \tag{68}$$

where for natural phase

$$U_r^m = \begin{pmatrix} -2^{-\frac{1}{2}} & -i2^{-\frac{1}{2}} & 0 \\ 0 & 0 & 1 \\ 2^{-\frac{1}{2}} & -i2^{-\frac{1}{2}} & 0 \end{pmatrix}. \tag{69}$$

The notation U_r^m emphasizes that $U_r^m = U_m^{r*} = \langle m \,|\, r\rangle$ can be regarded as a matrix element of the unit tensor U in "mixed" representation, and thus has the properties

$$\begin{aligned} U_r^m U_{m'}^r &= \delta_{m'}^m, \\ U_r^m U_m^{r'} &= \delta_r^{r'}, \\ U_m^r U_{m'}^r &= (-1)^m \delta_{m,-m'}, \end{aligned} \tag{70}$$

which are "mixed" representations of $\mathsf{U}\cdot\mathsf{U} = \mathsf{U}$. Also, $\mathsf{E}^{(j)}$ can be expressed in "mixed" representation,

$$\langle m_1 \cdots m_j |\, \mathsf{E}^{(j)} \,| r \cdots r_j\rangle = \sum_t c_t (U_r^m)^{j-2t}(\delta_{r,r})^t(-1)^m(\delta_{m,-m})^t, \tag{71}$$

which is just the transformation from Cartesian representation to the spherical vector representation. [In Eq. (71), c_t is given by Eq. (22), and the symmetrization convention is used.] The spherical tensor basis set in Cartesian form is obtained by combining the U_m^r transformation with Eq. (64) to give

$$\mathbf{e}_m(r_i \cdots r_j) = \binom{2j}{j-m}^{-\frac{1}{2}} \sum_{m_i \cdots m_j} \delta_m^{\Sigma_i m_i} 2^{\frac{1}{2}(j-\mu)}(U_m^r)^j, \tag{72}$$

which satisfies[47]

$$g_{mm'} = \mathbf{e}_m(j) \odot^j \mathbf{e}_{m'}(j) = (-1)^m \delta_{m,-m'} \tag{73}$$

and

$$\sum_m \mathbf{e}_m^{(j)}(-1)^m \mathbf{e}_{-m}(j) = \mathsf{E}^{(j)}. \tag{74}$$

Since the dual basis is $\mathbf{e}^m(j) = (-1)^m \mathbf{e}_{-m}(j)$, the last equation is just the resolution of the natural projection $\mathsf{E}^{(j)}$ according to the basis $\mathbf{e}_m(j)$. In general, the $\mathbf{e}_m(j)$'s are quite complicated, although particular cases are easy to write out and can be found in literature.[48] Of particular simplicity is

$$\mathbf{e}_0(j) = \frac{1}{j!}\left(\frac{(2j)!}{2^j}\right)^{\frac{1}{2}} \bar{Y}_j(\hat{z}) \tag{75}$$

from which all the other $\mathbf{e}_m(j)$'s may be obtained by raising or lowering operators.

The relation between natural tensors and natural spinors can be obtained through their spherical tensor components. As discussed earlier, the two sets of components give rise to different metrics [Eqs. (63) and (73)]. These can, of course, be brought into coincidence by introducing a factor $(i)^j$ or $(-i)^j$ into the Cartesian tensor to spherical tensor correspondence. If a factor $+i$ is introduced into the definitions of $\mathbf{e}_m(1)$ the vector $\mathbf{a}$ and the symmetric second-order spinor $A^{\mu\nu}$ are related by

$$a^r = U_{\mu\nu}^r A^{\mu\nu}, \tag{76}$$

with

$$U_{\mu\nu}^r = \frac{-i}{(2)^{\frac{1}{2}}} \begin{pmatrix} -1 & 0 & 0 & 1 \\ i & 0 & 0 & i \\ 0 & 1 & 1 & 0 \end{pmatrix}, \tag{77}$$

where the labeling on the four spinor components is $A^{11}, A^{12}, A^{21}, A^{22}$ where $A^{12} = A^{21}$, and a^r is a^x, a^y, a^z. Higher-ordered tensors and spinors can be built up with the aid of this transformation, two spinor indices being equivalent to one Cartesian index. The following correspondences between invariant spinors and invariant tensors are particularly important:

$$\delta_{rs} \leftrightarrow (\epsilon_{\mu\nu})^2, \tag{78}$$

$$\epsilon_{rst} \leftrightarrow +2^{\frac{1}{2}}\epsilon_{\mu\nu}\epsilon_{\nu\tau}\epsilon_{\tau\mu}. \tag{79}$$

Here the indical correspondences are $r \leftrightarrow \mu_1, \mu_2$, $s \leftrightarrow \nu_1, \nu_2$ and $t \leftrightarrow \tau_1, \tau_2$, with the symmetrization convention being used on the repeated unlabeled spinor indices. Since all invariant tensors can be expressed in terms of the elementary invariant tensors δ_{rs} and ϵ_{rst}, Eqs. (78) and (79) give a prescription for transcribing any invariant tensor into an invariant even-ordered spinor.

The whole of Cartesian tensor analysis could have been formulated in terms of the extraction and reembedding of spherical tensors—in fact, this is one interpretation that can be put on the basis tensors of Sec. II. However, this is only one possible basis set and others may be simpler to work with, if indeed one needs a basis set at all. It has been our general experience that the irreducible Cartesian tensors can be handled very conveniently as entities by themselves without the complicated and usually unnecessary transformation to spherical components and back again.

Note added in proof: The basis tensors $\mathbf{e}_m(j)$ for the spherical components (Sec. VII) can be obtained in a simpler form by projecting out the traceless symmetric part of the product $(\mathbf{e}_1)^m(\mathbf{e}_0)^{j-m}$. For $m > 0$ the result must be of the general form

$$\mathbf{e}_m(r_1 \cdots r_j) = N \sum_t a_t[\mathbf{e}_1(r)^m \mathbf{e}_0(r)^{j-m-2t}(\delta_{rr})^t],$$

where the symmetrization convention is understood and where the coefficients a_t can be determined, in a way similar to those of the projections $E^{(j)}$ (Sec. IV), by requiring the expression to be traceless. The resulting recursion formula yields

$$a_t = \frac{(-1)^t(j-m)!\,j!\,(2j-2t)!}{(2j)!\,(j-m-2t)!\,t!\,(j-t)!},$$

where $a_0 = 1$. With the normalization

$$N = 2^{\frac{1}{2}(m-j)}\binom{2j}{j-m}^{\frac{1}{2}},$$

the $\mathbf{e}_m$ have the normalization and phase of Eqs. (64) and (73). For $m < 0$, replace m by $|m|$ in all these equations and also replace $\mathbf{e}_1$ by $\mathbf{e}_{-1}$ in the first equation. Apart from the normalization these are the coefficients of the associated Legendre polynomials, a result which is consistent with the fact that $\mathbf{e}^m(j) \odot^j (\hat{u})^j$ is the spherical harmonic $\overline{Y}_j^m(\hat{u})$:

$$\mathbf{e}^m(j) \odot^j (\hat{u})^j.$$

VIII. IRREDUCIBILITY UNDER OTHER GROUPS

The considerations of Sec. II apply to groups other than the rotation group, but the elementary invariant tensors are changed. Irreducible tensors can still be represented as *reduced* or *natural* tensors of minimal order embedded in a tensor space of higher order. However, the minimal orders for distinct symmetries are not necessarily distinct. One must first determine the "natural projections" $\mathsf{E}^{(j)}(m \mid m)$ for the irreducible subspaces of symmetry j in the space $\mathfrak{X}^m$ of minimal order. Unlike the case of the rotation groups, these in general will not be all of the same type. Some examples are discussed in the remainder of this section.

In general, the higher the symmetry, the fewer the number of invariant tensors. This is exemplified by the general linear group $GL(3)$ where the only invariant tensor is δ_s^t (there being no invariant metric) so that the only invariant mappings are permutations. In particular, there can be no invariant reduction of order, and so tensors of different order have different group symmetry. Thus the irreducible representations can be classified according to the irreducible representations of the symmetric group, the well-known result.[30,49]

The improper rotation group $O(3)$ in three dimensions is the direct product of the proper rotation group $SO(3)$ and the inversion group $\{1, I\}$ so that the irreducible representations can be labeled by (j, π) according to the rotational symmetry j and the parity π. Since the inversion I changes the sign of any vector $\mathbf{r} \in \mathfrak{X}$, i.e., $I\mathbf{r} = -\mathbf{r}$, any element $\mathsf{T}(n) \in \mathfrak{X}^n$ obeys $I^n\mathsf{T}(n) = (-1)^n\mathsf{T}(n)$. In particular, $\boldsymbol{\epsilon}$ considered as an element of $\mathfrak{X}^3$ changes sign, $I^3\boldsymbol{\epsilon} = -\boldsymbol{\epsilon}$ and so is no longer an invariant tensor. Thus the only elementary invariant tensors are δ_r^s and the metric δ_{rs} together with its inverse δ^{rs}. Consequently not all the natural forms are symmetric traceless tensors; those for the irreducible representations $(j, (-1)^j)$ actually being symmetric traceless, while those for $(j, (-1)^{j+1})$ are not. In fact, the minimal tensorial order is 3 for the $(0, -1)$ representation and $j + 1$ for $(j, (-1)^{j+1})$, $j \neq 0$. However, it is convenient (and common practice) to reduce to symmetric traceless form in the following invariant way: Define $\boldsymbol{\epsilon}'$ as having the same components as $\boldsymbol{\epsilon}$ but transforming differently under inversion, namely

$$\boldsymbol{\epsilon}' \xrightarrow{I} \boldsymbol{\epsilon}', \tag{80}$$

thus being invariant. Clearly $\boldsymbol{\epsilon}'$ is not an element of $\mathfrak{X}^3$ but can be considered as an element of $\mathfrak{Y}^3$, the space of third-order pseudotensors. In general, the space of nth-order pseudotensors $\mathfrak{Y}^n$ can be defined as $\mathfrak{Y}^n = \mathfrak{p}\mathfrak{X}^n$ where $\mathfrak{p}$ is a unit pseudoscalar having the basic properties $I\mathfrak{p} = -\mathfrak{p}$, and $\mathfrak{p}^2 = 1$. The second property implies that the product of two pseudotensors is a proper tensor, i.e., is an element of $\mathfrak{X}^m$ for some integer m. In particular $\boldsymbol{\epsilon}'$ is given by $\boldsymbol{\epsilon}' = \mathfrak{p}\boldsymbol{\epsilon}$. In literature, the usual meaning of $\boldsymbol{\epsilon}$ is, in fact, really $\boldsymbol{\epsilon}'$ rather than $\boldsymbol{\epsilon}$! By means of $\boldsymbol{\epsilon}'$, all irreducible (proper) tensors can be reduced to symmetric traceless form, those belonging to the representations $(j, (-1)^j)$ being tensors, and those belonging to the representations $(j, (-1)^{j+1})$, pseudotensors.

For the group C_∞ of proper rotations about the z axis of a three-dimensional space, U and $\boldsymbol{\epsilon}$ can be reduced into the more elementary invariant tensors $\mathsf{U}^{(2)} = \mathsf{U} - \hat{z}\hat{z}$, $\boldsymbol{\epsilon} \cdot \hat{z}$, and $\hat{z}$. Thus the reduction problem is equivalent to the analogous problem for the rotation group $SO(2)$ in two-dimensional space. The elementary invariant tensors for $SO(2)$ are $\delta_{\mu\nu}$ and $\epsilon_{\mu\nu}$ [see Eq. (46)]. Because of the presence of $\delta_{\mu\nu}$ (implying that covariant and contravariant Cartesian indices need not be distinguished), $\epsilon_{\mu\nu}$, in contrast to the case of $SU(2)$, also has the significance of ϵ_μ^ν that is of a map of vectors into vectors. This is not the identity map, and this is fundamental in studying the tensor representations of $SO(2)$. The two eigenvectors

of the map ϵ^{ν}_{μ}, satisfying $\sum_{\mu} \epsilon^{\nu}_{\mu} a^{\mu} = \lambda a^{\mu}$, cannot be mapped into one another by any invariant tensor $\mathsf{T}^{(0)}(1 \mid 1)$, since any such tensor is merely a linear combination of ϵ^{ν}_{μ} and δ^{ν}_{μ}, and they therefore have different symmetries, which are both one dimensional and of multiplicity one. Thus the eigenvectors $\mathbf{e}_{+} = 2^{-\frac{1}{2}}(1, i)$ and $\mathbf{e}_{-} = 2^{-\frac{1}{2}}(1, -i)$, belonging to rotational symmetry $m = +1$ and -1, respectively, are irreducible under $SO(2)$. The mapping of an nth-order irreducible tensor to natural form is attained by contracting with $\delta_{\mu\nu}$ and/or $\epsilon_{\mu\nu}$ to obtain a symmetric traceless tensor of lower order $|m|$. The result must be an eigentensor of ϵ^{ν}_{μ} *with respect to each index*, for the two distinct eigentensors cannot be mapped into one another by any invariant map, being easily seen to be in fact eigentensors of all invariant automorphisms, and are therefore of different symmetry, each symmetry being of dimension one and having multiplicity one. The eigenvalues λ of ϵ^{ν}_{μ} are either $+i$ for all indices, or all $-i$, since the tensors are symmetric. Thus for each order $|m|$ there exist two traceless symmetric natural forms which can be labeled by $\{\lambda, |m|\}$, or by the product $m = i\lambda |m| = \pm |m|$, coinciding with the usual symmetry designation for $SO(2)$. The natural projections into representations $\pm |m|$ are

$$\mathsf{E}^{\pm|m|}(\mu_1 \cdots \mu_m, \nu_1 \cdots \nu_m) = 2^{-m}(\delta^{\mu}_{\nu} \mp i\epsilon^{\mu}_{\nu})^m. \quad (81)$$

When reflection is introduced giving $C_{\infty v}$ [or $O(2)$], then $\epsilon_{\mu\nu}$ is no longer invariant, and the irreducible representations are changed in an essential way. For $m \neq 0$ the m and $-m$ representations of C_{∞} coalesce to form two-dimensional irreducible representations, whereas there are two representations for $m = 0$ which can be labeled $(0, +)$ and $(0, -)$ in analogy to the improper rotation group, or by the spectroscopic symbols Σ_g and Σ_u. The natural forms for $|m| > 0$ and for the Σ_g state are the symmetric traceless tensors of order $|m|$ with natural projections,

$$\mathsf{E}^{|m|}(r_1 \cdots r_{|m|} \mid s_1 \cdots s_{|m|}) = \sum_{t=0}^{\frac{1}{2}|m|} c'_t (\delta_{r,s})^{|m|-2t} (\delta_{r,r})^t (\delta_{s,s})^t, \quad (82)$$

where

$$c'_t = (-1)^t |m| \frac{(|m| - t - 1)!}{2^{2t} t! (|m| - 2t)!}, \quad |m| > 1,$$
$$c'_t = \delta_{t,0}, \quad |m| = 0, 1. \quad (83)$$

For Σ_u, the natural form is the antisymmetric second-order tensor[50] with natural projection

$$\mathsf{E}^{(0,-)}(r_1 r_2 \mid s_1 s_2) = \tfrac{1}{2}(\delta_{r_1 s_1}\delta_{r_2 s_2} - \delta_{r_1 s_2}\delta_{r_2 s_1}). \quad (84)$$

Table V gives the reduction of fourth-order tensors under C_{∞} in the present formalism.

Finally,[51] for the finite rotation groups C_n, there are two additional invariant tensors, $(\mathbf{e}_{+})^n$ and $(\mathbf{e}_{-})^n$, besides $\delta_{\mu\nu}$ and $\epsilon_{\mu\nu}$. These extra invariant tensors map symmetric traceless tensors of order $n + r$ to order r so that there is only a finite number of inequivalent irreducible representations.

APPENDIX A

The reduction of second-order tensors is extremely simple since all multiplicities are unity, and well known, but it serves to illustrate the notation. Table II summarizes the reduction according to our formalism by elementary embeddings. For symmetry $j = 2$, $\mathsf{T}^{(0)}_1(2 \mid 2) = \mathsf{T}^{(0)1}(2 \mid 2)$ is the natural projection $\mathsf{E}^{(2)}$ so that the reduced and embedded forms coincide. For symmetry 1, $\mathsf{T}^{(0)}_1(2 \mid 1)$ is chosen as $\boldsymbol{\epsilon}$. Then

$$\mathsf{T}^{(0)}_1(2 \mid 1) \circ^2 \mathsf{T}^{(0)}_1(2 \mid 1) = \sum_{r_1 r_2} \epsilon_{r_1 r_2 s} \epsilon_{r_1 r_2 s'} = 2\delta_{ss'} = 2\mathsf{E}^{(1)} \quad (A1)$$

TABLE V. Reduction of 4th-order two-dimensional tensors under $C_{\infty v}$ by elementary contractions.

Symmetry m	Contraction	g^{pq}	g_{pq}
4	none	1	1
2	$\delta_{r_1 r_2}$	$\begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$	$\frac{1}{5}\begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix}$
	$\delta_{r_2 r_3}$		
	$\delta_{r_3 r_4}$		
	$\delta_{r_1 r_4}$		
$(0, +)$	$\delta_{r_1 r_2}\delta_{r_3 r_4}$	$2\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$	$\frac{1}{8}\begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}$
	$\delta_{r_1 r_3}\delta_{r_2 r_4}$		
	$\delta_{r_1 r_4}\delta_{r_2 r_3}$		
$(0, -)$	$\delta_{r_2 r_3}$	$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$	$\frac{1}{4}\begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}$
	$\delta_{r_2 r_4}$		
	$\delta_{r_3 r_4}$		

implies that $g_{11} = 2$, $g^{11} = \frac{1}{2}$, and $\mathsf{T}^{(0)1}(1 \mid 2) = \frac{1}{2}\boldsymbol{\epsilon}$. The corresponding second-order projection is

$$\begin{aligned}\boldsymbol{\Pi}_1(r_1r_2 \mid r_1'r_2') &= \mathsf{T}^{(0)}(r_1r_2 \mid 1) \odot^1 \mathsf{T}^{(0)}(1 \mid r_1'r_2') \\ &= \tfrac{1}{2}\sum_s \epsilon_{r_1r_2s}\epsilon_{r_1'r_2's} \\ &= \tfrac{1}{2}(\delta_{r_1r_1'}\delta_{r_2r_2'} - \delta_{r_1r_2'}\delta_{r_1'r_2}). \quad \text{(A2)}\end{aligned}$$

The effect of these tensors on the tensor product $\mathbf{a} \otimes \mathbf{b}$ is given by

$$\mathsf{T}^{(0)1}(1 \mid 2) \odot^2 \mathbf{a} \otimes \mathbf{b} = \mathbf{a} \times \mathbf{b}, \quad \text{(A3)}$$

$$\mathsf{T}_1^{(0)}(2 \mid 1) \odot^1 \mathbf{a} \times \mathbf{b} = \tfrac{1}{2}(\mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}), \quad \text{(A4)}$$

where the natural form is $\mathbf{a} \times \mathbf{b}$, the cross product of $\mathbf{a}$ and $\mathbf{b}$ and the re-embedded form is the antisymmetric part of $\mathbf{a} \otimes \mathbf{b}$. Similarly for $j = 0$, $\mathsf{T}^{(0)1}(0 \mid 2) = \mathsf{U}$ contracts (takes the trace or dot product), while $\mathsf{T}_1^{(0)}(2 \mid 0)$ re-embeds the trace, e.g.,

$$\mathsf{T}^{(0)1}(0 \mid 2) \odot^2 \mathbf{a} \otimes \mathbf{b} = \mathbf{a} \cdot \mathbf{b}, \quad \text{(A5)}$$

$$\mathsf{T}_1^{(0)}(2 \mid 0) \odot^0 \mathbf{a} \cdot \mathbf{b} = \tfrac{1}{3}(\mathbf{a} \cdot \mathbf{b})\mathsf{U}. \quad \text{(A6)}$$

The reduction of third-rank tensors is the simplest case involving multiplicities. Table III gives the complete reduction by elementary contractions and illustrates the simplicity and compactness of the present formalism, in comparison with the more lengthy calculations in Part I, Sec. IV. Consider the symmetry $j = 1$ case as an example. There are three ways of contracting a third-order tensor to first-order: $\delta_{s_2s_3}$, $\delta_{s_1s_3}$, and $\delta_{s_1s_2}$, conveniently equal to the multiplicity $N_3^{(1)} = 3$. Thus we set $\mathsf{T}^{(0)1}(1 \mid 3) = \delta_{rs_1}\delta_{s_2s_3}$, etc. The metric g^{-1} is obtained by considering products of the form

$$\mathsf{T}^{(0)1}(r \mid 3) \odot^3 \mathsf{T}^{(0)1}(r' \mid 3) = 3\delta_{r,r'} = 3\mathsf{E}^{(1)}(r \mid r'), \quad \text{(A7)}$$

$$\mathsf{T}^{(0)1}(r \mid 3) \odot^3 \mathsf{T}^{(0)2}(r' \mid 3) = \delta_{r,r} = \mathsf{E}^{(1)}(r \mid r'). \quad \text{(A8)}$$

The inverse metric gives the dual tensors, e.g.,

$$\mathsf{T}_1^{(0)}(s \mid r) = \tfrac{1}{10}[4\mathsf{T}^{(0)1}(r \mid s) - \mathsf{T}^{(0)2}(r \mid s) - \mathsf{T}^{(0)3}(r \mid s)]. \quad \text{(A9)}$$

Reductions based on elementary embeddings or contractions are quite simple for 4th- and 5th-order tensors as well as the 6th-order invariant tensors. These and other convenient choices of basis tensors will be discussed in Part IV.

* Supported in part by the National Research Council of Canada.

[1] See, e.g., E. P. Wigner, *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959) or L. C. Biedenharn and H. Van Dam, *Quantum Theory of Angular Momentum* (Academic Press Inc., New York, 1965).

[2] In the kinetic theory of gases with nonspherical molecules, irreducible Cartesian tensors were introduced by Y. Kagan and L. Maksimov, Zh. Eksp. Teor. Fiz. **41**, 842 (1961) [Sov. Phys.—JETP **14**, 604 (1962)]. W. J. Meath and J. O. Hirschfelder, J. Chem. Phys. **44**, 3197 (1966), have recently used Cartesian tensors for studying intermolecular forces. In the latter case, the spherical tensors are expressed in Cartesian form.

[3] B. Barsella and E. Fabri, Nuovo Cimento Suppl. (Ser. I) **2**, 293 (1964).

[4] C. Zemach, Phys. Rev. **140B**, 97 (1965).

[5] H. Grad, Commun. Pure Appl. Math. **3**, 325 (1949).

[6] J. A. R. Coope, R. F. Snider, and F. R. McCourt, J. Chem. Phys. **43**, 2269 (1965). This is referred to as Part I in the text.

[7] J. A. R. Coope, J. Math. Phys. (to be published). (Part III).

[8] J. A. R. Coope (to be published). (Part IV).

[9] W. H. Greub, *Linear Algebra* (Academic Press Inc., New York, 1963).

[10] This is the common notation, see, e.g., L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton University Press, Princeton, N.J., 1947).

[11] Both of the terms "rank" and "order" are commonly used for the number of indices on a tensor and we used the term rank in Part I. As the term rank also has a meaning (see Ref. 9) as the dimension of the image space of a given linear mapping, which is used in Sec. IV, we will here follow Greub in using the word order for the number of indices.

[12] An nth-order tensor has n indices $r_1, r_2, \cdots, r_n$; however, for most purposes it is completely unnecessary to explicitly write down these indices; hence, the symbol (n) is used which merely implies that there are n indices. In contracting an nth-order contravariant tensor with an nth-order covariant tensor, the n-fold contraction is denoted by $\odot^n$. However, some convention, as to which of the $2n$ indices are to be contracted, must be followed when doing the contraction. In Part I, the convention that nearest indices are contracted was used, while T. B. Drew [*Handbook of Vector and Polyadic Analysis* (Reinhold Publ. Corp., New York, 1961)] uses the convention of contracting left-hand index with left-hand index and so on, down the line. This latter convention is somewhat more convenient for higher-order tensors and is used in the following when an explicit convention is required. However, the formal development is independent of what convention is adopted.

[13] See, e.g., H. Boerner, *Representations of Groups* (North-Holland Publ. Co., Amsterdam, 1963).

[14] For the rotation group, j is the highest weight (see Ref. 13) of the irreducible representation and was simply called the weight in Part I. Here, j will be considered as a composite label to distinguish the different irreducible representations of the arbitrary group G, hence the name "symmetry j" or "symmetry type j."

[15] It is assumed that G is a decomposable group.

[16] In general, a superscript in brackets designates the symmetry, whereas without brackets, it will denote a tensorial product, e.g., $\mathfrak{T}^n$ or Eq. (23). Again, (n) stands for the n indices (see Ref. 12).

[17] Since the space of symmetry j is only D_j dimensional, there can be at most D_j independent tensors of this symmetry.

[18] For the group $GL(v)$, the invariant mappings are sometimes called tensorial mappings (see Ref. 9).

[19] Care must be exercised in applying this theorem to assure that the domains of the mappings are only the relevant *irreducible* subspaces.

[20] Ref. 1, p. 77.

[21] See Sec. VIII for a discussion of the improper rotation group $O(3)$ where $\boldsymbol{\epsilon}$ is not invariant.

[22] This can be considered as the multiplication rule for determinants.

[23] This does not mean that quantum-mechanical problems involving half-integral *spin* cannot be handled by these methods since, in general, it is the quantum-mechanical operators which are required, and these can be expressed in terms of the quantum-mechanical vector operator $\mathbf{J}$.

[24] Correctly, the infinitesimal generators are anti-Hermitian, being $i\mathfrak{J}_x(n \mid n)$, etc.

[25] $\otimes$ denotes the tensor product (see Ref. 9).

[26] For example, $n = 2$: $\mathfrak{J}^2(r_1r_2 \mid s_1s_2) = 4\delta_{r_1s_1}\delta_{r_2s_2} + 2(\delta_{r_1s_2}\delta_{r_2s_1} - \delta_{r_1r_2}\delta_{s_1s_2})$.

[27] 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). The reduction of third-order tensors is explicitly obtained by them with this method.

[28] P.-O. Löwdin, Rev. Mod. Phys. **36**, 966 (1964).

[29] This is equivalent to Dirac's formula $P_{12} = \frac{1}{2}(1 + \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2)$.

[30] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, N.J., 1939); see also M. Hamermesh, *Group Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1962).

[31] These can also be written as $\bar{\mathsf{Y}}^{(j)}(\hat{r}) = j!\, r^{j+1}(-\nabla)^j/r$, a form appearing in multipole expansions.

[32] The expansion coefficients are tensors $\mathsf{C}^{(j)}(j)$ satisfying $\mathsf{E}^{(j)} \bigcirc^j \mathsf{C}^{(j)}(j) = \mathsf{C}^j(j)$.

[33] H. A. Kramers, Koninkl. Ned. Akad. Wetenschap. **34**, 956 (1931); see also J. Schwinger, "On Angular Momentum," in *Quantum Theory of Angular Momentum* (Ref. 1).

[34] The normalization is chosen so that Eq. (37) takes a simple form. For a fixed basis $\{\mathbf{e}_\sigma(j)\}$ of H_j^j with dual basis $\{\mathbf{e}^\sigma(j)\}$, then $\mathbf{e}_{p\sigma}(n) \equiv \mathsf{T}_p(n \mid j) \bigcirc^j \mathbf{e}_\sigma(j)$ and $\mathbf{e}^{q\tau}(n) \equiv \mathbf{e}^\tau(j) \bigcirc^j \mathsf{T}^q(j \mid n)$ are dual basis sets in H_j^n and $\overline{H}_j^n$, respectively, i.e.,

$$\mathbf{e}^{q\tau}(n) \bigcirc^n \mathbf{e}_{p\sigma}(n) = \delta_p^q \mathbf{e}^\tau(j) \bigcirc^j \mathbf{e}_\sigma(j) = \delta_p^q \delta_\sigma^\tau .$$

[35] H. A. Kramers, Koninlk. Ned. Akad. Wetenschap. Proc. **33**, 953 (1930); **34**, 956 (1931); H. C. Brinkman, *Applications of Spinor Invariants in Atomic Physics* (North-Holland Publ. Co., Amsterdam, 1956).

[36] The sign agrees with that of L. D. Landau and E. M. Lifshitz, *Quantum Mechanics Non-relativistic Theory* (Pergamon Press, Ltd., London, 1959).

[37] See Chapter I, Sec. 4 of Ref. 35.

[38] The μ and ν indices in Eqs. (55) and (56) are not symmetrized according to the convention of Sec. IV—hence the subscript NS. On the other hand, the σ and τ indices are symmetrized.

[39] Also called the Heitler–London–Slater–Pauling method. See, e.g., L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics* (McGraw-Hill Book Co., New York, 1935).

[40] For first-order spinors, $v_i = 1$ or 2 corresponds to $m = \pm\frac{1}{2}$, so that the δ term just equates m and $\sum_i m_i$.

[41] H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1949), p. 145; E. P. Wigner, Ref. 1, p. 163. The factor $[(2j)!]^{\frac{1}{2}}$ is not usually used.

[42] P. Güttinger, Z. Physik **73**, 169 (1932).

[43] B. L. Van der Waerden, *Die Gruppentheoretische Methode in der Quantenmechanik* (Springer-Verlag, Berlin, 1932).

[44] Some of the 6-j coefficients were derived by Kramers (Ref. 35) by spinor methods. These being scalars are essentially the same for spherical tensors, spinors, and Cartesian tensors.

[45] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge Univ. Press, London, 1935).

[46] All the transformations in this section are unitary. The only change in orthogonality relations is therefore in the representation of the natural projection. For example, the change from the Cartesian $\overline{\mathsf{Y}}^{(j)}$ of Sec. IV to the $\overline{Y}_j^m$ here only changes the orthogonality relation (34) from $\mathsf{E}^{(j)}$ to $(-1)^m \delta_{m,-m'}$.

[47] See the discussion following Eq. (64).

[48] Two recent articles containing examples are R. McWeeny, J. Chem. Phys. **42**, 1717 (1965); W. J. Meath and J. O. Hirschfelder, *ibid.* **44**, 3197 (1966).

[49] A discussion of the irreducible representations multiplicity problem for $GL(n)$ is given by D. R. Tompkins, Phys. Rev. Letters **16**, 1058 (1966).

[50] As in the case of the rotation group one could use the non-invariant $\varepsilon_{\mu\nu}$ to devise a space of pseudoscalars.

[51] The invariant tensors for the low-order crystallographic groups are given by G. F. Smith, J. Math. Phys. **5**, 1612 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On the Ising Model with Long-Range Interaction. II. Critical-Region Analysis

COLIN J. THOMPSON*
Mathematics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

ARNOLD J. F. SIEGERT†
Physics Department, Northwestern University, Evanston, Illinois 60201

AND

DAVID J. VEZZETTI‡
Physics Department, University of Illinois at Chicago Circle, Chicago, Illinois 60680

(Received 19 May 1969)

The high-temperature expansion for the free energy in powers of γ (the reciprocal of the range of interaction) developed in a previous paper is studied in the critical region. In terms of the graphology introduced in the previous paper, it is proved that the ring diagrams give the dominant contribution in the critical region, if and only if the integral

$$R(\nu\gamma) = \left(\frac{1}{2\pi}\right)^D \int_0^{2\pi} \cdots \int \frac{g(\omega)\, d^D\omega}{1 - \nu\gamma g(\omega)}$$

diverges no worse than logarithmically at [the Curie–Weiss (CW) point] $\nu_{\mathrm{CW}} = J/kT_{\mathrm{CW}} = [\gamma g(0)]^{-1}$, where D denotes dimensionality and $\gamma g(\omega)$ the Fourier transform of the interaction potential. The results are in agreement with various model results, and it is conjectured that the above condition is also a necessary and sufficient condition for the existence of a phase transition.

1. INTRODUCTION

In a recent paper[1] (hereafter referred to as I) a functional-integral representation for the partition function of an Ising model with long-range interaction was used to develop high- and low-temperature expansions for the free energy in powers of γ, the reciprocal of the range of interaction, with the classical Curie–Weiss (CW) theory as leading term. Previous graphical results of Brout[2] and others[3,4] are obtained very easily by this method, and by a combination perturbation–variational technique (which we call the "α trick") it was shown that, as a first approximation, one obtains the "sphericalized" graphical results of Brout[2] and Mühlschlegel and Zittartz.[5] The limitations of these results obtained via the α trick are discussed in Paper I (Sec. 8), and are summarized briefly at the end of the present Sec. 3.

A parallel development in the study of systems with long-range interaction was begun some time ago by Kac,[6] who proposed a number of models with long-range exponential interaction, for which explicit expansions could be worked out at temperatures above *and below* the classical Curie–Weiss point (CWP) as well as in the critical region around the CWP. This approach has been summarized in detail by Kac[7] in the 1966 Brandeis lectures, where detailed references can be found.

In a recent paper Kac and Thompson[8] made a detailed critical-region study of a 1-dimensional model with infinite-range interaction, a 2-dimensional model, and a 3-dimensional model, all with exponential interactions. The method is briefly as follows: One first develops expansions in powers of γ above and below the CWP and notes that for a range of temperatures around the CWP these expansions break down, e.g., for the 2-dimensional model, when $T - T_{\mathrm{CW}} \sim \gamma \log 1/\gamma$. This range of temperature defines the critical region, and in this region one sums the most-divergent terms to all orders in perturbation. One finds that in all cases the resummed series to lowest order are now completely analytic at the CWP, and that the singularity is shifted by a small amount to a modified critical point, e.g., for the 2-dimensional model to

$$\frac{J}{kT_C'} = \frac{J}{kT_{\mathrm{CW}}} + \frac{\gamma}{4\pi} \log\left(\frac{1}{\gamma}\right).$$

In principle, one can then investigate the resummed series to determine a new critical region and then resum to determine a second approximation to the critical point, and so on. In this way, one hopefully obtains a converging sequence of approximations to the true critical behavior. The procedure, however, as one might imagine, gets rapidly out of hand, and even for the first resummation it was necessary to conjecture the form of the most divergent terms in perturbation beyond second order. It is also not clear how the true critical behavior is approached by this method, e.g., in two dimensions the first resummation

gives a logarithmically divergent energy at the modified critical point, which is definitely incorrect (one would rather expect a logarithmically divergent specific heat!). In fact, the first resummation gives critical behavior which is identical with Brout's sphericalized results. Whether or not the next resummation will change this is not known at the moment.

In spite of these difficulties it is encouraging that the critical-region resummation moves the singularity away from the CWP in the direction of the true critical point, and for the models, at least, the essential information about the modified critical point and critical exponents can be deduced from the high-temperature expansion (above the CWP) alone, i.e., the resummed low-temperature expansions give the same critical behavior and match up with the resummed high-temperature expansions at the CWP.

The striking similarity of the critical region resummations for the 1-, 2-, and 3-dimensional models suggests that it may be possible to carry out the type of procedure described above for general lattice systems with long-range interaction. The purpose of this note is to show how this can be done for temperatures above the CWP. Moreover, we will prove the most divergent term conjectures of Ref. 8 in complete generality. We are not able as yet, however, to treat the problem below the CWP, but in view of the model results this does not seem to be a serious drawback of the method.

We take as our starting point the functional-integral reduction given in Ref. 1. This is summarized briefly in the following section, and in Sec. 3 we carry out the critical-region resummation and relate the resummation to the α trick. For those not interested in the technical details we have summarized and discussed the results in some detail in the final section.

2. FUNCTIONAL-INTEGRAL REDUCTION OF THE PARTITION FUNCTION

For simplicity we will carry through the analysis for a 1-dimensional chain of spins $\mu_i = \pm 1$, $i = 1, 2, \cdots, N$, with interaction

$$E = -\tfrac{1}{2}J\gamma \sum_{k \neq l=1}^{N} \rho_{k-l}(\gamma)\mu_k\mu_l, \tag{2.1}$$

where it will be assumed throughout that the interaction potential is periodic and ferromagnetic, i.e., $\rho_l(\gamma) > 0$ and for convenience $\rho_0(\gamma)$ will be taken to be unity. The results for higher dimensions can be obtained simply from the formulas below by interpreting the k's etc., to be vectors of appropriate dimension.

To calculate the partition function for the interaction (2.1)

$$Q_N = \sum_{\{\mu\}} \exp\left[\tfrac{1}{2}\nu\gamma \sum_{k \neq l} \rho_{k-l}(\gamma)\mu_k\mu_l\right], \tag{2.2}$$

where $\nu = J/kT$ (k is Boltzmann's constant and T the absolute temperature) and the summation $\{\mu\}$ denotes the sum over all configurations

$$\mu_i = \pm 1, \quad i = 1, 2, \cdots, N,$$

we use the identity[9]

$$\exp\left[-\tfrac{1}{2}\sum_{k,l} \xi_k(\alpha^{-1})_{kl}\xi_l\right] = (2\pi)^{-\frac{1}{2}N}(\det \alpha)^{\frac{1}{2}} \times \int_{-\infty}^{\infty}\cdots\int d^N x \exp\left(i\sum_k \xi_k x_k - \tfrac{1}{2}\sum_{k,l} x_k\alpha_{kl}x_l\right), \tag{2.3}$$

which is valid for any positive definite symmetric matrix α and any set of complex variables ξ_k.

To obtain the leading contribution (for small γ) to Q_N for high temperatures, we add and subtract, with $\rho_0(\gamma) = 1$,

$$-\tfrac{1}{2}N(1 - \nu\gamma) = -\tfrac{1}{2}\sum_{k,l} \delta_{k,l}(1 - \nu\gamma\rho_{k-l}(\gamma))$$

to the exponent in (2.2), and then apply the identity (2.3), obtaining

$$\begin{aligned} Q_N &= \exp\left[\tfrac{1}{2}N(1 - \nu\gamma)\right] \\ &\quad \times \sum_{\{\mu\}} \exp\left[-\tfrac{1}{2}\sum_{k,l} \mu_k(\delta_{k,l} - \nu\gamma\rho_{k-l}(\gamma))\mu_l\right] \\ &= \exp\left[\tfrac{1}{2}N(1 - \nu\gamma)\right](2\pi)^{-\frac{1}{2}N}[\det(I - \nu\gamma\rho)]^{-\frac{1}{2}} \\ &\quad \times \int_{-\infty}^{\infty}\cdots\int d^N x \exp\left(-\tfrac{1}{2}\sum_{k,l} x_k(I - \nu\gamma\rho)^{-1}_{k,l}x_l\right) \\ &\quad \times \sum_{\{\mu\}} \exp\left(i\sum_k x_k\mu_k\right) \\ &= Q_N^{(0)} q_N, \end{aligned} \tag{2.4}$$

where

$$Q_N^{(0)} = 2^N \exp(-\tfrac{1}{2}N\nu\gamma)[\det(I - \nu\gamma\rho)]^{-\frac{1}{2}} \tag{2.5}$$

represents the leading contribution[2] to Q_N for small γ and

$$q_N = (2\pi/e)^{-\frac{1}{2}N}\int_{-\infty}^{\infty}\cdots\int d^N x \times \exp\left(-\tfrac{1}{2}\sum_{k,l} x_k(I - \nu\gamma\rho)^{-1}_{k,l}x_l\right)\prod_{k=1}^{N}\cos x_k. \tag{2.6}$$

In this case, the reduction to (2.4) is valid only if the matrix $I - \nu\gamma\rho$ is positive definite, which implies, since we are assuming that ρ is periodic (and ferromagnetic), that the smallest eigenvalue of $I - \nu\gamma\rho$, viz., $1 - \nu\gamma \sum_k \rho_k$, is positive, i.e.,

$$\nu < \left(\gamma \sum_k \rho_k\right)^{-1} = \nu_{\mathrm{CW}}. \tag{2.7}$$

The right-hand side of (2.6) will be recognized immediately as the Curie–Weiss value of $\nu = J/kT$.

To simplify (2.7), we define the matrix $\tilde{\rho}$ by

$$I + \nu\gamma\tilde{\rho} = (I - \nu\gamma\rho)^{-1}. \tag{2.8}$$

The homogeneity of ρ carries over to $\tilde{\rho}$, i.e., $\tilde{\rho}_{kl} = \tilde{\rho}_{k-l}$, so that q_N, Eq. (2.6), can be written as

$$\begin{aligned} q_N &= (2\pi/e)^{-\frac{1}{2}N} \int_{-\infty}^{\infty} \cdots \int d^N x \\ &\quad \times \exp\left(-\tfrac{1}{2}\sum_k x_k^2(1 + \nu\gamma\tilde{\rho}_0) - \tfrac{1}{2}\nu\gamma \sum_{k\neq l} x_k \tilde{\rho}_{k-l} x_l\right) \\ &\quad \times \prod_k \cos x_k \\ &= (2\pi/e)^{-\frac{1}{2}N} u^N \int_{-\infty}^{\infty} \cdots \int d^N \eta \\ &\quad \times \exp\left(-\tfrac{1}{2}\sum_k \eta_k^2 - \tfrac{1}{2}\nu\gamma u^2 \sum_{k\neq l} \eta_k \tilde{\rho}_{k-l} \eta_l\right) \\ &\quad \times \prod_k \cos(u\eta_k), \end{aligned} \tag{2.9}$$

where

$$u = (1 + \nu\gamma\tilde{\rho}_0)^{-\frac{1}{2}} \tag{2.10}$$

and to obtain (2.9) we have made the change of variables $x_k = u\eta_k$.

The factor

$$(2\pi)^{-\frac{1}{2}N} \exp\left(-\tfrac{1}{2}\sum_k \eta_k^2\right)$$

in (2.9) can be considered as the probability density for independent Gaussian random variables η_k with zero mean and covariance unity, so we can write q_N, given by Eq. (2.9), in the form

$$q_N = q_N^{(1)} q_N^{(2)} \tag{2.11}$$

with

$$q_N^{(1)} = e^{\frac{1}{2}N} u^N \langle \cos u\eta \rangle^N \tag{2.12}$$

and

$$q_N^{(2)} = \left\langle \exp\left(-\tfrac{1}{2}\nu\gamma u^2 \sum_{k\neq l} \eta_k \tilde{\rho}_{k-l} \eta_l\right) \times \prod_k [\cos(u\eta_k)/\langle \cos u\eta_k \rangle] \right\rangle, \tag{2.13}$$

where

$$\langle \cos u\eta \rangle = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\eta^2} \cos(u\eta)\, d\eta = e^{-\frac{1}{2}u^2}. \tag{2.14}$$

Thus, from (2.4) and (2.11),

$$Q_N = q_N^{(1)} q_N^{(2)} Q_N^{(0)}, \tag{2.15}$$

which gives for the free energy per spin ψ_N the following:

$$\begin{aligned} -\psi_N/kT &= N^{-1} \log Q_N \\ &= \log 2 + \tfrac{1}{2}(1 - \nu\gamma) \\ &\quad - (2N)^{-1} \log \det(I - \nu\gamma\rho) \\ &\quad - \tfrac{1}{2} \log(1 + \nu\gamma\tilde{\rho}_0) \\ &\quad - \tfrac{1}{2}(1 + \nu\gamma\tilde{\rho}_0)^{-1} + N^{-1} \log q_N^{(2)}. \end{aligned} \tag{2.16}$$

Proceeding now to the limit $N \to \infty$, we obtain [for details see Paper I, Eqs. (4.21) to (4.27)]

$$\begin{aligned} -\psi/kT &= \log 2 - \tfrac{1}{2}\nu\gamma + f(\nu) \\ &\quad + \frac{1}{2}\left(\frac{\nu\gamma R(\nu\gamma)}{1 + \nu\gamma R(\nu\gamma)} - \log[1 + \nu\gamma R(\nu\gamma)]\right) \\ &\quad + \lim_{N\to\infty} N^{-1} \log q_N^{(2)}, \end{aligned} \tag{2.17}$$

where

$$f(\nu) = -\frac{1}{4\pi} \int_0^{2\pi} \log[1 - \nu\gamma g(\omega)]\, d\omega, \tag{2.18}$$

$$g(\omega) = \sum_{k=-\infty}^{\infty} e^{ik\omega} \rho_k(\gamma), \tag{2.19}$$

and

$$\begin{aligned} R(\nu\gamma) &= \lim_{N\to\infty} \tilde{\rho}_0 \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{g(\omega)}{1 - \nu\gamma g(\omega)}\, d\omega. \end{aligned} \tag{2.20}$$

The reduction to Eq. (2.17) was carried out in I, where it was shown by a method discussed in the following section, that for $\nu < \nu_{\mathrm{CW}}$ [see Eq. (2.7)]

$$\begin{aligned} q &= \lim_{N\to\infty} N^{-1} \log q_N^{(2)} \\ &= (\nu\gamma)^4 \left\{\tfrac{1}{4}[R(\nu\gamma)]^2 R_{22}(\nu\gamma) + \frac{1}{2 \times 3!} R_{42}(\nu\gamma)\right\} + O(\gamma^4), \end{aligned} \tag{2.21}$$

where

$$R_{22}(\nu\gamma) = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{g(\omega)}{1 - \nu\gamma g(\omega)}\right)^2 d\omega \tag{2.22}$$

and

$$\begin{aligned} R_{42}(\nu\gamma) &= \left(\frac{1}{2\pi}\right)^3 \int_0^{2\pi}\int\int\int d\omega_1\, d\omega_2\, d\omega_3 \prod_{j=1}^{3} \left(\frac{g(\omega_j)}{1 - \nu\gamma g(\omega_j)}\right) \\ &\quad \times \left(\frac{g(\omega_1 + \omega_2 + \omega_3)}{1 - \nu\gamma g(\omega_1 + \omega_2 + \omega_3)}\right). \end{aligned} \tag{2.23}$$

It is also to be noted that $\gamma R_{22}(\nu\gamma)$ and $\gamma R_{42}(\nu\gamma)$ are bounded in the limit $\gamma \to 0$ provided $\nu < \nu_{CW}$, so that Eq. (2.21) gives the order γ^3 contribution to ψ [Eq. (2.17)] above the CWP. The motivation behind the trickery leading to Eq. (2.17) is now clear when we note that $f(\nu)$ and the square bracketed term in (2.17) are of order γ and γ^2, respectively, so above the CWP (2.17) and (2.21) they give the first three terms in the high-temperature γ expansion for ψ.

We have given the above reduction in some detail in order to point out the (formal) identity of Eqs. (2.17) and (2.21), and the perturbation results for the models discussed in Ref. 8. Thus, if one compares (2.17) and (2.21) above with Eqs. (2.13) and (2.16), combined with (2.43) of Ref. 8 for the 1-dimensional model [cf. Eqs. (4.13) and (5.9) of Ref. 8 for the 2- and 3-dimensional models,respectively], one will note that the formulas are identical; that is, if one identifies $R(\nu\gamma)$, $\gamma R_{22}(\nu\gamma)$, and $\gamma R_{42}(\nu\gamma)$ above with ρ, T_2, and T_4, respectively, defined in Ref. 8, so that, formally, at least, one can immediately carry over the model critical-region analysis to the general formulas given above. Before going into this in some detail, let us first consider two examples which will illustrate what we have in mind:

(a) the 1-dimensional Kac model[6]

$$\rho_t(\gamma) = e^{-\gamma|t|} \tag{2.24}$$

and

(b) the 1-dimensional infinite-range model discussed in Ref. 8,

$$\rho_t(\gamma) = \int_0^1 \lambda^a e^{-\lambda\gamma|t|}\, d\lambda, \quad 0 < a < 1. \tag{2.25}$$

The formulas for $R(\nu\gamma)$, $R_{22}(\nu\gamma)$, etc., for (a) are given in Appendix C of Paper I, so we will simply summarize the results:

$$g(\omega) = \sum_{k=-\infty}^{\infty} e^{ik\omega-\gamma|k|} = \frac{\sinh\gamma}{\cosh\gamma - \cos\omega}, \tag{2.26}$$

where the CWP is given from Eqs. (2.7) and (2.19) by

$$\nu_{CW} = [\gamma g(0)]^{-1} = \gamma^{-1}\tanh\tfrac{1}{2}\gamma = \tfrac{1}{2} + O(\gamma^2); \tag{2.27}$$

$R(\nu\gamma)$ and $f(\nu)$ defined by Eqs. (2.20) and (2.18), respectively, are given by

$$\begin{aligned} R(\nu\gamma) &= \sinh\gamma\{(\cosh\gamma - \nu\gamma\sinh\gamma)^2 - 1\}^{-\frac{1}{2}} \\ &= (1-2\nu)^{-\frac{1}{2}} + O(\gamma^2) \end{aligned} \tag{2.28}$$

and

$$\begin{aligned} f(\nu) &= -\cosh^{-1}(\cosh\gamma - \nu\gamma\sinh\gamma) + \gamma \\ &= -\gamma[(1-2\nu)^{\frac{1}{2}} - 1] + O(\gamma^3); \end{aligned} \tag{2.29}$$

and the "third-order terms" [Eqs. (2.21), (2.22), and (2.23)] are to leading order

$$(\nu\gamma)^4\tfrac{1}{4}[R(\nu\gamma)]^2 R_{22}(\nu\gamma) \sim \gamma^3(1-2\nu)^{-\frac{5}{2}} \tag{2.30}$$

and

$$(\nu\gamma)^4\frac{1}{2\times 3!}R_{42}(\nu\gamma) \sim \tfrac{1}{2}\gamma^3(1-2\nu)^{-\frac{5}{2}}, \tag{2.31}$$

respectively.

We see then that the "first-order" terms in ψ are given by

$$-\tfrac{1}{2}\nu\gamma + f(\nu) \sim -\gamma[(1-2\nu)^{\frac{1}{2}} - 1 + \tfrac{1}{2}\nu] + O(\gamma^3); \tag{2.32}$$

the "second-order" terms, by

$$\begin{aligned} \frac{1}{2}\left(\frac{\nu\gamma R(\nu\gamma)}{1+\nu\gamma R(\nu\gamma)} - \log[1+\nu\gamma R(\nu\gamma)]\right) \\ = \tfrac{1}{4}[\nu\gamma R(\nu\gamma)]^2 + O(\gamma^3) \\ = (\tfrac{1}{2}\nu\gamma)^2(1-2\nu)^{-1} + O(\gamma^3); \end{aligned} \tag{2.33}$$

and the third-order terms, by (2.28) and (2.29). It is then obvious that the expansions break down when

$$1 - 2\nu \sim \gamma^{\frac{2}{3}}, \tag{2.34}$$

since in this region there are terms of order $\gamma^{\frac{4}{3}}$ in first, second, third, and, as is not difficult to convince oneself, all orders. Equation (2.34) defines the critical region, and the idea is to resum the series for ψ [Eq. (2.17)], typically by picking out the most divergent terms in all orders of perturbation. This is clearly a hopeless task in the above case, since from (2.30) and (2.31) *the two third-order terms contribute equally in the critical region. This situation, as we will show in the following section, occurs in general when* $R(\nu\gamma)$ *diverges algebraically at* ν_{CW} [cf. Eq. (2.28)], and certainly for model (a) we know from the outset that there is no phase transition (for finite γ). We conjecture that this is in general the case when $R(\nu\gamma)$ diverges algebraically at ν_{CW}. The situation is completely different for model (b), Eq. (2.25). In this case, from (2.25) and (2.26),

$$g(\gamma\omega) = \int_0^1 \frac{\lambda^a \sinh(\lambda\gamma)\, d\lambda}{\cosh(\lambda\gamma) - \cos(\omega\gamma)} \sim \frac{2}{\gamma}h(\omega), \tag{2.35}$$

where

$$h(x) = \int_0^1 \frac{\lambda^{a+1}}{\lambda^2 + x^2}\, d\lambda \tag{2.36}$$

and

$$R(\nu\gamma) = \frac{1}{2\pi}\int_0^{2\pi}\frac{g(\omega)}{1-\nu\gamma g(\omega)}\, d\omega \sim \frac{1}{\pi}\int_0^{\infty}\frac{h(x)\, dx}{1-2\nu h(x)}, \tag{2.37}$$

which is precisely the form for ρ given in Ref. 8 [see in particular Eqs. (A1) and (A11) of Appendix A].

The CWP from (2.7) and (2.35) is given by

$$\nu_{\mathrm{CW}} = \tfrac{1}{2}a, \tag{2.38}$$

and, for $0 < a < 1$ [see Eqs. (3.5) and (3.6), respectively, of Ref. 8; note that in general $df/d\nu = \frac{1}{2}\gamma R$],

$$-\frac{2}{\gamma a^2} f(\nu) \sim A(a - 2\nu) + B(a - 2\nu)^{1/a} + \cdots, \tag{2.39}$$

$$R(\nu\gamma) \sim 2a^2[A + (B/a)(a - 2\nu)^{(1/a)-1}] + \cdots; \tag{2.40}$$

and, from (2.22) and (2.20) [cf. Eq. (3.11) of Ref. 8],

$$\gamma R_{22}(\nu\gamma) = \frac{\partial}{\partial \nu} R(\nu\gamma) \sim -4B(1 - a)(a - 2\nu)^{(1/a)-2}. \tag{2.41}$$

It is now evident that the expansion (2.17), as for model (a), will break down for that range of ν around ν_{CW}, such that

$$f(\nu) \sim \tfrac{1}{4}[\nu\gamma R(\nu\gamma)]^2; \tag{2.42}$$

i.e., from (2.39) and (2.40), when

$$a - 2\nu \sim \gamma. \tag{2.43}$$

In this case, terms of order $\gamma^{(1/a)+1}$ occur in first and second order and also, from (2.41) and (2.21), in third order; moreover, it is not difficult to convince oneself that such terms will be present in all orders. The essential difference, however, between models (a) and (b) becomes clear when we note that the other third-order term in (2.21) [cf. Eq. (3.12) of Ref. 8]

$$\frac{1}{2 \times 3!}(\nu\gamma)^4 R_{42}(\nu\gamma) \sim \gamma^{2/a}, \quad \text{when} \quad a - 2\nu \sim \gamma, \tag{2.44}$$

which, since $a < 1$, is of higher order than $\gamma^{(1/a)+1}$.

For model (b), then, the "most divergent terms" in the critical region [defined by Eq. (2.43)] in first, second, and third order are, respectively, $f(\nu)$,

$$\tfrac{1}{4}[\nu\gamma R(\nu\gamma)]^2 = \tfrac{1}{2}[\nu^2\gamma R(\nu\gamma)]\frac{df}{d\nu}$$

and

$$\tfrac{1}{4}(\nu\gamma)^4[R(\nu\gamma)]^2 R_{22}(\nu\gamma) = \tfrac{1}{2}[\nu^2\gamma R(\nu\gamma)]^2 \frac{d^2 f}{d\nu^2}; \tag{2.45}$$

and the conjecture stated in Ref. 8 for the models may be now stated (in general) as follows:

The most divergent term in $(n + 1)$th order is given by

$$[-\nu^2\gamma R(\nu\gamma)]^n \frac{1}{n!}\frac{d^n f}{d\nu^n}, \quad n > 2; \tag{2.46}$$

and all other terms in the critical region are of higher order, if $R(\nu\gamma)$ is finite or diverges no worse than logarithmically at ν_{CW}.

This conjecture will be proved in the following section.

3. CRITICAL-REGION RESUMMATION

To prove the result (2.46), we use the graphical representation of q [Eq. (2.21)] given in Paper I, which is briefly as follows:

Expanding the exponential in (2.13), we obtain

$$q_N^{(2)} = \sum_{n=0}^{\infty} \frac{1}{n!}(-\tfrac{1}{2}\nu\gamma u^2)^n \left\langle \left(\sum_{k,l}{}' \tilde{\rho}_{kl}\eta_k\eta_l \right)^n \prod_j \varphi(\eta_j) \right\rangle, \tag{3.1}$$

where

$$\varphi(\eta_j) = \cos(u\eta_j)/\langle \cos u\eta \rangle = e^{\frac{1}{2}u^2} \cos(u\eta_j) \tag{3.2}$$

and $\prod_j$ extends over all lattice points j. If we now represent a pair of points (p_α', p_α'') by p_α and denote $\tilde{\rho}_{p_\alpha' p_\alpha''}$ by $\tilde{\rho}_{p_\alpha}$, we can write (3.1) as

$$q_N^{(2)} = \sum_{N=0}^{\infty} \frac{1}{N!}(-\tfrac{1}{2}\nu\gamma u^2)^N \times \sum_{p_1, p_2, \cdots, p_N}{}' \prod_{\alpha=1}^{N} \tilde{\rho}_{p_\alpha} F_N(p_1, p_2, \cdots, p_N), \tag{3.3}$$

where the sum is over all pairs of distinct lattice points and

$$F_N(p_1, p_2, \cdots, p_N) = \left\langle \prod_{\alpha=1}^{N} \eta_{p_\alpha'}\eta_{p_\alpha''} \prod_j \varphi(\eta_j) \right\rangle. \tag{3.4}$$

The form (3.3) is now completely analogous to a grand canonical partition function of a "gas," the pairs corresponding to the position vectors of the gas molecules and the sum over pairs to the integral over coordinates. We can then represent each term in the sum over pairs by a graph, if each pair p_α is represented by a line joining the two lattice points p_α' and p_α'' and by the standard cluster argument[10]

$$N^{-1} \log q_N^{(2)} = \sum_{\lambda=2}^{\infty} (-\nu\gamma u^2)^\lambda b_\lambda \tag{3.5}$$

with cluster coefficients b_λ given by

$$N\lambda!\, b_\lambda = \sum_{p_1, \cdots, p_\lambda} S_\lambda(p_1, \cdots, p_\lambda) \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha}, \tag{3.6}$$

where $S_\lambda(p_1, \cdots, p_\lambda)$ are Ursell cluster functions associated with the weights $F_N(p_1, \cdots, p_N)$. Moreover, from the cluster property of the S_λ's and Eq. (3.4), only even connected graphs contribute to (3.6).

The weights $F_N(p_1, \cdots, p_N)$ can be calculated very easily as follows. If the graph has m_f points of degree f (the degree of a point is the number of lines entering

the point), it is obvious that

$$F_\lambda(p_1, \cdots, p_\lambda) = \prod_f [\langle \eta^f \cos u\eta \rangle / \langle \cos u\eta \rangle]^{m_f}, \quad (3.7)$$

where

$$\lambda = \tfrac{1}{2} \sum_f f m_f; \quad (3.8)$$

and, from (2.14), it follows that

$$\langle \eta^f \cos u\eta \rangle / \langle \cos u\eta \rangle = e^{\frac{1}{2}u^2}(-1)^{\frac{1}{2}f}\left(\frac{\partial}{\partial u}\right)^f e^{-\frac{1}{2}u^2}$$
$$= (-1)^{\frac{1}{2}f} H_f(u), \quad (3.9)$$

where $H_f(u)$ is the Hermite polynomial of degree f. For example, from (2.10),

$$H_2(u) = u^2 - 1$$
$$= -\nu\gamma\tilde{\rho}_0(1 + \nu\gamma\tilde{\rho}_0)^{-1}. \quad (3.10)$$

Usually, the easiest class of graphs to sum is the rings, i.e., all points are of degree 2. In this case the ring contribution to $N\lambda!\, b_\lambda$ is clearly

$$(N\lambda!\, b_\lambda)_{\text{rings}}$$
$$= \tfrac{1}{2}(\lambda - 1)!\, [-H_2(u)]^\lambda \sum_{k, l_1, \cdots, l_{\lambda-1}} \tilde{\rho}_{kl_1}\tilde{\rho}_{l_1l_2}, \cdots, \tilde{\rho}_{l_{\lambda-1}k}$$
$$= \tfrac{1}{2}(\lambda - 1)!\, [-H_2(u)]^\lambda \sum_k (\tilde{\rho}^\lambda)_{kk}. \quad (3.11)$$

If we now proceed to the limit $N \to \infty$, recalling from (2.8) and (2.19) that

$$\lim_{N\to\infty} (\tilde{\rho}^\lambda)_0 = \frac{1}{2\pi}\int_0^{2\pi} \left(\frac{g(\omega)}{1 - \nu\gamma g(\omega)}\right)^\lambda d\omega$$
$$= \frac{2\gamma^{-\lambda}}{(\lambda - 1)!}\frac{d^\lambda f}{d\nu^\lambda}, \quad (3.12)$$

where $f(\nu)$ is defined by (2.18), we see that the ring contribution to q [Eq. (2.21)] is simply

$$q_{\text{rings}} = \sum_{\lambda=2}^{\infty} [-\nu\gamma u^2(1 - u^2)]^\lambda \frac{\gamma^{-\lambda}}{\lambda!}\frac{d^\lambda f}{d\nu^\lambda}, \quad (3.13)$$

where (in the limit $N \to \infty$)

$$u^2 = [1 + \nu\gamma R(\nu\gamma)]^{-1} = 1 - \nu\gamma R(\nu\gamma) + \cdots. \quad (3.14)$$

Thus, since $u^2(1 - u^2) \sim \nu\gamma R(\nu\gamma)$ for small γ, (3.13) is the sum of the "most divergent terms" in Eq. (2.46). In other words, the critical-region analysis in the functional-integral formalism boils down to a sum over ring diagrams. Our work will be complete, then, once we have proved that the rings give the dominant contribution to q in the critical region. We will prove in fact that this is the case if and only if $R(\nu\gamma)$ diverges no worse than logarithmically at ν_{CW}.

Let us first consider the case where $R(\nu\gamma)$ is finite at ν_{CW} and write [cf. model (b) in the previous section, Eq. (2.40) in particular]

$$R(\nu\gamma) = A + B(\nu_{\text{CW}} - \nu)^{\alpha-1} + \cdots, \quad \alpha > 1. \quad (3.15)$$

The high-temperature expansion (2.17) clearly breaks down when

$$\nu_{\text{CW}} - \nu \sim \gamma, \quad (3.16)$$

since in this region (the critical region), noting from (2.45), (3.15), and (3.14), respectively, that

$$\gamma^{-\lambda}\frac{d^\lambda f}{d\nu^\lambda} \sim \gamma^{1-\lambda}(\nu_{\text{CW}} - \nu)^{\alpha-\lambda} \sim \gamma^{\alpha-2\lambda+1} \quad (3.17)$$

and

$$1 - u^2 \sim \gamma, \quad (3.18)$$

the rings, contribute a term of order $\gamma^{\alpha+1}$ to the free energy ψ. To estimate the sum over diagrams excluding rings, we essentially follow the argument given in Appendix B of Paper I. Thus, consider first graphs of λ lines containing one point of degree $f > 2$, and from (3.8), $\lambda - \frac{1}{2}f$ points of degree 2. From (3.9) and (3.10) [also (3.14)], the point of degree $f > 2$ has weight $H_f(u)$, which is finite when $u \to 1$, $\gamma \to 0$, and each point of degree 2 has weight $H_2(u)$ which contributes order γ in the critical region. The total weight in the critical region is, therefore, of order $\gamma^{\lambda-\frac{1}{2}f}$; and, from (3.6), the contribution to $N\lambda!\, b_\lambda$ from these graphs is bounded by

$$(N\lambda!\, b_\lambda)' \le C^\lambda \gamma^{\lambda-\frac{1}{2}f} \sum_{p_1, \cdots, p_\lambda} \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha}, \quad (3.19)$$

where C is a (suitably chosen) constant, independent of γ, and the sum is over all graphs with λ lines containing one point of degree $f > 2$ and $\lambda - \frac{1}{2}f$ points of degree 2. The sum in (3.19) can be estimated as follows: Since the graphs under consideration are contained in the class of graphs consisting of $\frac{1}{2}f$ chains leaving and returning to the point k, with no restriction on the remaining points, we have

$$\sum_{p_1, \cdots p_\lambda} \prod_{\alpha=1}^{\lambda} \tilde{\rho}_{p_\alpha} \le \sum'_{\lambda_1, \cdots, \lambda_{\frac{1}{2}f}} \sum_k \prod_{\sigma=1}^{f/2} \left(\sum_{\{l\}} \tilde{\rho}_{kl_1}\tilde{\rho}_{l_1l_2} \cdots \tilde{\rho}_{l_\sigma k}\right), \quad (3.20)$$

where σ labels the $\frac{1}{2}f$ chains and λ_σ denotes the number of $\tilde{\rho}$ bonds in the chain, so that

$$\sum_{\sigma=1}^{\frac{1}{2}f} \lambda_\sigma = \lambda. \quad (3.21)$$

The right-hand side of (3.20) is easily seen to be

$$\sum'_{\lambda_1, \cdots, \lambda_{\frac{1}{2}f}} \sum_k \prod_{\sigma=1}^{\frac{1}{2}f} (\tilde{\rho}^{\lambda_\sigma})_{kk} = N \sum'_{\lambda_1, \cdots, \lambda_{\frac{1}{2}f}} \prod_{\sigma=1}^{\frac{1}{2}f} (\tilde{\rho}^{\lambda_\sigma})_0; \quad (3.22)$$

and since in the limit $N \to \infty$, from (2.20) and (3.15),

$$(\tilde{\rho}^n)_0 = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{g(\omega)}{1-\nu\gamma g(\omega)}\right)^n d\omega$$
$$= \gamma^{1-n}\frac{d^{n-1}R(\nu\gamma)}{d\nu^{n-1}}$$
$$\sim \gamma^{1-n}(\nu_{CW}-\nu)^{\alpha-n}$$
$$\sim \gamma^{\alpha-2n+1} \text{ in the critical region,} \quad (3.23)$$

Eqs. (3.19)–(3.23) give

$$(N\lambda!\, b_\lambda)' \leqslant \gamma^{\lambda-\frac{1}{2}f}\gamma^{\frac{1}{2}f(\alpha+1)}\gamma^{-2\lambda} = \gamma^{\frac{1}{2}f\alpha-\lambda} \quad (3.24)$$

for the bound on the critical-region contribution from graphs with one point of degree $f > 2$ and all other points of degree 2. The contribution to the free energy, from Eqs. (3.5) and (2.17), is thus at least of order $\gamma^{\frac{1}{2}f\alpha}$, which, since $\alpha > 1$ and $f > 2$, is of higher order than the ring contribution $\gamma^{\alpha+1}$.

The above proof can be easily extended to graphs containing more than one point of degree higher than two, but we leave this for the reader.

The next case to consider is when $R(\nu\gamma)$ diverges logarithmically at ν_{CW}, i.e.,

$$R(\nu\gamma) \sim -A\log(\nu_{CW}-\nu), \quad \text{as} \quad \nu \to \nu_{CW}^{-}. \quad (3.25)$$

By comparing the first two terms in the high-temperature expansion (i.e., $f(\nu)$ and $\frac{1}{4}[\nu\gamma R(\nu\gamma)]^2$) for ψ, we see that the critical region is of order

$$\nu_{CW}-\nu \sim \gamma\log(\gamma)^{-1}, \quad (3.26)$$

and in this region it is trivially verified from (3.13) that the rings contribute a term of order

$$\gamma^2\log(\gamma)^{-1}. \quad (3.27)$$

To estimate the contribution from other graphs, we proceed exactly as in Eqs. (3.19)–(3.24). Thus, for graphs containing one point of degree $f > 2$ and the remainder of degree 2, the weight is of order $(\gamma\log(\gamma)^{-1})^{\lambda-\frac{1}{2}f}$ and the bound (3.20) is of order

$$[\gamma(\nu_{CW}-\nu)]^{\frac{1}{2}f-\lambda} \sim [\gamma^2\log(\gamma)^{-1}]^{\frac{1}{2}f-\lambda}, \quad (3.28)$$

giving a contribution to $N\lambda!\, b_\lambda$ at least of order $\gamma^{\frac{1}{2}f-\lambda}$, which, combined with Eqs. (3.5) and (2.17), gives a contribution to the free energy ψ at least of order $\gamma^{\frac{1}{2}f}$. The next order correction (after rings) to q is, therefore, at least of order γ^2 ($f = 4$)! The proof, as before, may be easily extended to graphs containing more than one point of degree higher than 2.

It remains to consider the case when $R(\nu\gamma)$ diverges algebraically at ν_{CW} [typified by the 1-dimensional model (a) of the previous section], i.e.,

$$R(\nu\gamma) \sim (\nu_{CW}-\nu)^{-\alpha}, \quad \alpha > 0. \quad (3.29)$$

In this case, by comparing the first two terms in the high-temperature expansion, we see that the critical region is of order

$$\nu_{CW}-\nu \sim \gamma^{1/(\alpha+1)}, \quad (3.30)$$

and in this region the rings contribute a term of order $\gamma^{2/(\alpha+1)}$. Notice now, however, from the above arguments, that the contribution from graphs with at least one point of degree $f > 2$ is of order $\gamma^{f/2(1+\alpha)}$, so that when $f = 4$, these graphs are of the same order as the rings! In fact, it is not difficult to convince oneself that a large class of graphs will contribute terms of equal order in the critical region.

The analysis above is similar in many ways to the α trick analysis given in Sec. 7 of Paper I, which is briefly as follows.

Going back to the beginning, Eq. (2.1), one adds and subtracts

$$-\tfrac{1}{2}N(\alpha-\nu\gamma) = -\tfrac{1}{2}\sum_{k,l}\delta_{k,l}(\alpha-\nu\gamma\rho_{k-l}(\gamma)) \quad (3.31)$$

to the exponent in (2.2), instead of $-\frac{1}{2}N(1-\nu\gamma)$. The resulting modification of Eq. (2.4) is then legitimate even below the CWP, if α is chosen to be larger than $\nu\gamma g(0)$ [cf. Eq. (2.7)]. The free energy per spin is then obtained in I, Eqs. (7.10)–(7.13) and (7.18):

$$\psi = -kT(\mathcal{F}_1+\mathcal{F}_2), \quad (3.32)$$

where

$$\mathcal{F}_1 = \log 2u + \tfrac{1}{2}(\alpha-\nu\gamma-u^2\alpha) + \int_0^{\nu\gamma/\alpha}R(t)\,dt, \quad (3.33)$$

with

$$R(t) = \left(\frac{1}{2\pi}\right)^D\int_0^{2\pi}\cdots\int\frac{g(\omega)\,d^D\omega}{1-tg(\omega)} \quad (3.34)$$

(D is dimensionality) and

$$u = [1+(\nu\gamma/\alpha)R(\nu\gamma/\alpha)]^{-\frac{1}{2}} \quad (3.35)$$

and $\mathcal{F}_1$ corresponds to the first five terms in Eq. (2.16). The term $\mathcal{F}_2$ is a diagram series analogous to Eq. (3.5), which can be obtained from Eq. (7.9) of Paper I.

The free energy ψ given by Eq. (3.32) is, of course, independent of α, while $\mathcal{F}_1$ and $\mathcal{F}_2$ depend on α. The best approximation to ψ by $-kT\mathcal{F}_1$ alone is, therefore, obtained by choosing the value of α for which $\mathcal{F}_1$ is stationary (this could also give a locally worst approximation, but other arguments in I suggest that this is very unlikely). This leads to the choice of α as a function of ν, given by Eq. (7.16) of Paper I,

$$\alpha - 1 = (\nu\gamma/\alpha)R(\nu\gamma/\alpha), \quad (3.36)$$

or, equivalently, Eq. (8.13) of Paper I,

$$1 = \left(\frac{1}{2\pi}\right)^D \int_0^{2\pi} \cdots \int \frac{d^D\omega}{\alpha - \nu\gamma g(\omega)}. \tag{3.37}$$

One easily sees that $d\alpha/d\nu > 0$, $d(\alpha/\nu)/d\nu < 0$, and that $\alpha \to 1$ as $\nu \to 0$.

The function $\mathcal{F}_1$ and any diagram in $\mathcal{F}_2$ are analytic functions of ν not only above the CWP, but also, at least, in a small temperature region below the CWP. The failure of the previous expansion at the CWP is thus avoided.

As was pointed out in I, Sec. 7, the diagram series $\mathcal{F}_2$ for the above choice of α contains no diagrams with points of degree less than four. The result of the resummation in the present paper, therefore, agrees with the dominant terms in $\mathcal{F}_1$.

In I, we also compared $\mathcal{F}_1$ with the results of Brout and Mühlschlegel and Zittartz and found formal agreement. In terms of the models of Ref. 8, the α trick is equivalent to using, in the Schrödinger-type equation as zeroth-order approximation, not the naturally occurring oscillator potential, but a properly chosen oscillator potential, which does not become flat in any direction when the temperature is lowered through the CWP. The total potential ceases to have only one minimum there, but the ground-state wavefunction still has only one maximum, if the temperature is sufficiently close to the CWP. In this temperature region (the critical region) the use of an oscillator potential as zeroth approximation is, therefore, reasonable, but one expects that such an approximation, as well as the α trick, will fail at still lower temperatures. This expectation is borne out by the remarks in I, Sec. 8: $\mathcal{F}_1$ can become singular only at a temperature defined by

$$\alpha(\nu_0) = \nu_0\gamma g(0). \tag{3.38}$$

This, by Eq. (3.37), leads to

$$\nu_0 = \left(\frac{1}{2\pi}\right)^D \int_0^{2\pi} \cdots \int \frac{d^D\omega}{\gamma[g(0) - g(\omega)]}. \tag{3.39}$$

If, for example, ρ_k has a second moment so that

$$g(0) - g(\omega) \sim \omega^2, \quad \text{for small } \omega, \tag{3.40}$$

one has $\nu_0 = 0$ for the 1- and 2-dimensional models, which is in agreement with the fact that the results for the 2-dimensional case are quite similar to those for the 2-dimensional spherical model which does not have a phase transition. For the 3-dimensional case one finds that while $\nu_0 > 0$, the specific heat per spin remains finite at the temperature corresponding to ν_0. As we pointed out in I, Sec. 8, Eq. (3.38) also contradicts the initial assumption $\alpha > \nu\gamma g(0)$, made to justify the representation [Eq. (7.4) of Paper I] from which Eqs. (3.32)–(3.37) are derived.

While these arguments show that the approximation

$$\psi \approx -kT\mathcal{F}_1, \tag{3.41}$$

with α determined by Eqs. (3.36) and (3.37), cannot correctly represent the phase transition, the estimates of the diagrams given above for temperatures in the critical region *above the* CWP hold *a fortiori* for the diagrams in $\mathcal{F}_2$ in the critical region *above and below* the CWP. The approximation given by Eq. (3.41) is thus the dominant term in the critical region, except where $R(\nu\gamma)$ diverges algebraically at ν_{CW}. The exception accounts for the discrepancies obtained for the 1-dimensional model (a), discussed in the previous section, and at the end of Sec. 9 in I.

The approximation (3.41) thus agrees with the result of the resummation above to leading order and extends it to the critical region below the CWP. Specifically, one has to leading order,

$$\alpha(\nu_{CW}) = 1 + \nu_{CW}\gamma A, \tag{3.42}$$

when $R(\nu_{CW}\gamma) = A$ is finite, and

$$\alpha(\nu_{CW}) = 1 + \nu_{CW}A\gamma \log(\gamma)^{-1}, \tag{3.43}$$

when $R(\nu\gamma) \sim -A \log(\nu_{CW} - \nu)$, as $\nu \to \nu_{CW}^-$.

While it is clear that the α trick as well as the resummation will fail when the temperature approaches the true critical temperature, the temperature at which this failure occurs can be expected to be a better approximation to the true critical temperature than the CW temperature. One can obtain an improved estimate of the critical temperature under the assumption that $\mathcal{F}_1$ is still a reasonable approximation, but taking for α not the value obtained from Eq. (3.36) or (3.37), but the value obtained in the critical region. One then obtains the estimates given by Eqs. (4.13) in the following section.

4. SUMMARY AND DISCUSSION

We now summarize and discuss the main results, which for simplicity are given for the 1-dimensional chain of spins $\mu_i = \pm 1$, $i = 1, 2, \cdots, N$, with periodic and ferromagnetic potential $J\gamma\rho_{k-l}(\gamma)$; i.e., the interaction energy is given by [Eq. (2.1)]

$$E = -\tfrac{1}{2}J\gamma \sum_{k\neq l=1}^{N} \rho_{k-l}(\gamma)\mu_k\mu_l. \tag{4.1}$$

For temperatures above the classical Curie–Weiss temperature T_{CW} ($\nu_{CW} = J/kT_{CW}$), i.e. [Eq. (2.7)],

$$\nu < \nu_{CW} = \left(\gamma \sum_{k=1}^{N} \rho_k\right)^{-1}, \tag{4.2}$$

the free energy per spin ψ in the limit $N \to \infty$ is given by [Eq. (2.17)]

$$-\psi/kT = \log 2 - \tfrac{1}{2}\nu\gamma + f(\nu) + \frac{1}{2}\left(\frac{\nu\gamma R(\nu\gamma)}{1 + \nu\gamma R(\nu\gamma)} - \log\left(1 + \nu\gamma R(\nu\gamma)\right)\right) + q, \tag{4.3}$$

where

$$f(\nu) = -\tfrac{1}{2}\int_0^{2\pi} \log\left(1 - \nu\gamma g(\omega)\right)\frac{d\omega}{2\pi}, \tag{4.4}$$

$$g(\omega) = \sum_{k=-\infty}^{\infty} e^{ik\omega}\rho_k(\gamma), \tag{4.5}$$

$$R(\nu\gamma) = \int_0^{2\pi} \frac{g(\omega)}{1 - \nu\gamma g(\omega)}\frac{d\omega}{2\pi}, \tag{4.6}$$

and q is defined by Eqs. (2.21) and (2.13). (In D dimensions the k's, l's, ω's, etc., above are D-dimensional vectors and $\int_0^{2\pi} d\omega/2\pi \cdots$ is replaced by

$$\int_0^{2\pi}\cdots\int d^D\omega/(2\pi)^D \cdots$$

etc.)

The reduction to Eq. (4.3) is given in I, where a graphical prescription for computing terms in the γ expansion for q can also be found. Our concern here has been the nature of this expansion in the critical region (where the expansion breaks down). In the graph formulation of I we have been able to prove that in the critical region the ring diagrams give the dominant contribution to q (and hence ψ), if and only if $R(\nu\gamma)$ is finite or diverges no worse than logarithmically at ν_{CW}. The ring contribution to q is given by [Eq. (3.13)]

$$\begin{aligned} q_{\text{rings}} &= \sum_{\lambda=2}^{\infty}[-\nu u^2(1-u^2)]^{\lambda}\frac{1}{\lambda!}\frac{d^{\lambda}f}{d\nu^{\lambda}} \\ &= f(\nu[1 - u^2(1-u^2)]) \\ &\quad + \nu u^2(1-u^2)\frac{df}{d\nu} - f(\nu), \end{aligned} \tag{4.7}$$

where

$$u^2 = [1 + \nu\gamma R(\nu\gamma)]^{-1}. \tag{4.8}$$

We list three cases:

(i) *$R(\nu_{CW}\gamma) = A$ is finite.*

In this case the critical region is of order γ [Eqs. (3.15) and (3.16)]. When $\nu_{CW} - \nu \sim \gamma$, since to leading order

$$u^2(1-u^2) \sim \nu\gamma R(\nu\gamma) \sim \nu_{CW}\gamma A, \tag{4.9}$$

the dominant contribution to ψ [Eqs. (4.3) and (4.7)] is given by

$$-\frac{\psi}{kT} \sim \log 2 - \tfrac{1}{2}\nu\gamma + \tfrac{1}{4}[\nu\gamma A]^2 - \tfrac{1}{2}\int_0^{2\pi} \log\left[1 - \nu\gamma(1 - \nu_{CW}\gamma A)g(\omega)\right]\frac{d\omega}{2\pi}. \tag{4.10}$$

(ii) *$R(\nu\gamma)$ diverges logarithmically, as $\nu \to \nu_{CW}^-$;*

i.e.,

$$R(\nu\gamma) \sim -A \log(\nu_{CW} - \nu), \quad \text{as} \quad \nu \to \nu_{CW}^-. \tag{4.11}$$

(One actually only requires that the divergence be weaker than any power.)

In this case the critical region is of order $\gamma \log \gamma^{-1}$ [Eq. (3.26)] and to leading order, when $\nu_{CW} - \nu \sim \gamma \log \gamma^{-1}$, ψ *is given by Eq.* (4.10) *with A replaced by $A \log \gamma^{-1}$.*

(iii) *$R(\nu\gamma)$ diverges algebraically as $\nu \to \nu_{CW}^-$;*

i.e.,

$$R(\nu\gamma) \sim A(\nu_{CW} - \nu)^{-\alpha}, \quad \alpha > 0, \quad \text{as} \quad \nu \to \nu_{CW}^-. \tag{4.12}$$

In this case the critical region is of order $\gamma^{1/(\alpha+1)}$ [Eqs. (3.29) and (3.30)], and when $\nu_{CW} - \nu \sim \gamma^{1/(\alpha+1)}$, the rings no longer give the dominant contribution to ψ. Thus, a simple resummation cannot be effected.

The α trick discussed in I and at the end of the previous section gives essentially the same results as above *in the critical region and if $R(\nu\gamma)$ diverges no worse than logarithmically at ν_{CW}*. Thus, if one compares (3.32) with (2.17) and notes the critical-region values for α [Eqs. (3.42) and (3.43), corresponding to cases (i) and (ii) above, respectively], it is clear that the α trick in the critical region is equivalent, to lowest order, to a summation on rings. The advantage of the α trick is that it can be easily extended to higher orders; thus, there is hope that the free energy minimization (with respect to α) to next order will give the next order critical region, and so on. Work along these lines is proceeding, and we hope to report on this at a later date.

Returning to (iii) above, where a resummation cannot be effected and where the α trick fails, one might expect that the systems in this case do not have phase transitions for any finite value of γ [which is in fact true for model (a) in Sec. 2, Eq. (2.24)].

In cases (i) and (ii) the resummations shift the critical point from ν_{CW} to ν'_C, which, from Eq. (4.10), is given by (to the accuracy of the resummation)

$$\nu'_C = \nu_{CW}(1 + \nu_{CW}\gamma A), \quad \text{for (i)},$$
$$\nu'_C = \nu_{CW}(1 + \nu_{CW}A\gamma \log \gamma^{-1}), \quad \text{for (ii)}, \quad (4.13)$$

so that in these cases we might expect that the system has a phase transition. We, in fact, state the following:

Conjecture I: An Ising model with interaction (4.1) has a phase transition if and only if (a) $R(\nu\gamma)$ given by (4.6) is finite or diverges no worse than logarithmically at ν_{CW}, or, equivalently, (b) the rings in the functional integral formulation of Paper I give the dominant contribution in the critical region.

Conjecture I(b) is clearly more general than (a), since the formalism of Paper I can be applied to almost any classical system with long-range interaction. One might hope then that, in general, an examination of the rings will correctly diagnose the existence or nonexistence of a phase transition.

Note that the integral $R(\nu\gamma)$ (Eq. 4.6) occurs in the equation determining the saddle point for the spherical model,[11] and, as is well known, the spherical model with interaction (4.1) has a phase transition if and only if $R(\nu\gamma)$ is finite at ν_{CW}. Conjecture I(a) is, then, partly related to a conjecture of Kac,[12] which states that an Ising model with a periodic and ferromagnetic potential will have a phase transition if the corresponding spherical model has a phase transition.

The logarithmic case (ii) occurs typically in two dimensions and case (i) in three dimensions. Since it is known from the work of Griffiths[13] that Ising models in two and three dimensions with interaction (4.1) have phase transitions, it is clear that Conjecture I(a) is true for dimensionality higher than one. The question then arises: What happens in one dimension? An almost complete answer to this question has been given recently by Dyson,[14] whose results are as follows: For the 1-dimensional Ising model with interaction (4.1) such that ρ_n is positive and monotonically decreasing with

$$\sum_{n=1}^{\infty} \rho_n < \infty, \quad (4.14)$$

the system has a phase transition if

$$\sum_{n=1}^{\infty} \log \log (n + 4)[n^3\rho_n]^{-1} < \infty \quad (4.15)$$

and does not have a phase transition if

$$\lim_{N\to\infty} (\log \log N)^{-1} \sum_{n=1}^{\infty} n\rho_n = 0. \quad (4.16)$$

Equation (4.16) is slightly stronger than a previous result of Ruelle[15] which states that the model (4.1) does not have a phase transition if $\sum_{n=1}^{\infty} n\rho_n$ is finite. Dyson's result (4.16) is sufficient to disprove a conjecture of Kac,[7] which states that the model (4.1) has a phase transition if and only if $\sum_{n=1}^{\infty} n\rho_n$ is unbounded. Note that this conjecture is not equivalent to Conjecture I(a), which in terms of the moment $\sum_{n=1}^{\infty} n\rho_n$ may be stated as follows:

Conjecture II: The 1-dimensional Ising model with interaction (4.1) has a phase transition if and only if $\sum_{n=1}^{N} n\rho_n$ diverges at least logarithmically as $N \to \infty$.

This is most easily seen for the particular case

$$\rho_n = (n^{1+a})^{-1}, \quad (4.17)$$

which is essentially equivalent to model (b) Eq. (2.25). Thus, from the definitions (4.5) and (4.6), it is clear that $R(\nu\gamma)$ is finite at $\nu = \nu_{CW}$ if $0 < a < 1$, diverges logarithmically as $\nu \to \nu_{CW}^-$ if $a = 1$, and diverges algebraically as $\nu \to \nu_{CW}^-$ if $a > 1$. Therefore, Conjecture I(a) implies that the 1-dimensional model (4.1) and (4.17) has a phase transition when $0 < a \leq 1$ and does not have a phase transition when $a > 1$. This conclusion also follows from Conjecture II. From Dyson's theorem, on the other hand [Eqs. (4.15) and (4.16)], we conclude that the 1-dimensional model (4.17) does, in fact, have a phase transition when $0 < a < 1$ and does not have a phase transition when $a > 1$. The case $a = 1$, however, is undecided at the moment. In essence, then, we may consider our Conjecture I(a) to be proved in all cases except for the 1-dimensional model (4.1) with ρ_n such that $\sum_{n=1}^{\infty} n\rho_n$ diverges logarithmically, or simply stated:

Conjecture III: The 1-dimensional Ising model with interaction $\rho_n = n^{-2}$ has a phase transition.

This model, then, is an important test case for our resummation procedure or, rather, our diagnostic Conjecture I(b).

Two independent results which support Conjecture III are the following:

(a) Joyce[16] has developed high-temperature susceptibility expansions for the n^{-2} model and concludes, using series extrapolation techniques, that the susceptibility diverges at a finite temperature T_c like $(T - T_c)^{-3}$.

(b) Rapaport and Frankel[17] have analyzed numerically the model

$$\rho_n = n^{-(1+a)}, \quad n \leq L,$$
$$\rho_n = 0, \quad n > L, \tag{4.18}$$

for various values of L (up to $L = 20$) and find that for $0 < a \leq 1$ the specific heat maxima appear to diverge with increasing L, and in particular for $a = 1$, it appears that $C_{\max}$ is proportional to $\log L$, a result which is very reminiscent of the finite 2-dimensional Ising model.[18] In fact, as we have seen, the $a = 1$ [logarithmically divergent $R(v\gamma)$] case is very similar to normal 2-dimensional systems.

In conclusion, we make the trivial observation that for nearest-neighbor Ising models the above expansions may be thought of as expansions in the inverse of the lattice coordination number q. For the f.c.c. lattice $q = 12$, so that an expansion in inverse powers of q may be of some use in this case. In particular, in three dimensions, from Eqs. (4.9) and (4.13), the modified critical point is given by

$$v'_C = v_{CW}[1 + v_{CW}\gamma R(v_{CW}\gamma)], \tag{4.19}$$

where the square bracketed term [from Eq. (4.6)] is simply the Watson integral[19]

$$W = \left(\frac{1}{2\pi}\right)^3 \iiint_0^{2\pi} \frac{d^3\omega}{1 - v_{CW}\gamma g(\omega)}$$
$$= 1.3446610732 \cdots \text{f.c.c.} \quad (q = 12)$$
$$= 1.3932039297 \cdots \text{b.c.c.} \quad (q = 8)$$
$$= 1.5163860591 \cdots \text{s.c.} \quad (q = 6). \tag{4.20}$$

Since for nearest neighbor systems $v_{CW} = q^{-1}$ [Eq. (4.2)], we have

$$v'_C = 0.11 \quad \text{f.c.c.} \quad (0.102),$$
$$= 0.17 \quad \text{b.c.c.} \quad (0.157),$$
$$= 0.25 \quad \text{s.c.} \quad (0.222), \tag{4.21}$$

with the numbers in parentheses denoting the series estimates of Domb *et al.*[20] The agreement, as one would expect, is best for the largest value of $q(12)$. Note also that the values for v'_C are essentially the critical values for the spherical model.[11,21]

For 2-dimensional Ising models, from (4.6),

$$1 + v\gamma R(v\gamma) = \left(\frac{1}{2\pi}\right)^2 \iint_0^{2\pi} \frac{d^2\omega}{1 - v\gamma g(\omega)}$$
$$\sim -\log(v_{CW} - v). \tag{4.22}$$

As we have seen, to leading order in the critical region ($v_{CW} - v \sim \gamma \log \gamma^{-1}$) it is appropriate to replace $v_{CW} - v$ in (4.22) by γ, or, in this case, by q^{-1}, giving for the modified critical point, by analogy with (4.19),

$$v'_C = (\log q)/q. \tag{4.23}$$

For the triangular ($q = 6$) and the square ($q = 4$) lattices, this gives the estimates

$$v'_C = 0.29 \text{ triangular} \quad (\text{exact}^{20} = 0.2746 \cdots)$$
$$= 0.35 \text{ square} \quad (\text{exact}^{20} = 0.4407 \cdots), \tag{4.24}$$

and again, as one would expect, the agreement is best for the larger q value.

Thus, it seems that our resummation method locates the critical point quite accurately, at least for small γ. However, the first resummation, resulting in Eq. (4.10), predicts a logarithmically divergent energy in two dimensions and a square root specific heat divergence in three dimensions; so the outstanding problem remains: how is the correct critical behavior approached (if at all) by this method? We are hopeful that the next order critical-region analysis will throw some light on this difficult problem.

* Permanent address from September 1969: Northwestern University, Evanston, Illinois 60680.

† Supported by the National Science Foundation and the Office of Naval Research.

‡ Supported in part by NSF Grants GP7328 and GP8734.

[1] A. J. F. Siegert and D. J. Vezzetti, J. Math. Phys. **9**, 2173 (1968).
[2] R. Brout, Phys. Rev. **118**, 1009 (1960).
[3] G. Horwitz and H. B. Callen, Phys. Rev. **124**, 1757 (1961).
[4] J. L. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. **6**, 1282 (1965); G. Stell, J. L. Lebowitz, S. Baer, and W. Theumann, *ibid.* **7**, 1532 (1966).
[5] B. Mühlschlegel and H. Zittartz, Z. Physik **175**, 553 (1963).
[6] M. Kac, Phys. Fluids **2**, 8 (1959).
[7] M. Kac, in *Brandeis Summer Lectures 1966* (Gordon and Breach, Science Publishers, Inc., New York, 1968).
[8] M. Kac and C. J. Thompson, J. Math. Phys. **10**, 1373 (1969).
[9] H. Cramer, *Mathematical Methods in Statistics* (Princeton University Press, Princeton, N. J., 1951), p. 118.
[10] 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.
[11] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[12] E. H. Lieb and C. J. Thompson, J. Math. Phys. **10**, 1403 (1969).
[13] R. B. Griffith, J. Math. Phys. **8**, 478 (1967).
[14] F. J. Dyson, Commun. Math. Phys. **12**, 91, 212 (1969).
[15] D. Ruelle, Commun. Math. Phys. **9**, 267 (1968).
[16] G. S. Joyce, reported by C. Domb, Nat. Bur. Std. (U.S.), Misc. Publ. **273**, 29 (1966).
[17] D. Rapaport and N. E. Frankel, Phys. Letters **28A**, 405 (1968).
[18] L. Onsager, Phys. Rev. **65**, 117 (1944).
[19] G. N. Watson, Quart. J. Math. **10**, 266 (1939).
[20] M. E. Fisher, J. Math. Phys. **4**, 278 (1963).
[21] M. Lax, Phys. Rev. **97**, 629 (1955).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Space–Temperature Correlations in Quantum-Statistical Mechanics*†

WILLIAM KUNKIN‡
Department of Chemistry, State University of New York at Albany, Albany, New York
AND
JEROME K. PERCUS
Courant Institute of Mathematical Sciences, New York University, New York, New York

(Received 3 April 1969)

The method of functional differentiation, used in classical statistical mechanics to obtain approximate integral equations for the pair distribution function, is extended to quantum systems obeying Maxwell–Boltzmann statistics. The grand partition function is written as a path integral, and space–temperature distributions are generated from it by successive functional differentiation. Distributions can be expanded in powers of an external potential via a functional Taylor series. When the series is truncated so that no distributions higher than second order enter and the external potential is specialized to one arising from a fixed particle in the system, coupled integral equations result for the two kinds of pair distributions. The first-order Ursell function calculated from these equations yields the Montroll–Ward ring summation for the grand potential. This approximate Ursell function determines also the Fourier transform of the momentum density (useful in calculating the $\mathbf{p} = 0$ occupation number). Although divergences appear in the low-temperature limit, they can be removed by a simple modification of the independent functional. An extension to Fermi–Dirac and Böse–Einstein statistics is indicated at appropriate places in the text.

1. INTRODUCTION

Numerous authors[1–4] have investigated approximation methods for classical statistical mechanics based upon functional differentiation and the associated functional Taylor expansion. Thus, by proper choice of independent and dependent functionals taken in the Taylor series, Percus was able to derive simply a wide variety of approximate integral equations for the pair distribution function, including the Debye–Hückel, Percus–Yevick, and convolution hypernetted chain equations.[5] Lebowitz and Percus later applied the method in several other directions; for example, to obtain rigorous bounds on distributions of any order.[6]

This work is the beginning of an extension of the foregoing into quantum-statistical mechanics. It is not difficult to adapt the classical techniques once the grand partition function has been given a classical-like form, i.e., written as a Feynman path integral. The distribution functions dealt with depend not only on position or momentum (the latter are most profitably defined by means of a phase-space path-integral representation of the partition function), but on temperature parameters which can vary between infinity and the true system temperature (see Sec. 2). No particular physical interpretation has been given to this dependence on temperature and, in fact, the quantities of physical interest are found from the pair distribution in which temperature intervals have been set to zero. These distributions satisfy functional derivative relations very similar to those for Green's functions[7] and they are, in the Fermi–Dirac or Böse–Einstein case, diagonal elements of the Green's functions.

We will find here (Sec. 3), just as in the classical theory, that there is a crucial relationship connecting a nonuniform density to the pair distribution for a uniform system. The nonuniformity comes about by "fixing" a particle so that its pair interaction with the remaining particles of the system now becomes an external potential. We shall see that this offers no difficulties even in the quantum case: when properly formulated, there are no recoil effects.

The number and kinds of integral equations derivable by the methods of this work (we derive one set only) seem to be severely limited. There are at least two causes for this. First, we require that no path integrals appear in the final equations. Second, there appears to be no analog of the Boltzmann factor $e^{-\beta\phi(\mathbf{x})}$, which correlates only two particles. Thus, except briefly in Sec. 4, we shall treat only the basic density-potential expansion in which both limitations are avoidable. The resulting equations may be termed the quantum Debye–Hückel approximation.

2. COORDINATE AND MOMENTUM DISTRIBUTIONS

Consider a system of a large number of particles in a volume V interacting through a pair potential and obeying Maxwell–Boltzmann (classical) statistics. Denoting the N-particle Hamiltonian by

$$H_N = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m} + \sum_{i<j} \varphi(\mathbf{x}_i - \mathbf{x}_j),$$

the grand canonical partition function will be

$$Z = \sum_{N=0}^{\infty} \frac{z^N}{N!} \operatorname{Tr} e^{(-\beta H_N)},$$

in which no symmetry restrictions are observed in computing the trace. The fugacity is denoted by z, while $\beta = 1/kT$.

We shall find it more convenient to employ the Lagrangian path-integral representation of Z,[8] namely,

$$Z = \sum_N \frac{z^N}{N!} C^{NM} \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_1 \, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \cdots d\mathbf{x}_N \, d\mathbf{x}_{N\beta_1} \cdots d\mathbf{x}_{N\beta_{M-1}} \times \exp\left[-\sum_{k=1}^{M} \epsilon \left(\sum_{i=1}^{N} \frac{m}{2\hbar^2} \dot{\mathbf{x}}_{i\beta_k}^2 + \sum_{i<j} \varphi(\mathbf{x}_{i\beta_k} - \mathbf{x}_{j\beta_k}) \right)\right], \tag{2.1}$$

in which

$$\beta_k - \beta_{k-1} = \epsilon, \quad \beta_M = M\epsilon = \beta, \quad \beta_0 = 0,$$

and

$$\mathbf{x}_{i\beta} = \mathbf{x}_{i0} \equiv \mathbf{x}_i, \quad i = 1, \cdots, N.$$

It is to be understood, if not explicitly stated, that the limit $\epsilon \to 0$ is taken in (2.1). The ϵ-dependent normalization factor C insures that the integrals converge.

The expectation value of any path-dependent function of coordinates is found by inserting the function into the integrand of (2.1) and dividing by Z. We shall denote such expectation values by $\langle \ \rangle_L$. If, however, the function depends on momenta as well, we must use the Hamiltonian or phase-space path-integral representation of Z,[9,10] i.e.,

$$Z = \sum_N \frac{z^N}{N!} (2\pi\hbar)^{-3MN} \int_{-\infty}^{\infty} \cdots \int d\mathbf{p}_{1\beta_1} \cdots d\mathbf{p}_{1\beta_M} \cdots d\mathbf{p}_{N\beta_1} \cdots d\mathbf{p}_{N\beta_M} \, d\mathbf{x}_1 \times d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \cdots d\mathbf{x}_N \, d\mathbf{x}_{N\beta_1} \cdots d\mathbf{x}_{N\beta_{M-1}} \times \exp\left[\sum_{k=1}^{M} \epsilon \left(\sum_{j=1}^{N} \frac{i}{\hbar} \mathbf{p}_{j\beta_k} \dot{\mathbf{x}}_{j\beta_k} - \sum_{j=1}^{N} \frac{\mathbf{p}_{j\beta_k}^2}{2m} - \sum_{j<l} \varphi(\mathbf{x}_{j\beta_k} - \mathbf{x}_{l\beta_k}) \right)\right]. \tag{2.2}$$

Expectations employing (2.2) will be denoted by $\langle \ \rangle_H$.

We define the one-particle coordinate distribution $n_1(\mathbf{x}', \beta')$ by

$$n_1(\mathbf{x}', \beta') = \langle N\delta(\mathbf{x}_{1\beta'} - \mathbf{x}') \rangle_L, \quad 0 < \beta' < \beta. \tag{2.3}$$

More conventionally, this function can be written in terms of a trace:

$$n_1(\mathbf{x}', \beta') = \sum_{N=1}^{\infty} \frac{z^N N}{ZN!} \operatorname{Tr} [\delta(\mathbf{x}_1 - \mathbf{x}') e^{[-\beta H_N]}].$$

(The calculation is facilitated by factoring the path integrals into two groups, one containing integrals occurring "earlier" than β' and the other containing integrals occurring "later" than β', then replacing each group by its equivalent "thermal" propagator.[11]) Thus (2.3) is the ordinary density function; it is independent of β' in the range 0–β and is a constant, ρ, for a uniform system.

The two-particle (pair) distribution is defined as the expectation value of a product of two δ functions, but since the arguments of the δ functions can refer to the same path or two different paths, there exists two such distributions, namely, a "same-particle" distribution,

$$n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'') = \langle N\delta(\mathbf{x}_{1\beta'} - \mathbf{x}')\delta(\mathbf{x}_{1\beta''} - \mathbf{x}'') \rangle_L, \tag{2.4}$$

and a "different-particle" distribution,

$$n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta'') = \langle N(N-1)\delta(\mathbf{x}_{1\beta'} - \mathbf{x}')\delta(\mathbf{x}_{2\beta''} - \mathbf{x}'') \rangle_L. \tag{2.5}$$

(A semicolon will separate arguments belonging to different particles.) Just as with Eq. (2.3), both distributions can be written in conventional notation. The functions are symmetric and satisfy

$$\begin{aligned} n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'') &= Z^{-1} \sum_{N=1}^{\infty} \frac{z^N}{N!} N \operatorname{Tr} [e^{-(\beta-\beta')H_N} \delta(\mathbf{x}_1 - \mathbf{x}') \\ &\quad \times e^{-(\beta'-\beta'')H_N} \delta(\mathbf{x}_1 - \mathbf{x}'') e^{-\beta'' H_N}], \quad \text{for} \quad \beta' > \beta'', \\ &= n_1(\mathbf{x}', \beta')\delta(\mathbf{x}' - \mathbf{x}''), \quad \text{for} \quad \beta' = \beta'', \end{aligned}$$

and

$$\begin{aligned} n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta'') &= Z^{-1} \sum_{N=2}^{\infty} \frac{z^N}{N!} N(N-1) \operatorname{Tr} [e^{-(\beta-\beta')H_N} \delta(\mathbf{x}_1 - \mathbf{x}') \\ &\quad \times e^{-(\beta'-\beta'')H_N} \delta(\mathbf{x}_2 - \mathbf{x}'') e^{-\beta'' H_N}], \quad \text{for} \quad \beta' > \beta'' \\ &= Z^{-1} \sum_{N=2}^{\infty} \frac{z^N}{N!} N(N-1) \\ &\quad \times \operatorname{Tr} [\delta(\mathbf{x}_1 - \mathbf{x}')\delta(\mathbf{x}_2 - \mathbf{x}'') e^{-\beta H_N}], \\ &\quad \text{for} \quad \beta' = \beta. \end{aligned}$$

The different-particle pair distribution at equal temperatures is the physical pair distribution. While the same-particle distribution has no direct physical significance, it is useful in evaluating momentum, kinetic energy or other one-particle expectations.[12]

We shall occasionally meet with three- or four-particle distributions in the course of this work, but their definitions proceed along the lines already indicated for the one- and two-particle functions.

We turn now to momentum expectations. The formalism developed here permits us to calculate the expectation of any function of momentum by inserting the function into the integrand of (2.2), and integrating and dividing by Z. To illustrate, we calculate the expectation value of the kinetic energy

$$(2m)^{-1}\langle N\mathbf{p}_{1\beta'}^2\rangle_H .$$

From (2.2)

$$\begin{aligned}
&\frac{1}{2m}\langle N\mathbf{p}_{1\beta'}^2\rangle_H \\
&= Z^{-1}\sum_{N=1}^{\infty}\frac{z^N}{N!}N\frac{(2\pi\hbar)^{-3MN}}{2m} \\
&\quad\times\int_{-\infty}^{\infty}\cdots\int d\mathbf{p}_{1\beta_1}\cdots d\mathbf{p}_{1\beta'}\cdots d\mathbf{p}_{N\beta_M} \\
&\quad\times\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_1\cdots d\mathbf{x}_N\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_{1\beta_1}\cdots d\mathbf{x}_{N\beta_{M-1}}\mathbf{p}_{1\beta'}^2 \\
&\quad\times\exp\sum_{k=1}^{M}\epsilon\left(\sum_{j=1}^{N}\frac{i}{\hbar}\mathbf{p}_{j\beta_k}\dot{\mathbf{x}}_{j\beta_k}-\sum_{j=1}^{N}\frac{\mathbf{p}_{j\beta_k}^2}{2m}\right. \\
&\qquad\left.-\sum_{j<l}^{N}\phi(\mathbf{x}_{j\beta_k}-\mathbf{x}_{l\beta_k})\right).
\end{aligned}$$

We complete the square, which gives

$$\begin{aligned}
&-\frac{\mathbf{p}_{j\beta_k}^2}{2m}+\frac{i}{\hbar}\mathbf{p}_{j\beta_k}\dot{\mathbf{x}}_{j\beta_k} \\
&\qquad=-\frac{1}{2m}\left(\mathbf{p}_{j\beta_k}-\frac{i}{\hbar}m\dot{\mathbf{x}}_{j\beta_k}\right)^2-\frac{m}{2\hbar^2}\dot{\mathbf{x}}_{j\beta_k}^2 \\
&\qquad\equiv-\frac{1}{2m}\mathbf{p}_{j\beta_k}'^2-\frac{m}{2\hbar^2}\dot{\mathbf{x}}_{j\beta_k}^2,
\end{aligned}$$

and all the momentum integrals can be done, which leaves

$$\begin{aligned}
&\frac{1}{2m}\langle N\mathbf{p}_{1\beta'}^2\rangle_H \\
&\quad= Z^{-1}\sum_{N=1}^{\infty}\frac{z^N}{N!}NC^{MN} \\
&\qquad\times\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_1\cdots d\mathbf{x}_N \\
&\qquad\times\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_{1\beta_1}\cdots d\mathbf{x}_{1\beta'}\cdots d\mathbf{x}_{N\beta_{M-1}} \\
&\qquad\times\left(-\frac{m}{2\hbar^2}\dot{\mathbf{x}}_{1\beta'}^2+\frac{3}{2\epsilon}\right) \\
&\qquad\times\exp\left[-\sum_{k=1}^{M}\epsilon\left(\sum_{i=1}^{N}\frac{m}{2\hbar^2}\dot{\mathbf{x}}_{i\beta_k}^2-\sum_{i<j}^{N}\phi(\mathbf{x}_{i\beta_k}-\mathbf{x}_{j\beta_k})\right)\right] \\
&\quad= N\left\langle-\frac{m}{2\hbar^2}\dot{\mathbf{x}}_{1\beta'}^2+\frac{3}{2\epsilon}\right\rangle_L .
\end{aligned}\tag{2.6}$$

Curiously, the first term on the right is not by itself the average kinetic energy. The average square velocity does not exist but becomes infinite as $\epsilon\to 0$. The expectation of a function which depends on two space points on the same path involves the "same-particle" pair distribution; thus,

$$\begin{aligned}
&\frac{1}{2m}\langle N\mathbf{p}_{1\beta'}^2\rangle_H \\
&\quad=\lim_{\epsilon\to 0}-\frac{m}{2\hbar^2\epsilon^2} \\
&\qquad\times\int_{-\infty}^{\infty}d\mathbf{x}\,d\mathbf{y}\left((\mathbf{x}-\mathbf{y})^2-\frac{3\hbar^2\epsilon}{m}\right) \\
&\qquad\times n_{2s}(\mathbf{x},\beta',\mathbf{y},\beta'-\epsilon) \\
&\quad=\lim_{\epsilon\to 0}-\frac{mV}{2\hbar^2\epsilon^2}\int_{-\infty}^{\infty}d\mathbf{z}\left(\mathbf{z}^2-\frac{3\hbar^2\epsilon}{m}\right)n_{2s}(\mathbf{z},\epsilon).
\end{aligned}\tag{2.7}$$

The last line follows from the translational invariance of uniform distributions with respect to space and temperature variables, so we may drop the β'.

Precisely as in (2.3), the one-particle momentum distribution is given by

$$n_1(\mathbf{p},\beta')=\langle N\delta(\mathbf{p}_{1\beta'}-\mathbf{p})\rangle_H . \tag{2.8}$$

By a calculation similar to that preceding (2.7), one obtains the relation between this momentum distribution and the "same-particle" pair distribution, viz.,

$$\begin{aligned}
n_1(\mathbf{p},\beta')&=\lim_{\epsilon\to 0}\left(\frac{\epsilon}{2\pi m}\right)^{\frac{3}{2}}\int_{-\infty}^{\infty}d\mathbf{x}\,d\mathbf{y}\,n_{2s}(\mathbf{x},\beta',\mathbf{y},\beta'-\epsilon) \\
&\quad\times\exp\left[-\frac{\epsilon}{2m}\left(\mathbf{p}-\frac{im}{\hbar\epsilon}(\mathbf{x}-\mathbf{y})\right)^2\right]
\end{aligned}$$

or

$$\begin{aligned}
n_1(\mathbf{p})&=\lim_{\epsilon\to 0}\left(\frac{\epsilon}{2\pi m}\right)^{\frac{3}{2}}V\int_{-\infty}^{\infty}d\mathbf{z}\,n_{2s}(\mathbf{z},\epsilon) \\
&\quad\times\exp\left[-\frac{\epsilon}{2m}\left(\mathbf{p}-\frac{im}{\hbar\epsilon}\mathbf{z}\right)^2\right].
\end{aligned}\tag{2.9}$$

Manifestly real versions of this relation may also be obtained; e.g.,

$$\begin{aligned}
\tilde{n}(\mathbf{q})&\equiv\int_{-\infty}^{\infty}d\mathbf{p}\exp\left(-\frac{i}{\hbar}\mathbf{q}\cdot\mathbf{p}\right)n_1(\mathbf{p})\,d\mathbf{p} \\
&=\lim_{\epsilon\to 0}V\exp\left(-\frac{m\mathbf{q}^2}{2\epsilon\hbar^2}\right)\int_{-\infty}^{\infty}d\mathbf{z}\,n_{2s}(\mathbf{z},\epsilon) \\
&\quad\times\exp\left(\frac{m\mathbf{q}\cdot\mathbf{z}}{\epsilon\hbar^2}\right).
\end{aligned}\tag{2.10}$$

It is easy to see that when the $\mathbf{p}=0$ state is important, $\tilde{n}(\mathbf{q})$ has the property

$$\lim_{|\mathbf{q}|\to\infty}\tilde{n}(\mathbf{q})=\text{number of particles having zero momentum,}$$

so that if n_{2s} could be bounded for some system of interacting particles, the extent to which the system condenses in momentum space at $\beta = \infty$ (Boson condensation) could be investigated. At absolute zero, Boltzmann and Boson systems are identical and the number of particles having zero momentum is in that case $\lim \lim \tilde{n}(\mathbf{q})$, as $|\mathbf{q}|$, $\beta \to \infty$. The $\beta \to \infty$ limit must be taken first.

The momentum pair distributions (there are two of them) involve four-particle coordinate distributions since each momentum factor is effectively the difference of two space points slightly separated in time. Thus the momentum pair distribution for two different particles,

$$n_2(\mathbf{p}', \beta'; \mathbf{p}'', \beta'') = \langle N(N-1)\delta(\mathbf{p}_{1\beta'} - \mathbf{p}')\delta(\mathbf{p}_{2\beta''} - \mathbf{p}'')\rangle_H, \quad (2.11)$$

is readily shown to be

$$\lim_{\epsilon \to 0} \left(\frac{\epsilon}{2\pi m}\right)^3 \int_{-\infty}^{\infty} d\mathbf{w}\, d\mathbf{x}\, d\mathbf{y}\, d\mathbf{z} \times n_{4s}(\mathbf{w}, \beta', \mathbf{x}, \beta' - \epsilon; \mathbf{y}, \beta'', \mathbf{z}, \beta'' - \epsilon) \times \exp - \frac{\epsilon}{2m}\left(\mathbf{p}' - (\mathbf{w} - \mathbf{x})\frac{im}{\epsilon\hbar}\right)^2 \times \exp - \frac{\epsilon}{2m}\left(\mathbf{p}'' - (\mathbf{y} - \mathbf{z})\frac{im}{\epsilon\hbar}\right)^2,$$

where the four-particle distribution is a mixed type: The arguments $(\mathbf{w}, \beta')$ and $(\mathbf{x}, \beta' - \epsilon)$ refer to one particle, and $(\mathbf{y}, \beta'')$ and $(z, \beta'' - \epsilon)$ refer to a second particle.

All of the formalism developed thus far is immediately extendable to quantum statistics. One simply permits the particles, which are located at $(\mathbf{x}_1, \cdots, \mathbf{x}_N)$ at "time" zero, to complete their paths at any permutation of the $(\mathbf{x}_1, \cdots, \mathbf{x}_N)$ at time β because of indistinguishability. One then performs the usual sum over all permutations in computing Z.[13] Our definitions of the distribution functions can be taken over without further change.

3. INTEGRAL EQUATIONS FOR THE PAIR DISTRIBUTION FUNCTIONS

Let us append to the exponent of (2.1) the external potential

$$U = \sum_{k=1}^{M} \epsilon \sum_{i=1}^{N} U(\mathbf{x}_{i\beta_k}, \beta_k) = \int_0^{\beta} \sum_{i=1}^{N} U(\mathbf{x}_{i\gamma}, \gamma)\, d\gamma,$$

and let us compute the first two functional derivatives of $\ln Z[U]$ with respect to the external potential. From the modified (2.1), we have

$$\frac{\delta \ln Z[U]}{-\delta U(\mathbf{x}', \beta')} = \left\langle \epsilon \sum_{k=1}^{M} \sum_{i=1}^{N} \frac{\delta U(\mathbf{x}_{i\beta_k}, \beta_k)}{\delta U(\mathbf{x}', \beta')} \right\rangle_{L[U]} = \langle N\delta(\mathbf{x}_{i\beta'} - \mathbf{x}')\rangle_{L[U]} = n_1(\mathbf{x}', \beta' \mid U), \quad (3.1)$$

by (2.3), where we have used, for discrete temperatures,

$$\frac{\delta U(\mathbf{x}, \gamma)}{\delta U(\mathbf{x}', \gamma')} = \delta(\mathbf{x} - \mathbf{x}')\delta_{\gamma,\gamma'}\frac{1}{\epsilon}.$$

Differentiating once more, we have

$$\frac{\delta^2 \ln Z[U]}{\delta U(\mathbf{x}', \beta')\delta U(\mathbf{x}'', \beta'')} = -\frac{\delta n_1(\mathbf{x}', \beta' \mid U)}{\delta U(\mathbf{x}'', \beta'')} = \left\langle N\delta(\mathbf{x}_{1\beta'} - \mathbf{x}')\epsilon \sum_{k=1}^{M} \sum_{i=1}^{N} \frac{\delta U(\mathbf{x}_{i\beta_k}, \beta_k)}{\delta U(\mathbf{x}'', \beta'')} \right\rangle_{L[U]} - n_1(\mathbf{x}', \beta' \mid U)n_1(\mathbf{x}'', \beta'' \mid U) = n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta'' \mid U) + n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'' \mid U) - n_1(\mathbf{x}', \beta' \mid U)n_1(\mathbf{x}'', \beta'' \mid U), \quad (3.2)$$

by Eqs. (2.4) and (2.5).

The functional Taylor expansions[14] for $\ln Z[U]$ vs. U, and $n_1(\mathbf{x}', \beta' \mid U)$ vs. U now follow with the aid of (3.1) and (3.2). They are, through terms quadratic and linear in U, respectively $(\epsilon \to 0)$[15]:

$$\ln Z[U] = \ln Z - \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}'\, d\gamma' n_1(\mathbf{y}', \gamma')U(\mathbf{y}', \gamma') + \tfrac{1}{2}\int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}'\, d\mathbf{y}''\, d\gamma'\, d\gamma''[n_2(\mathbf{y}', \gamma'; \mathbf{y}'', \gamma'') + n_{2s}(\mathbf{y}', \gamma', \mathbf{y}'', \gamma'') - \rho^2] \times U(\mathbf{y}', \gamma')U(\mathbf{y}'', \gamma'') + \cdots \quad (3.3)$$

and

$$n_1(\mathbf{x}', \beta' \mid U) = \rho - \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}'\, d\gamma'[n_2(\mathbf{x}', \beta'; \mathbf{y}', \gamma') + n_{2s}(\mathbf{x}', \beta', \mathbf{y}', \gamma') - \rho^2]U(\mathbf{y}', \gamma') + \cdots, \quad (3.4)$$

in which all of the distribution appearing on the right-hand sides are for uniform systems. Higher terms in the series introduce distributions for three or more particles.

We are now very close to the desired approximate integral equations for the two pair distributions. [Note that (3.4) is the derivative of (3.3), so that the equations are consistent.] What is needed are identities relating the nonuniform distributions appearing on the left of (3.3) and (3.4) to uniform distributions. These can be found if a special kind of external

potential is used, viz., one in which a fixed particle, identical to the remaining $N-1$ in the system, provides the external potential through its pair interaction with the rest. Such an external potential is, fixing particle "one,"

$$U_\phi = \sum_{k=1}^{M} \epsilon \sum_{i=2}^{N} \phi(\mathbf{x}_{i\beta_k} - \mathbf{x}_{1\beta_k}).$$

The modified (2.1) is in detail

$$\begin{aligned} Z[U_\phi] \equiv \sum_{N=1}^{\infty} \frac{z^{N-1}}{(N-1)!} C^{(N-1)M} \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_2 \cdots d\mathbf{x}_N \\ \times \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_{2\beta_1} \cdots d\mathbf{x}_{N\beta_{M-1}} \\ \times \exp\left[-\sum_{k=1}^{M} \epsilon \left(\sum_{i=2}^{N} \frac{m}{2\hbar^2} \dot{\mathbf{x}}_{i\beta_k}^2 \right.\right. \\ \left.\left. + \sum_{1<i<j}^{N} \phi(\mathbf{x}_{1\beta_k} - \mathbf{x}_{j\beta_k}) \right) \right] e^{-U\phi}, \end{aligned} \tag{3.5}$$

where all integrations over the first particle have been omitted. No recoil effects occur as a result of fixing a particle: its kinetic energy is a constant with respect to the other path integrations. But $Z[U_\phi]$ converts easily to $n_1(\mathbf{x}', \beta')$ for a *uniform* system through

$$\begin{aligned} zC^M \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_1 \, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \\ \times \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') Z[U_\phi] e^{-K\{\mathbf{x}_{1\gamma}\}} = n_1(\mathbf{x}', \beta') Z, \end{aligned} \tag{3.6}$$

where

$$K\{\mathbf{x}_{1\gamma}\} \equiv \sum_{k=1}^{M} \epsilon \frac{m}{2\hbar^2} \dot{\mathbf{x}}_{1\beta_k}^2,$$

since

$$zC^M \cdot \frac{z^{N-1}}{(N-1)!} C^{(N-1)M} = \frac{z^N}{N!} N C^{NM}.$$

[Compare with Eq. (2.3).]

Similarly,

$$\begin{aligned} zC^M \int_{-\infty}^{\infty} \cdots d \int \mathbf{x}_1 \, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \\ \times \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') \delta(\mathbf{x}_{1\beta''} - \mathbf{x}'') e^{-K\{\mathbf{x}_{1\gamma}\}} Z[U_\phi] \\ = n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'') Z, \end{aligned} \tag{3.7}$$

and any higher "same-particle" distribution can be generated in like manner by including more δ-functions. Finally, starting with $n_1(\mathbf{x}'', \beta'' \mid U_\phi)$, we obtain

$$\begin{aligned} zC^M \int_{-\infty}^{\infty} \cdots d\mathbf{x}_1 \int d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \\ \times \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') n_1(\mathbf{x}'', \beta'' \mid U_\phi) Z[U_\phi] e^{-K\{x_{1\gamma}\}} \\ = n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta'') Z. \end{aligned} \tag{3.8}$$

It is in these relationships (3.6)–(3.8) that the major differences between this treatment and a quantum-statistical one appear. It can be shown that the analogous relations in the latter case can only be obtained by fixing a number of particles say t, rather than one, subject to the endpoint conditions

$$\mathbf{x}_{10} = \mathbf{x}_{t\beta}, \; \mathbf{x}_{20} = \mathbf{x}_{1\beta}, \; \cdots, \; \mathbf{x}_{t0} = \mathbf{x}_{t-1\beta},$$

forming a closed loop. (Compare this with the concept of the "toron." [16]) It is then necessary to sum over all values of t.

Let us return to the functional Taylor expansions (3.3) and (3.4), and replace U by U_φ and $U(\mathbf{y}', \gamma')$ by $\varphi(\mathbf{y}' - \mathbf{x}_{1\gamma'})$, etc. Exponentiate (3.3), multiply by zC^M and perform the integrations indicated in (3.7). One obtains the expansion (after linearizing)

$$\begin{aligned} n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'') \\ = zC^M \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_1 \, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \\ \times \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') \delta(\mathbf{x}_{1\beta''} - \mathbf{x}'') \, e^{-K\{\mathbf{x}_{1\gamma}\}} \\ \times \left(1 - \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}' \, d\gamma' \rho \phi(\mathbf{y}' - \mathbf{x}_{1\gamma'}) \right. \\ + \tfrac{1}{2} \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}' \, d\mathbf{y}'' \, d\gamma' \, d\gamma'' [n_2(\mathbf{y}', \gamma'; \mathbf{y}'', \gamma'') \\ + n_{2s}(\mathbf{y}', \gamma', \mathbf{y}'', \gamma'') - \rho^2] \phi(\mathbf{y}' - \mathbf{x}_{1\gamma'}) \\ \times \left. \phi(\mathbf{y}'' - \mathbf{x}_{1\gamma''}) + \cdots \right). \end{aligned} \tag{3.9}$$

Observe that for free particles (denoted by 0) (3.6) reads

$$\begin{aligned} zC^M \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_1 \, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') \\ \times e^{-K\{\mathbf{x}_{1\gamma}\}} = n_1^0(\mathbf{x}', \beta'), \end{aligned} \tag{3.10}$$

and similarly for higher "same-particle" distributions. Using (3.10), Eq. (3.9) now becomes

$$\begin{aligned} n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'') \\ = n_{2s}^0(\mathbf{x}', \beta', \mathbf{x}'', \beta'') \left(1 - \rho\beta \int_{-\infty}^{\infty} \phi(\mathbf{y}') \, d\mathbf{y}' \right) \\ + \tfrac{1}{2} \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}' \, d\mathbf{y}'' \, d\mathbf{z}' \, d\mathbf{z}'' \, d\gamma' \, d\gamma'' \\ \times [n_2(\mathbf{y}', \gamma'; \mathbf{y}'', \gamma'') + n_{2s}(\mathbf{y}', \gamma', \mathbf{y}'', \gamma'') - \rho^2] \\ \times \phi(\mathbf{y}' - z') \phi(\mathbf{y}'' - z'') \\ \times n_{4s}^0(\mathbf{x}', \beta', \mathbf{x}'', \beta'', \mathbf{z}', \gamma', \mathbf{z}'', \gamma''), \end{aligned} \tag{3.11}$$

where n_{4s}^0 is the "same-particle" distribution for four free particles.

We have one integral equation for the pair distributions. The second is obtained by multiplying (3.4) by $zC^M Z[U_\phi]$ and integrating by means of (3.6)–(3.8). We obtain

$$\begin{aligned} n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta'') &= \rho^2 - zC^M \int_{-\infty}^{\infty} d\mathbf{x}_1\, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \\ &\quad \times \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') Z[U_\phi] Z^{-1} e^{-K\{\mathbf{x}_{1\gamma}\}} \\ &\quad \times \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}'\, d\gamma' [n_2(\mathbf{x}'', \beta''; \mathbf{y}', \gamma') \\ &\quad + n_{2s}(\mathbf{x}'', \beta'', \mathbf{y}', \gamma') - \rho^2] \phi(\mathbf{y}' - \mathbf{x}_{1\gamma'}) + \cdots \\ &= \rho^2 - \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}'\, d\mathbf{y}''\, d\gamma' [n_2(\mathbf{x}'', \beta''; \mathbf{y}', \gamma') \\ &\quad + n_{2s}(\mathbf{x}'', \beta'', \mathbf{y}', \gamma') - \rho^2] \phi(\mathbf{y}' - \mathbf{y}'') \\ &\quad \times n_{2s}(\mathbf{x}', \beta', \mathbf{y}'', \gamma'). \end{aligned} \tag{3.12}$$

When truncated as indicated, Eqs. (3.12) and (3.11) are coupled integral equations for the pair distributions.

4. EXAMPLES AND COMMENTS

We define the two-particle Ursell function

$$\begin{aligned} F(\mathbf{x}', \beta', \mathbf{x}', \beta'') &\equiv n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta'') \\ &\quad + n_{2s}(\mathbf{x}', \beta', \mathbf{x}'', \beta'') - \rho^2, \\ \tilde{F}(\mathbf{k}, \beta', \beta'') &= \int_{-\infty}^{\infty} d\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} F(\mathbf{x}, \beta', \beta''). \end{aligned} \tag{4.1}$$

An equation for the Ursell function through terms linear in the potential is, by (3.12) and (3.11),

$$\begin{aligned} F(\mathbf{x}', \beta', \mathbf{x}'', \beta'') &= n_{2s}^0(\mathbf{x}', \beta', \mathbf{x}'', \beta'') \\ &\quad - \int_{-\infty}^{\infty} \int_0^{\beta} d\mathbf{y}'\, d\mathbf{y}''\, d\gamma' F(\mathbf{x}'', \beta'', \mathbf{y}', \gamma') \\ &\quad \times \phi(\mathbf{y}' - \mathbf{y}'') n_{2s}^0(\mathbf{x}', \beta', \mathbf{y}'', \gamma'). \end{aligned} \tag{4.2}$$

Taking the Fourier transform, we obtain

$$\begin{aligned} \tilde{F}(\mathbf{k}, \beta', \beta'') &= \tilde{n}_{2s}^0(\mathbf{k}, \beta', \beta'') - \int_0^{\beta} d\gamma' \tilde{F}(\mathbf{k}, \beta'', \gamma') \\ &\quad \times \tilde{\phi}(\mathbf{k}) n_{2s}^0(\mathbf{k}, \beta', \gamma'). \end{aligned} \tag{4.3}$$

The kernel of (4.2) is a one-particle path integral,

$$\begin{aligned} n_{2s}^0(\mathbf{x}', \beta', \mathbf{x}'', \beta'') &= zC^M \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}_1\, d\mathbf{x}_{1\beta_1} \cdots d\mathbf{x}_{1\beta_{M-1}} \\ &\quad \times \delta(\mathbf{x}_{1\beta'} - \mathbf{x}') \delta(\mathbf{x}_{1\beta''} - \mathbf{x}'') e^{-K\{\mathbf{x}_{1\gamma}\}}, \end{aligned}$$

and it is easy to show that

$$n_{2s}^0(\mathbf{k}, \beta', \beta'') = \rho \left[\exp\left(-\frac{\hbar^2 \mathbf{k}^2}{2m} \right) \right] \left(|\beta' - \beta''| - \frac{|\beta' - \beta''|^2}{\beta} \right). \tag{4.4}$$

This function, which is the kernel of the integral equation (4.3), has the property

$$\tilde{n}_{2s}^0(\mathbf{k}, \beta - \gamma) = \tilde{n}_{2s}^0(\mathbf{k}, \gamma), \quad \text{for } \gamma = |\beta' - \beta''|, \tag{4.5a}$$

and we define, for temperatures larger than β,

$$\tilde{n}_{2s}^0(\mathbf{k}, \beta + \gamma) = \tilde{n}_{2s}^0(\mathbf{k}, \gamma). \tag{4.5b}$$

It follows that the solution of (4.3) in terms of the eigenfunctions and eigenvalues of the kernel, $\phi_n(\beta')$ and $\lambda_n(\mathbf{k})$, is

$$\tilde{F}(\mathbf{k}, \beta', \beta'') = \sum_{n=-\infty}^{\infty} \frac{\lambda_n \lambda}{\lambda - \lambda_n} \phi_n^*(\beta'') \phi_n(\beta'), \tag{4.6}$$

where

$$\begin{aligned} \phi_n(\beta') &= \beta^{-\frac{1}{2}} e^{-(2\pi i n \beta')/\beta}, \\ \lambda_n(\mathbf{k}) &= \int_0^{\beta} d\gamma \tilde{n}_{2s}^0(\mathbf{k}, \gamma) e^{(2\pi i n \gamma)/\beta}, \end{aligned} \tag{4.7}$$

and

$$\lambda = -1/\tilde{\phi}(\mathbf{k}).$$

In the ground state

$$\tilde{n}_{2s}^0(\mathbf{k}, \beta', \beta'') = \rho \exp[-T(\mathbf{k}) |\beta' - \beta''|],$$

where

$$T(\mathbf{k}) = \hbar^2 \mathbf{k}^2 / 2m,$$

and $\tilde{F}$ satisfies the equivalent differential equation

$$\begin{aligned} \frac{d^2 \tilde{F}(\mathbf{k}, \beta', \beta'')}{d\beta''^2} - T(\mathbf{k})[T(\mathbf{k}) + 2\rho\tilde{\phi}(\mathbf{k})] \tilde{F}(\mathbf{k}, \beta', \beta'') \\ = -\rho 2T(\mathbf{k}) \delta(\beta' - \beta''). \end{aligned} \tag{4.8}$$

This has the solution

$$\tilde{F}(\mathbf{k}, \beta', \beta'') = \frac{\rho T(\mathbf{k}) \exp\left(-\{T(\mathbf{k})[T(\mathbf{k}) + 2\rho\tilde{\phi}(\mathbf{k})]\}^{\frac{1}{2}} |\beta' - \beta''|\right)}{\{T(\mathbf{k})[T(\mathbf{k}) + 2\rho\tilde{\phi}(\mathbf{k})]\}^{\frac{1}{2}}}. \tag{4.9}$$

Once F is known we may reinsert it into (3.11) and (3.12) to find corrections to n_{2s} and n_2. For example, calling $g \equiv n_2 - \rho^2$, we obtain

$$\tilde{g}(\mathbf{k}, \beta', \beta'') = \sum_{n=-\infty}^{\infty} \frac{\lambda_n^2}{\lambda - \lambda_n} \phi_n^*(\beta'') \phi_n(\beta'), \quad \text{for } \beta < \infty. \tag{4.10}$$

Now that $n_2(\mathbf{x}', \beta'; \mathbf{x}'', \beta')$ has been found, $\ln Z$ is easily calculated via some thermodynamic relationship. Since n_2 is an approximation we know that different thermodynamic relations will, in general, yield different results.[17] The relation used here seems most natural in this work and consists in requiring that the partition function satisfy

$$\frac{\delta \ln Z[\phi]}{-\beta \delta \phi(\mathbf{x}, \mathbf{y})} = n_2(\mathbf{x}; \mathbf{y}), \tag{4.11}$$

where ϕ is the interparticle potential, which is of course an identity if exact distributions are used.[18]

The integral of (4.11), turning on $\phi(\mathbf{x}, \mathbf{y}, \alpha) \equiv \alpha\phi(\mathbf{x}, \mathbf{y})$ from $\alpha = 0$ to $\alpha = 1$, is

$$\ln Z - \ln Z^0 = -\tfrac{1}{2}\beta\int_0^1\int_{-\infty}^{\infty} d\alpha\, d\mathbf{x}\, d\mathbf{y} n_2(\mathbf{x}; \mathbf{y}, \alpha)\phi(\mathbf{x}, \mathbf{y}) = -\frac{\beta V}{2(2\pi)^3}\int_0^1\int_{-\infty}^{\infty} d\alpha\, d\mathbf{k}\tilde{n}_2(\mathbf{k}, \alpha)\tilde{\phi}(\mathbf{k}). \tag{4.12}$$

Placing the equal-temperature value of (4.10) into (4.12), there results

$$\ln Z - \ln Z^0 = \frac{V}{2(2\pi)^3}\int_{-\infty}^{\infty} d\mathbf{k} \sum_n [\tilde{\phi}(\mathbf{k})\lambda_n - \ln(1 + \tilde{\phi}(\mathbf{k})\lambda_n)]. \tag{4.13}$$

This is also found by Montroll and Ward[19] in their theory of the electron gas and is equivalent to a sum of "ring" diagrams in that work. The high-temperature, or classical, limit of (4.10),

$$\tilde{g}(\mathbf{k}) = -\beta\rho^2\tilde{\phi}(\mathbf{k})/[1 + \beta\rho\tilde{\phi}(\mathbf{k})]$$

(because $\lambda_n \to \beta\rho\delta_{n,0}$), yields the Debye–Hückel equation of state for a Coulomb potential with neutralizing background.[20]

We now show how the correction to the transformed momentum density $\tilde{n}(\mathbf{q})$ can also be related to the (approximate) Ursell function (4.6). By (2.10) and (3.7), we have

$$\begin{aligned}
\tilde{n}(\mathbf{q}) - \tilde{n}^0(\mathbf{q}) &= \lim_{\epsilon\to 0} \exp\left(-\frac{m\mathbf{q}^2}{2\epsilon\hbar^2}\right)\int_{-\infty}^{\infty}[n_{2s}(\mathbf{z}', \beta', \mathbf{z}'', \beta' - \epsilon) - n_{2s}^0(\mathbf{z}', \beta', \mathbf{z}'', \beta' - \epsilon)] \exp\left(\frac{m\mathbf{q}\cdot(\mathbf{z}' - \mathbf{z}'')}{\epsilon\hbar^2}\right) d\mathbf{z}'\, d\mathbf{z}'' \\
&= \lim_{\epsilon\to 0} \exp\left(-\frac{m\mathbf{q}^2}{2\epsilon\hbar^2}\right)\int_{-\infty}^{\infty} d\mathbf{z}'\, d\mathbf{z}''\left(Z^{-1}zC^M\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_1\, d\mathbf{x}_{1\beta_1}\cdots d\mathbf{x}_{1\beta_{M-1}}\delta(\mathbf{x}_{1\beta'} - \mathbf{z}')\delta(\mathbf{x}_{1\beta'-\epsilon} - \mathbf{z}'')\right. \\
&\quad \left.\times e^{-K\{\mathbf{x}_{1\gamma}\}}Z[U_\phi] - n_{2s}^0(\mathbf{z}', \beta', \mathbf{z}'', \beta' - \epsilon)\right) \exp\left(\frac{m\mathbf{q}\cdot(\mathbf{z}' - \mathbf{z}'')}{\epsilon\hbar^2}\right) \\
&= \lim_{\epsilon\to 0} zC^M\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_1\, d\mathbf{x}_{1\beta_1}\cdots d\mathbf{x}_{1\beta_{M-1}} \exp(-K\{\mathbf{x}_{1\gamma}\}_{\gamma\neq\beta'}) \exp\left(-\frac{m}{2\hbar^2\epsilon}(\mathbf{x}_{1\beta'} - \mathbf{x}_{1\beta'-\epsilon} - \mathbf{q})^2\right) \\
&\quad \times \tfrac{1}{2}\int_{-\infty}^{\infty}\int_0^{\beta} d\mathbf{y}'\, d\mathbf{y}''\, d\gamma'\, d\gamma'' F(\mathbf{y}', \gamma', \mathbf{y}'', \gamma'')\phi(\mathbf{y}' - \mathbf{x}_{1\gamma'})\phi(\mathbf{y}'' - \mathbf{x}_{1\gamma''}),
\end{aligned} \tag{4.14}$$

where we have used a linearized version of (3.3), and the $\mathbf{z}'$ and $\mathbf{z}''$ integrations have been performed in the last line. When a change of variables is made,

$$\mathbf{x}_{1\gamma} = \mathbf{y}_{1\gamma}, \quad \gamma = 0, \cdots, \beta' - \epsilon,$$
$$\mathbf{x}_{1\gamma} - \mathbf{q} = \mathbf{y}_{1\gamma}, \quad \gamma = \beta', \cdots, \beta, \quad \mathbf{y}_{1\beta} = \mathbf{y}_{10} - \mathbf{q},$$

one sees that (4.14) becomes

$$\begin{aligned}
\tilde{n}(\mathbf{q}) - \tilde{n}^0(\mathbf{q}) &= -\tilde{n}^0(\mathbf{q})\rho\beta\tilde{\phi}(0) + \tfrac{1}{2}\int_{-\infty}^{\infty}\int_0^{\beta} d\mathbf{y}'\, d\mathbf{y}''\, d\mathbf{z}'\, d\mathbf{z}''\, d\gamma'\, d\gamma'' \\
&\quad \times F(\mathbf{y}', \gamma', \mathbf{y}'', \gamma'')\phi(\mathbf{y}' - \mathbf{z}')\phi(\mathbf{y}'' - \mathbf{z}'' + \mathbf{q}) \\
&\quad \times n_{2s}^{0*}(\mathbf{z}', \gamma', \mathbf{z}'', \gamma'') \\
&= -\tilde{n}^0(\mathbf{q})\rho\beta\tilde{\phi}(0) + \frac{1}{2}\frac{V}{(2\pi)^6}\int_{-\infty}^{\infty}\int_0^{\beta} d\mathbf{k}\, d\gamma'\, d\gamma'' \\
&\quad \times \tilde{F}(\mathbf{k}, \gamma', \gamma'')\tilde{\phi}(\mathbf{k})^2\tilde{n}_{2s}^{0*}(\mathbf{k}, \gamma', \gamma'')e^{i\mathbf{k}\cdot\mathbf{q}},
\end{aligned} \tag{4.15}$$

where the linear term in (3.3) has been restored. The star denotes a modified same-particle distribution in which the final space–time point differs from the initial point by $-\mathbf{q}$, a distinction which may be ignored in the ground state. Both terms on the right side of (4.15) diverge as $\beta \to \infty$ [see (4.9)], so the functional expansion must be modified before it will apply to Boson ground states. Once such modification is to expand not in the potential but in a "quantum f factor" defined by

$$\begin{aligned}
f(\mathbf{y}', \beta' \mid \{\mathbf{x}_\gamma\}) \equiv zC^M\int_{-\infty}^{\infty}\cdots\int d\mathbf{y}\, d\mathbf{y}_{\beta_1}\cdots d\mathbf{y}_{\beta_{M-1}} \\
\times \delta(\mathbf{y}_{\beta'} - \mathbf{y}')e^{-K\{\mathbf{y}_\gamma\}} \\
\times \exp\left(-\sum_{k=1}^{M}\epsilon\phi(\mathbf{x}_{\beta_k} - \mathbf{y}_{\beta_k})\right) - \rho.
\end{aligned} \tag{4.16}$$

For example, the divergent linear term in the n_{2s} expansion (3.11) now becomes

$$\begin{aligned}
-\rho\beta\tilde{\phi}(0)n_{2s}^0(\mathbf{x}', \beta', \mathbf{x}'', \beta'') \\
\to zC^M\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}\, d\mathbf{x}_{\beta_1}\cdots d\mathbf{x}_{\beta_{M-1}} \\
\times\int_{-\infty}^{\infty} d\mathbf{y}' f(\mathbf{y}', \gamma' \mid \{\mathbf{x}_\gamma\}) \\
\times \delta(\mathbf{x}_{\beta'} - \mathbf{x}')\delta(\mathbf{x}_{\beta''} - \mathbf{x}'')e^{-K\{\mathbf{x}_\gamma\}} \\
- \rho V n_{2s}^0(\mathbf{x}', \beta', \mathbf{x}'', \beta'').
\end{aligned} \tag{4.17}$$

(The right-hand side of (4.17) reduces to the left-hand side when the potential is very weak.) When (4.17) is inserted into the first line of (4.14) it contributes the correction

$$\tilde{n}(\mathbf{q}) - \tilde{n}^0(\mathbf{q}) = zC^M \int_{-\infty}^{\infty} \cdots \int d\mathbf{x}\, d\mathbf{x}_{\beta_1} \cdots d\mathbf{x}_{\beta_{M-1}} \times \int_{-\infty}^{\infty} d\mathbf{y}' f(\mathbf{y}', \gamma' \mid \{\mathbf{x}_\gamma\}) \exp\left(-K\{\mathbf{x}_\gamma\}_{\gamma \neq \beta'}\right) \times \exp\left(-\frac{m}{2\hbar^2\epsilon}(\mathbf{x}_{\beta'} - \mathbf{x}_{\beta'-\epsilon} - \mathbf{q})^2\right) - \rho V \tilde{n}^0(\mathbf{q}). \tag{4.18}$$

We use the same change of variables as in (4.14) and then shift to relative and center-of-mass variables in the path integral. Performing the center-of-mass integrations we are left with

$$\tilde{n}(\mathbf{q}) - \tilde{n}^0(\mathbf{q}) = \rho V 2^{\frac{3}{2}} \exp\left(-\frac{m\mathbf{q}^2}{\hbar^2\beta}\right) \times z \int_{-\infty}^{\infty} d\mathbf{x} K(\mathbf{x} + \mathbf{q}, \beta; \mathbf{x}, 0) - \rho V \tilde{n}^0(\mathbf{q}), \tag{4.19}$$

where K is a propagator for a single particle of mass $\frac{1}{2}m$ moving in an external potential identical to the pair interaction of the many-body system. The explicit fugacity dependence can be eliminated by evaluating (4.19) at $\mathbf{q} = 0$. One obtains finally

$$\tilde{n}(\mathbf{q}) - \tilde{n}^0(\mathbf{q}) = \langle N \rangle^2 \left[\exp\left(-\frac{m\mathbf{q}^2}{\hbar^2\beta}\right) \int_{-\infty}^{\infty} d\mathbf{x} K(\mathbf{x} + \mathbf{q}, \beta; \mathbf{x}, 0) \Big/ \int_{-\infty}^{\infty} d\mathbf{x} K(\mathbf{x}, \beta; \mathbf{x}, 0) - 1 \right]. \tag{4.20}$$

Unfortunately, as $\beta \to \infty$ the $\mathbf{q}$ dependence in the integral becomes negligible, so that while the temperature divergence has been removed, the present approximation predicts that all the particles condense into the $\mathbf{p} = 0$ state. We reserve for a future paper improvements upon (4.20) to be found by examination of higher terms in the f-factor expansion.

* Supported in part by AEC contract No. AT(30-1)-1480.
† Supported by NSF Grant No. GP-8449.
‡ Present address: Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass.

[1] J. Yvon, Nuovo Cimento **9**, 144 (1958).
[2] J. K. Percus, Phys. Rev. Letters **8**, 462 (1962).
[3] J. L. Lebowitz and J. K. Percus, Phys. Rev. **122**, 1675 (1961).
[4] N. N. Bogoliubov, in *Studies in Statistical Mechanics*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publ. Co., Amsterdam, 1962), Vol. 1, Chap. 1.
[5] J. K. Percus, in *Equilibrium Theory of Classical Fluids*, H. L. Frisch and J. L. Lebowitz, Eds. (W. A. Benjamin and Sons, Inc., New York, 1964), Sec. II, p. 323.
[6] J. L. Lebowitz and J. K. Percus, J. Math. Phys. **4**, 1495 (1963).
[7] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (W. A. Benjamin and Sons, Inc., New York, 1962), Chap. 1.
[8] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Co., New York, 1965), Chap. 2.
[9] R. P. Feynman, Phys. Rev. **84**, 108 (1951).
[10] C. Garrod, Rev. Mod. Phys. **38**, 483 (1966).
[11] Reference 8, Chaps. 4 and 10.
[12] Mixed distributions are possible for three or more particles.
[13] See Ref. 8, Chap. 10.
[14] See Ref. 5.
[15] The linear term in (3.3) will often be omitted for simplicity in writing.
[16] E. W. Montroll and J. C. Ward, Phys. Fluids **1**, 55 (1958).
[17] See Ref. 5.
[18] S. A. Rice and P. Gray, *Statistical Mechanics of Simple Liquids* (John Wiley and Sons, Inc., New York, 1965), Chap. 2.
[19] See Ref. 16.
[20] See Ref. 5, Sec. II, p. 73.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

A Superposition Principle in Physics

STANLEY P. GUDDER
Department of Mathematics, University of Denver, Denver, Colorado 80210

(Received 18 July 1968; Revised Manuscript Received 7 July 1969)

A superposition principle is considered both in classical mechanics and in the quantum logic approach to quantum mechanics. It is shown, roughly speaking, that in classical mechanics the only type of superposition of states is a mixture and that no pure state is a nontrivial superposition of other pure states. In quantum mechanics it is shown that, if a superposition principle holds, then the quantum logic is a complete atomic lattice.

1. INTRODUCTION

According to Dirac,[1] the principle of superposition of states is one of the most fundamental properties of quantum mechanics. In fact, this principle contains the essential properties of quantum systems and forms the basis of Dirac's quantum mechanical formulation. In this paper we consider superpositions of states and a superposition principle as they occur in classical mechanics and in the quantum logic approach[2] to axiomatic quantum mechanics. Since Dirac's superposition principle is formulated in the vector-space model for quantum mechanics, our first problem is to decide what form this principle should take in the quantum logic approach. One of the goals of this last approach is to find physically plausible axioms[3] that can be added to the quantum logic axioms so that the resulting axiomatic system becomes isomorphic to the usual von Neumann–Hilbert space model for quantum mechanics.[4] We show that a superposition principle itself is just what is needed. More precisely, we show that a quantum logic in which a superposition principle holds is a complete atomic lattice.

2. NOTATION AND DEFINITIONS

Let L be an orthocomplemented partially ordered set or *logic*. Precisely, L is a partially ordered set with first and last elements 0 and 1, respectively, a complementation $a \to a'$ which satisfies $(a')' = a$, $a \leq b$ implies $b' \leq a'$, $a \vee a' = 1$, and if a_i is a sequence of mutually disjoint elements, then $\bigvee a_i$ exists. We call the elements of L *propositions*. A map $m: L \to [0, 1]$ which satisfies $m(1) = 1$, and $m(\bigvee a_i) = \sum_{i=1}^{\infty} m(a_i)$ if the a_i are disjoint, (i.e., $a_i \leq a_j'$, $i \neq j$) is a *state* on L. If m is a state which cannot be written in the form $m = cm_1 + (1 - c)m_2$, where $0 < c < 1$ and m_1 and m_2 are distinct states, then m is called a *pure state*. We denote the set of states on L by M, and the set of pure states on L by P. If $a \in L$, $m \in P$, we define $P_a = \{m \in P: m(a) = 1\}$, $L_m = \{a \in L: m(a) = 1\}$. If $P_a \subset P_b$ implies $a \leq b$ and $L_{m_1} \subset L_{m_2}$ implies $m_1 = m_2$, we call (L, M) a *quantum logic*. A proposition a is an *atom* if $a \neq 0$ and $b \leq a$ implies $b = 0$ or $b = a$. A logic is a *lattice* if finite suprema exist and is *complete* if arbitrary suprema exist.

If $S \subset M$, $a \in L$, we write $S(a) = 0$ if $m(a) = 0$ for all $m \in S$. If $S \subset M$, $m_0 \in M$, then m_0 is a *superposition of states in* S if $S(a) = 0$ implies $m_0(a) = 0$. We would now like to formulate a superposition principle for the quantum logic (L, M). In his book,[5] Jauch formulated the superposition principle as follows: for any pair of distinct atoms $l_1, l_2 \in L$ there is a third atom $l_3 \in L$ such that $l_1 \vee l_2 = l_1 \vee l_3 = l_2 \vee l_3$. However, this formulation is not acceptable in our present system, since we do not even know atoms exist *a priori* and, if we did, since L need not be a lattice, we would not know if their suprema existed. We will follow Varadarajan's definition[6] of the superposition principle, which is more in line with the principle of Dirac. If $S \subset P$, we denote by $\bar{S}$ the set of all pure states that are superpositions of states in S. Of course, $S \subset \bar{S}$. We call $S \subset P$ *closed* if $S = \bar{S}$, and denote by $\mathcal{M}$ the class of all closed subsets of P. Under set inclusion, $\mathcal{M}$ becomes a partially ordered set. We say that a *superposition principle holds* in (L, M) if $\mathcal{M}$ is isomorphic to L, i.e., if there exists a 1-to-1 map from $\mathcal{M}$ onto L that preserves order. Roughly speaking, this may be interpreted to mean that given a class of pure states S, there is a proposition which is true if and only if the system is in a superposition of states in S. In the usual von Neumann–Hilbert-space model for quantum mechanics, the superposition principle is easily seen to hold using Gleason's theorem.[7] In the next section we will show that a superposition principle holds in certain classical mechanic systems as well.

3. MEASURABLE SUPERPOSITIONS IN CLASSICAL MECHANICS

Let us now consider classical mechanics and classical statistical mechanics. Let X be a set called *phase space* and let L be a σ-algebra of subsets of X. L

corresponds to the set of mechanical events. Usually, X is taken to be a subset of a finite-dimensional Euclidean space and L the Borel subsets of X. If measure-theoretic technicalities are of no consequence, one usually takes L to be the collection of all subsets of X. As we shall see, it is in this latter case that the superposition principle holds. The set of states M on L is the collection of probability measures on L, and (L, M) is a quantum logic under the usual definition of order and complementation. We call (L, M) a *classical logic.*

Now let (Ω, F, μ) be a probability space and let $\omega \to m_\omega$ be a map from Ω into M such that $\omega \to m_\omega(a)$ is measurable for all $a \in L$. We then call $\{m_\omega : \omega \in \Omega\}$ a *measurable set of states* on L. If $S = \{m_\omega : \omega \in \Omega\}$ is a measurable set of states on L and $m_0 \in M$, then m_0 is a *measurable superposition of states* in S if $m_\omega(a) = 0$ for almost every $\omega \in \Omega$ implies $m_0(a) = 0$. For a superposition of a countable number of states, we can take $\Omega = \{1, 2, 3, \cdots\}$, F the class of all subsets of Ω, and μ to be, for example, $\mu(\{n\}) = 2^{-n}$. In this case, "almost everywhere" is the same as "everywhere" and a measurable superposition is the same as a superposition. We now give an example of a measurable superposition. Let $\{m_\omega : \omega \in \Omega\}$ be a measurable set of states on L and let f be a nonnegative measurable function on (X, L) such that

$$\int_X f \, dm_\omega = 1$$

for every $\omega \in \Omega$. If we define m by

$$m(a) = \int_\Omega \left(\int_a f \, dm_\omega \right) \mu(d\omega),$$

then it can be shown that m is well defined and is a state on L. We call m a *classical superposition* or *mixture* of the m_ω. In the countable case, this becomes

$$m(a) = \sum c_i \int_a f \, dm_i \ ,$$

where $c_i \geq 0$ and $\sum c_i = 1$, or, if $f \equiv 1$, we have the more usual $m(a) = \sum c_i m_i(a)$. Notice that a mixture is a measurable superposition. We now show that a mixture is the only kind of measurable superposition possible for a classical system. This gives one of the important distinctions between classical and quantum mechanics, since in quantum theory there are other types of superpositions which are purely quantal in nature, having no classical counterpart.

Theorem 1: Let $S = \{m_\omega : \omega \in \Omega\}$ be a measurable set of states on L. Then $m_0 \in M$ is a measurable superposition of states in S, if and only if m_0 is a mixture of states in S.

Proof: Suppose m_0 is a measurable superposition of states in S. Let

$$\nu(a) = \int_\Omega m_\omega(a) \mu(d\omega).$$

Then, it is easily verified that ν is a state on L. If $\nu(a) = 0$, then $m_\omega(a) = 0$ for almost every $\omega \in \Omega$ and, hence, $m_0(a) = 0$. Therefore, m_0 is absolutely continuous relative to ν. By the Radon–Nikodym theorem there is a nonnegative measurable function f on (X, L) such that

$$m_0(a) = \int_a f \, d\nu = \int_a f(\lambda) \int_\Omega m_\omega(d\lambda) \mu(d\omega)$$
$$= \int_\Omega \left(\int_a f \, dm_\omega \right) \mu(d\omega),$$

where the verification of the interchange of integration order is left to the reader.

The set of mechanical events L on X is called *separable* if L is the smallest σ-algebra containing some countable collection of events. In this we call the classical logic (L, M) *separable.* In the usual case, where X is n-dimensional Euclidean space, the class of Borel subsets is separable. A state m on L is *concentrated* at the point $x \in X$, if $m(a) = 1$ if $x \in a$, and if $m(a) = 0$ if $x \notin a$. Notice that we need not assume $\{x\} \in L$. Our next theorem tells us that, in a separable classical logic, a pure state cannot be a nontrivial superposition of other pure states. This theorem is a generalization of a result of Varadarajan,[8] and is proved using similar techniques.

Theorem 2: Let (L, M) be a separable classical logic. Then a state is pure, if and only if it is concentrated at a point. If S is a collection of pure states, then m_0 is a superposition of states of S, if and only if $m_0 \in S$.

Proof: If m is concentrated at a point, then, clearly, it is pure. Now suppose m is pure. If, for some $a \in L$, $0 < m(a) < 1$, then

$$m(b) = m(a)[m(b \cap a)/m(a)] + [1 - m(a)][m(b \cap a')/m(a')],$$

for every $b \in L$, which contradicts the fact that m is pure. Therefore, m has only the values 0 and 1. Suppose a_i, $i = 1, 2, \cdots$, generates L. By replacing a_n by a_n' if necessary, we may assume that $m(a_i) = 1$ for all i. Let $b = \bigcap a_i$. Then $m(b) = 1$. Let $x \in b$. Now, the collection of events $a \in L$, such that $b \subset a$

or $b \cap a = \varnothing$, is a σ-algebra containing all a_i and, hence, coincides with L. Thus, $m(a) = 1$ if $x \in a$ and $m(a) = 0$ if $x \notin a$; so m is concentrated at x. Now suppose m_0 is pure and a superposition of a collection of pure states S. Suppose $m \in S$ is concentrated at $x_m \in X$ and m_0 is concentrated at $x_0 \in X$. As before, there is a $b \in L$, such that $x_0 \in b$, and every $a \in L$ satisfies $b \subset a$ or $b \cap a = \varnothing$. Now suppose $m_0 \notin S$. Then, $x_m \notin b$ for any $m \in S$, since, otherwise, m_0 would equal this m. Thus, $m(b) = 0$ for every $m \in S$, but $m_0(b) = 1$, a contradiction. Hence, $m_0 \in S$.

Now suppose (L, M) is a separable classical logic. If $\mathcal{M}$ is the class of subsets of the pure states P defined in Sec. 2, it follows from Theorem 2 that $\mathcal{M}$ is the class of *all* subsets of P. Since $\mathcal{M}$ is naturally isomorphic to the class of all subsets 2^X of X, it follows that the superposition principle holds on (L, M), if and only if L is 2^X. Thus, the superposition principle does not hold, if L is the class of Borel subsets of n-dimensional Euclidean space. However, it is easy to construct examples of classical logics in which a superposition principle does hold.

4. SUPERPOSITION PRINCIPLE FOR A QUANTUM LOGIC

In this section we consider a general quantum logic (L, M) and use the notation introduced in Sec. 2. In the following lemma we collect some preliminary results.

Lemma 1: (i) If $S \subset P$ then $S(a) = 0$, if and only if $\bar{S}(a) = 0$ and $\bar{S} = \bar{\bar{S}}$. The smallest closed set in P containing S is $\bar{S}$. (ii) If $S_1 \subset S_2 \subset P$, then $\bar{S}_1 \subset \bar{S}_2$. (iii) If $S_\alpha \subset P$, $\alpha \in A$, then $\bigcap \bar{S}_\alpha$ is closed and $\bigcap \bar{S}_\alpha \supset \overline{\bigcap S_\alpha}$. (iv) If $S \subset P$ and $S = \varnothing$ or S contains one state, then $S = \bar{S}$.

Proof: (i) Clearly, $S(a) = 0$, if $\bar{S}(a) = 0$, and if $S(a) = 0$, then, by definition, $\bar{S}(a) = 0$. Clearly, $S \subset \bar{S}$. If $m \in \bar{\bar{S}}$ and $S(a) = 0$, then $\bar{S}(a) = 0$, which implies $m(a) = 0$, and, hence, $m \in \bar{S}$. For the last statement, $\bar{S}$ is closed, contains S, and is contained in any closed set containing S. (ii) If $S_1 \subset S_2 \subset P$, then $S_1 \subset \bar{S}_2$, and since $\bar{S}_2$ is closed, $\bar{S}_1 \subset \bar{S}_2$. (iii) Suppose $m \in \overline{\bigcap \bar{S}_\alpha}$ and $S_\alpha(a) = 0$ for some $\alpha \in A$. Then $\bar{S}_\alpha(a) = 0$ and $(\bigcap \bar{S}_\alpha)(a) = 0$. Hence, $m(a) = 0$ and $m \in \bar{S}_\alpha$. Thus, $m \in \bigcap \bar{S}_\alpha$, and $\bigcap \bar{S}_\alpha$ is closed. Since $\bigcap \bar{S}_\alpha$ is closed and $\bigcap \bar{S}_\alpha \supset \bigcap S_\alpha$, we have $\bigcap \bar{S}_\alpha \supset \overline{\bigcap S_\alpha}$. (iv) If $S = \{m\}$ and $m_1 \in \bar{S}$, then $L_m \subset L_{m_1}$, and, hence, $m_1 = m$. Thus, $S = \bar{S}$.

Notice that the operation $S \to \bar{S}$ satisfies the axioms of a Kuratowski closure operation[9] except $\overline{\bar{S}_1 \cup \bar{S}_2}$ need not equal $\bar{S}_1 \cup \bar{S}_2$, although clearly, $\bar{S}_1 \cup \bar{S}_2 \subset \overline{\bar{S}_1 \cup \bar{S}_2}$.

Theorem 3: $\mathcal{M}$ is a complete atomic lattice. If a superposition principle holds on (L, M), then L is a complete atomic lattice.

Proof: It follows from Lemma 1 (iv) and (iii) that $\mathcal{M}$ is atomic and contains arbitrary infima. Now suppose $S_\alpha \in \mathcal{M}$, $\alpha \in A$, and let $S = \overline{\bigcup \bar{S}_\alpha}$. Then $S \in \mathcal{M}$ by Lemma 1 (i) and $S \supset S_\alpha$, for all $\alpha \in A$. Now if $S_0 \in \mathcal{M}$ and $S_0 \supset S_\alpha$, for all $\alpha \in A$, then $S_0 \supset \overline{\bigcup \bar{S}_\alpha} = S$. Hence, S is the supremum of the S_α, and the theorem is proved.

We have seen that, for a separable classical logic, $S = \bar{S}$ for every $S \subset P$, and that $\mathcal{M}$ is the class of all subsets of P. We also have a kind of converse to this result.

Lemma 2: If a superposition principle holds on the quantum logic (L, M), and if $S = \bar{S}$ for all $S \subset P$, then (L, M) is a classical logic.

We thus see that classical logics are roughly characterized by the fact that superpositions of pure states do not produce new pure states.

If a superposition principle holds on a quantum logic, it is not clear what explicit form the isomorphism between $\mathcal{M}$ and L takes. Also, it is not clear if there is a complementation on $\mathcal{M}$ for which $\mathcal{M}$ and L are isomorphic as logics. However, if we add one more axiom, we can construct a logic isomorphism explicitly.

Axiom 1: If $a_\alpha \in L$, $\alpha \in A$, and $m(a_\alpha) = 1$, then $m(\bigwedge a_\alpha) = 1$.

This axiom is postulated by Jauch and Piron[2] in their formulation for quantum mechanics. In the sequel, we suppose that a superposition principle and Axiom 1 hold on (L, M). If $S \subset P$, let $a_S \in L$ be defined by $a_S = \bigwedge \{a \in L : S(a) = 1\}$. It is clear that $S(a_S) = 1$, and that a_S is the smallest proposition with this property.

Lemma 3: (i) If $a \in L$, then $P_a \in \mathcal{M}$. (ii) If $S \in \mathcal{M}$, then $S = P_{a_S}$. (iii) If S, $T \in \mathcal{M}$, then $S \subset T$, if and only if $a_S \leq a_T$.

Proof: (i) If $m \in \overline{P_a}$, then, since $P_a(a') = 0$, we have $m(a') = 0$ and, hence, $m \in P_a$. (ii) Clearly, $S \subset P_{a_S}$. Now suppose $m_0 \in P_{a_S}$. If $S(a) = 0$, then $S(a') = 1$ and, hence, $a' \geq a_S$. Therefore, $m_0(a') = 1$, which implies $m_0(a) = 0$. Hence, $m_0 \in \bar{S} = S$. (iii) By (ii) $S \subset T$, if and only if $P_{a_S} \subset P_{a_T}$, but the latter is equivalent to $a_S \leq a_T$.

If $m_1, m_2 \in P$, we say that m_1 and m_2 are *orthogonal* and write $m_1 \perp m_2$, if there is an $a \in L$ such that $m_1(a) = 0$ and $m_2(a) = 1$. If $m_0 \in P$, $S \subset P$, we write $m_0 \perp S$ if $m_0 \perp m$ for every $m \in S$. If $S \subset P$, we define $S' = \{m \in P : m \perp S\}$. It will follow from Theorem 3 that $S \to S'$ is a complementation on $\mathcal{M}$.

Lemma 4: Suppose $m_0 \in P$, $S \subset P$, and $m_0 \perp S$. Then (i) there is an $a \in L$ such that $m_0(a) = 1$ and $S(a) = 0$ and (ii) $m_0 \perp \bar{S}$.

Proof: (i) Let $S = \{m_\alpha : \alpha \in A\}$, and suppose $m_0(a_\alpha) = 1$, $m_\alpha(a_\alpha) = 0$. If $a = \bigvee \{a_\alpha : \alpha \in A\}$, then $m_0(a) = 1$ and $S(a) = 0$. (ii) Applying (i), there is an $a \in L$ such that $m_0(a) = 1$ and $S(a) = 0$, and, by Lemma 1(i), $\bar{S}(a) = 0$. Hence, $m_0 \perp \bar{S}$.

Theorem 4: The map $S \to S'$ is a complementation on $\mathcal{M}$, and the map $S \to a_S$ is a logic isomorphism from $\mathcal{M}$ onto L.

Proof: To show $S \to a_S$ is 1-to-1, suppose $S, T \in \mathcal{M}$ and $a_S = a_T$. Then, by Lemma 3(ii), $S = P_{a_S} = P_{a_T} = T$. To show $S \to a_S$ is onto, let $a \in L$. Then if $S = P_a$, we have $a = a_S$. Applying Lemma 3(iii), it follows that $S \to a_S$ preserves order. Let $S \in \mathcal{M}$ and $T = \{m \in P : m(a_S) = 0\}$. We now show $S' = T$. Clearly, $T \subset S'$. Suppose $m \in S'$ and $m(a_S) \neq 0$. Now, by Lemma 4(i), there is a $b \in L$ such that $S(b) = 1$ and $m(b) = 0$. Thus, $a_S \leq b$ so $m(b) \neq 0$, a contradiction. Applying Lemma 3(i) and (ii), we get $S' = P_{a_{S'}} \in \mathcal{M}$ and $P_{a_{S'}} = S' = P_{a_{S'}}$. Hence, $a_S = a'_S$. Therefore, the complementation is preserved, and the proof is complete.

[1] P. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, London, 1958).
[2] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936); G. Mackey, *The Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1963); C. Piron, Helv. Phys. Acta **37**, 439 (1964); J. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1968); V. Varadarajan, *Geometry of Quantum Theory* (D. Van Nostrand Co., Inc., Princeton, N.J., 1968), Vol. 1.
[3] N. Zierler, Pacific J. Math. **12**, 1151 (1961); G. Bodiou, *Théorie dialectique des probabilities* (Gauthier-Villars, Paris, 1964); J. Gunson, Commun. Math. Phys. **6**, 262 (1967); C. Piron, Ref. 2.
[4] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1955).
[5] J. Jauch, Ref. 2.
[6] V. Varadarajan, Ref. 2.
[7] A. Gleason, J. Math. & Mech. **6**, 885 (1957).
[8] Theorem 6.6 in V. Varadarajan, Ref. 2.
[9] K. Kuratowski, *Topologie* (Warsaw, 1948).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

On Some Classes of Solutions of the Wave Equation $\partial_t^2 f - \Delta f = 0$

H. A. KASTRUP AND D. H. MAYER
Universität München, Sektion Physik, München, Germany

(Received 30 June 1969)

We investigate certain classes of solutions of the wave equation for which $Rf(x) = -(1/x^2)f(+x/x^2)$ is a well-defined C^∞ solution for all points of the Minkowski space, if $f(x)$ is a C^∞ solution. We mainly exploit the fact that the transformation R is a generalized Hankel transformation in momentum space. We make use of several recent results obtained by Zemanian in connection with the Hankel transformation and construct a self-adjoint representation of the Lie algebra of the group $O(2, 4)$ which is an invariance group of the wave equation. Finally, we construct and discuss the eigenfunctions of R in Minkowski space.

I. INTRODUCTION

If $f(x)$, $x = (x^0 = t, \mathbf{x})$, $x^2 = (x^0)^2 - (\mathbf{x})^2$, is a solution of the wave equation

$$\partial_t^2 f - \Delta f = 0, \tag{1}$$

then it is known[1] that the function

$$Rf(x) = -(1/x^2)f(Rx),$$
$$Rx = -x/x^2 \tag{2}$$

is—at least formally—another solution. The length inversion R is not well defined in the Minkowski space M^4, because the light cone $\{x : x^2 = 0\}$ has no image in M^4. However, it may happen that $Rf(x)$ is a well-defined function on M^4. We shall indeed see that such functions exist.

In the following we shall deal only with those solutions of Eq. (1) which have the representation

$$f(x) = (2\pi)^{-\frac{3}{2}} \int_{V_+} d\hat{\Omega}\varphi(p)e^{ip\cdot x}, \tag{3}$$

where $d\hat{\Omega}$ is the Lorentz-invariant measure $d^3\mathbf{p}/2p_0$, $p_0 = +(\mathbf{p}^2)^{\frac{1}{2}}$, on the closed positive light cone $V_+ = \{p : p^2 = 0, p_0 \geq 0\}$. The functions $\varphi(p)$ belong to different spaces to be specified later.

In an earlier paper[2] it was shown that the length inversion R, acting on the functions $\varphi(p)$, is given by the integral kernel

$$R(p, h) = (2\pi)^{-1}J_0[(2p \cdot h)^{\frac{1}{2}}]$$
$$= (2\pi)^{-1}(p_0h_0)^{-\frac{1}{2}}$$
$$\times \sum_{l=0}^{\infty}(2l + 1)J_{2l+1}[2(p_0h_0)^{\frac{1}{2}}]P_l(z), \tag{4}$$

where $z = \cos(\mathbf{p}, \mathbf{h})$ and J_n is the Bessel function of order n.

Thus the length inversion R is a generalized Hankel transformation in momentum space. However, the arguments given in Ref. 2 were not completely rigorous mathematically. We shall try to improve this situation in the present paper which is organized in the following way: In Sec. II we shall consider the functions as elements of $L^2(\hat{\Omega})$ and shall show $R(p, h)$ to be an unitary and self-adjoint transformation of this space onto itself. The result is a generalization of a corresponding one by Titchmarsh[3] for the Hankel transformation on the positive real line.

In Sec. III we expand $\varphi(p)$ in terms of spherical harmonics $Y_l^m(\hat{\mathbf{p}})$ and consider its radial part $\varphi_l(p_0)$ to be an element of the space $\mathfrak{H}_{2l+1}$ introduced by Zemanian.[4] This space plays the same role for the Hankel transform as the space S of fast-decreasing functions at infinity does for the Fourier transform.

Using the results of Sec. III, we construct in Sec. IV a representation of the Lie algebra of the Liouville group $O(2, 4)$ in terms of self-adjoint operators in $L^2(\hat{\Omega})$ [the Liouville or conformal group $O(2, 4)$ is an invariance group of the wave equation[1]]. Here we exploit mainly the facts that the eigenfunctions of $R(p, h)$ are elements of the spaces $L^2(\hat{\Omega})$ and $\mathfrak{H}_{2l+1}$ simultaneously, that they are dense in $L^2(\hat{\Omega})$, and that the elements of the Lie algebra $\mathfrak{L}[O(2, 4)]$ may be constructed by the generators P^μ, $\mu = 0, 1, 2, 3$, of the translations in M^4 and the length inversion R alone.[5]

In Sec. V we calculate and discuss the explicit forms of the eigenfunctions of the operator R in M^4 by Fourier transforming the eigenfunctions of $R(p, h)$ in $L^2(\hat{\Omega})$.

Finally, in Sec. VI we show that any function $f(x)$ which can be represented as

$$f_{lm}(x) = \int_{V_+} d\hat{\Omega}\varphi_l(p_0)Y_l^m(\hat{\mathbf{p}})e(x, p),$$
$$e(x, p) = (2\pi)^{-\frac{3}{2}} \exp(ip \cdot x), \tag{5}$$
$$\tfrac{1}{2}(2p_0)^{\frac{3}{2}}\varphi_l(p_0) \in \mathfrak{H}_{2l+1}$$

can also be represented in the form

$$f_{lm}(x) = \int_{V_+} d\hat{\Omega} R[\varphi_l(h_0) Y_l^m(\hat{\mathbf{h}})] g(x, h),$$
$$g(x, h) = Re(x, h). \tag{6}$$

The equality holds in the usual sense (i.e., not only in the mean!). Since the Hankel transformation represents an automorphism of $\mathfrak{H}_{2l+1}$ onto itself,[4] we see that with any solution of the wave equation $f_{lm}(x)$, given by Eq. (5), the function $Rf_{lm}(x)$ is defined everywhere on M^4 and is a smooth solution of the wave equation, too. The same holds for any solution which can be represented as a finite linear combination of the functions $f_{lm}(x)$.

II. THE $L^2(\hat{\Omega})$ THEORY OF THE LENGTH INVERSION R

The Hilbert space $L^2(\hat{\Omega})$ consists of all equivalence classes—the elements differ at most on a set of measure zero—of functions $\varphi(p)$ for which

$$(\varphi, \varphi) = \int_{V_+} d\hat{\Omega} \, |\varphi(p)|^2 < \infty. \tag{7}$$

Since the rotations commute with R, it is convenient to expand all functions $\varphi(p) \in L^2(\hat{\Omega})$ in terms of spherical harmonics:

$$\varphi(p) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} \varphi_{lm}(p_0) Y_l^m(\hat{\mathbf{p}}), \tag{8}$$

where the equality means convergence on $L^2(\hat{\Omega})$. Accordingly, we decompose the space $L^2(\hat{\Omega})$ into a direct sum

$$L^2(\hat{\Omega}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} \oplus L^2_{lm}(\hat{\Omega}).$$

The projection operator onto the space

$$L^2_l(\hat{\Omega}) = \sum_{m=-l}^{+l} \oplus L^2_{lm}(\hat{\Omega})$$

is the kernel

$$\Pi_l(\hat{\mathbf{h}}, \hat{\mathbf{p}}) = \sum_{m=-l}^{+l} Y_l^m(\hat{\mathbf{h}}) Y_l^{m*}(\hat{\mathbf{p}}) = (2l + 1)/4\pi \, P_l(z).$$

With these definitions we can write

$$R(p, h) = \sum_{l=0}^{\infty} R_l(p_0 h_0) \Pi_l(\hat{\mathbf{h}}, \hat{\mathbf{p}}),$$
$$R_l(p_0 h_0) = 2(p_0 h_0)^{-\frac{1}{2}} J_{2l+1}[2(p_0 h_0)^{\frac{1}{2}}]. \tag{9}$$

For $\tilde{\varphi}_{lm}(\mathbf{p}) = \varphi_{lm}(p_0) Y_l^m(\hat{\mathbf{p}})$ we have

$$R\tilde{\varphi}_{lm}(\mathbf{h}) = Y_l^m(\hat{\mathbf{h}}) \int_0^{\infty} d\hat{\omega} R_l(p_0 h_0) \varphi_{lm}(p_0), \tag{10}$$

where $d\hat{\omega}$ is the measure $\frac{1}{2} p_0 \, dp_0$ on the positive real line.

The eigenfunctions $e_{nl}(p_0)$ of the integral operator (10) are known[2]:

$$e_{nl}(p_0) = a_{nl}^{-\frac{1}{2}} e^{-p_0/2} (2p_0)^l L_n^{2l+1}(2p_0), \tag{11}$$

where

$$L_n^{\alpha}(2p_0) = \sum_{k=0}^{n} \binom{n + \alpha}{n - k} (-2p_0)^k / k!$$

are Laguerre's polynomials and

$$a_{nl} = 8^{-1}(2l + 1 + n)!/n!$$

is a normalization factor. The eigenfunctions (11) belong to the eigenvalues $(-1)^n$.

The functions $e_{nl}(p_0)$ form a complete orthogonal system in the Hilbert space $L^2(\hat{\omega})$, because the functions $(p_0/2)^{\frac{1}{2}} e_{nl}(p_0)$ form a complete system[6] with respect to the positive real line and measure dp_0.

Since the spherical harmonics form a dense orthogonal system on the sphere, the eigenfunctions

$$e_{nlm}(\mathbf{p}) = e_{nl}(p_0) Y_l^m(\hat{\mathbf{p}}) \tag{12}$$

of $R(p, h)$ in $L^2(\hat{\Omega})$ form a complete orthogonal system in this space.

Because of

$$(Re_{nlm}, e_{nlm}) = (e_{nlm}, Re_{nlm}) = (-1)^n,$$
$$(Re_{nlm}, Re_{nlm}) = (e_{nlm}, e_{nlm}),$$

the transformation R is symmetric and bounded on a dense set of $L^2(\hat{\Omega})$ and can therefore[7] be extended to a bounded self-adjoint operator with $L^2(\hat{\Omega})$ itself as its domain of definition. The latter of the above relations shows in addition that R is unitary.

The spectral representation of $R_l(p_0 h_0)$ is given by

$$\begin{aligned} R_l(p_0 h_0) &= 2(p_0 h_0)^{-\frac{1}{2}} J_{2l+1}[2(p_0 h_0)^{\frac{1}{2}}] \\ &= \sum_{n=0}^{\infty} (-1)^n e_{nl}(h_0) e_{nl}(p_0) \\ &= 8 e^{-h_0/2 - p_0/2} (4 p_0 h_0)^l \\ &\quad \times \sum_{n=0}^{\infty} (-1)^n \frac{n!}{(2l + 1 + n)!} \\ &\quad \times L_n^{2l+1}(2h_0) L_n^{2l+1}(2p_0). \end{aligned} \tag{13}$$

The series on the right-hand side of Eq. (13) converges even in the usual sense.[8]

The spectral representation (9) of $R(p, h)$ itself also converges uniformly and absolutely as an expansion[9] in terms of Legendre functions $P_l(z)$ for all finite p_0 and h_0, because the Bessel function $J_0[(2p_0 h_0)^{\frac{1}{2}}(1 - z)^{\frac{1}{2}}]$ is an entire function in z for all fixed finite p_0 and h_0.

The action of R on an arbitrary element

$$\varphi(p) = \sum_{n,l,m} c_{nlm} e_{nlm}(\mathbf{p})$$

of $L^2(\hat{\Omega})$ is given by

$$R\varphi(h) = \sum_{n,l,m} (-1)^n c_{nlm} e_{nlm}(\mathbf{h}),$$

$$c_{nlm} = \int d\hat{\Omega} e^*_{nlm}(\mathbf{p})\varphi(p).$$

III. THE SPACE $\tilde{\mathfrak{H}}_{2l+1}$ FOR THE RADIAL FUNCTIONS $\varphi_{lm}(p_0)$

In this section we shall consider only the radial Hilbert space $L_l^2(\hat{\omega})$ and its elements $\varphi_l(p_0)$ (we drop the spherical index "m" in this section). We shall translate some of the results obtained by Zemanian[4] for our purpose. We first introduce the one-to-one mapping

$$\begin{aligned} &u \to p_0 = 2^{-1}u^2, \quad p_0 \to u = +(2p_0)^{\frac{1}{2}}; \\ &\psi_l(u) \to \varphi_l(p_0) = 2[u(p_0)]^{-\frac{3}{2}}\psi_l[u(p_0)], \\ &\varphi_l(p_0) \to \psi_l(u) = 2^{-1}u^{\frac{3}{2}}\varphi_l(u^2/2); \qquad (14) \\ &\int_0^\infty d\hat{\omega}\,|\varphi_l(p_0)|^2 = \int_0^\infty du\,|\psi_l(u)|^2. \end{aligned}$$

The mapping $\varphi_l \to R_l\varphi_l$ now takes the form $\psi_l \to \mathcal{K}_{2l+1}\psi_l$, where

$$\mathcal{K}_{2l+1}\psi_l(v) = \int_0^\infty du(vu)^{\frac{1}{2}}J_{2l+1}(uv)\psi_l(u). \qquad (15)$$

The subspace D_I of $L_l^2(u)$ consists of all infinitely differentiable ("smooth") functions with compact support on the interval $I = \{u: 0 < u < \infty\}$ with the usual topology of uniform convergence.

The subspace $\mathfrak{H}_{2l+1}(\psi_l)$ consists of all smooth functions on I for which the quantities

$$\gamma^{2l+1}_{j,k}(\psi_l) = \sup_I \left| u^j \left(u^{-1}\frac{d}{du}\right)^k [u^{-2l-\frac{3}{2}}\psi_l(u)] \right|, \qquad (16)$$

$j, k = 0, 1, \cdots$, are finite. Zemanian proved that the mapping (15) is a topological automorphism of $\mathfrak{H}_{2l+1}$, where the topology is defined by the seminorms (16). For the eigenfunctions $e_{nl}(p_0)$ we have

$$e_{nl}(p_0) \to E_{nl}(u) = 2^{-1}a_{nl}^{-\frac{1}{2}}e^{-u^2/2}u^{2l+\frac{3}{2}}L_n^{2l+1}(u^2).$$

Thus the functions $E_{nl}(u)$ are elements of $\mathfrak{H}_{2l+1}(\psi_l)$.

The mapping (14) maps the space $\mathfrak{H}_{2l+1}(\psi_l)$ on a space $\tilde{\mathfrak{H}}_{2l+1}(\varphi_l)$ with an induced topology defined by the seminorms

$$\tilde{\gamma}^{2l+1}_{j,k}(\varphi_l) = \sup_{0<p_0<\infty} \left| p_0^{j/2}\left(\frac{d}{dp_0}\right)^k (p_0^{-l}\varphi_l(p_0)) \right| < \infty, \quad j, k = 0, 1, \cdots. \qquad (17)$$

The mapping (14) is continuous in both directions [in the topologies (16) and (17)].

Since the eigenfunctions $e_{nl}(p_0)$ are elements of $\tilde{\mathfrak{H}}_{2l+1}(\varphi_l)$ and are dense in $L_l^2(\hat{\omega})$, we see that $\tilde{\mathfrak{H}}_{2l+1}(\varphi_l)$ is dense in $L_l^2(\hat{\omega})$. This property will be very helpful in the next section. [As the elements of $\tilde{\mathfrak{H}}_{2l+1}$ are of fast decrease at infinity, and since they are bounded, they belong to all $L^p(p_0)$, $p = 1, 2, \cdots$.]

We finally translate some of the lemmata of Zemanian for $\mathfrak{H}_{2l+1}$ into corresponding ones for $\tilde{\mathfrak{H}}_{2l+1}(\varphi_l)$:

(1) $\tilde{\mathfrak{H}}_{2l_2+1} \subset \tilde{\mathfrak{H}}_{2l_1+1}$ for $l_2 > l_1$;

(2) $R_l(p_0\varphi_l) = A_lR_l\varphi_l$, $R_l(A_l\varphi_l) = p_0R_l\varphi_l$, (18)

$$A_l = -\left(\frac{d^2}{dp_0^2}p_0 - \frac{l(l+1)}{p_0}\right).$$

The operator A_l corresponds to the operator $-M_{2l+1}N_{2l+1}$ of Zemanian in $\mathfrak{H}_{2l+1}$.

(3) The mapping $\varphi_l \to p_0\varphi_l$ is a continuous mapping of $\tilde{\mathfrak{H}}_{2l+1}$ into itself.

IV. CONSTRUCTION OF SELF-ADJOINT GENERATORS OF THE GROUP $O(2, 4)$ IN THE SPACE $L^2(\hat{\Omega})$

We now use the results of the preceding sections in order to construct the self-adjoint generators of a representation of the Lie algebra of the group $O(2, 4)$ in $L^2(\hat{\Omega})$. We mainly exploit the fact[5] that all 15 elements of the Lie algebra $\mathfrak{L}[O(2, 4)]$ can be constructed by means of the generators P^μ, $\mu = 0, 1, 2, 3$, of the translations (in M^4) and the length inversion R (which itself is not an element of $\mathfrak{L}[O(2, 4)]$) in the following way: We define the action of P^μ on $\varphi(p)$ as the multiplication of $\varphi(p)$ by p^μ. The generators K^μ of the special Liouville transformations[10] are given by

$$K^\mu = RP^\mu R, \quad \mu = 0, 1, 2, 3, \qquad (19)$$

and the remaining elements of the Lie algebra by the commutators

$$[K^\mu, P^\nu] = 2i(g^{\mu\nu}D - M^{\mu\nu}), \qquad (20)$$

where D is the generator of the dilatations ($p_\mu \to \rho^{-1}p_\mu$) and the $M^{\mu\nu}$ generate the proper orthochronous Lorentz transformations. Furthermore, $g^{00} = -g^{11} = -g^{22} = -g^{33} = 1$; all other $g^{\mu\nu} = 0$.

We start the construction by defining the operators P_0 and K_0 in $L_l^2(\hat{\omega})$: The operation $\varphi_l(p_0) \to p_0\varphi_l(p_0)$ is defined for all elements $\varphi_l \in L_l^2(\hat{\omega})$ for which

$$\int_0^\infty d\hat{\omega}\,|p_0\varphi_l(p_0)|^2 < \infty.$$

We denote this domain by $\mathfrak{D}_l[P_0]$. Since P_0 is an unbounded operator, we have $\mathfrak{D}_l[P_0] \neq L_l^2(\hat{\omega})$. As a

multiplication operator, P_0 is self-adjoint in $L_l^2(\hat{\omega})$. Furthermore, the results of the last section imply that

$$\tilde{\mathfrak{H}}_{2l+1} \subset \mathfrak{D}_l[P_0], \quad P_0\tilde{\mathfrak{H}}_{2l+1} \subset \tilde{\mathfrak{H}}_{2l+1}. \tag{21}$$

We now define K_0 by

$$K_0 = R_l P_0 R_l.$$

As R_l is unitary, K_0 is also self-adjoint, has the same spectrum as P_0, and its domain of definition is $\mathfrak{D}_l[K_0] = R_l\mathfrak{D}_l[P_0]$. Because of $R_l\tilde{\mathfrak{H}}_{2l+1} = \tilde{\mathfrak{H}}_{2l+1}$, P_0 and K_0 have the common dense domain $\tilde{\mathfrak{H}}_{2l+1}$. The Eqs. (18) show that

$$K_0 = -\left(\frac{d^2}{dp_0^2}p_0 - \frac{l(l+1)}{p_0}\right) \tag{22}$$

on $\tilde{\mathfrak{H}}_{2l+1}$.

Since $K_0\tilde{\mathfrak{H}}_{2l+1} \subset \tilde{\mathfrak{H}}_{2l+1}$ [see Eq. (21)], we can multiply and add P_0 and K_0 on $\tilde{\mathfrak{H}}_{2l+1}$ without restrictions. Making use of this fact, from Eqs. (20) and (22) we obtain

$$(2i)^{-1}[K_0, P_0] = \hat{D} = i\left(p_0\frac{d}{dp_0} + 1\right) \tag{23}$$

on $\tilde{\mathfrak{H}}_{2l+1}$. $\hat{D}$ is symmetric in $L_l^2(\hat{\omega})$. Since the equations $\hat{D}^*\varphi_l^{\pm} = \pm i\varphi_l^{\pm}$ do not have any solutions unequal to zero which lie in $L_l^2(\hat{\omega})$, the deficiency indices[11] of $\hat{D}$ are (0, 0). This can be seen as follows: The linear hull of the functions

$$h_m = p_0^m \exp(-p_0), \quad m = m_0, m_0 + 1, \cdots, \infty,$$

is dense in $L_l^2(\hat{\omega})$ for each fixed $m_0 > -1$. This follows, for instance, from the fact that Laguerre's functions form a complete basis in $L_l^2(\hat{\omega})$. If

$$b_m^{\pm} = (\varphi^{\pm}, h_m),$$

then the equations

$$(\hat{D}^*\varphi^{\pm}, h_m) = \mp i(\varphi^{\pm}, h_m) = (\varphi^{\pm}, \hat{D}h_m)$$

imply the recursion formulas

$$b_{m+1}^{+} = (m+2)b_m^{+}, \quad b_{m+1}^{-} = mb_m^{-}.$$

It then follows for the expansion coefficients $c_n^{\pm} = (e_{n1}, \varphi^{\pm})$ that $c_n^{-} = 0$ for all n and that the sum

$$\sum_{n=0}^{\infty} |c_n^{+}|^2$$

does not converge. Thus, $\hat{D}$ has a unique self-adjoint extension D in $L_l^2(\hat{\omega})$.

All the above results remain unchanged if we go over from $L_l^2(\hat{\omega})$ to $L_{lm}^2(\hat{\Omega})$. On $\tilde{\mathfrak{H}}_{2l+1}Y_l^m(\hat{\mathfrak{p}})$, we may write for K_0

$$K_0 = -p_0\Delta, \tag{24}$$

where Δ is the Laplacian in momentum space.

We now have an irreducible representation[12] of the subalgebra $\mathfrak{L}(P_0, K_0, D)$ by means of self-adjoint operators in $L_{lm}^2(\hat{\Omega})$.

We next turn to the operators $P^j = p^j$, $j = 1, 2, 3$. It is convenient to introduce spherical coordinates

$$\mathfrak{p} = (p^1, p^2, p^3) = p_0(\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta).$$

With

$$p^{\pm} = (p^1 \pm ip^2) = p_0\sin\vartheta e^{\pm i\varphi},$$

we have[13]

$$p^{+}\varphi_{lm}(p_0)Y_l^m(\hat{\mathfrak{p}}) = p_0\varphi_{lm}(p_0)[a_1(l,m)Y_{l+1}^{m+1}(\hat{\mathfrak{p}}) + a_2(l,m)Y_{l-1}^{m+1}(\hat{\mathfrak{p}})], \tag{25}$$

$$p^{-}\varphi_{lm}(p_0)Y_l^m(\hat{\mathfrak{p}}) = p_0\varphi_{lm}(p_0)[a_3(l,m)Y_{l+1}^{m-1}(\hat{\mathfrak{p}}) + a_4(l,m)Y_{l-1}^{m-1}(\hat{\mathfrak{p}})], \tag{26}$$

$$p^{3}\varphi_{lm}(p_0)Y_l^m(\hat{\mathfrak{p}}) = p_0\varphi_{lm}(p_0)[a_5(l,m)Y_{l+1}^{m}(\hat{\mathfrak{p}}) + a_6(l,m)Y_{l-1}^{m}(\hat{\mathfrak{p}})], \tag{27}$$

where $a_\beta(l, m)$, $\beta = 1, \cdots, 6$ are constants depending on l and m.

We see that the operators P^j, $j = 1, 2, 3$, mix the spaces $L_{lm}^2(\hat{\Omega})$; and from repeated application of P^j it follows that we need all l-values in order to be able to represent the operators P^j. We now combine the transformations R and P^j. We have $K^j = RP^jR$, and the K^j are self-adjoint in $L^2(\hat{\Omega})$ because the P^j are self-adjoint.

We next define the space $\mathfrak{H}$ which consists of all finite linear combinations of functions

$$\hat{\varphi}_{lm}(\mathfrak{p}) = \varphi_{lm}(p_0)Y_l^m(\hat{\mathfrak{p}}), \quad l = 0, 1, \cdots; |m| \leq l,$$

where $\varphi_{lm}(p_0) \in \tilde{\mathfrak{H}}_{2l+1}$. The space $\mathfrak{H}$ can be considered as the inductive limit[14] of the spaces $\tilde{\mathfrak{H}}_{2l+1}$. However, we shall not need this property in the following.

The space $\mathfrak{H}$ which is dense in $L^2(\hat{\Omega})$ is mapped onto itself by the length inversion R, because each $\tilde{\mathfrak{H}}_{2l+1}$ is mapped onto itself. Clearly, $\mathfrak{H}$ belongs to the domain of definition $\mathfrak{D}[P^j]$ of P^j. Since $\mathfrak{D}[K^j] = R\mathfrak{D}[P^j]$, we have $\mathfrak{H}$ as a common dense domain of P^μ and K^μ, $\mu = 0, 1, 2, 3$.

As $p_0\varphi_{lm}(p_0) \in \tilde{\mathfrak{H}}_{2l+3}$ if $\varphi_{lm}(p_0) \in \tilde{\mathfrak{H}}_{2l+1}$ [see Eq. (17)], and because of Eqs. (25)–(27), we have $P^\mu\mathfrak{H} \subset \mathfrak{H}$, $K^\mu\mathfrak{H} \subset \mathfrak{H}$. Thus we can multiply and add the operators P^μ and K^μ on $\mathfrak{H}$ without restrictions. In particular, we

can calculate the commutators

$$M^{\mu\nu} = i2^{-1}[K^\mu, P^\nu], \quad \mu \neq \nu.$$

The operators $M^{\mu\nu}$ are symmetric on $\mathfrak{H} \subset L^2(\hat{\Omega})$ with respect to the metric (7) and can be extended uniquely to self-adjoint operators in $L^2(\hat{\Omega})$, as is well known from the construction of unitary representations[15] of the Poincaré group.

The operators K^j have the explicit form[16]

$$K^j = -(2\partial^j + 2p^k\partial_k\partial^j + p^j\Delta), \quad \partial_j = \frac{\partial}{\partial p^j}.$$

We remark that

$$K^\mu R(p, h) = h^\mu R(p, h), \quad \mu = 0, 1, 2, 3,$$

i.e., the kernel $R(p, h)$, is a common nonnormalizable eigenfunction of the operators K^μ.

Finally, we mention that $R\varphi \in \mathfrak{D}[D] \subset L^2(\hat{\Omega})$, if $\varphi \in \mathfrak{D}[D]$. This property follows from the relation

$$R(D\varphi) = -D(R\varphi), \quad D = i\left(\sum_{j=1}^{3} p^j\partial_j + 1\right),$$

which obviously holds on $\mathfrak{H}$. However, it holds for all $\varphi \in \mathfrak{D}[D]$ because $\mathfrak{H}$ is dense in $L^2(\hat{\Omega})$, R unitary, and D closed. Thus the length inversion R maps $\mathfrak{D}[D]$ onto itself.

V. THE EIGENFUNCTIONS $f_{nlm}(x)$ OF R IN THE MINKOWSKI SPACE

In this section we shall calculate the functions

$$f_{nlm}(x) = (2\pi)^{-\frac{3}{2}}\int_{V_+} d\hat{\Omega}e_{nlm}(\mathbf{p})e^{ip\cdot x},$$

$$e_{nlm}(\mathbf{p}) = a_{nl}^{-\frac{1}{2}}e^{-p_0}(2p_0)^l L_n^{2l+1}(2p_0)Y_l^m(\hat{\mathbf{p}}). \quad (28)$$

The integral exists because $e_{nlm}(\mathbf{p}) \in L^1(\hat{\Omega})$. This follows from the fact that

$$\int_0^\infty dp_0 p_0^j e^{-p_0} < \infty$$

for all $j = 0, 1, 2, \cdots$. The integral over the sphere gives

$$\int d\Omega e^{-ip_0 r\cos(\hat{\mathbf{p}},\hat{\mathbf{x}})}Y_l^m(\hat{\mathbf{p}}) = (-i)^l 4\pi(\pi/2p_0r)^{\frac{1}{2}}J_{l+\frac{1}{2}}(rp_0)Y_l^m(\hat{\mathbf{x}}). \quad (29)$$

This can be seen immediately from the well-known expansion of $\exp(-i\mathbf{p}\cdot\mathbf{x})$ in terms of spherical harmonics.[17] Thus we have

$$f_{nlm}(x) = (-i)^l 2^{l-1}a_{nl}^{-\frac{1}{2}}r^{-\frac{1}{2}}Y_l^m(\hat{\mathbf{x}})I_{nl}(x),$$

$$I_{nl}(x) = \int_0^\infty dp_0 e^{-p_0(1-ix^0)}p_0^{l+\frac{1}{2}}L_n^{2l+1}(2p_0)J_{l+\frac{1}{2}}(p_0r). \quad (30)$$

It is convenient to calculate the integral $I_{nl}(x)$ by means of the generating function[18]

$$F(z, p_0) = (1 - z)^{-\alpha-1}\exp\left(\frac{2p_0z}{z-1}\right) = \sum_{n=0}^{\infty} L_n^\alpha(2p_0)z^n, \quad |z| < 1,$$

$$L_n^\alpha(2p_0) = (1/n!)\partial_z^n F(z, p_0)\big|_{z=0}. \quad (31)$$

Since

$$\left|\partial_z^n F(z, p_0)\right| \leq \sum_{\nu=0}^{n} b_\nu(n, \epsilon)p_0^\nu \quad \text{for} \quad 0 \leq z \leq \epsilon < 1,$$

where the coefficients $b_\nu(n, \epsilon)$ are independent of p_0, the integral

$$\int_0^\infty dp_0\partial_z^n F(z, p_0)e^{-p_0(1-ix^0)}p_0^{l+\frac{1}{2}}J_{l+\frac{1}{2}}(p_0r) \quad (32)$$

converges uniformly for $0 \leq z \leq \epsilon < 1$ and we are allowed to interchange integration and differentiation. For $n = 0$ we get[19] for this integral

$$G(x, z) = \int_0^\infty dp_0 F(z, p_0)e^{-p_0(1-ix^0)}p_0^{l+\frac{1}{2}}J_{l+\frac{1}{2}}(p_0r) = \pi^{-\frac{1}{2}}l!\,(2r)^{l+\frac{1}{2}}w(x, z)^{-l-1},$$

$$w(x, z) = [z(1 + ix^0) + 1 - ix^0]^2 + (z - 1)^2r^2. \quad (33)$$

The integral $I_{nl}(x)$ is then given by

$$I_{nl}(x) = (1/n!)\partial_z^n G(x, z)\big|_{z=0}.$$

Thus we arrive at the result

$$f_{nlm}(x) = A_{nl}Y_l^m(\hat{\mathbf{x}})r^l\partial_z^n w(x, z)^{-l-1}\big|_{z=0},$$

$$A_{nl} = \pi^{-\frac{1}{2}}(-i)^l 2^{2l+1}l!\,[n!\,(2l + 1 + n)!]^{-\frac{1}{2}}. \quad (34)$$

If we define

$$g_{nl}(x^0, r) = r^l\partial_z^n w(x, z)^{-l-1}\big|_{z=0},$$

we get

$$g_{0l}(x^0, r) = r^l(1 - 2ix^0 - x^2)^{-l-1},$$

$$g_{1l}(x^0, r) = -2(l + 1)r^l(1 + x^2)(1 - 2ix^0 - x^2)^{-l-2}. \quad (35)$$

Furthermore, it is not difficult to find[20] the following recursion formula for $n \geq 2$:

$$g_{nl}(x^0, r) = -2(l + 1)r^{-1}(1 + x^2)g_{n-1,l+1}(x^0, r) + 2(l + 1)(1 - n)r^{-1}(1 + 2ix^0 - x^2)g_{n-2,l+1}(x^0, r). \quad (36)$$

We list the following rather obvious properties of the functions $f_{nlm}(x)$ without the details of the proof. They can be verified by explicit calculations[20] in x-space:

(1) The functions $f_{nlm}(x)$ are defined and smooth for all $x \in M^4$.

(2) We have

$$Rf_{nlm}(x) = -(1/x^2)f_{nlm}(Rx) = (-1)^n f_{nlm}(x) \tag{37}$$

for all $x \in M^4$. Therefore we can represent $f_{nlm}(x)$ in the form

$$f_{nlm}(x) = (2\pi)^{-\frac{3}{2}}(-1)^{n+1}(1/x^2)\int d\hat{\Omega} e_{nlm}(h)e^{-ih\cdot x/x^2}.$$

(3) For $x^2 \neq 0$ and r fixed, we have for large $|x^0|$

$$f_{nlm}(x) \sim (x_0^2)^{-l-1}.$$

(4) For $x^2 \neq 0$ and x^0 fixed, we have for large r

$$f_{nlm}(x) \sim r^{-l-2}.$$

(5) For $x^2 = 0$, we have for large $|x^0| = r$,

$$f_{nlm}(x) \sim |x^0|^{-1} \text{ for even } n,$$
$$f_{nlm}(x) \sim |x^0|^{-2} \text{ for odd } n.$$

Consider now the bilinear form

$$\langle f_1, f_2\rangle = -i\int_{x^0=\tau} d^3\mathbf{x}(f_1^*\partial_0 f_2 - \partial_0 f_1^* f_2). \tag{38}$$

The left-hand side is independent of τ, if f_1 and f_2 are solutions of the wave equation. It is convenient to put $\tau = 0$ in the following.

The above results show that the integral (28) exists uniformly for all x^0. Thus we are allowed to differentiate Eq. (28) with respect to x^0 under the integral and put $x^0 = 0$ afterwards. We get

$$f_{nlm}(x)\big|_{x^0=0} = (2\pi)^{-\frac{3}{2}}\int \frac{d^3p}{2p_0} e_{nlm}(\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{x}},$$

$$\partial_0 f_{n'l'm'}(x)\big|_{x^0=0} = (2\pi)^{-\frac{3}{2}} i\int d^3\mathbf{p} 2^{-1} e_{n'l'm'}(\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{x}}.$$

If we apply Parceval's theorem for Fourier transforms to these functions, we have

$$\langle f_{n'l'm'}, f_{nlm}\rangle = (e_{n'l'm'}, e_{nlm}) = \delta_{n'n}\delta_{l'l}\delta_{m'm}.$$

Thus the functions $f_{nlm}(x)$ form an orthogonal set with respect to the bilinear form (38), which is positive definite for these functions.

VI. PROPERTIES OF THE FOURIER TRANSFORMS OF THE FUNCTIONS $\varphi(p) \in \mathfrak{H}$

We want to show that any function

$$f(x) = \int_{V_+} d\hat{\Omega}\varphi(p)e(x, p),$$
$$\varphi(p) \in \mathfrak{H}, \quad e(x, p) = (2\pi)^{-\frac{3}{2}}\exp(ip\cdot x), \tag{39}$$

can also be represented in the form

$$f(x) = \int_{V_+} d\hat{\Omega} R\varphi(h)g(x, h),$$
$$g(x, h) = Re(x, h). \tag{40}$$

The equality holds in the strict sense, not only in the mean: In other words, if $f(x)$ in Eq. (39) is a solution of the wave equation, then the function $Rf(x)$ is a solution, too. This is so, because $R\varphi \in \mathfrak{H}$, if $\varphi \in \mathfrak{H}$.

First we notice that $f(x)$ in Eq. (39) exists uniformly for all $x \in M^4$, because $\varphi(p)e(x, p) \in L^1(\hat{\Omega})$ and $|e(x, p)|$ is independent of x. As $p^\mu\varphi \in \mathfrak{H}$ if $\varphi \in \mathfrak{H}$, we can compute arbitrary derivatives of $f(x)$ by differentiating under the integral sign.

In the following it will suffice to consider only the basis elements $\hat{\varphi}_{lm}(\mathbf{p}) = \varphi_{lm}(p_0)Y_l^m(\hat{\mathbf{p}})$ of $\mathfrak{H}$, because all the other elements are finite linear combinations of them.

Since the integral (39) exists, we are allowed to calculate the integral over the sphere first. This gives [see Eq. (29)]

$$f_{lm}(x) = Y_l^m(\hat{\mathbf{x}})\int_0^\infty d\hat{\omega}\varphi_l(p_0)B_l(rp_0)e^{ip_0x_0},$$
$$B_l(rp_0) = (-i)^l 4\pi(\pi/2p_0r)^{\frac{1}{2}}J_{l+\frac{1}{2}}(p_0r). \tag{41}$$

We notice that $|B_l(rp_0)\exp(ip_0x_0)|$ is uniformly bounded for all r and p_0. We now expand $\varphi_l(p_0)$ in terms of the eigenfunctions $e_{nl}(p_0)$:

$$\varphi_l(p_0) = \sum_{n=0}^\infty c_n e_{nl}(p_0). \tag{42}$$

The equality holds in the $L_l^2(\hat{\omega})$ topology. However, inserting it into Eq. (41) yields a pointwise convergence of the resulting series (see Appendix). Thus the expansion

$$f_{lm}(x) = Y_l^m(\hat{\mathbf{x}})\sum_{n=0}^\infty c_n\int_0^\infty d\hat{\omega}e_{nl}(p_0)B_l(rp_0)e^{ip_0x_0} = \sum_{n=0}^\infty c_n f_{nlm}(x) \tag{43}$$

is an equality in the usual sense.

Consider now the function

$$\tilde{f}_{lm}(x) = \int_{V_+} d\hat{\Omega} R\varphi_{lm}(p)e(p, x). \tag{44}$$

The same reasoning as above shows that

$$\tilde{f}_{lm}(x) = \sum_{n=0}^\infty c_n Rf_{nlm}(x) \tag{45}$$

holds in the usual sense, too. Thus we have $\tilde{f}_{lm}(x) = Rf_{lm}(x)$. From this the validity of Eq. (40) follows

immediately. As $R\varphi \in \mathfrak{H}$, we see that $\check{f} = Rf$ is defined everywhere and its derivatives of any order exist for all $x \in M^4$.

The above result holds, of course, for all subspaces of $\mathfrak{H}$ or $\tilde{\mathfrak{H}}_{2l+1}$—in particular, for the spaces $\tilde{\mathcal{D}}_I$, $\tilde{\mathcal{B}}_{2l+1}$, and $\tilde{\mathcal{Y}}_{2l+1}$ (the arguments of its elements restricted to the positive real line) defined by Zemanian.[4]

ACKNOWLEDGMENTS

We thank Professor K. Jörgens and Dr. M. Rinke for interesting discussions. One of us (D. H. M.) is indebted to the Deutsche Forschungsgemeinschaft for financial support.

APPENDIX

We want to prove that the series (43) converges pointwise, even though the expansion (42) holds only in the mean.

Since the functions $e_{nl}(p_0)$ form a complete orthogonal system in $L_l^2(\hat{\omega})$, we have for any $\varphi_l^{(1)}$, $\varphi_l^{(2)} \in L_l^2(\hat{\omega})$, Parceval's equation

$$\int_0^\infty d\hat{\omega}\varphi_l^{(2)*}\varphi_l^{(1)} = \sum_{n=0}^\infty c_n \int_0^\infty d\hat{\omega}\varphi_l^{(2)*} e_{nl},$$

$$c_n = \int_0^\infty d\hat{\omega} e_{nl}(p_0)\varphi_l^{(1)}(p_0), \qquad \text{(A1)}$$

where the right-hand side converges in the usual sense.[21] The function $B_l(rp_0)\exp(ip_0x_0)$ is not an element of $L_l^2(\hat{\omega})$; but, if we multiply it by a factor $\exp(-\lambda p_0)$, it is one for all $\lambda > 0$. Furthermore, the integrals

$$\int_0^\infty d\hat{\omega}\varphi_l(p_0)B_l(rp_0)e^{-\lambda p_0}e^{ip_0x_0},$$

$$\int_0^\infty d\hat{\omega}e_{nl}(p_0)B_l(rp_0)e^{-\lambda p_0}e^{ip_0x_0}$$

exist and are uniformly convergent for all $\lambda \geq 0$. Combining this fact with Eq. (A1), we get an expansion of the form

$$b(\lambda) = \sum_{n=0}^\infty b_n(\lambda), \quad \lambda > 0,$$

where $b(\lambda)$ and $b_n(\lambda)$ exist and are uniformly convergent for all $\lambda \geq 0$ and all $x \in M^4$. From this it follows that

$$b(0^+) = \sum_{n=0}^\infty b_n(0^+). \qquad \text{Q.E.D.}$$

[1] E. Cunningham, Proc. London Math. Soc. **8**, 77 (1910); H. Bateman, *ibid.* **8**, 223 (1910).

[2] H. A. Kastrup, Phys. Rev. **140**, B183 (1965).

[3] E. C. Titchmarsh, J. London Math. Soc. **1**, 195 (1926); see also C. M. Wing, Pacific J. Math. **1**, 313 (1951).

[4] A. H. Zemanian, SIAM J. Appl. Math. **14**, 561; **14**, 678 (1966).

[5] H. A. Kastrup, Ann. Physik (Leipzig) **9**, 388 (1962).

[6] See, for instance, G. Sansone, *Orthogonal Functions* (Interscience Publ. Inc., New York, 1959), p. 349.

[7] F. Riesz and B. Sz. Nagy, *Funktionalanalysis* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1956), p. 282.

[8] A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill Book Co., New York, 1953), Vol. II, p. 189, formula (20). One has to perform Abel's limit $z \to -1$ in this formula.

[9] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, 1965), 4th ed., p. 322.

[10] For the motivation of the name "Liouville transformation" see H. A. Kastrup, "Gauge Properties of the Minkowski Space," in *Particles, Currents, Symmetries*, P. Urban, Ed. (Springer-Verlag, Berlin, 1968), p. 452.

[11] See Ref. 7, p. 310.

[12] The irreducibility follows from an example discussed by M. A. Naimark, *Normed Rings* (P. Noordhoff Ltd., Groningen, 1964), p. 381.

[13] Ref. 8, Vol. I, p. 160.

[14] A. Friedman, *Generalized Functions and Partial Differential Equations* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963), p. 20.

[15] E. P. Wigner, Ann. Math. **39**, 1 (1939).

[16] H. A. Kastrup, Phys. Rev. **142**, 1060 (1966).

[17] Ref. 8, Vol. I, p. 98.

[18] Ref. 8, Vol. II, p. 189, formula (17).

[19] A. Erdélyi, Ed., *Tables of Integral Transforms* (McGraw-Hill Book Co. Inc., New York, 1964), Vol. I, p. 182.

[20] D. H. Mayer, diplom thesis, University of Munich, 1969.

[21] F. G. Tricomi, *Vorlesungen über Orthogonalreihen* (Springer-Verlag, Berlin, 1955), p. 242.

Extremum Principles for the Föppl–Hencky Equation

N. ANDERSON AND A. M. ARTHURS
Department of Mathematics, University of York, York, England

(Received 12 September 1969)

Maximum and minimum principles are obtained for the nonlinear Föppl–Hencky differential equation which arises in the theory of elasticity. Approximate solutions are obtained from a simple class of trial functions.

1. INTRODUCTION

In a recent series of papers,[1–3] variational principles associated with a wide class of boundary-value problems have been derived which under certain circumstances lead to both maximum and minimum (complementary) principles. These complementary variational principles provide a systematic method of obtaining approximate solutions to various linear and nonlinear problems. In this paper we apply the theory to the nonlinear Föppl–Hencky equation

$$\frac{d^2\phi}{dx^2} + \frac{3}{x}\frac{d\phi}{dx} = -\frac{2}{\phi^2}, \quad 0 \le x \le 1, \tag{1}$$

with

$$\phi'(0) = 0, \quad \phi(1) = \lambda > 0, \quad \phi' \equiv d\phi/dx. \tag{2}$$

This equation arises in elastic membrane theory,[4–6] where ϕ is essentially the dimensionless stress which develops in a circular membrane when subjected to a constant normal pressure.

2. COMPLEMENTARY VARIATIONAL PRINCIPLES

Consider the boundary-value problem

$$-\frac{d}{dx}\left(\gamma(x)\frac{d\Phi}{dx}\right) = f(\Phi), \quad 0 \le x \le 1, \tag{3}$$

$$\Phi'(0) = 0, \quad \Phi(1) = \lambda > 0, \tag{4}$$

where f is a real-valued function which possesses an inverse. The exact solution of this problem will be denoted by ϕ. Following Noble,[1] we write (3) in the canonical form

$$\frac{d\Phi}{dx} = \frac{1}{\gamma}U, \tag{5}$$

$$-\frac{dU}{dx} = f(\Phi), \tag{6}$$

and introduce the functionals

$$I(u, \phi) = \int_0^1 [\tfrac{1}{2}(\phi')^2 - F(\phi)]\,dx, \tag{7}$$

$$J(\Phi) = \int_0^1 [\tfrac{1}{2}(\Phi')^2 - F(\Phi)]\,dx, \quad \Phi(1) = \lambda, \tag{8}$$

$$G(U) = \int_0^1 [-U'f^{-1}(-U') - (2\gamma)^{-1}U^2 - F(f^{-1}(-U'))]\,dx + \lambda U(1), \quad \gamma^{-1}U = 0 \text{ at } x = 0, \tag{9}$$

where

$$F(\phi) = \int^{\phi} f(\theta)\,d\theta. \tag{10}$$

Then, if

$$\gamma \ge 0 \quad \text{and} \quad \frac{df(\phi)}{d\phi} \le 0, \quad 0 \le x \le 1, \tag{11}$$

we have the complementary variational principles[3]

$$G(U) \le I(u, \phi) \le J(\Phi) \tag{12}$$

for all function pairs (U, Φ) which are sufficiently close to the exact solution (u, ϕ) of (5) and (6).

3. FÖPPL–HENCKY EQUATION

Provided λ is sufficiently large, the solution of the Föppl–Hencky equation (1), subject to (2), exists and is unique.[6] For our purposes, it is convenient to rewrite Eq. (1) in the form

$$-\frac{d}{dx}\left(x^3\frac{d\Phi}{dx}\right) = \frac{2x^3}{\Phi^2}. \tag{13}$$

Comparison of (3) and (13) then gives

$$\gamma = x^3, \quad f(\Phi) = 2x^3/\Phi^2. \tag{14}$$

Since

$$\gamma = x^3 \ge 0, \quad \frac{df}{d\Phi} = -\frac{4x^3}{\Phi^3} \le 0, \quad 0 \le x \le 1, \tag{15}$$

for positive stress Φ, Eq. (11) is satisfied and the complementary variational principles (12) can be applied to this problem.

The basic functionals in this case are

$$I(u, \phi) = \int_0^1 \left(\tfrac{1}{2}x^3(\phi')^2 + \frac{2x^3}{\phi}\right)dx, \tag{16}$$

$$J(\Phi) = \int_0^1 \left(\tfrac{1}{2}x^3(\Phi')^2 + \frac{2x^3}{\Phi}\right)dx, \quad \Phi = \lambda \text{ at } x = 1, \tag{17}$$

$$G(U) = \int_0^1 \left(2(2x^3)^{\frac{1}{2}}(-U')^{\frac{1}{2}} - \frac{1}{2x^3} U^2 \right) dx + \lambda U(1),$$

$$\frac{1}{x^3} U = 0 \quad \text{at} \quad x = 0. \quad (18)$$

Since, by (5) and (14), the exact u is related to ϕ by

$$u = x^3 \phi', \quad (19)$$

we shall choose a trial U of the form

$$U = x^3 \Psi', \quad \text{where} \quad \Psi'(0) = 0. \quad (20)$$

From (18), we have

$$G(x^3\Psi') = \int_0^1 \left[2(2x^3)^{\frac{1}{2}} \left(-\frac{d}{dx}(x^3\Psi') \right)^{\frac{1}{2}} - \tfrac{1}{2} x^3 (\Psi')^2 \right] \times dx + \lambda \Psi'(1). \quad (21)$$

Hence, by (12),

$$G(x^3\Psi') \leq I(u, \phi) \leq J(\Phi), \quad (22)$$

for trial functions Φ and Ψ sufficiently close to the solution ϕ of (1) and (2).

4. CALCULATIONS

To obtain a variational solution of (1) and (2) by means of these complementary principles we have performed calculations with trial functions

$$\Phi = \alpha_1 + (\lambda - \alpha_1)x^{\beta_1}, \quad \Psi = \alpha_2 + (\lambda - \alpha_2)x^{\beta_2} \quad (23)$$

for the case $\lambda = 1$. The parameters α_1, α_2, β_1, and β_2 were determined by optimizing the functionals G and J and the results are given in Table I. We see from the table that, in terms of the metric $I(u, \phi)$, the trial functions (23) are accurate to within 0.4 percent. This accuracy can be improved of course by taking more elaborate trial functions.

TABLE I. Optimum parameters and bounds for the case $\lambda = 1$.

α_1	β_1	J	α_2	β_2	G	$(J - G)/J$
1.238	1.70	0.48311	1.236	1.93	0.48125	0.00385

Finally, we remark that the results of Sec. 2 have possible applications to generalized forms of the Föppl-Hencky equation such as those considered by Dickey.[7]

[1] B. Noble, University of Wisconsin, Mathematics Research Centre, Report No. 473, 1964.
[2] L. B. Rall, J. Math. Anal. Appl. **14**, 174 (1966).
[3] A. M. Arthurs, Proc. Cambridge Phil. Soc. **65**, 803 (1969).
[4] A. Föppl, *Vorlesungen über Technische Mechanik* (G. Teubner, Leipzig, 1907), Vol. 5.
[5] H. Hencky, Z. Math. Phys. **63**, 311 (1915).
[6] R. W. Dickey, Arch. Ratl. Mech. Anal. **26**, 219 (1967).
[7] R. W. Dickey, J. Differential Equations **4**, 399 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Clebsch–Gordan Coefficients for the Coupling of $SL(2, C)$ Principal-Series Representations

R. L. ANDERSON,* R. RĄCZKA,† M. A. RASHID,‡ AND P. WINTERNITZ§
International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy

(Received 9 September 1969)

An explicit expression is obtained for the Clebsch–Gordan coefficients which appear in the specification of a basis corresponding to the decomposition of the tensor product of two arbitrary $SL(2, C)$ principal-series representations. Special cases of physical interest are treated in detail.

I. INTRODUCTION

The knowledge of the Clebsch–Gordan (CG) coefficients for the homogeneous Lorentz group $SO(3, 1)$ [or the covering group $SL(2, C)$] is of fundamental importance in various areas of elementary particle physics. Indeed, the group $SO(3, 1)$ is being intensively used in scattering theory to furnish invariant expansions of relativistic amplitudes. It serves as the group of motions of the space of independent kinematical variables in one approach,[1] as the invariance group of the zero-angle elastic scattering amplitude in the second approach,[2,3] and as a group generating a complete set of basis functions in the third.[4] In all cases, the CG coefficients are necessary to write matrix elements and especially to express experimental quantities in terms of the amplitudes. Further, the CG coefficients for $SL(2, C)$ play an important role in the theory of relativistic wave equations[5] and in various theories involving infinite-particle multiplets, in which it is necessary to construct invariant vertex functions, Lagrangians, etc.[6]

The problem of reducing the tensor products of two irreducible unitary $SL(2, C)$ representations has been completely solved by Naimark[7] in a series of papers (for a brief summary consult Ref. 8). However, the calculation of the explicit form of the associated CG coefficients has been carried out only in certain special cases by Dolginov and Toptygin,[9] Bisiacchi and Fronsdal,[10] and Bamberg.[11]

We present here the solution to the problem of the CG coefficients for arbitrary unitary $SL(2, C)$ representations of the principal series. Our method depends, in general, on the knowledge of matrix elements for irreducible representations and, in particular, for the purposes of this paper, on those for the principal series of $SL(2, C)$ representations. These have been calculated by various authors[12] and are now completely known because of the work of Ström[12b] and Duc and Hieu.[12d]

II. INTEGRAL REPRESENTATION FOR THE CG COEFFICIENTS IN TERMS OF $SL(2, C)$ TRANSFORMATION MATRIX ELEMENTS

A. Derivation

First, we shall sketch a method for constructing a basis consisting of generalized eigenvectors, for a space carrying an irreducible unitary representation, by means of operator-valued distributions[13]

Let G be a locally compact type-I Lie group with $H = L^2(G)$ the Hilbert space of square-integrable functions with respect to the Haar measure on G. Let $\Phi \subset H \subset \Phi'$ denote a Gel'fand triplet. For the sake of definiteness, we set $\Phi = \mathfrak{D}(G)$, where $\mathfrak{D}(G)$ is the Schwartz space: the space of infinitely differentiable functions defined on G equipped by Schwartz with a locally convex topology stronger than that of $L^2(G)$, which makes it a nuclear space. Then $\Phi' = \mathfrak{D}'(G)$ = space of continuous linear functionals on Φ.

Let $\{D^{\lambda}_{p,q}(g); g \in G\}$ be a set of matrix elements which are, in general, of a subclass of unitary irreducible representations of G; the latter constitute a complete set of generalized eigenvectors for a maximal set of commuting, essentially self-adjoint operators from the regular representation of the enveloping algebra of the Lie algebra of G. The symbol λ denotes a set of eigenvalues which are invariant for an irreducible representation T^{λ} of G, while $p(q)$ denote the remaining sets of noninvariant eigenvalues corresponding to the left (right) representations. We assume for simplicity that the sets of numbers $p(q)$ are discrete and the set λ is continuous. This set of matrix elements $\{D^{\lambda}_{p,q}(g)\}$ then provides a complete set of functions for any function $\varphi(g) \in \mathfrak{D}(G)$ and satisfies the following orthogonality and completeness relations:

$$\int_G dg D^{\lambda}_{pq}(g) D^{\lambda'*}_{p'q'}(g) = \frac{1}{\mu(\lambda)} \delta(\lambda - \lambda')\delta_{pp'}\delta_{qq'}, \quad (1)$$

$$\int \mu(\lambda)\, d\lambda \sum_{pq} D^{\lambda}_{pq}(g) D^{\lambda*}_{pq}(g') = \delta(g - g'), \quad (2)$$

where the asterisk denotes complex conjugation. Here

we have assumed that the spectral measure $d\mu(\lambda)$ for the invariant operators is absolutely continuous relative to the Lebesque measure $d\lambda$; i.e., $d\mu(\lambda) = \mu(\lambda)\, d\lambda$. The functions $D^{\lambda}_{pq}(g)$ also satisfy the following unitarity condition and composition law:

$$D^{\lambda}_{pq}(g^{-1}) = D^{\lambda *}_{qp}(g), \tag{3}$$

$$D^{\lambda}_{pq}(gg') = \sum_{s} D^{\lambda}_{ps}(g) D^{\lambda}_{sq}(g'). \tag{4}$$

Specifically, any element $\varphi(g) \in \mathfrak{D}(G)$ has the following expansion in terms of the $D^{\lambda}_{pq}(g)$ functions:

$$\varphi(g) = \int d\mu(\lambda) \sum_{pq} \hat{\varphi}_{pq}(\lambda) D^{\lambda}_{pq}(g), \tag{5}$$

where

$$\hat{\varphi}_{pq}(\lambda) = \mu(\lambda) \int_{G} dg \varphi(g) D^{\lambda *}_{pq}(g). \tag{6}$$

Now set

$$P^{\lambda}_{pq} = \mu(\lambda) \int_{G} dg D^{\lambda *}_{pq}(g) T_g, \tag{7}$$

where T_g is the right regular representation of G in H, i.e., $(T_{g_0}\varphi)(g) = \varphi(gg_0)$. Then

$$\begin{aligned} P^{\lambda}_{pq}\varphi(g) &= \mu(\lambda) \int_{G} dg' D^{\lambda *}_{pq}(g') \varphi(gg') \\ &= \mu(\lambda) \int_{G} d\tilde{g} D^{\lambda *}_{pq}(g^{-1}\tilde{g}) \varphi(\tilde{g}) \\ &= \sum_{s} D^{\lambda}_{ps}(g^{-1}) \mu(\lambda) \int_{G} D^{\lambda *}_{sq}(\tilde{g}) \varphi(\tilde{g})\, d\tilde{g} \\ &= \sum_{s} D^{\lambda}_{sp}(g) \hat{\varphi}_{sq}(\lambda) \in \mathfrak{D}'(G). \end{aligned} \tag{8}$$

The interchange of the processes of integration and summation in the third step follows from Tonelli's theorem.[13] Thus, we see that P^{λ}_{pq} represents a map from $\mathfrak{D}(G)$ into $\mathfrak{D}'(G)$. Consequently, it represents an operator-valued distribution (cf. Ref. 13, Sec. 2). Using the formalism of operator-valued distributions one proves the following properties[13] of P^{λ}_{pq}:

$$P^{\lambda}_{pq} P^{\lambda'}_{p'q'} = \delta(\lambda - \lambda') \delta_{qp'} P^{\lambda}_{pq'}, \tag{9a}$$

$$(P^{\lambda}_{pq})^{\dagger} = P^{\lambda}_{qp}, \tag{9b}$$

$$T_g P^{\lambda}_{pq} = \sum_{s} D^{\lambda}_{sp}(g) P^{\lambda}_{sq}. \tag{9c}$$

This adjoint in Eq. (9b) is to be understood in the sense of operator-valued distributions.

If $H = L^2(X)$, where X is a certain homogeneous space of G, and if T_g is a quasiregular representation of G in $L^2(X)$ [i.e., $(T_g\varphi)(x) = \varphi(xg)$], the quantity P^{λ}_{nm} defined by Eq. (7) still possesses the properties (a)–(c) of Eq. (9).[13] If $u(x) \in \Phi \subset H$, then it follows from (9c) that the element

$$|\lambda;\, p\rangle = P^{\lambda}_{pq} u \tag{10}$$

transforms according to an irreducible unitary representation T^{λ} of G. Further, the linear space of all vectors for a fixed λ and q forms a carrier space for an irreducible representation T^{λ} of G (cf. Ref. 14).

If the representation T_g in formula (7) is a tensor-product representation of two irreducible unitary representations, i.e., $T_g = T^{\lambda_1}_g \otimes T^{\lambda_2}_g$, and $\Phi \subset H^{\lambda_1} \otimes H^{\lambda_2} \subset \Phi'$ is a Gel'fand triplet for the product space, then for $u \in \Phi$ formula (10) again represents a map of the space Φ into a subspace of Φ which is a carrier space of an irreducible representation T^{λ} of G (for details, cf. Ref. 14).

In the case of the Lorentz group, the explicit form of the matrices $\{D^{\nu\rho}(g)\}$ have been calculated recently by various authors.[12] Thus, we can explicitly construct the operators (7) and the basis vectors (10) and utilize them for obtaining the CG coefficients.

In fact, let us consider the tensor product of two irreducible representations $D^{\nu_1\rho_2}$ and $D^{\nu_2\rho_2}$ (ν_i being integers or half-integers, $-\infty \leq \rho_i \leq \infty$, $i = 1, 2$) which are realized in Hilbert spaces $H^{\nu_1\rho_1}$ and $H^{\nu_2\rho_2}$ and spanned by the sets of canonical basis vectors $\{|\nu_1\rho_1 J_1 M_1\rangle\}$ and $\{|\nu_2\rho_2 J_2 M_2\rangle\}$, respectively. From Naimark's results[7] we already know that the decomposition of the tensor product $D^{\nu_1\rho_1} \otimes D^{\nu_2\rho_2}$ has the following form:

$$D^{\nu_1\rho_1} \otimes D^{\nu_2\rho_2} = \sum_{\nu} \int_{-\infty}^{\infty} D^{\nu\rho}\, d\mu(\nu, \rho), \tag{11}$$

where the summation extends over all ν's such that

$$\nu + \nu_1 + \nu_2 = \text{nonnegative integer} \tag{12}$$

and each representation appears at most once.[7,15]

In the tensor-product space $H = H^{\nu_1\rho_1} \otimes H^{\nu_2\rho_2}$ we can construct two sets of orthogonal basis vectors. The first one consists of the Kronecker product of the original basis vectors

$$|\nu_1\rho_1 J_1 M_1,\, \nu_2\rho_2 J_2 M_2\rangle = |\nu_1\rho_1 J_1 M_1\rangle\, |\nu_2\rho_2 J_2 M_2\rangle, \tag{13}$$

while the second one contains the basis vectors

$$|\nu\rho JM;\, \nu_1\rho_1\nu_2\rho_2\rangle \tag{14}$$

which span an irreducible space $H^{\nu\rho}$ contained in the direct integral of the Hilbert spaces corresponding to the decomposition (11).

In what follows we shall omit indices $\nu_1\rho_1\nu_2\rho_2$ in the ket vector (14).

The normalization properties of the basis vectors (13) and (14) are

$$\begin{aligned} &\langle \nu_1\rho_1 J_1 M_1,\, \nu_2\rho_2 J_2 M_2 \,|\, \nu_1\rho_1 J_1' M_1',\, \nu_2\rho_2 J_2' M_2' \rangle \\ &\qquad = \delta_{J_1 J_1'} \delta_{J_2 J_2'} \delta_{M_1 M_1'} \delta_{M_2 M_2'} \end{aligned} \tag{15}$$

and

$$\langle \nu\rho JM \,|\, \nu'\rho' J' M' \rangle = \frac{\kappa \delta(\rho - \rho')}{4\nu^2 + \rho^2} \delta_{\nu\nu'} \delta_{JJ'} \delta_{MM'}, \tag{16}$$

where κ is a numerical constant determined by Eqs. (1), (2), (24), and (24').

In Naimark's notations, the vector $|\nu\rho JM\rangle$ corresponds to a function $f^{\nu\rho}_{JM}(z, \bar{z})$ belonging to the Hilbert space $H(R^2)$ of functions $f(z, \bar{z})$ with the domain R^2 (Ref. 16, p. 162). Here z and $\bar{z}$ are understood as independent variables ($z = x + iy$, $\bar{z} = x - iy$).

The basis vectors (13) and (14), as stated previously, are related by

$$|\nu\rho JM\rangle = (N^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{J'M'J_1M_1J_2M_2})^{-1} \times \int dg D^{\nu\rho *}_{JM,J'M'}(g) T_g |\nu_1\rho_1 J_1 M_1\rangle |\nu_2\rho_2 J_2 M_2\rangle. \quad (17)$$

In this formula for the state $|\nu_1\rho_1 J_1 M_1\rangle |\nu_2\rho_2 J_2 M_2\rangle$ we may take any vector (13) from the tensor-product space $H^{\nu_1\rho_1} \otimes H^{\nu_2\rho_2}$. N is some normalization coefficient. The operator T_g acts on the basis vector (13) as

$$T_g |\nu_1\rho_1 J_1 M_1\rangle |\nu_2\rho_2 J_2 M_2\rangle = \sum_{J_1'M_1'J_2'M_2'} D^{\nu_1\rho_1}_{J_1'M_1',J_1M_1}(g) \times |\nu_1\rho_1 J_1' M_1'\rangle D^{\nu_2\rho_2}_{J_2'M_2',J_2M_2}(g) |\nu_2\rho_2 J_2' M_2'\rangle. \quad (18)$$

Using the invariance of the measure dg and the group properties of the matrices $D^{\nu\rho}(g)$ we can easily check that the basis vectors (17) have the correct transformation properties.

The CG coefficients are the matrix elements of the transition matrix between the basis vectors (13) and (14). Utilizing (17) and (18) we get

$$\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 | \nu\rho JM\rangle = (N^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{J'M'J_1'M_1'J_2'M_2'})^{-1} \times \int dg D^{\nu\rho *}_{JM,J'M'}(g) D^{\nu_1\rho_1}_{J_1M_1,J_1'M_1'}(g) D^{\nu_2\rho_2}_{J_2M_2,J_2'M_2'}(g). \quad (19)$$

In order to find the normalization coefficient $N^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{JMJ_1M_1J_2M_2}$, we calculate the square of a basis vector (17). From the group properties of the matrices $D^{\nu\rho}(g)$ [Eq. (4)] and the normalization (16), we find that

$$N^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{JMJ_1M_1J_2M_2} = \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 | \nu\rho JM\rangle^*. \quad (20)$$

Note that the normalization (16) is chosen so as to cancel the factor

$$A^{\nu\rho} = \frac{\kappa}{4\nu^2 + \rho^2} \quad (21)$$

which appears in the integration defining normalization of the $D^{\nu\rho}_{JM,J'M'}(g)$ functions [see Eqs. (24a) and (24b)].

The group element g can be represented in the form

$$g = u_1(\eta, \psi, 0) g(a) u_2(\varphi_1, \theta, \varphi_2), \quad (22)$$

with $0 \le \eta < 4\pi$, $0 \le \varphi_1, \varphi_2 < 2\pi$, $0 \le \psi, \theta \le \pi$, and $0 \le a < \infty$, where u_1 and u_2 belong to $SU(2)$ and $g(a)$ is an element of a 1-parameter noncompact subgroup of $SL(2, C)$. The decomposition (22) induces the following form for the $D^{\nu\rho}(g)$ matrices:

$$D^{\nu\rho}_{JM,J'M'}(g) = \sum_{\lambda=-\min(J,J')}^{\min(J,J')} D^{J}_{M\lambda}(\eta, \psi, 0) \times D^{\nu\rho}_{JJ'\lambda}(a) D^{J'}_{\lambda M'}(\varphi_1, \theta, \varphi_2). \quad (23)$$

With respect to the parametrization (22) the invariant measure on $SL(2, C)$ has the form

$$dg = d(\cos\psi)\, d\eta\, d\varphi_1\, d(\cos\theta)\, d\varphi_2 \sinh^2 a\, da \quad (24a)$$

and

$$\mu(\lambda) = (4\nu^2 + \rho^2)/\kappa. \quad (24b)$$

Using formulas (19), (20), (23), (24a), and (24b), carrying out the elementary integrations over the variables ψ, η, φ_1, θ, φ_2, we arrive at

$$\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 | \nu\rho JM\rangle \times \langle \nu_1\rho_1 J_1' M_1', \nu_2\rho_2 J_2' M_2' | \nu\rho J'M'\rangle^* = \frac{64\pi^3}{(2J+1)(2J'+1)} (J_1M_1, J_2M_2 | JM) \times (J_1'M_1', J_2'M_2' | J'M') \sum_{\lambda_1\lambda_2} (J_1\lambda_1, J_2\lambda_2 | J\,\lambda_1 + \lambda_2) \times (J_1'\lambda_1, J_2'\lambda_2 | J'\,\lambda_1 + \lambda_2) \times \int_0^\infty D^{\nu\rho *}_{JJ'\lambda_1+\lambda_2}(a) D^{\nu_1\rho_1}_{J_1J_1'\lambda_1}(a) D^{\nu_2\rho_2}_{J_2J_2'\lambda_2}(a) \sinh^2 a\, da, \quad (25)$$

where $(a, b, c, d | e, f)$ is a CG coefficient for the rotation group $O(3)$.

The last integration will be carried out in Sec. III.

B. Orthogonality and Completeness Relations

The CG coefficients as defined by (19) figure in the decomposition

$$|\nu\rho JM\rangle = \sum_{J_1=\nu_1}^{\infty} \sum_{J_2=\nu_2}^{\infty} \sum_{M_1M_2} \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 | \nu\rho JM\rangle \times |\nu_1\rho_1 J_1 M_1\rangle |\nu_2\rho_2 J_2 M_2\rangle, \quad (26)$$

where the canonical bases are normalized according to (15) and (16).

Calculating the norms of both sides in (26) we obtain the orthogonality relation

$$\sum_{J_1J_2M_1M_2} \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 | \nu\rho JM\rangle \times \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 | \nu'\rho' J'M'\rangle^* = \frac{\delta(\rho - \rho')\kappa}{\rho^2 + 4\nu^2} \delta_{\nu\nu'} \delta_{JJ'} \delta_{MM'}. \quad (27)$$

Using (27) we can easily check that the formula inverse to (26) is

$$|\nu_1\rho_1 J_1 M_1\rangle\,|\nu_2\rho_2 J_2 M_2\rangle = \sum_\nu \int \frac{(\rho^2+4\nu^2)}{\kappa}\, d\rho \sum_{JM} \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \,|\, \nu\rho JM\rangle^* \times |\nu\rho JM\rangle. \quad (28)$$

Combining (26) and (28), we obtain

$$\sum_\nu \int \frac{(\rho^2+4\nu^2)}{\kappa}\, d\rho \sum_{JM} \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \,|\, \nu\rho JM\rangle \times \langle \nu_1\rho_1 J_1' M_1', \nu_2\rho_2 J_2' M_2' \,|\, \nu\rho JM\rangle^* = \delta_{J_1 J_1'}\delta_{J_2 J_2'}\delta_{M_1 M_1'}\delta_{M_2 M_2'}, \quad (29)$$

which is the completeness relation.

C. Explicit Expression for the $SL(2, C)$ Transformation Matrices

In the following we shall need the explicit form of the transformation matrices $D^{\nu\rho}(a)$. For our purpose the most convenient one is that of Ström,[12b] which is calculated for a canonical basis as defined by Naimark,[16]

$$D^{\nu\rho}_{JJ'\lambda}(a) = \sum_{d=\max(0,-\nu-\lambda)}^{\min(J-\nu,J-\lambda)} \sum_{d'=\max(0,-\nu-\lambda)}^{\min(J'-\nu,J'-\lambda)} B^{\nu\rho}_{JJ'\lambda dd'} e^{-a(2d'+\lambda+\nu+1-\frac{1}{2}i\rho)} \times {}_2F_1(1+J'-\tfrac{1}{2}i\rho, \nu+\lambda+d+d'+1, J+J'+2, 1-e^{-2a}), \quad (30)$$

where

$$B^{\nu\rho}_{JJ'\lambda dd'} = \alpha^{\nu\rho *}_J \alpha^{\nu\rho}_{J'} (-1)^{J+J'-2\lambda+d+d'} \frac{1}{(J+J'+1)!} \times [(2J+1)(J-\nu)!\,(J+\nu)!\,(J-\lambda)!\,(J+\lambda)!\,(2J'+1)(J'-\nu)!\,(J'+\nu)!\,(J'-\lambda)!\,(J'+\lambda)!]^{\frac{1}{2}} \times \frac{(\nu+\lambda+d+d')!\,(J+J'-\nu-\lambda-d-d')!}{d!\,(J-\lambda-d)!\,(J-\nu-d)!\,(\lambda+\nu+d)!\,d'!\,(J'-\lambda-d')!\,(J-\nu-d)!\,(\lambda+\nu+d')!} \quad (31)$$

and

$$\alpha^{\nu\rho}_J = \prod_{s=|\nu|}^{J} \frac{-2s+i\rho}{(4s^2+\rho^2)^{\frac{1}{2}}} = \frac{\Gamma(-|\nu|+\frac{1}{2}i\rho+1)\,|\Gamma(-J+\frac{1}{2}i\rho)|}{|\Gamma(-|\nu|+\frac{1}{2}i\rho+1)|\,\Gamma(-J+\frac{1}{2}i\rho)}.$$

The choice of the phase $\alpha^{\nu\rho}_J$ is such that the vectors $|\nu\rho JM\rangle$ form a canonical basis in the sense of Naimark,[16] i.e., the operation of all generators on $|\nu\rho JM\rangle$ is completely prescribed.

III. THE CG COEFFICIENT

A. General Expression

In this section we carry out the integration appearing in the formula (25) for the CG coefficients. Substituting (30) in (25) and introducing a new variable $x = 1 - e^{-2a}$, we reduce the integral in (25) to a sum of terms of the type

$$\int_0^1 (1-x)^{d'+d_1'+d_2'+\lambda_1+\lambda_2+\frac{1}{2}(\nu+\nu_1+\nu_2)+\frac{1}{2}i(\rho-\rho_1-\rho_2)-\frac{1}{2}} \times {}_2F_1(1+J'+\tfrac{1}{2}i\rho, \nu+\lambda_1+\lambda_2+d+d'+1, J+J'+2, x) \times {}_2F_1(1+J_1-\tfrac{1}{2}i\rho_1, \nu_1+\lambda_1+d_1+d_1'+1, J_1+J_1'+2, x) \times {}_2F_1(1+J_2-\tfrac{1}{2}i\rho_2, \nu_2+\lambda_2+d_2+d_2'+1, J_2+J_2'+2, x) x^2\, dx.$$

To perform this integration, we develop each of the hypergeometric functions into a power series. These series converge uniformly for $x < 1$, and we can perform the integration term by term from 0 to $1-\epsilon$ and take the limit $\epsilon \to 0$. Since this limit exists, Abel's theorem on the value of a function on the boundary of its region of convergence ensures that the obtained value is the correct answer.[17] Performing the integration (see, e.g.,

Ref. 18, p. 284), we obtain

$$\begin{aligned}
&\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \,|\, \nu\rho JM\rangle\langle \nu_1\rho_1 J_1' M_1', \nu_2\rho_2 J_2' M_2' \,|\, \nu\rho J'M'\rangle^* \\
&\quad = \frac{8\pi^3}{(2J+1)(2J'+1)} (J_1M_1, J_2M_2 \,|\, JM)(J_1'M_1', J_2'M_2' \,|\, J'M') \\
&\quad\quad \times \sum_{\lambda_1\lambda_2} (J_1\lambda_1, J_2\lambda_2 \,|\, J\,\lambda_1+\lambda_2)(J_1'\lambda_1, J_2'\lambda_2 \,|\, J'\,\lambda_1+\lambda_2) \\
&\quad\quad \times \sum_{dd'd_1d_1'd_2d_2'} B^{\nu\rho^*}_{JJ'\,\lambda_1+\lambda_2 dd'} B^{\nu_1\rho_1}_{J_1J_1'\lambda_1 d_1 d_1'} B^{\nu_2\rho_2}_{J_2J_2'\lambda_2 d_2 d_2'} \\
&\quad\quad \times \sum_{nn_1n_2} \frac{(1+J'+\frac{1}{2}i\rho)_n (\nu+\lambda_1+\lambda_2+d+d'+1)_n}{n!\,(J+J'+2)_n} \\
&\quad\quad \times \frac{(1+J_1'-\frac{1}{2}i\rho_1)_{n_1}(\nu_1+\lambda_1+d_1+d_1'+1)_{n_1}}{n_1!\,(J_1+J_1'+2)_{n_1}} \frac{(1+J_2'-\frac{1}{2}i\rho_2)_{n_2}(\nu_2+\lambda_2+d_2+d_2'+1)_{n_2}}{n_2!\,(J_2+J_2'+2)_{n_2}} \\
&\quad\quad \times \frac{\Gamma(d'+d_1'+d_2'+\lambda_1+\lambda_2+\frac{1}{2}(\nu+\nu_1+\nu_2)+\frac{1}{2}i(\rho-\rho_1-\rho_2)+\frac{1}{2})(n+n_1+n_2+2)!}{\Gamma(d'+d_1'+d_2'+\lambda_1+\lambda_2+\frac{1}{2}(\nu+\nu_1+\nu_2)+\frac{1}{2}i(\rho-\rho_1-\rho_2)+\frac{7}{2}+n+n_1+n_2)}. \qquad (32)
\end{aligned}$$

The sums with respect to λ's, d's, and d''s are finite, their limits being

$$-\min(J_i, J_i') \le \lambda_i \le \min(J_i, J_i'), \quad i = 1, 2,$$
$$\min(0, -\lambda-\nu) \le d \le \min(J-\lambda, J-\nu), \quad \text{etc.}$$

The sums over n, n_1, and n_2 are from 0 to ∞, and the triple sum is convergent. We use the notation

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}.$$

Let us remark that the triple infinite sum $\sum_{nn_1n_2}$ figuring in (32) and all subsequent expressions can be rearranged in such a way that two of the sums can be expressed as generalized hypergeometric functions of unit argument. The CG coefficient can thus be expressed using a single infinite and many finite sums over a product of Γ functions, a terminating ${}_4F_3$ function and an ${}_3F_2$ function. Since this does not simplify calculations, we shall not give the explicit expression here.

The number of finite sums in formula (32) can be reduced to, at most, four by a suitable choice of the superfluous parameters J_i' and M_i'. Using the symmetry properties of the $D^{\nu\rho}_{J'\lambda}(a)$ we can always arrange the coefficient to be such that

$$\nu_1 \ge \nu_2 \ge \nu. \qquad (33)$$

(These symmetry relations are discussed in detail in the next paper.) Choosing

$$J_1' = \nu_1, \quad J_2' = \nu_2 \quad (\text{and, e.g., } M_1' = \nu_1, M_2' = -\nu_2), \qquad (33')$$

we eliminate the sums over d_1' and d_2' and can perform the d_1 and d_2 summations with the help of

$$\sum_{d_i} (-1)^{d_i} \frac{(\nu_i+\lambda_i+d_i+n_i)!}{d_i!\,(J_i-\nu_i-d_i)!\,(\lambda_i+\nu_i+d_i)!} = (-1)^{J_i-\nu_i} \frac{(\nu_i+\lambda_i+n_i)!\,n_i!}{(J_i-\nu_i)!\,(J_i+\lambda_i)!\,(\nu_i-J_i+n_i)!}. \qquad (34)$$

Thus, we obtain

$$\begin{aligned}
&\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \,|\, \nu\rho JM\rangle\langle \nu_1\rho_1\nu_1\nu_1, \nu_2\rho_2\nu_2 \,-\nu_2 \,|\, \nu\rho J'\,\nu_1-\nu_2\rangle^* \\
&\quad = \frac{8\pi^3}{(2J+1)(2J'+1)} (J_1M_1, J_2M_2 \,|\, JM)(\nu_1\nu_1, \nu_2\,-\nu_2 \,|\, J'\,\nu_1-\nu_2) \\
&\quad\quad \times \sum_{\lambda_1\lambda_2} A^{\nu_1\rho_1}_{J_1\lambda_1} A^{\nu_2\rho_2}_{J_2\lambda_2} (J_1\lambda_1, J_2\lambda_2 \,|\, J\,\lambda_1+\lambda_2)(\nu_1\lambda_1, \nu_2\lambda_2 \,|\, J'\,\lambda_1+\lambda_2) \\
&\quad\quad \times \sum_{dd'} B^{\nu\rho^*}_{JJ'\,\lambda_1+\lambda_2 dd'} \sum_{nn_1n_2} \frac{(1+J'+\frac{1}{2}i\rho)_n(\nu+\lambda_1+\lambda_2+d+d'+1)_n}{n!\,(J+J'+2)_n} \\
&\quad\quad \times \frac{(1+\nu_1-\frac{1}{2}i\rho_1)_{n_1}(\nu_1+\lambda_1+n_1)!}{(J_1+\nu_1+1+n_1)!\,(-J_1+\nu_1+n_1)!} \frac{(1+\nu_2-\frac{1}{2}i\rho_2)_{n_2}(\nu_2+\lambda_2+n_2)!}{(J_2+\nu_2+1+n_2)!\,(-J_2+\nu_2+n_2)!} \\
&\quad\quad \times \frac{\Gamma(d'+\lambda_1+\lambda_2+\frac{1}{2}(\nu+\nu_1+\nu_2)+\frac{1}{2}i(\rho-\rho_1-\rho_2)+\frac{1}{2})(n+n_1+n_2+2)!}{\Gamma(d'+\lambda_1+\lambda_2+\frac{1}{2}(\nu+\nu_1+\nu_2)+\frac{1}{2}i(\rho-\rho_1-\rho_2)+\frac{7}{2}+n+n_1+n_2)}, \qquad (35)
\end{aligned}$$

where

$$A^{\nu\rho}_{J\lambda} = \alpha^{\nu\rho^*}_J \alpha^{\nu\rho}_\nu \left(\frac{(2J+1)(2\nu+1)!\,(J+\nu)!\,(J-\lambda)!}{(J-\nu)!\,(J+\lambda)!\,(\nu+\lambda)!\,(\nu-\lambda)!} \right)^{\frac{1}{2}} \tag{36}$$

and $B^{\nu\rho}_{JJ'\lambda\, dd'}$ was defined in Eq. (31).

In order to fix the CG coefficient completely, it is still necessary to choose the value of J' and the phase, for instance, by taking [see Eq. (33)]

$$J' = \nu_1 - \nu_2 \tag{33''}$$

and defining the coefficient

$$\langle \nu_1\rho_1\nu_1\nu_1,\ \nu_2\rho_2\nu_2\ -\nu_2 \mid \nu\rho\ \nu_1 - \nu_2\ \nu_1 - \nu_2 \rangle$$

to be real and positive.

B. Simplification when ν, ν_1, and ν_2 Satisfy a Triangular Relation

If the invariant parameters ν, ν_1, and ν_2 satisfy the relation

$$|\nu_1 - \nu_2| \leq \nu \leq \nu_1 + \nu_2,$$

we can put [cf. (33'')]

$$J' = \nu \tag{33'''}$$

in Eq. (35). Thus, we achieve a further simplification:

$$\begin{aligned}
&\langle \nu_1\rho_1 J_1 M_1,\ \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM \rangle \langle \nu_1\rho_1\nu_1\nu_1,\ \nu_2\rho_2\nu_2\ -\nu_2 \mid \nu\rho\nu\ \nu_1 - \nu_2 \rangle^* \\
&\quad = \frac{8\pi^3}{(2J+1)(2\nu+1)} (J_1 M_1, J_2 M_2 \mid JM) \langle \nu_1\nu_1,\ \nu_2\ -\nu_2 \mid \nu\ \nu_1 - \nu_2 \rangle \\
&\qquad \times \sum_{\lambda_1\lambda_2} (J_1\lambda_1, J_2\lambda_2 \mid J\ \lambda_1 + \lambda_2)(\nu_1\lambda_1, \nu_2\lambda_2 \mid \nu\ \lambda_1 + \lambda_2) A^{\nu\rho^*}_{J\lambda_1+\lambda_2} A^{\nu_1\rho_1}_{J_1\lambda_1} A^{\nu_2\rho_2}_{J_2\lambda_2} \\
&\qquad \times \sum_{n n_1 n_2} \frac{(1 + \nu + \frac{1}{2}i\rho)_n (\nu + \lambda_1 + \lambda_2 + n)!}{(J + \nu + 1 + n)!\,(-J + \nu + n)!} \\
&\qquad \times \frac{(1 + \nu_1 - \frac{1}{2}i\rho_1)_{n_1} (\nu_1 + \lambda_1 + n_1)!}{(J_1 + \nu_1 + 1 + n_1)!\,(-J_1 + \nu_1 + n_1)!} \frac{(1 + \nu_2 - \frac{1}{2}i\rho_2)_{n_2} (\nu_2 + \lambda_2 + n_2)!}{(J_2 + \nu_2 + 1 + n_2)!\,(-J_2 + \nu_2 + n_2)!} \\
&\qquad \times \frac{\Gamma(\lambda_1 + \lambda_2 + \frac{1}{2}(\nu + \nu_1 + \nu_2) + \frac{1}{4}i(\rho - \rho_1 - \rho_2) + \frac{1}{2})(n + n_1 + n_2 + 2)!}{\Gamma(\lambda_1 + \lambda_2 + \frac{1}{2}(\nu + \nu_1 + \nu_2) + \frac{1}{4}i(\rho - \rho_1 - \rho_2) + n + n_1 + n_2 + \frac{7}{2})}.
\end{aligned} \tag{37}$$

The d and d' summations have completely disappeared.

IV. CG COEFFICIENT FOR THE COUPLING OF DEGENERATE REPRESENTATIONS

A. Calculation of a "Normalization" CG Coefficient

If we consider the case $\nu = \nu_1 = \nu_2 = 0$, Eq. (37) reduces to

$$\begin{aligned}
&\langle 0\rho_1 J_1 M_1,\ 0\rho_2 J_2 M_2 \mid 0\rho JM \rangle \langle 0\rho_1 00,\ 0\rho_2 00 \mid 0\rho 00 \rangle^* \\
&\quad = 8i\pi^3 \alpha^{0\rho}_J \alpha^{0\rho_1^*}_{J_1} \alpha^{0\rho_2^*}_{J_2} \left(\frac{(2J_1+1)(2J_2+1)}{(2J+1)} \right)^{\frac{1}{2}} \frac{(J_1 M_1, J_2 M_2 \mid JM)(J_1 0, J_2 0 \mid J0)}{\Gamma(1 + \frac{1}{2}i\rho)\Gamma(1 - \frac{1}{2}i\rho_1)\Gamma(1 - \frac{1}{2}i\rho_2)} \\
&\qquad \times \sum_{n n_1 n_2} \frac{\Gamma(1 + \frac{1}{2}i\rho + n) n!}{(J + n + 1)!\,(-J + n)!} \frac{\Gamma(1 - \frac{1}{2}i\rho_1 + n_1) n_1!}{(J_1 + n_1 + 1)!\,(-J_1 + n_1)!} \\
&\qquad \times \frac{\Gamma(1 - \frac{1}{2}i\rho_2 + n_2) n_2!}{(J_2 + n_2 + 1)!\,(-J_2 + n_2)!} \frac{\Gamma(\frac{1}{4}i(\rho - \rho_1 - \rho_2) + \frac{1}{2})(n + n_1 + n_2 + 2)!}{\Gamma(\frac{1}{4}i(\rho - \rho_1 - \rho_2) + n + n_1 + n_2 + \frac{7}{2})}.
\end{aligned} \tag{38}$$

In this case, the "normalization CG coefficient" is especially simple and can be calculated explicitly. In fact, from Eq. (38), by manipulating the triple sum

$$\begin{aligned}
&\langle 0\rho_1 00,\ 0\rho_2 00 \mid 0\rho 00 \rangle \\
&\quad = \frac{8\pi^2}{(\rho\rho_1\rho_2)^{\frac{1}{2}}} \left(\frac{\sinh \frac{1}{2}\pi\rho \sinh \frac{1}{2}\pi\rho_1 \sinh \frac{1}{2}\pi\rho_2}{\cosh \frac{1}{4}\pi(\rho + \rho_1 + \rho_2) \cosh \frac{1}{4}\pi(\rho - \rho_1 - \rho_2) \cosh \frac{1}{4}\pi(\rho + \rho_1 - \rho_2) \cosh \frac{1}{4}\pi(\rho - \rho_1 + \rho_2)} \right)^{\frac{1}{2}}.
\end{aligned} \tag{39}$$

B. Explicit Formulas for the CG Coefficient

Inserting (39) into (38), we obtain

$$\begin{aligned}
&\langle 0\rho_1 J_1 M_1, 0\rho_2 J_2 M_2 \mid 0\rho J M\rangle \\
&\quad = i\pi(\rho\rho_1\rho_2)^{\frac{1}{2}}\left(\frac{(2J_1+1)(2J_2+1)}{(2J+1)}\right)^{\frac{1}{2}}\alpha_J^{0\rho}\alpha_{J_1}^{0\rho_1 *}\alpha_{J_2}^{0\rho_2 *}(J_1M_1, J_2M_2 \mid JM)(J_10, J_20 \mid J0) \\
&\qquad \times \frac{\Gamma(\frac{1}{2}+\frac{1}{4}i(\rho-\rho_1-\rho_2))}{\Gamma(1+\frac{1}{2}i\rho)\Gamma(1-\frac{1}{2}i\rho_1)\Gamma(1-\frac{1}{2}i\rho_2)} \\
&\qquad \times \left(\frac{\cosh\frac{1}{4}\pi(\rho+\rho_1+\rho_2)\cosh\frac{1}{4}\pi(\rho-\rho_1-\rho_2)\cosh\frac{1}{4}\pi(\rho+\rho_1-\rho_2)\cosh\frac{1}{4}\pi(\rho-\rho_1+\rho_2)}{\sinh\frac{1}{2}\pi\rho\sinh\frac{1}{2}\pi\rho_1\sinh\frac{1}{2}\pi\rho_2}\right)^{\frac{1}{2}} \\
&\qquad \times \sum_{nn_1n_2}\frac{\Gamma(1+\frac{1}{2}i\rho+n)n!}{(J+n+1)!\,(-J+n)!}\,\frac{\Gamma(1-\frac{1}{2}i\rho_1+n_1)n_1!}{(J_1+n_1+1)!\,(-J_1+n_1)!} \\
&\qquad \times \frac{\Gamma(1-\frac{1}{2}i\rho_2+n_2)n_2!}{(J_2+n_2+1)!\,(-J_2+n_2)!}\,\frac{(n+n_1+n_2+2)!}{\Gamma(\frac{1}{4}i(\rho-\rho_1-\rho_2)+n+n_1+n_2+\frac{7}{2})}\,.
\end{aligned} \tag{40}$$

In the evaluation of (39) we have assumed the coefficient to be real and positive.

From the presence of $(J_10, J_20 \mid J0)$ in the above formula, it is clear that the only nonzero CG coefficients are the ones with $J_1 + J_2 + J =$ even integer in this special case.

So far we have made only a partial comparison of (40) with the results of Dolginov[9] and Domokos.[19] This comparison shows that our expressions (38)–(40) agree with their results.

V. ALTERNATIVE METHOD FOR CALCULATING NUMERICAL VALUES

For concrete calculations of specific CG coefficients, it may prove advantageous to use a slightly different approach.

So far we have made use of $D^{\nu\rho}$ matrices for the $SL(2, C)$ group expressed in terms of hypergeometric functions. However, they can also be expressed in terms of elementary functions. Indeed, using an integral representation for the hypergeometric function, we have

$$\begin{aligned}
D^{\nu\rho}_{JJ'\lambda}(a) = \sum_{dd'} B^{\nu\rho}_{JJ'\lambda dd'} e^{-a(2d'+\lambda+\nu+2+J)}\,&\frac{(J+J'+1)!}{(\nu+\lambda+d+d')!\,(J+J'-\nu-\lambda-d-d')!} \\
&\times \int_0^\infty \left(\frac{e^{-a}(1+t)}{e^{-2a}+t}\right)^{-1-J-\frac{1}{2}i\rho} t^{J+J'-\nu-\lambda-d-d'}(e^{-2a}+t)^{-2-J-J'}\,dt.
\end{aligned}$$

Putting $x = e^{-a}(1+t)/(e^{-2a}+t)$ and expanding the positive integer powers of $(1-e^a x)$ and $(x-e^a)$ into (finite) binomial series, we obtain (this is a slight modification of the results in Ref. 12c)

$$D^{\nu\rho}_{JJ'\lambda}(a) = \sum_{dd'pq} \tilde{B}^{\nu\rho}_{JJ'\lambda dd'pq} e^{-a(\nu+\lambda+2d'+p-q-J')}(\sinh a)^{-J-J'-1}\sinh a(-J+p+q-\tfrac{1}{2}i\rho), \tag{41}$$

where

$$\tilde{B}^{\nu\rho}_{JJ'\lambda dd'pq} = \frac{(-1)^{p+q-J-J'+2\lambda}}{2^{J+J'}(-J-\frac{1}{2}i\rho+p+q)}\,\frac{(J+J'+1)!\,B^{\nu\rho}_{JJ'\lambda dd'}}{p!\,q!\,(\nu+\lambda+d+d'-q)!\,(J+J'-\nu-\lambda-d-d'-p)!}\,.$$

Using this D matrix to calculate the CG coefficient, we get immediately

$$\begin{aligned}
&\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM\rangle\langle \nu_1\rho_1 J_1' M_1', \nu_2\rho_2 J_2' M_2' \mid \nu\rho J'M'\rangle^* \\
&\quad = \frac{64\pi^3}{(2J+1)(2J'+1)}(J_1M_1, J_2M_2 \mid JM)(J_1'M_1', J_2'M_2' \mid J'M')\sum_{\lambda_1\lambda_2}(J_1\lambda_1, J_2\lambda_2 \mid J\,\lambda_1+\lambda_2)(J_1'\lambda_1, J_2'\lambda_2 \mid J'\,\lambda_1+\lambda_2) \\
&\qquad \times \sum_{dd'pq}\sum_{d_1d_1'p_1q_1}\sum_{d_2d_2'p_2q_2}\tilde{B}^{\nu\rho}_{JJ'\lambda_1+\lambda_2 dd'pq}\tilde{B}^{\nu_1\rho_1}_{J_1J_1'\lambda_1 d_1d_1'p_1q_1}\tilde{B}^{\nu_2\rho_2}_{J_2J_2'\lambda_2 d_2d_2'p_2q_2} \\
&\qquad \times \int_0^\infty e^{-a[-J'-J_1'-J_2'+\nu+\nu_1+\nu_2+2(\lambda_1+\lambda_2+d'+d_1'+d_2')+p+p_1+p_2-q-q_1-q_2]} \\
&\qquad \times (\sinh a)^{-J-J'-J_1-J_1'-J_2-J_2'-1}\sinh a(-J+p+q+\tfrac{1}{2}i\rho) \\
&\qquad \times \sinh a(-J_1+p_1+q_1-\tfrac{1}{2}i\rho_1)\sinh a(-J_2+p_2+q_2-\tfrac{1}{2}i\rho_2)\,da.
\end{aligned} \tag{42}$$

The integrals in (42) are, in general, divergent, and in explicit calculations they must be combined in such a way that the divergences cancel.

Using (42), we can calculate one CG coefficient for each set of invariant parameters $\nu_1\rho_1$, $\nu_2\rho_2$, and $\nu\rho$. Then, as shown in the next paper, this "starting" coefficient and recursion relations allow one to calculate any arbitrary CG coefficient corresponding to this set of invariant parameters.

Let us demonstrate the calculation of a "starting" coefficient with two examples:

(a) $\langle 0\rho_1 00, 0\rho_2 00 \mid 0\rho 00\rangle$

$$= \frac{16(2\pi^3)^{\frac{1}{2}}}{(\rho\rho_1\rho_2)^{\frac{1}{2}}}\left(\int_0^\infty \frac{\sin\frac{1}{2}\rho a \sin\frac{1}{2}\rho_1 a \sin\frac{1}{2}\rho_2 a\, da}{\sinh a}\right)^{\frac{1}{2}}$$

$$= \frac{8\pi^2}{(\rho\rho_1\rho_2)^{\frac{1}{2}}}\left(\frac{\sinh\frac{1}{2}\pi\rho \sinh\frac{1}{2}\pi\rho_1 \sinh\frac{1}{2}\pi\rho_2}{\cosh\frac{1}{4}\pi(\rho+\rho_1+\rho_2)\cosh\frac{1}{4}\pi(\rho-\rho_1-\rho_2)\cosh\frac{1}{4}\pi(\rho+\rho_1-\rho_2)\cosh\frac{1}{4}\pi(\rho-\rho_1+\rho_2)}\right)^{\frac{1}{2}}, \tag{43}$$

agreeing with (39), obtained previously by a different method.

(b) A less trivial CG coefficient which can be computed by performing tedious, though elementary calculations is

$$\langle \tfrac{1}{2}\rho_1 \tfrac{1}{2} M_1, \tfrac{1}{2}\rho_2 \tfrac{1}{2} -M_2 \mid 0\rho 00\rangle^2$$

$$= \frac{2(32)^2\pi^3}{\rho}\langle \tfrac{1}{2}M_1, \tfrac{1}{2} -M_1 \mid 00\rangle^2 \langle \tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2} -\tfrac{1}{2} \mid 00\rangle^2$$

$$\times \operatorname{Re}\left(\frac{1}{(1+i\rho_1)(1+i\rho_2)}\int_0^\infty \frac{\sin\frac{1}{2}\rho a}{(\sinh a)^3}\sinh\tfrac{1}{2}a(1+i\rho_1)\sinh\tfrac{1}{2}a(1+i\rho_2)\,da\right.$$

$$\left. - \frac{1}{(1+i\rho_1)(1-i\rho_2)}\int_0^\infty \frac{\cosh a \sin\frac{1}{2}\rho a}{(\sinh a)^3}\sinh\tfrac{1}{2}a(1+i\rho_1)\sinh\tfrac{1}{2}a(1-i\rho_2)\,da\right)$$

or

$$\langle \tfrac{1}{2}\rho_1 \tfrac{1}{2}\tfrac{1}{2}, \tfrac{1}{2}\rho_2 \tfrac{1}{2} -\tfrac{1}{2} \mid 0\rho 00\rangle$$

$$= 4\pi^2\left(\frac{(\rho+\rho_1+\rho_2)(\rho-\rho_1-\rho_2)\sinh\frac{1}{2}\pi\rho\cosh\frac{1}{2}\pi\rho_1\cosh\frac{1}{2}\pi\rho_2}{\rho(1+\rho_1^2)(1+\rho_2^2)\sinh\frac{1}{4}\pi(\rho+\rho_1+\rho_2)\sinh\frac{1}{4}\pi(\rho-\rho_1-\rho_2)\cosh\frac{1}{4}\pi(\rho+\rho_1-\rho_2)\cosh\frac{1}{4}\pi(\rho-\rho_1+\rho_2)}\right)^{\frac{1}{2}}. \tag{44}$$

In both cases we have chosen the CG coefficients to be real and positive.

Incidentally, the coupling of representation $\nu = \frac{1}{2}$ and $\rho = 0$ is of special interest, since this and the representation $\nu = 0$ with $\rho = \frac{1}{2}i$, belonging to the supplementary series, are the only unitary irreducible representations of the $SL(2, C)$ group, for which the basis functions satisfy an invariant wave equation[16] (these are the Majorana representations).

VI. CONCLUSIONS

We have considered the problem of the CG coefficients for the $SL(2, C)$ group and have obtained a general expression for these coefficients for the principal series representations. The merit of our derivation is that it generalizes to other series and other locally compact groups. We have been able to achieve it through the knowledge of the matrix elements of the irreducible representations. Previous attempts by Dolginov *et al.* for calculating special cases were based on the use of recursion relations and the observation that Fano functions satisfied the same recursion relations. Effectively, this was an analytic continuation approach. (In the next paper, we arrive at some concrete results on analytic continuation from the finite- to the infinite-dimensional case.)

Our methods can be generalized to the calculation of CG coefficients in cases where multiplicity appears. Also, one might apply them for the couplings between principal and supplementary, supplementary and supplementary, or unitary and finite-dimensional nonunitary representations.

In the next paper, we shall discuss a set of recursion relations which suffice, after calculating one "starting" coefficient by the methods of this paper, for any set of invariant parameters $\nu_1\rho_1$, $\nu_2\rho_2$, and $\nu\rho$, to determine

any arbitrary coefficient corresponding to this set. Further, the implications of this result for the possibility of analytically continuing the CG coefficient for the coupling of finite-dimensional representations to those for the infinite-dimensional case will be discussed.

ACKNOWLEDGMENTS

We are indebted to Professors A. Salam and P. Budini and to the International Atomic Energy Agency for hospitality extended to us at the International Centre for Theoretical Physics, Trieste. We are grateful to Professor A. Salam also for suggesting the problem and for continued encouragement. We are grateful for very profitable discussions with Dr. J. Strathdee.

* Present address: Institute for Nuclear Research, Warsaw, Poland (Visiting scientist, PAN-NAS Exchange Program).

† On leave of absence from Institute for Nuclear Research, Warsaw, Poland.

‡ On leave of absence from Pakistan Atomic Energy Commission, Lahore, Pakistan.

§ On leave of absence from Institute of Nuclear Research, Řež, Czechoslovakia. Present address: Department of Physics, Carnegie–Mellon University, Pittsburgh, Pa.

[1] J. A. Smorodinsky and N. J. Vilenkin, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Soviet Phys.—JETP **19**, 1209 (1964)]; M. Sheftel, J. A. Smorodinsky, and P. Winternitz, Yad. Fiz. **7**, 1325 (1968) [Sov. J. Nucl. Phys. **7**, 785 (1968)]. See also Phys. Letters **26B**, 241 (1968).

[2] A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967); M. Toller, Nuovo Cimento **53A**, 671 (1968); **54A**, 295 (1968).

[3] J. F. Boyce, R. Delbourgo, A. Salam, and J. Strathdee, ICTP, Trieste, Preprint IC/67/9.

[4] R. Delbourgo, A. Salam, and J. Strathdee, Phys. Letters, **25B**, 230 (1967).

[5] E.g., D. T. Stoyanov and I. T. Todorov, ICTP, Trieste, Preprint IC/67/58; C. Fronsdal, Phys. Rev. **156**, 1653 (1967).

[6] C. Fronsdal and R. White, Phys. Rev. **151**, 1287 (1966).

[7] M. A. Naimark, Tr. Mosk. Mat. Obshch. **8**, 121 (1959); **9**, 237 (1960); **10**, 181 (1961) [Amer. Math. Soc. Transl., Ser. 2, **36**, 101–229 (1964)].

[8] M. A. Naimark, Lecture, in *Proceedings of the International School for Theoretical Physics* (Yalta, 1966), V. Shelest, Ed. (Kiev, 1967).

[9] A. Z. Dolginov and I. N. Toptygin, Zh. Eksp. Teor. Fiz. **35**, 794 (1958); **37**, 1441 (1959) [Soviet Phys.—JETP **8**, 550 (1959); **10**, 1022 (1960)].

[10] G. Bisiacchi and C. Fronsdal, Nuovo Cimento **41**, 35 (1965).

[11] P. G. Bamberg, Proc. Roy. Soc. (London) **300A**, 337 (1967).

[12] (a) L. D. Eskin, Izv. Vysshik Uchebn. Zavedenii, Mat. **6**, 179 (1961). (b) S. Ström, Arkiv Fysik **29**, 467 (1965); Arkiv Fysik **33**, 465 (1966). (c) A. Sciarrino and M. Toller, Ref. 2. (d) D. W. Duc and N. V. Hieu, Ann. Inst. Henri Poincaré **6**, 17 (1967). (e) I. A. Verdiev and L. A. Dadashev, Yad. Fiz. **6**, 1094 (1967) [Sov. J. Nucl. Phys. **6**, 795 (1968)].

[13] R. Rączka, Göteborg, Preprint, 1969.

[14] R. L. Anderson and R. Rączka, Institut Badan Jadrowych, Warsaw, Preprint, 1969.

[15] G. W. Mackey, Ann. Math. (N.Y.) **55**, 101 (1952).

[16] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, London, 1964).

[17] See, e.g., K. Knopp, *Theory and Application of Infinite Series* (Blackie and Son Ltd., London, 1964), 2nd English edition.

[18] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic Press, New York, 1965).

[19] G. Domokos, Phys. Rev. **159**, 1387 (1967).

Recursion and Symmetry Relations for the Clebsch–Gordan Coefficients of the Homogeneous Lorentz Group

R. L. ANDERSON,* R. RĄCZKA,† M. A. RASHID,‡ AND P. WINTERNITZ§
International Atomic Energy Agency, International Centre for Theoretical Physics Trieste, Italy

(Received 9 September 1969)

Recursion and symmetry relations are obtained for the Clebsch–Gordan coefficients associated with the coupling of two $SL(2, C)$ principal-series representations. An explicit procedure based on the recursion relations which are not restricted to the case of principal-series representations is given for generating a set of coefficients from a single "initial" coefficient. A program for the study of a possible connection, through analytic continuation, between the coupling coefficients for finite-dimensional representations and those for the infinite-dimensional case is presented.

I. INTRODUCTION

In this paper we extend our treatment of the determination of the Clebsch–Gordan (CG) coefficients associated with the coupling of two $SL(2, C)$ principal-series representations. In our previous paper[1] we derived a general expression (convergent infinite series) for these coefficients and presented a feasible method for calculating in closed form several coefficients for a given set of invariant parameters. Each set of invariant parameters corresponds to one irreducible representation in the decomposition of a tensor-product representation. Here, we shall derive a set of recursion relations (RR) for these coefficients such that given one "initial" coefficient we can generate in a simple manner all others for a given set of invariant parameters. As will be evident by the method of derivation, this result is not restricted only to the case of principal-series representations. Further, we shall use the integral representation of the CG coefficients in terms of $SL(2, C)$ transformation matrix elements in order to determine some symmetry properties of these coefficients. Finally, we shall outline a future program dealing with the possible connection through analytic continuation of the CG coefficients for the coupling of two finite-dimensional representations to those for the infinite-dimensional case.

II. REVIEW

We shall consider irreducible unitary representations of the principal series operating in the Hilbert space $\mathfrak{L}^2(z, \bar{z}) = \mathfrak{L}^2(R^2)$ of all functions $f(z, \bar{z})$, satisfying the condition

$$\int |f(z, \bar{z})|^2 \, dz \, d\bar{z} < \infty, \tag{1}$$

where $dz\, d\bar{z} = dx\, dy$. Following Naimark,[2] we shall label the representations by an integer or half-integer number ν and a real number ρ and, henceforth, suppressing $\bar{z}$ in the notation, choose a canonical basis $f^{\nu\rho}_{JM}(z)$ in $\mathfrak{L}^2(z)$, with $\nu \leq J < \infty$ and $-J \leq M \leq J$, such that the infinitesimal operators of $SL(2, C)$ operate on $f^{\nu\rho}_{JM}(z)$ in a prescribed manner:

$$H_\mu f^{\nu\rho}_{JM}(z) = [J(J+1)]^{\frac{1}{2}} \times (JM, 1\mu \mid J\, M+\mu) f^{\nu\rho}_{J\,M+\mu}(z), \tag{2}$$

$$F_\mu f^{\nu\rho}_{JM}(z) = -\sum_k R^{\nu\rho J}_{J+k}(-1)^k \times (JM, 1\mu \mid J+k\, M+\mu) f^{\nu\rho}_{J+k\, M+\mu}(z). \tag{3}$$

Here μ and k attain the values $-1, 0, 1$, and we have

$$H_{-1} = 2^{-\frac{1}{2}}(iA_1 + A_2), \quad H_0 = iA_3,$$
$$H_{+1} = -2^{-\frac{1}{2}}(iA_1 - A_2),$$
$$F_{-1} = 2^{-\frac{1}{2}}(iB_1 + B_2), \quad F_0 = iB_3,$$
$$F_{+1} = -2^{-\frac{1}{2}}(iB_1 - B_2),$$

where A_i and B_i are the rotation and pure Lorentz generators as defined in Ref. 2. The quantities $(JM,\ 1\mu \mid J+k,\ M+\mu)$ are CG coefficients of the $O(3)$ group and

$$R^{\nu\rho J}_{J-1} = \frac{1}{2}\left(\frac{J}{2J-1}\right)^{\frac{1}{2}} R^{\nu\rho J}, \quad R^{\nu\rho J}_{J} = \frac{\nu\rho}{2[J(J+1)]^{\frac{1}{2}}},$$
$$R^{\nu\rho J}_{J+1} = \frac{1}{2}\left(\frac{J+1}{2J+3}\right)^{\frac{1}{2}} R^{\nu\rho\, J+1}, \tag{4}$$
$$R^{\nu\rho J} = -\frac{2i}{J}[(J^2 - \nu^2)(J^2 + \tfrac{1}{4}\rho^2)]^{\frac{1}{2}}.$$

The normalization of the basis functions is such that

$$\int f^{\nu\rho}_{J'M'}(z)\overline{f^{\nu\rho}_{JM}(z)}\, dz = \delta_{J'J}\delta_{M'M}. \tag{5}$$

Now let us consider the tensor product of two irreducible unitary representations of the principal series $T^{\nu_1\rho_1}$ and $T^{\nu_2\rho_2}$ realized in the Hilbert spaces $\mathfrak{L}^2(z_1)$ and $\mathfrak{L}^2(z_2)$, spanned by the sets of canonical basis vectors $f^{\nu_1\rho_1}_{J_1M_1}(z_1)$ and $f^{\nu_2\rho_2}_{J_2M_2}(z_2)$, respectively.

Naimark shows[3] that the decomposition of the tensor product $T^{\nu_1\rho_1} \otimes T^{\nu_2\rho_2}$ has the following form:

$$T^{\nu_1\rho_1} \otimes T^{\nu_2\rho_2} = \sum_{\nu} \int_{-\infty}^{\infty} T^{\nu\rho}\, d\mu(\rho), \tag{6}$$

where the summation extends over all ν's such that

$$\nu + \nu_1 + \nu_2 = \text{nonnegative integer}. \tag{7}$$

For the tensor-product space $\mathfrak{L} = \mathfrak{L}_1^2 \otimes \mathfrak{L}_2^2$, we can construct two sets of orthogonal basis vectors. The first one is an orthonormal basis consisting of the set of all products of the original basis vectors, one factor from each component space:

$$f^{\nu_1\rho_1}_{J_1M_1}(z_1) f^{\nu_2\rho_2}_{J_2M_2}(z_2). \tag{8}$$

The second one is an orthogonal basis which spans the tensor-product space but lies in the space of all continuous linear functions[3]

$$f^{\nu\rho,\nu_1\rho_1,\nu_2\rho_2}_{JM}(z_1, z_2), \tag{9}$$

which we normalize in the following manner:

$$\int f^{\nu'\rho'\nu_1\rho_1\nu_2\rho_2}_{J'M'}(z_1, z_2) \overline{f^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{JM}}(z_1, z_2)\, dz_1\, dz_2 = \frac{\delta(\rho' - \rho)}{4\nu^2 + \rho^2} \delta_{\nu'\nu}\delta_{J'J}\delta_{M'M}, \tag{10}$$

where we introduce the Plancherel measure $4\nu^2 + \rho^2$.

The two sets of basis vectors (8) and (9) are related by the formula

$$f^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{JM}(z_1, z_2) = \sum_{J_1=\nu_1}^{\infty} \sum_{J_2=\nu_2}^{\infty} \sum_{M_1=-J_1}^{J_1} \sum_{M_2=-J_2}^{J_2} \times \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM\rangle \times f^{\nu_1\rho_1}_{J_1M_1}(z_1) f^{\nu_2\rho_2}_{J_2M_2}(z_2), \tag{11}$$

which serves as the definition of the CG coefficients for the homogeneous Lorentz group.

A calculation of the norm of each side of (11) [using (5) and (10)] yields the orthogonality relations for the CG coefficients[1]

$$\sum_{\substack{J_1J_2\\M_1M_2}} \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM\rangle \times \overline{\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu'\rho' J' M'\rangle} = \frac{\delta(\rho - \rho')}{4\nu^2 + \rho^2} \delta_{\nu\nu'}\delta_{JJ'}\delta_{MM'}. \tag{12}$$

The inverse formula to (11) is

$$f^{\nu_1\rho_1}_{J_1M_1}(z_1) f^{\nu_2\rho_2}_{J_2M_2}(z_2) = \sum_{\nu} \int_{-\infty}^{\infty} (\rho^2 + 4\nu^2)\, d\rho \times \sum_{JM} \overline{\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM\rangle} \times f^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{JM}(z_1, z_2), \tag{13}$$

which can easily be verified by substituting (13) into the right-hand side of (11) and applying (12). Conversely, a substitution of (11) into the right-hand side of (13) yields the completeness relation for the CG coefficients[1]:

$$\sum_{\nu} \int_{-\infty}^{\infty} (\rho^2 + 4\nu^2)\, d\rho \sum_{JM} \langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM\rangle \times \overline{\langle \nu_1\rho_1 J_1' M_1', \nu_2\rho_2 J_2' M_2' \mid \nu\rho JM\rangle} = \delta_{J_1J_1'}\delta_{J_2J_2'}\delta_{M_1M_1'}\delta_{M_2M_2'}. \tag{14}$$

In (13) and (14) the summation is over all ν satisfying (7).

III. RECURSION RELATIONS FOR THE CG COEFFICIENTS

A. Derivation of Recursion Relations

The action of the generators H_μ and F_μ of the $SL(2, C)$ group on an arbitrary canonical basis vector is given by (2) and (3), and these can be used to obtain recursion relations (RR) for the CG coefficients. The procedure is to apply H_μ and F_μ separately to Eq. (13), then to expand the left-hand side of the result in terms of the irreducible basis by using (13) again. The RR can now be read off by equating the coefficients of the orthonormal basis functions. This procedure applied to the generators H_μ of the $SU(2)$ subgroup gives the recursion relation for the CG coefficients of $SU(2)$. This proves that the CG coefficient of the $SL(2, C)$ group can be factorized as

$$\langle \nu_1\rho_1 J_1 M_1, \nu_2\rho_2 J_2 M_2 \mid \nu\rho JM\rangle = (J_1M_1, J_2M_2 \mid JM) X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J), \tag{15}$$

where $(J_1M_1, J_2M_2 \mid JM)$ is an $SU(2)$ CG coefficient.

Applying the above procedure to the generators F_μ, we obtain an RR for $\bar{X}\ (\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J)$ and, taking the complex conjugate of all terms, we obtain

$$\begin{aligned}
&\sum_{k=-1,0,1} R_J^{\nu\rho\, J+k}(J + k\, M, 1\mu \mid J\, M + \mu) \\
&\quad \times (J_1M_1, J_2M_2 \mid J + k\, M) \\
&\quad \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J + k) \\
&= \sum_{k=-1,0,1} [R^{\nu_1\rho_1 J_1}_{J_1+k}(J_1M_1, 1\mu \mid J_1 + k\, M_1 + \mu) \\
&\quad \times (J_1 + k\, M_1 + \mu, J_2M_2 \mid J\, M + \mu) \\
&\quad \times X(\nu_1\rho_1\, J_1 + k, \nu_2\rho_2 J_2, \nu\rho J) \\
&\quad + R^{\nu_2\rho_2 J_2}_{J_2+k}(J_2M_2, 1\mu \mid J_2 + k\, M_2 + \mu) \\
&\quad \times (J_1M_1, J_2 + k\, M_2 + \mu \mid J\, M + \mu) \\
&\quad \times X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 + k, \nu\rho J)],
\end{aligned} \tag{16}$$

where the coefficients $R^{\nu\rho J}_{J+K}$ are defined by (4).

In the above equation the variable μ takes the values $-1, 0, 1$. In order to derive recursion relations which can be used practically, we first eliminate the M-dependence. We multiply both sides of (16) by $(J_1M_1, J_2M_2 \mid J + l\,M)$ and sum over M_1, M_2, keeping M fixed. Cancelling the factor

$$(J + l\,M, 1\mu \mid J\,M + \mu)$$

we arrive at

$$\begin{aligned} R_J^{\nu\rho\,J+l} X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J + l) \\ = \sum_{k=-1,0,1} [(-1)^{k+l} R_{J_1+k}^{\nu_1\rho_1 J_1} U(1J_1 J J_2; J_1 + k\,J + l) \\ \times X(\nu_1\rho_1\, J_1 + k, \nu_2\rho_2 J_2, \nu\rho J) \\ + R_{J_2+k}^{\nu_2\rho_2 J_2} U(1J_2 J J_1; J_2 + k\,J + l) \\ \times X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 + k, \nu\rho J)], \end{aligned} \tag{17a}$$

with $l = -1, 0, 1$ and

$$U(abcd; ef) = [(2e + 1)(2f + 1)]^{\frac{1}{2}} W(abcd; ef),$$

where the W's are the Racah coefficients.

In exactly the same fashion we can also derive

$$\begin{aligned} R_{J_1+l}^{\nu_1\rho_1 J_1} X(\nu_1\rho_1\, J_1 + l, \nu_2\rho_2 J_2, \nu\rho J) \\ = \sum_{k=-1,0,1} [(-1)^{k+l} R_J^{\nu\rho\, J+k} U(1JJ_1J_2; J + k\,J_1 + l) \\ \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J + k) \\ + (-1)^k R_{J_2+k}^{\nu_2\rho_2 J_2} U(1J_2J_1J; J_2 + k\,J_1 + l) \\ \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2 + k, \nu\rho J)], \end{aligned} \tag{17b}$$

and

$$\begin{aligned} R_{J_2+l}^{\nu_2\rho_2 J_2} X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 + l, \nu\rho J) \\ = \sum_{k=-1,0,1} [(-1)^l R_{J_1+k}^{\nu_1\rho_1 J_1} U(1J_1J_2J; J_1 + k\,J_2 + l) \\ \times X(\nu_1\rho_1\, J_1 + k, \nu_2\rho_2 J_2, \nu\rho J) \\ + R_J^{\nu\rho\, J+k} U(1JJ_2J_1; J + k\,J_2 + l) \\ \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J + k)]. \end{aligned} \tag{17c}$$

The U's appearing in Eqs. (17) are known in closed form.[4] Recursion relations of type (17a) for the special case $\nu_1 = \nu_2 = \nu = 0$ were already derived in Ref. 5.

Putting $l = 0$ in (17a) and bearing in mind that $U(abcd; ef) = 0$ unless (abe), (cde), (acf), and (bdf) satisfy triangular relations, we obtain

$$X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J) = 0, \quad (J_1, J_2, J) \text{ nontriangular.} \tag{18}$$

The recursion relations (17) are in a simple and elegant form which is useful for studying general properties of the CG coefficients; in particular, we plan to use them for a study of the analytic structure. However, they involve increases in at least two of the J's simultaneously, so that they are not convenient for practical applications.

B. Simplification and Discussion of the Recursion Relations

We shall now proceed to obtain three RR's in which only J, J_1, or J_2, respectively, increases. For instance, let us eliminate $J_1 + 1$ and $J_2 + 1$. To do this, take (17b) for $l = 0$ and $l = -1$. These two equations do not contain $J_1 + 1$ and using them together we can eliminate $J_2 + 1$. Substituting the explicit values of the Racah coefficients (e.g., using Ref. 4) into the equation obtained, we obtain after simple but tedious algebraic calculations

$$\begin{aligned} &\frac{R^{\nu\rho J+1}}{2J+1} [(J + J_1 + J_2 + 2)(J + J_1 + J_2 + 1)(J + J_1 - J_2 + 1)(J - J_1 + J_2 + 1)]^{\frac{1}{2}} \\ &\quad \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J + 1) \\ &= [(J + J_1 + J_2 + 1)(-J + J_1 + J_2)]^{\frac{1}{2}} \left(\frac{\nu\rho(J_1 - J_2)}{J(J+1)} + \frac{\nu_2\rho_2}{J_2} - \frac{\nu_1\rho_1}{J_1} \right) X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J) \\ &\quad + \frac{R^{\nu\rho J}}{2J+1} [(J + J_1 - J_2)(J - J_1 + J_2)(-J + J_1 + J_2 + 1)(-J + J_1 + J_2)]^{\frac{1}{2}} X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J - 1) \\ &\quad + R^{\nu_1\rho_1 J_1} \left(\frac{(2J_1 + 1)(J - J_1 + J_2 + 1)(J + J_1 - J_2)}{2J_1 - 1} \right)^{\frac{1}{2}} X(\nu_1\rho_1\, J_1 - 1, \nu_2\rho_2 J_2, \nu\rho J) \\ &\quad + R^{\nu_2\rho_2 J_2} \left(\frac{(2J_2 + 1)(J + J_1 - J_2 + 1)(J - J_1 + J_2)}{2J_2 - 1} \right)^{\frac{1}{2}} X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 - 1, \nu\rho J). \end{aligned} \tag{19a}$$

Similarly, using (17c) for $l = 0$ and $l = -1$, we eliminate the function $X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J + 1)$ and obtain

$$R^{\nu_1\rho_1 J_1+1}\left(\frac{(J + J_1 + J_2 + 2)(J + J_1 + J_2 + 1)(-J + J_1 + J_2 + 1)(J + J_1 - J_2 + 1)}{(2J_1 + 1)(2J_1 + 3)}\right)^{\frac{1}{2}}$$
$$\times X(\nu_1\rho_1 J_1 + 1, \nu_2\rho_2 J_2, \nu\rho J)$$
$$= -[(J + J_1 + J_2 + 1)(J - J_1 + J_2)]^{\frac{1}{2}}\left(\frac{\nu\rho}{J} - \frac{\nu_1\rho_1(J - J_2)}{J_1(J_1 + 1)} - \frac{\nu_2\rho_2}{J_2}\right)X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J)$$
$$+ R^{\nu\rho J}[(J + J_1 - J_2)(-J + J_1 + J_2 + 1)]^{\frac{1}{2}}X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J - 1)$$
$$+ R^{\nu_1\rho_1 J_1}\left(\frac{(J + J_1 - J_2)(J - J_1 + J_2 + 1)(J - J_1 + J_2)(-J + J_1 + J_2)}{(2J_1 + 1)(2J_1 - 1)}\right)^{\frac{1}{2}} X(\nu_1\rho_1\, J_1 - 1, \nu_2\rho_2 J_2, \nu\rho J)$$
$$- R^{\nu_2\rho_2 J_2}\left(\frac{(2J_2 + 1)(J + J_1 - J_2 + 1)(-J + J_1 + J_2)}{2J_2 - 1}\right)^{\frac{1}{2}}X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 - 1, \nu\rho J) \tag{19b}$$

and, using (17a), we eliminate the function $X(\nu_1\rho_1 J_1 + 1, \nu_2\rho_2 J_2, \nu\rho J)$ to obtain

$$R^{\nu_2\rho_2 J_2+1}\left(\frac{(J + J_1 + J_2 + 2)(J + J_1 + J_2 + 1)(J - J_1 + J_2 + 1)(-J + J_1 + J_2 + 1)}{(2J_2 + 1)(2J_2 + 3)}\right)^{\frac{1}{2}}$$
$$\times X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 + 1, \nu\rho J)$$
$$= [(J + J_1 + J_2 + 1)(J + J_1 - J_2)]^{\frac{1}{2}}\left(\frac{\nu\rho}{J} - \frac{\nu_1\rho_1}{J_1} - \frac{(J - J_1)\nu_2\rho_2}{J_2(J_2 + 1)}\right)X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J)$$
$$+ R^{\nu\rho J}[(J - J_1 + J_2)(-J + J_1 + J_2 + 1)]^{\frac{1}{2}}X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho\, J - 1)$$
$$- R^{\nu_1\rho_1 J_1}\left(\frac{(2J_1 + 1)(J - J_1 + J_2 + 1)(-J + J_1 + J_2)}{2J_1 - 1}\right)^{\frac{1}{2}}X(\nu_1\rho_1\, J_1 - 1, \nu_2\rho_2 J_2, \nu\rho J)$$
$$+ R^{\nu_2\rho_2 J_2}\left(\frac{(J + J_1 - J_2 + 1)(J + J_1 - J_2)(J - J_1 + J_2)(-J + J_1 + J_2)}{(2J_2 + 1)(2J_2 - 1)}\right)^{\frac{1}{2}}X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_2 - 1, \nu\rho J). \tag{19c}$$

Equation (19a) is valid whenever the triad $(J_1 J_2 J)$ forms a triangle. However, it is useful only when $(J_1 J_2\, J + 1)$ also forms a triangle, since otherwise both sides of the equation vanish identically. If it is applicable, one can utilize it for calculating an X with larger J in terms of those with smaller J's. We can apply the same consideration to Eqs. (19b) and (19c). The only case when none of the equations (19) will work is if $(J_1 J_2 J)$ forms a triangle but none of the triads $(J_1 - 1\, J_2 J)$, $(J_1\, J_2 - 1\, J)$, or $(J_1 J_2\, J - 1)$ do. It is easy to check that this is only possible if the equations

$$J_1 = |J_2 - J|, \quad J_2 = |J_1 - J|, \quad J = |J_1 - J_2|, \tag{20}$$

are satisfied simultaneously. This is only possible if one of the J's is zero and the other two are equal to each other.

Thus, the only CG coefficients which we are not able to express in terms of those with lower values of the J's are coefficients of the type

$$X(\nu_1\rho_1 J, \nu_2\rho_2 J, 0\rho 0). \tag{21}$$

Fortunately, however, we can find other RR's which involve just these coefficients.

Let us return to Eq. (17a) and consider it for $l = 1$, $J = 0$, and $\nu = 0$ [we have not so far utilized Eqs. (17) for $l = 1$]. For $J_2 = J_1$, since $U(10JJ; 1J) = 1$, this immediately leads to

$$R^{0\rho 1}X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_1, 0\rho 1)$$
$$= [J_1(J_1 + 1)]^{-\frac{1}{2}}(\nu_2\rho_2 - \nu_1\rho_1)X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_1, 0\rho 0) \tag{22}$$

and, since

$$U(1J0\, J \pm 1; J \pm 1\ 1) = U(1\, J \pm 1\ 0J; J1) = 1,$$

we obtain for $J_2 = J_1 + 1$

$$R^{0\rho 1}X(\nu_1\rho_1 J_1, \nu_2\rho_2\, J_1 + 1, 0\rho 1)$$
$$= (J_1 + 1)^{\frac{1}{2}}\left(\frac{R^{\nu_1\rho_1 J_1+1}}{(2J_1 + 3)^{\frac{1}{2}}}\right.$$
$$\times X(\nu_1\rho_1\, J_1 + 1, \nu_2\rho_2\, J_1 + 1, 0\rho 0)$$
$$\left. + \frac{R^{\nu_2\rho_2 J_1+1}}{(2J_1 + 1)^{\frac{1}{2}}}X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_1, 0\rho 0)\right) \tag{23}$$

and for $J_2 = J_1 - 1$

$$R^{0\rho 1}X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_1 - 1, 0\rho 1) = J_1^{\frac{1}{2}}\left(\frac{R^{\nu_2\rho_2 J_1}}{(2J_1+1)^{\frac{1}{2}}}X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_1, 0\rho 0) + \frac{R^{\nu_1\rho_1 J_1}}{(2J_1-1)^{\frac{1}{2}}}X(\nu_1\rho_1 J_1 - 1, \nu_2\rho_2 J_1 - 1, 0\rho 0)\right). \quad (24)$$

Substituting Eqs. (22)–(24) in (19c) with $\nu = 0$, $J_1 = J_2$, $J = 1$ and simplifying, we find

$$\frac{2(J_1+1)}{[(2J_1+1)(2J_1+3)]^{\frac{1}{2}}}R^{\nu_1\rho_1 J_1+1}R^{\nu_2\rho_2 J_1+1} \times X(\nu_1\rho_1 J_1 + 1, \nu_2\rho_2 J_1 + 1, 0\rho 0) = \left(8J_1^2 + 8J_1 + 4 - 4\nu_1^2 - 4\nu_2^2 - \rho^2 + \rho_1^2 + \rho_2^2 - 2\frac{\nu_1\rho_1\nu_2\rho_2}{J_1(J_1+1)}\right) \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_1, 0\rho 0) - \frac{2J_1}{[(2J_1-1)(2J_1+1)]^{\frac{1}{2}}}R^{\nu_1\rho_1 J_1}R^{\nu_2\rho_2 J_1} \times X(\nu_1\rho_1 J_1 - 1, \nu_2\rho_2 J_1 - 1, 0\rho 0). \quad (25a)$$

A special case of this relation, namely, with $\nu_1 = \nu_2 = \nu = 0$, was given in Ref. 6.

For completeness, we also quote below two other relations of the above form:

$$\frac{2(J+1)}{2J+1}R^{\nu_2\rho_2 J+1}R^{\nu\rho J+1} \times X(0\rho_1 0, \nu_2\rho_2 J + 1, \nu\rho J + 1) = -\left(8J^2 + 8J + 4 - 4\nu^2 - 4\nu_2^2 - \rho_1^2 + \rho_2^2 + \rho^2 - 2\frac{\nu\rho\nu_2\rho_2}{J(J+1)}\right) \times X(0\rho 0, \nu_2\rho_2 J, \nu\rho J) - \frac{2J}{2J+1}R^{\nu_2\rho_2 J}R^{\nu\rho J} \times X(0\rho 0, \nu_2\rho_2 J - 1, \nu\rho J - 1); \quad (25b)$$

$$\frac{2(J+1)}{2J+1}R^{\nu_1\rho_1 J+1}R^{\nu\rho J+1} \times X(\nu_1\rho_1 J + 1, 0\rho_2 0, \nu\rho J + 1) = -\left(8J^2 + 8J + 4 - 4\nu^2 - 4\nu_1^2 + \rho_1^2 - \rho_2^2 + \rho^2 - 2\frac{\nu_1\rho_1\nu\rho}{J(J+1)}\right) \times X(\nu_1\rho_1 J, 0\rho_2 0, \nu\rho J) - \frac{2J}{2J+1}R^{\nu_1\rho_1 J}R^{\nu\rho J} \times X(\nu_1\rho_1 J - 1, 0\rho_2 0, \nu\rho J - 1). \quad (25c)$$

The procedure for calculating arbitrary CG coefficients from Eqs. (19) and (25) is now obvious. In view of (15), we can restrict our attention to the X functions. Starting from a set of coefficients with given $(\nu_1\rho_1, \nu_2\rho_2, \nu\rho)$, all of which have the same $J_1 + J_2 + J$, we utilize Eqs. (19) to raise $J_1 + J_2 + J$ by alternately increasing J, J_1, and J_2, using Eqs. (25) whenever necessary.

Let us introduce the concept of "minimal" CG coefficients. For a given set of invariant numbers $(\nu_1\rho_1, \nu_2\rho_2, \nu\rho)$, we can calculate an arbitrary CG coefficient in terms of the minimal ones, using only Eqs. (19) and (25). We shall now show that the "minimal" coefficients can always be expressed in terms of a single "initial" CG coefficient.

Let us make the following remark: The RR's which we have derived do not change if we change the sign of ν and ρ, simultaneously. This, together with the fact that an arbitrary CG coefficient can, as will be demonstrated below, be expressed in terms of a single "initial" one, shows that the coefficients $X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J)$ and $X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, -\nu - \rho J)$ can differ, at most, by a factor, independent of the J's. It then follows from (11) that

$$f^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{JM}(z_1, z_2) \quad \text{and} \quad f^{-\nu-\rho\nu_1\rho_1\nu_2\rho_2}_{JM}(z_1, z_2)$$

can differ, if at all, only by a phase factor. This, in combination with the results of Naimark[3] and Mackey,[7] shows that, at least when considering the reduction of the product of two $SL(2, C)$ representations of the principal series, each irreducible representation appearing in the decomposition of the tensor-product representations appears only once.

C. The "Minimal" CG Coefficients

Triangular Case

If the numbers ν_1, ν_2, and ν satisfy a triangular relation, it follows directly from Eqs. (19a) and (25a) and from the fact that $X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J) = 0$ for $J < \nu$, $J_1 < \nu_1$ or $J_2 < \nu_2$, that all CG coefficients can be expressed in terms of a single minimal one, the "initial" coefficient

$$X_{\text{initial}} = X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\nu). \quad (26)$$

Nontriangular Case

In the nontriangular case, we have to specify a whole set of minimal coefficients and we shall express them directly in terms of a single "initial" one.

Let us consider separate cases.

(i) $\nu \leq \nu_2 < \nu_1$, $\nu_1 > \nu + \nu_2$: In this case we have $X(J_1, J_2, J) = 0$ if $J_1 + J_2 + J < 2\nu_1$ (for brevity

we shall sometimes drop the ν's and ρ's in the CG coefficient). The minimal coefficients in this case are

$$X(\nu_1, \nu_2 + k, \nu_1 - \nu_2 - k), \quad 0 \leq k \leq \nu_1 - \nu_2 - \nu, \tag{27}$$

with $J_1 + J_2 + J = 2\nu_1$; it is easy to see that (19) and (25) will now give any $X(J_1, J_2, J)$ with $J_1 + J_2 + J > 2\nu_1$ in terms of (27).

Now let us find a separate RR for the minimal coefficients. Take (17b) for $l = -1$ and $X(J_1, J_2, J-1)$ on the left-hand side where $X(J_1, J_2, J)$ is a minimal coefficient. Remembering that these coefficients are equal to zero for $J_1 + J_2 + J < 2\nu_1$, we obtain

$$\frac{R^{\nu\rho J}}{2J-1}[(J_1 - J_2 + J)(J_1 - J_2 + J - 1)]^{\frac{1}{2}} X(J_1, J_2, J)$$
$$= R^{\nu_2\rho_2 J_2 + 1}\left(\frac{(J_1 + J_2 - J + 2)(J_1 + J_2 - J + 1)}{(2J_2 + 1)(2J_2 + 3)}\right)^{\frac{1}{2}}$$
$$\times X(J_1, J_2 + 1, J - 1). \tag{28}$$

This RR can be solved directly, and we obtain all the minimal coefficients in terms of the initial one:

$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2 + k, \nu\rho\,\nu_1 - \nu_2 - k)$$
$$= \left[\prod_{n=0}^{k-1} \frac{R^{\nu\rho\,\nu_1-\nu_2-n}}{R^{\nu_2\rho_2\nu_2+1+n}}\left(\frac{(\nu_2 + \frac{3}{2} + n)(\nu_1 - \nu_2 - n)}{(\nu_1 - \nu_2 - n - \frac{1}{2})(\nu_2 + n + 1)}\right)^{\frac{1}{2}}\right]$$
$$\times X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\,\nu_1 - \nu_2),$$
$$\text{for} \quad 0 \leq k \leq \nu_1 - \nu_2 - \nu. \tag{29}$$

A completely analogous precedure can be applied in all other cases and we shall just give the results, expressing the minimal coefficients in terms of the corresponding initial one. Thus,

(ii) $\nu_2 \leq \nu < \nu_1$, $\nu_1 > \nu_2 + \nu$:

$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\,\nu_1 - \nu - k, \nu\rho\,\nu + k)$$
$$= \left[\prod_{n=0}^{k-1} \frac{R^{\nu_2\rho_2\,\nu_1-\nu-n}}{R^{\nu\rho\,\nu+n+1}}\left(\frac{(\nu_1 - \nu - n)(\nu + \frac{1}{2} + n)}{(\nu_1 - \nu - n + \frac{1}{2})(\nu + n + 1)}\right)^{\frac{1}{2}}\right]$$
$$\times X(\nu_1\rho_1\nu_1, \nu_2\rho_2\,\nu_1 - \nu, \nu\rho\nu),$$
$$\text{for} \quad 0 \leq k \leq \nu_1 - \nu_2 - \nu; \tag{30}$$

(iii) $\nu \leq \nu_1 < \nu_2$, $\nu_2 > \nu_1 + \nu$:

$$X(\nu_1\rho_1\,\nu_1 + k, \nu_2\rho_2\nu_2, \nu\rho\,\nu_2 - \nu_1 - k)$$
$$= \left[\prod_{n=0}^{k-1} \frac{R^{\nu\rho\,\nu_2-\nu_1-n}}{R^{\nu_1\rho_1\nu_1+n+1}}\left(\frac{(\nu_2 - \nu_1 - n)(\nu_1 + n + \frac{3}{2})}{(\nu_2 - \nu_1 - n - \frac{1}{2})(\nu_1 + n + 1)}\right)^{\frac{1}{2}}\right]$$
$$\times X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\,\nu_2 - \nu_1),$$
$$\text{for} \quad 0 \leq k \leq -\nu_1 + \nu_2 - \nu; \tag{31}$$

(iv) $\nu_1 \leq \nu < \nu_2$, $\nu_2 > \nu_1 + \nu$:

$$X(\nu_1\rho_1\,\nu_2 - \nu - k, \nu_1\rho_1\nu_2, \nu\rho\,\nu + k)$$
$$= \left[\prod_{n=0}^{k-1} \frac{R^{\nu_1\rho_1\,\nu_2-\nu-n}}{R^{\nu\rho\,\nu+n+1}}\left(\frac{(\nu_2 - \nu - n)(\nu + n + \frac{1}{2})}{(\nu_2 - \nu - n + \frac{1}{2})(\nu + n + 1)}\right)^{\frac{1}{2}}\right]$$
$$\times X(\nu_1\rho_1\,\nu_2 - \nu, \nu_2\rho_2\nu_2, \nu\rho\nu),$$
$$\text{for} \quad 0 \leq k \leq -\nu_1 + \nu_2 - \nu; \tag{32}$$

(v) $\nu_1 \leq \nu_2 < \nu$, $\nu > \nu_1 + \nu_2$:

$$X(\nu_1\rho_1\,\nu - \nu_2 - k, \nu_2\rho_2\,\nu_2 + k, \nu\rho\nu)$$
$$= (-1)^k\left[\prod_{n=0}^{k-1} \frac{R^{\nu_1\rho_1\,\nu-\nu_2-n}}{R^{\nu_2\rho_2\,\nu_2+n+1}}\right.$$
$$\times \left.\left(\frac{(\nu - \nu_2 - n)(\nu_2 + n + \frac{3}{2})}{(\nu - \nu_2 - n + \frac{1}{2})(\nu_2 + n + 1)}\right)^{\frac{1}{2}}\right]$$
$$\times X(\nu_1\rho_1\,\nu - \nu_2, \nu_2\rho_2\nu_2, \nu\rho\nu),$$
$$\text{for} \quad 0 \leq k \leq -\nu_1 - \nu_2 + \nu; \tag{33}$$

(vi) $\nu_2 \leq \nu_1 < \nu$, $\nu > \nu_1 + \nu_2$:

$$X(\nu_1\rho_1\,\nu_1 + k, \nu_2\rho_2\,\nu - \nu_1 - k, \nu\rho\nu)$$
$$= (-1)^k\left[\prod_{n=0}^{k-1} \frac{R^{\nu_2\rho_2\,\nu-\nu_1-n}}{R^{\nu_1\rho_1\,\nu_1+n+1}}\right.$$
$$\times \left.\left(\frac{(\nu_1 + n + \frac{3}{2})(\nu - \nu_1 - n)}{(\nu_1 + n + 1)(\nu - \nu_1 - n + \frac{1}{2})}\right)^{\frac{1}{2}}\right]$$
$$\times X(\nu_1\rho_1\nu_1, \nu_2\rho_2\,\nu - \nu_1, \nu\rho\nu),$$
$$\text{for} \quad 0 \leq k \leq -\nu_1 - \nu_2 + \nu. \tag{34}$$

IV. THE PHASES OF THE CG COEFFICIENTS

The basis vectors in the product space are chosen so that they constitute a canonical basis, i.e., they satisfy (2) and (3) and they are normalized according to (10). Thus, the over-all phase for each set $(\nu_1\rho_1\nu_2\rho_2\nu\rho)$ can still be chosen arbitrarily. We shall do this in such a manner that all the CG coefficients are either real or pure imaginary (see Sec. VI).

The Triangular Case

If ν_1, ν_2, and ν satisfy a triangular relation we choose the initial CG coefficient to be real and positive:

$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\nu) > 0. \tag{35}$$

The Nontriangular Case

Let us again consider the separate cases:

(i) $\nu_1 > \max(\nu, \nu_2)$: We choose both possible initial coefficients to be real and positive, i.e.,

$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\,\nu_1 - \nu, \nu\rho\nu) > 0,$$
$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\,\nu_1 - \nu_2) > 0. \tag{36}$$

(ii) $\nu_2 > \max(\nu, \nu_1)$: Analogously, we choose

$$X(\nu_1\rho_1\,\nu_2 - \nu, \nu_2\rho_2\nu_2, \nu\rho\nu) > 0,$$
$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\,\nu_2 - \nu_1) > 0. \quad (37)$$

(iii) $\nu > \max(\nu_1, \nu_2)$: In this case, only one of the initial CG coefficients can, in general, be chosen to be positive. We choose

$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\,\nu - \nu_1, \nu\rho\nu) > 0. \quad (38)$$

Equations (29)–(34) give relations between the phases of various minimal coefficients, and the choices were made consistently. This is best checked by considering the case when the two nonmaximal ν's are equal. Equations (29)–(34) also determine the phases of all the minimal coefficients, namely:

(a) all the minimal coefficients are real;

(b) if we have $\nu_1 > \max(\nu_2, \nu)$ or $\nu_2 > \max(\nu_1, \nu)$, all the minimal CG coefficients are positive;

(c) if $\nu > \max(\nu_1\nu_2)$ and we postulate (38), then the other "initial" coefficient satisfies

$$\text{phase}\,\{X(\nu_1\rho_1\,\nu - \nu_2, \nu_2\rho_2\nu_2, \nu\rho\nu)\} = (-1)^{\nu-\nu_1-\nu_2}, \quad (39)$$

and the minimal coefficients satisfy

$$\text{phase}\,\{X(\nu_1\rho_1\,\nu - \nu_2 - k, \nu_2\rho_2\,\nu_2 + k, \nu\rho\nu)\} = (-1)^{\nu-\nu_1-\nu_2+k},$$
$$\text{phase}\,\{X(\nu_1\rho_1\,\nu_1 + k, \nu_2\rho_2\,\nu - \nu_1 - k, \nu\rho\nu)\} = (-1)^k. \quad (40)$$

V. REVIEW OF THE RELATION BETWEEN THE CG COEFFICIENTS AND THE $O(3, 1)$ TRANSFORMATION MATRIX ELEMENTS

Here we shall review some relations between the CG coefficients and the D functions which are essential for our purposes.[1]

The transformation matrix elements for finite Lorentz transformations in an angular momentum basis $D^{\nu\rho}_{JM\,J'M'}(g)$ have recently been given by a number of authors.[8] For our purposes, the most convenient one is essentially that of Duc and Hieu except that we have adopted the canonical angular momentum basis as defined by Naimark. For this choice, we have the following symmetry properties:

$$D^{-\nu-\rho}_{J_1J_2\lambda}(a) = D^{-\nu\rho}_{J_1J_2-\lambda}(a) = \overline{D^{\nu-\rho}_{J_1J_2\lambda}(a)} = D^{\nu\rho}_{J_1J_2\lambda}(a). \quad (41)$$

Note that these properties are simpler than those given in Refs. 8c and 9 because our D matrices contain an additional phase factor such that they transform one canonical (Naimark) basis into another.

Now consider formula (13) and let the Lorentz transformation operator T_g act on both sides. We obtain

$$\sum_{J_1'M_1'}\sum_{J_2'M_2'} D^{\nu_1\rho_1}_{J'M_1',J_1M_1}(g)D^{\nu_2\rho_2}_{J_2'M_2',J_2M_2}(g) \times f^{\nu_1\rho_1}_{J_1'M_1'}(z_1)f^{\nu_2\rho_2}_{J_2'M_2'}(z_2)$$
$$= \sum_\nu \int_{-\infty}^{\infty}(\rho^2 + 4\nu^2)\,d\rho \times \sum_{JM}\overline{\langle\nu_1\rho_1J_1M_1, \nu_2\rho_2J_2M_2 \mid \nu\rho JM\rangle} \times \sum_{J'M'} D^{\nu\rho}_{J'M',JM}(g)f^{\nu\rho\nu_1\rho_1\nu_2\rho_2}_{J'M'}(z_1, z_2). \quad (42)$$

Substituting (11) into the right-hand side and comparing the coefficients multiplying independent basis vectors, we obtain

$$D^{\nu_1\rho_1}_{J_1'M_1',J_1M_1}(g)D^{\nu_2\rho_2}_{J_2',M_2',J_2M_2}(g) = \sum_\nu\int_{-\infty}^{\infty}(\rho^2 + 4\nu^2)\,d\rho \times \sum_{JJ'}\overline{\langle\nu_1\rho_1J_1M_1, \nu_2\rho_2J_2M_2 \mid \nu\rho J\,M_1 + M_2\rangle} \times \langle\nu_1\rho_1J_1'M_1', \nu_2\rho_2J_2'M_2' \mid \nu\rho J'\,M_1' + M_2'\rangle \times D^{\nu\rho}_{J'\,M_1'+M_2',J\,M_1+M_2}(g). \quad (43)$$

Analogously, acting with T_g on (11) and again comparing the coefficients multiplying independent basis vectors, we obtain the inverse formula to (43)

$$D^{\nu\rho}_{J'M',JM}(g) = \sum_\nu\int_{-\infty}^{\infty}(4\nu^2 + \rho^2)\,d\rho \times \sum_{J_1J_2M_1M_2}\sum_{J_1'J_2'M_1'M_2'}\langle\nu_1\rho_1J_1M_1, \nu_2\rho_2J_2M_2 \mid \nu\rho JM\rangle \times \overline{\langle\nu_1\rho_1J_1'M_1', \nu_2\rho_2J_2'M_2' \mid \nu\rho J'M'\rangle} \times D^{\nu_1\rho_1}_{J_1'M_1',J_1M_1}(g)D^{\nu_2\rho_2}_{J_2'M_2',J_2M_2}(g). \quad (43')$$

Now multiplying (43) by $D^{\nu'\rho'}_{\tilde{J}'\tilde{M}',\tilde{J}\tilde{M}}(g)$ and integrating over the group manifold, we obtain our previous result[1]

$$\langle\nu_1\rho_1J_1M_1, \nu_2\rho_2J_2M_2 \mid \nu\rho JM\rangle \times \overline{\langle\nu_1\rho_1J_1'M_1', \nu_2\rho_2J_2'M_2' \mid \nu\rho J'M'\rangle} = \int dg\,\overline{D^{\nu\rho}_{JM,J'M'}(g)}D^{\nu_1\rho_1}_{J_1M_1,J_1'M_1'}(g)D^{\nu_2\rho_2}_{J_2M_2,J_2'M_2'}(g). \quad (44)$$

VI. SYMMETRY RELATIONS FOR THE CG COEFFICIENTS

A. The General Procedure

In this section we shall give some basic symmetry relations for the CG coefficients which follow directly

from the group theory. In view of formula (15), we can restrict ourselves to the "reduced" CG coefficients, i.e., to the functions $X(\nu_1\rho_1 J_1, \nu_2\rho_2 J, \nu\rho J)$. The relations we shall give are those relating

$$\overline{X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J)}, \quad X(\nu_2\rho_2 J_2, \nu_1\rho_1 J_1, \nu\rho J),$$
$$X(\nu\rho J, \nu_2\rho_2 J_2, \nu_1\rho_1 J_1) \quad \text{and} \quad X(\nu_1\rho_1 J_1, \nu\rho J, \nu_2\rho_2 J_2)$$

to the original CG coefficient. Other symmetries should, of course, also exist; in particular, more complicated ones of the type found by Regge[10] for the $O(3)$ group. However, no attempt to establish them is made in this paper.

The straightforward procedure that suggests itself to prove a symmetry relation between a given X and, say, an $\tilde{X}$ consists of the following steps:

(i) We find a substitution for X such that the RR's involving X go over into those involving $\tilde{X}$;

(ii) We establish the over-all phase of $\tilde{X}$ depending on the group invariants $\nu_1\rho_1, \nu_2\rho_2, \nu\rho$ so that it agrees with the conventions of Sec. IV;

(iii) The RR's only determine the CG coefficients and, hence, also their symmetries up to a factor depending on the group invariants. We determine this factor in the symmetry relations using formula (44).

It follows from these remarks and from the phase conventions that the symmetry relations can all be written in the form

$$X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J) = (-1)^{\alpha_1 J_1+\alpha_2 J_2+\alpha_3 J_3+\beta_1\nu_1+\beta_2\nu_2+\beta\nu} \times g(\nu_1\rho_1, \nu_2\rho_2, \nu\rho) f(J_1 J_2 J)\tilde{X}, \quad (45)$$

where the α's and β's are real constants and g and f are real and positive functions. In agreement with the above procedure, we obtain α_1, α_2, α_3, and $f(J_1J_2J)$ from the RR's β_1, β_2, and β by matching the phase conventions for the initial CG coefficients and the function g by making use of (44).

Let us now state the symmetry relations which we wish to prove.

B. The Symmetry Relations

For brevity, we shall sometimes drop the invariant parameters ν and ρ in the X's. An interchange of, e.g., J_1 and J_2, will, of course, mean a simultaneous interchange of $\nu_1\rho_1 J_1$ and $\nu_2\rho_2 J_2$.

Case (a)

Parameters ν_1, ν_2, and ν satisfy a triangular relation:

$$\overline{X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J)} = (-1)^{J_1+J_2+J+\nu_1+\nu_2+\nu} X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J), \quad (46)$$

$$X(\nu_2\rho_2 J_2, \nu_1\rho_1 J_1, \nu\rho J) = (-1)^{J_1+J_2+J+\nu_1+\nu_2+\nu} X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J), \quad (47)$$

$$X(\nu\rho J, \nu_2\rho_2 J_2, \nu_1\rho_1 J_1) = (-1)^{J_2-\nu_2}[(2J+1)/(2J_1+1)]^{\frac{1}{2}} \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J), \quad (48)$$

$$X(\nu_1\rho_1 J_1, \nu\rho J, \nu_2\rho_2 J_2) = (-1)^{J_1-\nu_1}[(2J+1)/(2J_2+1)]^{\frac{1}{2}} \times X(\nu_1\rho_1 J_1, \nu_2\rho_2 J_2, \nu\rho J). \quad (49)$$

Case (b)

Parameters ν_1, ν_2, and ν do not satisfy a triangular relation. We must again consider three separate cases:

(i) $\nu_1 > \max(\nu_2, \nu)$:

$$\overline{X(J_1, J_2, J)} = (-1)^{J_1+J_2+J+2\nu_1} X(J_1, J_2, J), \quad (50)$$

$$X(J_2, J_1, J) = (-1)^{J_1+J_2+J+2\nu_1} X(J_1, J_2, J), \quad (51)$$

$$X(J, J_2, J_1) = (-1)^{J_2+\nu-\nu_1}[(2J+1)/(2J_1+1)]^{\frac{1}{2}} X(J_1, J_2, J), \quad (52)$$

$$X(J_1, J, J_2) = (-1)^{J_1-\nu_1}[(2J+1)/(2J_2+1)]^{\frac{1}{2}} X(J_1, J_2, J); \quad (53)$$

(ii) $\nu_2 > \max(\nu_1, \nu)$:

$$\overline{X(J_1, J_2, J)} = (-1)^{J_1+J_2+J+2\nu_2} X(J_1, J_2, J), \quad (54)$$

$$X(J_2, J_1, J) = (-1)^{J_1+J_2+J+2\nu_2} X(J_1, J_2, J), \quad (55)$$

$$X(J, J_2, J_1) = (-1)^{J_2-\nu_2}[(2J+1)/(2J_1+1)]^{\frac{1}{2}} X(J_1, J_2, J), \quad (56)$$

$$X(J_1, J, J_2) = (-1)^{J_1-\nu_1}[(2J+1)/(2J_2+1)]^{\frac{1}{2}} X(J_1, J_2, J); \quad (57)$$

(iii) $\nu > \max(\nu_1, \nu_2)$:

$$\overline{X(J_1, J_2, J)} = (-1)^{J_1+J_2+J+2\nu} X(J_1, J_2, J), \quad (58)$$

$$X(J_2, J_1, J) = (-1)^{J_1+J_2+J+\nu_1+\nu_2+\nu} X(J_1, J_2, J), \quad (59)$$

$$X(J, J_2, J_1) = (-1)^{J_2-\nu+\nu_1}[(2J+1)/(2J_1+1)]^{\frac{1}{2}} X(J_1, J_2, J), \quad (60)$$

$$X(J_1, J, J_2) = (-1)^{J_1-\nu_1}[(2J+1)/(2J_2+1)]^{\frac{1}{2}} X(J_1, J_2, J). \quad (61)$$

C. Proof of the Symmetry Relations

All the relations (46)–(61) can be proved in a completely analogous manner which we shall exemplify by the single case (52) (which is one of the most involved cases). We proceed in the three steps sketched above starting from (45).

(a) The substitution

$$X(J, J_2, J_1) \to (-1)^{J_2}[(2J+1)/(2J_1+1)]^{\frac{1}{2}} X(J_1, J_2, J)$$

converts (17a) into (17b), (17b) into (17a), and (17c) into itself. This takes care of α_1, α_2, α, and $f(J_1J_2J)$ in (45).

(b) Formula (44) together with the explicit form[1] of $D^{\nu\rho}_{J_1J_2\lambda}(a)$ gives

$$|X(\nu_1\rho_1J_1, \nu_2\rho_2J_2, \nu\rho J)|^2 = \frac{32\pi^3}{(2J+1)^2} \sum_{\lambda_1\lambda_2} (J_1\lambda_1, J_2\lambda_2 \mid J\ \lambda_1+\lambda_2)^2 \times \int_0^\infty sh^2a\, da D^{\nu\rho}_{JJ\,\lambda_1+\lambda_2}(a) D^{\nu_1\rho_1}_{J_1J_1\lambda_1}(a) D^{\nu_2\rho_2}_{J_2J_2\lambda_2}(a). \tag{62}$$

Using this equation and the symmetries (41) of the D functions, one can easily check that

$$|X(\nu\rho J, \nu_2\rho_2J_2, \nu_1\rho_1J_1)|^2 = [(2J+1)/(2J_1+1)]\, |X(\nu_1\rho_1J_1, \nu_2\rho_2J_2, \nu\rho J)|^2. \tag{63}$$

Substituting

$$X(J, J_2, J_1) = (-1)^{J_2+\beta_1\nu_1+\beta_2\nu_2+\beta\nu} \times [(2J+1)/(2J_1+1)]^{\frac{1}{2}} g(\nu_1\rho_1, \nu_2\rho_2, \nu\rho) \times X(J_1, J_2, J) \tag{64}$$

into (63), we obtain in this case

$$g(\nu_1\rho_1, \nu_2\rho_2, \nu\rho) = 1.$$

Now let us consider the β's. According to (36) and (38), we have

$$X(\nu_1\rho_1\nu_1, \nu_2\rho_2\, \nu_1 - \nu, \nu\rho\nu) > 0,$$

and

$$X(\nu\rho\nu, \nu_2\rho_2\, \nu_1 - \nu, \nu_1\rho_1\nu_1) > 0.$$

Together with (64) this implies

$$(-1)^{\nu_1-\nu+\beta_1\nu_1+\beta_2\nu_2+\beta\nu} = 1$$

i.e.,

$$\beta_1 = -1, \quad \beta_2 = 0, \quad \beta = 1.$$

This completes the proof of (52).

Let us note that the symmetry relations of this section, as well as some further relations, can be proven directly using (44) and the phase conventions without using the RR's. In particular, it is easy to check symmetry relations

$$\begin{aligned} X(\nu_1\rho_1J_1, \nu_2\rho_2J_2, \nu\rho J) &= X(\nu_1\rho_1J_1, \nu_2\rho_2J_2, -\nu-\rho J) \\ &= X(\nu_1\rho_1J_1, -\nu_2-\rho_2J_2, \nu\rho J) \\ &= X(-\nu_1-\rho_1J_1, \nu_2\rho_2J_2, \nu\rho J) \\ &= X(-\nu_1-\rho_1J_1, -\nu_2-\rho J_2, \nu\rho J), \quad \text{etc.} \end{aligned} \tag{65}$$

VII. SPINOR REPRESENTATIONS AND ANALYTIC CONTINUATION

The RR's presented in Sec. III are applicable to all representations of the homogeneous Lorentz group and, in particular, to the finite-dimensional (nonunitary) representations or the so-called spinor representations. We pass between the unitary and spinor RR's with the substitutions

$$\nu + \tfrac{1}{2}i\rho - 1 \leftrightarrow k \quad \text{and} \quad -\nu + \tfrac{1}{2}i\rho - 1 \leftrightarrow n,$$

where k and n are integer of half-integer invariants which specify a spinor representation. Thus, the RR's for the unitary case are analytic continuations of the spinor cases.

This leads to the important question as to whether the unitary CG coefficients can be obtained by "analytically continuing" the known spinor CG coefficients. Indeed, they can be expressed in terms of a Fano function or 9-J symbol $\{\ \}$ as follows[11,12]

$$\langle k_1n_1J_1M_1, k_2n_2J_2M_2 \mid knJM\rangle = (-1)^{\frac{1}{2}(k+n)}[(2J_1+1)(2J_2+1)(k+1)(n+1)]^{\frac{1}{2}} \times (J_1M_1, J_2M_2 \mid JM) \begin{Bmatrix} \frac{1}{2}k_1 & \frac{1}{2}n_1 & J_1 \\ \frac{1}{2}k_2 & \frac{1}{2}n_2 & J_2 \\ \frac{1}{2}k & \frac{1}{2}n & J \end{Bmatrix}.$$

Here the conventions of Ref. 13 have been employed. We can, for instance, analytically continue the above coefficient by defining a new Fano function with complex parameters in terms of analytically continued Racah or 6-J coefficients, e.g., through the relation

$$\begin{Bmatrix} a_1 & b_1 & J_1 \\ a_2 & b_2 & J_2 \\ a_3 & b_3 & J_3 \end{Bmatrix} = \sum_\gamma (-1)^\gamma (2\gamma+1) \begin{Bmatrix} a_1 & a_2 & a_3 \\ b_3 & J_3 & \gamma \end{Bmatrix} \times \begin{Bmatrix} b_1 & b_2 & b_3 \\ a_2 & \gamma & J_2 \end{Bmatrix} \begin{Bmatrix} J_1 & J_2 & J_3 \\ \gamma & a_1 & b_1 \end{Bmatrix}.$$

In the above equation the a's and b's are complex while the J's are nonnegative integers and half-integers; thus, the γ summation is finite. This continuation can be performed explicitly because a Racah coefficient is expressible as a ${}_4F_3$ generalized hypergeometric function of unit argument,[14] which may then be analytically continued. The crucial question is then whether these analytically continued spinor coefficients correspond to the unitary ones.

We sketch here as an application of the RR's derived in this paper an approach to this question. It follows from the properties of the RR's that, if

one (e.g., spinor CG) coefficient analytically continues to a unitary one, up to a factor of the invariants, then all the coefficients do so. More specifically, this follows from the two properties that, once a starting coefficient is specified, then all others are uniquely determined through the RR's and that the RR's for the various cases analytically continue into each other. This then reduces the problem of analytic continuation to the comparison of initial coefficients, and we hope to return to this problem in a future publication.

VIII. RESULTS

Recursion relations for the CG coefficients have been derived and, for any given set of the invariant parameters $\nu_1\rho_1$, $\nu_2\rho_2$, $\nu\rho$, all CG coefficients can be calculated in terms of one single "initial" coefficient. The procedure is as follows:

(a) Let ν_1, ν_2, and ν satisfy a triangular relation, i.e., $|\nu_1 - \nu_2| \leq \nu \leq \nu_1 + \nu_2$. We choose $X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\nu)$ as the initial coefficient and use Eqs. (19) to raise the values of the J's, making use of Eqs. (25).

(b) If ν_1, ν_2, and ν are not triangular, the choice of the initial coefficients depends on the ordering of the ν's. For example, for $\nu \leq \nu_2 \leq \nu_1$, we choose $X(\nu_1\rho_1\nu_1, \nu_2\rho_2\nu_2, \nu\rho\, \nu_1 - \nu_2)$ as the initial coefficient, introduce a set of "minimal" coefficients $X(\nu_1\rho_1\nu_1, \nu_2\rho_2\, \nu_2 + k, \nu\rho\, \nu_1 - \nu_2 - k)$, with $0 \leq k \leq \nu_1 - \nu_2 - \nu$, which are expressed in terms of the initial coefficient by (29) and then again apply (19) and (25) to raise the J's. The procedure is analogous for all other orderings of ν_1, ν_2, and ν.

Certain phase conventions are introduced for the CG coefficients and some symmetry relations are established.

A program is given for the study of a possible connection, through analytic continuation, between the CG coefficients for the coupling of finite-dimensional representations and those for the infinite-dimensional case.

ACKNOWLEDGMENTS

We thank Professor A. Salam for suggesting the investigation and for continuous encouragement. We are grateful for very profitable discussions with Professor C. Fronsdal, Dr. H. D. Doebner, Dr. I. Marek, Dr. P. Pajas, Professor J. A. Smorodinsky, Dr. J. Strathdee, Dr. L. Trlifaj, and Dr. V. P. Zhelobenko.

We are indebted to Professors A. Salam and P. Budini, and to the IAEA for hospitality extended to us at the International Centre for Theoretical Physics, Trieste. One of the authors (P. W.) is indebted to UNESCO for financial support.

* Present address: Institute for Nuclear Research, Warsaw, Poland (Visiting scientist, PAN-NAS Exchange Program).

† On leave of absence from Institute for Nuclear Research, Warsaw, Poland.

‡ On leave of absence from Pakistan Atomic Energy Commission, Lahore, Pakistan.

§ On leave of absence from Institute of Nuclear Research, Řež, Czechoslovakia. Present address: Department of Physics, Carnegie-Mellon University, Pittsburgh, Pa.

[1] R. L. Anderson, R. Rączka, M. A. Rashid, and P. Winternitz, J. Math. Phys. **11**, 1050 (1970) (preceding article).

[2] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, London, 1964).

[3] M. A. Naimark, Tr. Mosk. Mat. Obsc. **8**, 121 (1959); **9**, 237 (1960); **10**, 181 (1961) [English translations, Amer. Math. Soc. Transl., Ser. 2, **36**, 101 (1964)].

[4] L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Rev. Mod. Phys. **24**, 249 (1952).

[5] A. Z. Dolginov and I. N. Toptygin, Zh. Eksp. Teor. Fiz. **35**, 793 (1958); **37**, 1441 (1959) [Soviet Phys.—JETP **8**, 550 (1959); **10**, 1022 (1960)].

[6] G. Bisiacchi and C. Fronsdal, Nuovo Cimento **41**, 35 (1965).

[7] G. W. Mackey, Ann. Math. **55**, 101 (1952).

[8] (a) L. D. Eskin, Izv. Vysshykh Uchebn. Zavedenii, Mat. **6**, 179 (1961); (b) S. Ström, Arkiv Fysik **29**, 467 (1965); (c) A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967); (d) D. W. Duc and N. V. Hieu, Ann. Inst. Henri Poincaré **6**, 17 (1967); (e) I. A. Verdiev and L. A. Dadashev, Yad. Fiz. **6**, 1094 (1967) [Sov. J. Nucl. Phys. **6**, 795 (1968)].

[9] J. F. Boyce, R. Delbourgo, A. Salam, and J. Strathdee, ICTP, Trieste, Preprint IC/67/9.

[10] T. Regge, Nuovo Cimento **10**, 545 (1958).

[11] J. M. Gel'fand, R. A. Minlos, and Z. J. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon Press, Oxford, 1963).

[12] R. Delbourgo, A. Salam, and J. Strathdee, Phys. Rev. **164**, 1981 (1967).

[13] A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic Press, New York, 1963).

[14] M. E. Rose, *Multipole Fields* (John Wiley and Sons, New York, 1955).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Random Operator Equations in Mathematical Physics. I

G. ADOMIAN
Department of Mathematics, University of Georgia, Athens, Georgia 30601

(Received 25 March 1967)

For stochastic differential equations arising in physical problems, the objectives, limitations, and restrictive assumptions of the various methods are studied and some promising new methods are derived which eliminate various limitations and allow treatment of a wide class of applications in physics. (Among these, will be an adequate treatment of the propagation of an electromagnetic wave in a random continuum or a random d'Alembertian operator without assumptions of "small randomness" and other restrictions.)

1. INTRODUCTION

Although considerable literature exists on random equations, most of it applies to first-order differential equations or to equations with constant coefficients, with the element of randomness arising from either a random forcing function or random boundary conditions. These cases are fundamentally simpler than the case of stochastic coefficients because of the deterministic relationship of the probabilistic or statistical properties of the solution to the statistical properties of the random quantity. (These cases are well covered in a survey by Syski[1] and will not be discussed here.) Studies of random operators on Banach spaces have usually been inapplicable to stochastic differential operators. Much of the related work in applications (propagation and scattering) has minimized the probabilistic aspects using methods whose validity is open to question and which are generally incorrect (averaging methods, sample-function approach, etc.). Finally, much interesting work has been restricted to special processes (white noise, etc.), and it would be desirable to consider the problem in greater generality. This paper will provide a basis for several papers concerned particularly with the linear random-operator equations

$$y = \mathcal{K}\{x\},$$
$$\mathfrak{L}\{y\} = x,$$

where

$$\mathfrak{L} = \sum_{\nu=0}^{n} a_\nu(t) \frac{d^\nu}{dt^\nu}$$

and the $a_\nu(t)$ and $x(t)$ are stochastic processes (SP's). The $x(t)$ or $x(t, \omega)$, where ω is an element of a probability space (Ω, F, μ), is to be taken as statistically independent of the $a_\nu(t)$ or $a_\nu(t, \omega)$ [although the $a_\nu(t)$ may be correlated with each other]. This, too, has been studied by many in the last few years; we refer particularly to the work of Samuels,[2] and Samuels and Eringen[3] (cf. also Grenander[4]). Some important matters will not be discussed here, e.g., the stability of systems described by the stochastic differential equations, and we will make no attempt to discuss the numerous possible applications discussed elsewhere.[5–7]

Keller has distinguished two methods of solving random differential equations of the form $\mathfrak{L}y = x$, where the operator $\mathfrak{L}$ involves a random parameter α. In the first method, the solution $y(r, t, \alpha)$ is determined for each value of the random parameter α. Thus, the randomness plays no real role in the process of finding y. Then the random nature of the solution is considered, and the expectation

$$\langle y \rangle = \int_A y(r, t, \alpha) P(\alpha)\, d\alpha$$

is computed as a weighting of each specific solution with its probability. Higher-order statistics are similarly found. This is a sample function approach and is generally not valid. In the second method, a direct determination is made of the equations satisfied by the various moments of the solution. To do this, the equation and a hierarchy derived from it (in a manner to be discussed) are averaged. This is also generally incorrect. The solution of the averaged equation is not always (or even usually) the average solution of the random equation. Furthermore, a so-called "closure approximation" is involved which cannot be justified and, in actuality, is equivalent to the use of perturbation theory to some order in the first place. The first method is right in principle but asks too much. It involves obtaining a solution first and then obtaining the statistics. This is not always possible; in fact, it is essentially limited to the case of a first-order equation (considered by Tikhonov,[8] Adomian,[5] Aström,[9] and others) since it involves the inversion of a stochastic matrix. Since only statistical information or ensemble averages are to be the end result, then elimination of the intermediate step, which requires more than this, should result in a cleaner and more useful method. The ultimate goal in such problems is the complete statistical description of the output y from the statistical knowledge of the coefficients and the input.

Often, however, it is sufficient to determine less complete "statistical measures" in the form of expectation, correlation function, or spectral density. (The term "statistical measures" will be rigorously defined elsewhere; here, it will be "defined" by example and will simply stand for the useful quantities which can be calculated for a random process.) Of course, by the use of various approximations based on assumptions of slow variation of parameters, small fluctuations, and restricted classes of processes (white noise), etc., solutions can be found to a wider class of differential equations than the first-order equation, so that the statistical properties can be determined (e.g., Samuels,[2] Chelpanov,[10] Keller[11]); however, we wish to eliminate these restrictions. To summarize, then, the first method usually requires the iterative approach; as used, it is often invalid. The second method involves unsupportable assumptions of separability and works only in highly restricted (Dirac measure) spaces or within a perturbation-theory framework. Our objective is a method for directly determining desired statistical measures of the dependent variable without averaging and unjustifiable assumptions. There are two possible ways of doing this: The first is a modification of Samuels' method and avoids the difficulty in separability; the second, a stochastic expansion method suggested by Adomian,[12] appears promising and some interesting results have been obtained. The question asked is whether one can find a stochastic Green's function for the desired statistical measure of the dependent variable analogous to finding a Green's function for an ordinary equation.

Most equations of mathematical physics and engineering science are assumed to be in the form

$$Lf(x) = s(x),$$

where L is a deterministic linear operator. (L may be a nth-order differential operator or a Hamiltonian operator, or the (self-adjoint) Sturm–Liouville operator

$$\frac{d}{dx}\left[p(x)\frac{d}{dx}\right] - q(x).$$

It is usually assumed that to L corresponds an integral kernel or Green's function $k(x, x')$ such that

$$f(x) = L^{-1}s(x) = \int_{-\infty}^{\infty} k(x, x')s(x')\,dx'$$

is the solution.

The analogous equations for a stochastic operator can be written

$$\mathfrak{L}f(x) = s(x) \tag{1.1a}$$

or

$$f(x) = \mathfrak{K}s(x). \tag{1.1b}$$

Both forms simply show a function being transformed into a new function by an operator, but it is convenient for both mathematical and physical reasons to consider the two equations separately. Equation (1.1a) involves a differential operator $\mathfrak{L}$ and a function $s(x)$ which may be an ordinary function and specifically given, or a random function whose statistics are known. (Generalization to random fields and partial differential operators follows but will not be discussed specifically here. See Refs. 5 and 7.) Then the solution $f(x)$ is a random function (stochastic process) whose statistics must be found. In the second form (1.1b), $s(x)$ is again given, either as an ordinary function if deterministic or through its statistics if random. $\mathfrak{K}$ is a given random transformation which transforms $s(x)$ into another random function $f(x)$. $\mathfrak{K}$ is a known operator, a measurement, or a processing of the signal $s(x)$ (by a known "black box" $\mathfrak{K}$). $\mathfrak{K}$ is stochastic by virtue of parameters, one or more of which are random, but whose statistics are known to the necessary degree. We can consider $\mathfrak{K}$ to be a "stochastic filter" and the resulting theory a generalization of ordinary filter theory.

Ability to treat these two possibilities would give us a two-pronged attack on most physical problems, in which we would either determine the characteristics of the solution to a (stochastic) differential equation, or the (statistical) characteristics of the random function resulting from a stochastic transformation. Since the result in either case is a stochastic process, our knowledge of the process will be specified in the form of expectations, spectral densities, higher moments, or, at most, in a distribution function. We will lump these terms together under the term "statistical measures" for the process (and use it rather than the unsatisfactory term "statistics" used above).[13] Thus, our objective will be to specify desired statistical measures for an output process in terms of the known statistical measures of the input process and an integral kernel depending upon the operator which we then call a "stochastic Green's function."[12]

Now, how do such stochastic or random operator equations arise? An operator may be fundamentally stochastic so such equations may conform more closely to the real problem than a deterministic analog. Or, it may simply be useful to assume a stochastic nature for $\mathfrak{L}$, because it sometimes provides more information to investigate a statistical solution of a stochastic operator equation than the corresponding deterministic model. Why? Because either the deterministic solution is too difficult or too complex to specify completely—a differential equation which we cannot solve or a many-body problem—or

because we want to specify simultaneously an ensemble of possible situations, measurements, or deterministic operators. The nature of such problems may be clarified by some conceptual examples.

Suppose we consider the propagation of waves in a continuous random medium where we have a linear stochastic (partial) differential equation. The coefficients in the equation characterize the propagation medium, e.g., by the refractive index $n(x)$, the density $p(x)$, the dielectric constant, etc. Thus, the random medium is an ensemble of media each labeled, e.g., with an $n(x)$ or $n(x, \alpha)$ with a probability $p(\alpha)$ for each value.[11]

The general inhomogeneous term, which we now replace with $g(x, t)$, is the source of the waves. It may be either random or deterministic. Similar problems arise in the propagation of acoustic waves in a turbulent medium, or the statistical mechanics of a continuous medium. A number of authors, (e.g., Keller,[11] Foldy,[14] Meecham,[15] Twersky[16]) have considered another type of random medium, one which consists of a random distribution of discrete scattering objects in a fixed medium in which waves can propagate, as in impurities distributed through a solid material, the Mie scattering problem or any matter on a molecular level. For example, in the molecular scattering of light, if we consider the propagation of waves in a gas or in air and assume that the molecules do not move since the waves pass through so rapidly, one could, in principle, determine how the molecules scatter the wave from the specified location of all the molecules. This is not only too complex, it is also wrong! If it was *only* too complex, we would still hope that the observable characteristics of the scattering would be found by replacing the actual distribution of molecular positions by a random distribution or a collection of distributions with a probability attached. However, in this case the situation is even more in favor of a stochastic solution; the physical situation *is* random, so that a stochastic solution corresponds more to the physics.

Most, if not all, of the equations of physics can be viewed as stochastic operator equations in this sense—either because the various parameters are subject to random fluctuations, or because of lack of *a priori* knowledge, or because it is simply convenient to view an equation as stochastic. As an example of the latter, suppose we are making a series of measurements where it is impossible to insure that the measurement is done in the same way each time. Each particular measurement is a deterministic operator; the ensemble is a stochastic operator. The flipping of a coin conceptually could be solved by knowledge of its launch-velocity vector, distribution of metals, and environmental influences. In actuality, it cannot because the problem is too complicated and because extremely small perturbations in the launch velocity—beyond the control of the experimenter—are sufficient to change the result from one to the other of the two allowed outcomes. A similar example in engineering is a statistical prediction of performance of an ensemble of systems which vary from one another in some uncontrollable fashion because of production tolerances or changes of some parameters of the system. (It is interesting to note that such statistical prediction sometimes is necessary even for a completely deterministic theory due to the finite velocity of propagation of physical influences.[17])

This work is being formulated by the author in the context of random operators on algebraic structures in a manner consistent with the work of the Prague school of probabilists or of Grenander.[18]

2. OPERATOR FORMULATION

Suppose a linear stochastic operator $\mathcal{K}_{\alpha\beta}\ldots$ (or $\mathcal{K}$, for brevity) acts upon a random function $\{x(\xi, \omega), \omega \in \Omega\}$, which we usually write simply as $x(\xi)$ and assume to be statistically known. Assume $\mathcal{K}$ is statistically independent of x. Physically, $\mathcal{K}$ may represent a filter, a scattering medium, a communication channel, or a measurement. $\alpha(\xi)$, $\beta(\xi)$, $\cdots$ represent system parameters or parameters of the operator $\mathcal{K}$, at least one of which we suppose to be random. $\mathcal{K}$ is then determined by the distributions of its parameters and represents an ensemble of operators H collectively represented by $\{H\}$ or $\mathcal{K}$. Writing the stochastic equation

$$y = \mathcal{K}_\alpha[x],$$

we see this represents an ensemble of equations depending upon the parameter α which ranges over an appropriate space A, in this case one in which a probability density $p(\alpha)$ is defined. $p(\alpha)$ determines the probability of a given value for α and, therefore, of the corresponding equation of the ensemble. We view $\mathcal{K}$ as mapping the random function x into the random function y. *Conceptually*, each H of $\{H\}$ has an associated Green's function $h(\xi, \eta)$ which represents the response at ξ of the representative system H to a unit impulse $\delta(\xi - \eta)$. Of course, one need never talk of δ functions. We can equally well use Stieltjes integrations or distributions in the sense of Schwartz. The latter approach is particularly interesting and leads to stochastic generalized functions or distributions. (See, e.g., Appendix 1 in A. M. Yaglom, *Theory of Stationary Random Functions*, translated by R. A.

Silverman, Prentice-Hall, Inc., Engelwood Cliffs, N. J., 1962.) Then the response of the H system to an input x is $y(\xi)$, where

$$y(\xi) = Hx(\xi) = \int_{-\infty}^{\infty} h(\xi, \eta)x(\eta)\, d\eta$$

if x is a defined continuous member of the x process (i.e., a realization) and H is a causal (physically realizable) system. This is not a sample-function approach. We can view y and x as random functions, i.e., output and input of a stochastic filter, and $h(\xi, \eta)$ as a *random* Green's function. For example, if ξ and η are the time coordinates t and τ, $h(t, \tau) = 0$ for $t < \tau$.

We define now the autocorrelation of the transformed process $y = \mathcal{K}[x]$ by

$$R_y(\xi_1, \xi_2) = E\{y(\xi_1)y^*(\xi_2)\}_{H,x}$$

or

$$R_y(\xi_1, \xi_2) = \langle y(\xi_1)y^*(\xi_2)\rangle_{H,x},$$

where the notation indicates an averaging over both ensembles involved. Quite generally the expectation value symbols can be moved inside the integral (Fubini's theorem). Hence, $R_y(\xi_1, \xi_2)$ is now

$$= \int_{-\infty}^{\infty}\int \langle h(\xi_1, \eta_1)h^*(\xi_2, \eta_2)x(\eta_1)x^*(\eta_2)\rangle\, d\eta_1\, d\eta_2.$$

Recognizing $\langle x(\eta_1)x^*(\eta_2)\rangle$ as the autocorrelation function $R_x(\eta_1, \eta_2)$ of the original process, we can separate the integrand into the product of ensemble averages for the system and the input by virtue of the statistical independence of $\mathcal{K}$ and x. Thus,

$$R_y(\xi_1, \xi_2) = \int_{-\infty}^{\infty}\int \langle h(\xi_1, \eta_1)h^*(\xi_2, \eta_2)\rangle R_x(\eta_1, \eta_2)\, d\eta_1\, d\eta_2$$

or, as a convenient general form,

$$R_{\mathcal{K}x}(\xi_1, \xi_2) = \int_{-\infty}^{\infty}\int G_{\mathcal{K}}(\xi_1, \xi_2, \eta_1, \eta_2)R_x(\eta_1, \eta_2)\, d\eta_1\, d\eta_2 \tag{2.1}$$

i.e., the autocorrelation of the transformed process $\mathcal{K}[x]$ is given generally in the terms of the autocorrelation of the original process $x(\xi)$ and a kernel function—in this case, $G_{\mathcal{K}}$, which is itself an autocorrelation, depending on the stochastic system $\mathcal{K}$. This kernel function is called the "stochastic Green's function." In general, x might be a function of time, position, or frequency, and the Green's functions are appropriate to the operator. We will discuss the representative case where $x = x(t)$ and $\mathcal{K}$ is a randomly time-varying operator which may represent a medium of transmission or of scattering, or an observation, experiment, or measurement:

$$R_{\mathcal{K}x}(t_1, t_2) = \int_{-\infty}^{\infty}\int G_{\mathcal{K}}(t_1, t_2, \tau_1, \tau_2)R_x(\tau_1, \tau_2)\, d\tau_1\, d\tau_2.$$

The limits of integration may be finite in a particular problem and may depend on the range of definition of the processes involved and on the existence of causality and memory.

We do not expect, of course, to find Green's functions for each and every realization, although this may be possible in the analysis of assumed mathematical models where we may generalize a Green's function for a deterministic problem to a stochastic Green's function for the corresponding stochastic problem.[19] Rather, we hope to calculate the stochastic Green's function from statistical knowledge which would ordinarily be available. We emphasize that we do not make a sample-function approach. This is a formal relation which occurs generally and which suggests an approach directly through the stochastic Green's function. If, instead of sample functions or realizations, we consider a matrix or state-space formulation, the equation $y = \mathcal{K}x$ relates the SP y to the SP x and the $h(\xi, \eta)$ becomes a random Green's function. Taking the correlation of the SP y, we immediately obtain the same relationship between correlation function of "output" and "input" and the identification of the stochastic Green's function for that particular statistical measure of input and output processes. Details are discussed in Sec. 5.

Let us now examine some forms of (2.1). Suppose $x(t)$ is a stationary SP. Then,

$$R_x(\tau_1, \tau_2) = R_x(\tau_1 - \tau_2) \doteq R_x(\sigma).$$

Suppose the transformation by $\mathcal{K}$ preserves stationarity in the transformed process. (In general, it does not, but it can under special conditions; we defer these questions for now. Precise conditions will be stated in a later article.)Then,

$$R_{\mathcal{K}x}(t_1, t_2) = R_{\mathcal{K}x}(t_1 - t_2) = R_{\mathcal{K}x}(\beta). \tag{2.2}$$

Consequently,

$$R_{\mathcal{K}x}(\beta) = \int g_{\mathcal{K}}(\beta, \sigma)R_x(\sigma)\, d\sigma,$$

where

$$g_{\mathcal{K}}(\beta, \sigma) = \int d\tau\langle h(t, \tau)h(t + \beta, \tau + \sigma)\rangle \tag{2.3}$$

is the stochastic Green's function for the autocorrelation for the stationary case. By appropriate

transformation, the spectral density can be found:

$$\Phi_v(f) = \int K_{\mathcal{H}}(s,f)\Phi_x(s)\,ds, \tag{2.4}$$

where[20]

$$K_{\mathcal{H}}(s,f) = \int_{-\infty}^{\infty} d\sigma \int_{-\infty}^{\infty} d\beta e^{2\pi i f\beta} e^{-2\pi i s\sigma} g_{\mathcal{H}}(\beta,\sigma).$$

Expression (2.3) must be satisfied for stationarity to hold, i.e., the kernel must not depend on time; so, the ensemble average must be time independent, which requires that the parameters of $\mathcal{H}$ must at least be stationary. A following paper will give necessary and sufficient conditions for stationarity under stochastic transformation. In general, the spectral density does not exist since the output is nonstationary. If $\mathcal{H}$ is ergodic, each member of the ensemble gives the same time average as every other member, so that it may be possible to find a single Green's function and, from it, the stochastic Green's function by suitable averaging. Thus,

$$\langle h(t,\tau)h(t+\beta,\tau+\sigma)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} du\, h(t+u,\tau)h^*(t+\beta+u,\tau+\sigma).$$

Since the $\alpha, \beta, \cdots$ enter into the h's, we have more explicitly

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} du \int_0^{\infty} d\alpha \int_0^{\infty} d\beta \cdots p(\alpha)p(\beta)\cdots h(\alpha,\beta,\cdots,t+u,\tau) \times h^*(\alpha,\beta,\cdots,t+\beta+u,\tau+\sigma).$$

3. PERTURBATION-THEORY APPROACH

A. Introduction

Let us suppose[21] first that the random variations are sufficiently small so corrections to the deterministic solution are of low order. Then, perturbation theory is useful. Consider a linear operator $\mathfrak{L}$ depending upon a parameter α which ranges over a measure space or set A. Suppose $p(\alpha)$ is a probability density defined over A (or appropriate subsets of A). Then $\mathfrak{L}_\alpha$ is a stochastic operator and we can consider the following stochastic equation:

$$\mathfrak{L}_\alpha u = g,$$

where, at first, g is a given nonrandom element of the linear space. Assuming a unique solution $u(\alpha)$ for each α, we see that the solution depends upon α. Thus, $u(\alpha)$ is a random solution and $p(\alpha)$ determines the probability density of $u(\alpha)$. The statistical measure of u which is of interest to us is specified to be the expectation $\langle u\rangle = \int_A u(\alpha)p(\alpha)\,d\alpha$.

Assume that $\mathfrak{L}_\alpha = \mathfrak{L}(\alpha,\epsilon)$ depends upon a small parameter ϵ and that, for $\epsilon = 0$, $\mathfrak{L}$ reduces to a deterministic operator L. Expanding $\mathfrak{L}_\alpha$ in powers of ϵ, we may write

$$\mathfrak{L} = L + \epsilon\mathfrak{L}_1(\alpha) + \epsilon^2\mathfrak{L}_2(\alpha) + O(\epsilon^3)$$

and

$$\mathfrak{L}u(\alpha,\epsilon) = g.$$

Thus, $\mathfrak{L}$ is given by the sum of the operator L which is deterministic, and $\mathfrak{L}_1$ and $\mathfrak{L}_2$ which are random and represent stochastic perturbations of the deterministic operator L. Suppose, for simplicity, that $\langle\mathfrak{L}_1\rangle = \langle\mathfrak{L}_2\rangle = 0$. If $\epsilon = 0$, the solution of the resulting deterministic equation $Lu = g$ is $u_0 = L^{-1}g$, assuming L^{-1} is defined. Now,

$$u = u_0 - \epsilon L^{-1}\mathfrak{L}_1 u - \epsilon^2 L^{-1}\mathfrak{L}_2 u + O(\epsilon^3).$$

If we average to get $\langle u\rangle$, we have quantities like $\langle\mathfrak{L}_1 u\rangle$ or $\langle\mathfrak{L}_2 u\rangle$, which ordinarily would not separate further since they involve the operator and the dependent variable or solution process. However, they do now because we have assumed perturbation theory to be applicable. Thus, by iteration or successive substitution, $u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \cdots$ becomes

$$u = u_0 - \epsilon L^{-1}\mathfrak{L}_1 u_0 + \epsilon^2 L^{-1}(\mathfrak{L}_1 L^{-1}\mathfrak{L}_1 - \mathfrak{L}_2)u_0 + O(\epsilon^3).$$

Taking the expectation, we have

$$\langle u\rangle = u_0 + \epsilon^2 L^{-1}\langle\mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle u_0 + O(\epsilon^3) = L^{-1}g(1 + \epsilon^2 L^{-1}\langle\mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle) + O(\epsilon^3).$$

Equivalently, since $u_0 = \langle u\rangle + O(\epsilon^2)$,

$$\langle u\rangle = L^{-1}g + \epsilon^2 L^{-1}\langle\mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle + O(\epsilon^3),$$
$$(L - \epsilon^2\langle\mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle)\langle u\rangle = g + O(\epsilon^3),$$

which is given by Keller[11] (and also found by Adomian[22]).

We note it is not necessary that g be deterministic. If g is stochastic, $L^{-1}g$ is stochastic and, in taking the expectation, we get $L^{-1}\langle g\rangle$. We find that averages involving $\mathfrak{L}_1$ or $\mathfrak{L}_2$ and g separate by only the statistical independence of $\mathfrak{L}$ and g. Thus, we need make no further assumptions of an artificial nature in an attempt to separate $\langle\mathfrak{L}u\rangle$ into $\langle\mathfrak{L}\rangle\langle u\rangle$. In general, $\mathfrak{L}$ cannot be assumed to be statistically independent of u; but $u_0 = L^{-1}g$ is deterministic in the first case, so that it separates out. And in the second case, $L^{-1}g$ is stochastic, but the statistical independence of $\mathfrak{L}$ and g is sufficient for the separations. To see this we proceed as follows:

$$[L + \epsilon\mathfrak{L}_1 + \epsilon^2\mathfrak{L}_2 + O(\epsilon^3)]u = g_0 + \epsilon g_1 + \epsilon^2 g_2 + O(\epsilon^3).$$

Let $u_0 = L^{-1}g_0$. Then,

$$Lu = g_0 + \epsilon g_1 + \epsilon^2 g_2 - \epsilon \mathcal{L}_1 u - \epsilon^2 \mathcal{L}_2 u + O(\epsilon^3),$$
$$u = u_0 + \epsilon L^{-1}(g_1 - \mathcal{L}_1 u_0) + \epsilon^2 L^{-1}(g_2 - \mathcal{L}_2 u_0 - \mathcal{L}_1 u_1).$$

Thus, the coefficient of the ϵ term is u_1 and the coefficient of the ϵ^2 term is u_2. We then have

$$\begin{aligned}\langle u\rangle &= u_0 + \epsilon L^{-1}(\langle g_1\rangle - \langle \mathcal{L}_1\rangle u_0) \\ &\quad + \epsilon^2 L^{-1}(\langle g_2\rangle - \langle \mathcal{L}_2\rangle u_0 - \langle \mathcal{L}_1 L^{-1} g_1\rangle \\ &\qquad + \langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle u_0) \\ &= L^{-1}\langle g\rangle - \epsilon L^{-1}\langle \mathcal{L}_1\rangle L^{-1} g_0 - \epsilon^2 L^{-1}\langle \mathcal{L}_2\rangle L^{-1} g_0 \\ &\quad - \epsilon^2 L^{-1}\langle \mathcal{L}_1\rangle L^{-1}\langle g_1\rangle - \epsilon^2 L^{-1}\langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle L^{-1} g_0.\end{aligned}$$

Combining the second and the fourth terms, we obtain

$$\begin{aligned}\langle u\rangle &= L^{-1}\langle g\rangle - \epsilon L^{-1}\langle \mathcal{L}_1\rangle L^{-1}(g_0 + \epsilon\langle g_1\rangle + \cdots) \\ &\quad - \epsilon^2 L^{-1}(\langle \mathcal{L}_2\rangle + \langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle) L^{-1} g_0 \\ &= L^{-1}\langle g\rangle - \epsilon^2 L^{-1}\langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle L^{-1}\langle g\rangle.\end{aligned}$$

if $\langle \mathcal{L}_1\rangle = \langle \mathcal{L}_2\rangle = 0$; this can be simplified further:

$$\langle u\rangle = L^{-1}\langle g\rangle(1 - \epsilon^2 L^{-1}\langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle)$$

or

$$(L + \epsilon^2\langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle)\langle u\rangle = \langle g\rangle.$$

Without the assumption of statistical independence of $\mathcal{L}$ and g, additional terms appear, as Chen[23] points out, which disappear only if either $\mathcal{L}_1$ and $\mathcal{L}_2$ or g_1 and g_2 are zero, which is physically reasonable. However, in the case of statistical independence, it is not true that there is a coupling term $\epsilon^2 L^{-1}\langle \mathcal{L}_1\rangle L^{-1}\langle g_1\rangle$. This term and the $\mathcal{L}_1 u_0$ part of $u_1 = L^{-1}(g_1 - \mathcal{L}_1 u_0)$ combine to give $\epsilon L^{-1}\langle \mathcal{L}_1\rangle L^{-1}\langle g\rangle$, just as $u_0 + \epsilon L^{-1}\langle g_1\rangle + \epsilon^2 L^{-1}\langle g_2\rangle$ becomes $L^{-1}\langle g\rangle$.

The use of the perturbation theory follows from the requirement that the random part of $\mathcal{L}$ be small, which, in turn, arises from the need to make only u_0 be involved in the $\langle \mathcal{L}u\rangle$ so that it will separate. Thus, we have seen that the desired "statistical measure" of u, in this case, the mean or expectation value, is given in terms of the same statistical measure of g (when g is random) and a functional involving only certain averages over L.[22]

B. Green's Function Method

Let[24] $u = u(x)$ be a n-component vector function of a vector variable x. Then, L, $\mathcal{L}_1, \mathcal{L}_2, \cdots$ are represented by nth-order matrices, each element of which is an operator (differential or integral operator). The inverse operator L^{-1} is also a nth-order matrix which we shall represent as an integral operator. The kernel $G(x, x')$ is the Green's matrix defined by

$$LG(x, x') = I\delta(x - x'),$$

where I is the unit matrix and δ is the Dirac δ. Now, in general,

$$L^{-1}f(x) = \int G(x, x')f(x')\,dx'.$$

We had $\mathcal{L}u(x) = g(x)$, so $\mathcal{L} = \mathcal{L}(x)$. Now, the equation for $\langle u\rangle$ derived earlier is written

$$L(x)\langle u(x)\rangle + \epsilon^2\left\langle \mathcal{L}_1(x)\int G(x, x')\mathcal{L}_1(x')\langle u(x')\rangle\,dx'\right\rangle = g(x) + O(\epsilon^3),$$

$$\langle u(x)\rangle = \int G(x, x')g(x')\,dx' + \epsilon^2\int G(x, x')\left\langle \mathcal{L}_1(x')\int G(x', x'')\mathcal{L}_1(x'') \times \langle u(x'')\rangle\,dx''\right\rangle.$$

Suppose we do not immediately assume that the random part is small and simply assume that we can separate the operator into the sum of two operators $L + R$, where L is deterministic and R is random. We will later assume that R is zero mean, since the random quantity will be assumed to be Gaussian and the Gaussian product will be simpler. Now, $\mathcal{L}u = g$ becomes

$$(L + R)u = g$$

or

$$u = L^{-1}g - L^{-1}Ru, \tag{3.1}$$

where we assume that L^{-1} is defined. Now, we proceed as before but assume, for simplicity, that $g_1 = g_2 = 0$ and $\langle \mathcal{L}_2\rangle = 0$; i.e., we assume "small randomness" by letting $u = u_0 + \epsilon u_1 + \epsilon^2 u_2$. Then, we get

$$\langle u\rangle = L^{-1}g_0(1 - \epsilon L^{-1}\langle \mathcal{L}_1\rangle + \epsilon^2 L^{-1}\langle \mathcal{L}_1 L^{-1}\mathcal{L}_1\rangle),$$

the perturbation-theory result. However, suppose we simply call $L^{-1}g = u_0$ and let $u = u_0 - u_1 + u_2 \cdots$; i.e., we say nothing about smallness of R. Now, since

$$\begin{aligned}u &= u_0 - L^{-1}Ru_0 + L^{-1}Ru_1 - L^{-1}Ru_2 \cdots \\ &= L^{-1}g - L^{-1}RL^{-1}g + L^{-1}RL^{-1}RL^{-1}g \cdots, \\ \langle u\rangle &= L^{-1}g - L^{-1}\langle R\rangle L^{-1}g + L^{-1}\langle RL^{-1}R\rangle L^{-1}g \cdots \\ &= L^{-1}g(1 - L^{-1}\langle R\rangle + L^{-1}\langle RL^{-1}R\rangle \cdots),\end{aligned}$$

we see that, if R is small, this is the same as perturbation theory. However, if it is not and the convergence can be established for cases of interest, the results may be of interest. Consequently, we will presently examine this in much more detail.

4. HIERARCHY EQUATIONS

A. Introduction

This approach,[11] which has received considerable use in theoretical physics,[25] involves obtaining a

hierarchy of equations and using some truncation procedure (closure approximation) to terminate the hierarchy. Thus, given a scalar function $u(x)$, the following stochastic equation is satisfied:

$$\mathfrak{L}u(x) = g(x), \tag{4.1}$$

where the stochastic operator $\mathfrak{L} = L_x + n(x)$ and $g(x)$ is a deterministic function. Here we assume that L_x is a deterministic operator; but $n(x)$ is random [so, of course, $u(x)$ is random]. If Eq. (4.1) is averaged in an attempt to find $\langle u(x)\rangle$, we find that

$$L_x\langle u(x)\rangle + \langle n(x)u(x)\rangle = g(x).$$

But, in the second term on the left side, $n(x)$ and $u(x)$ are not statistically independent and cannot be separated for the determination of $\langle u(x)\rangle$. If, to determine $\langle n(x)u(x)\rangle$, we multiply Eq. (4.1) by $n(x)$ and average again, we get the term $\langle n(x)L_x u(x)\rangle$; but L_x does not commute, in general, with $n(x)$. However, $n(x_1)$ can commute with L_x; therefore, we multiply by $n(x_1)$ and average to get

$$L_x\langle n(x_1)u(x)\rangle + \langle n(x_1)n(x)u(x)\rangle = \langle n(x_1)\rangle g(x). \tag{4.2}$$

If we could solve this for the first term, evaluate the result as it approaches $x_1 = x$ and substitute back, we would get $\langle u(x)\rangle$. Aside from the generally ignored question of the validity of letting $x_1 = x$, we find that we now need the new moment $\langle n(x_1)n(x)u(x)\rangle$; thus, it is necessary to repeat the procedure, multiplying by $n(x_2)$. We are led to an infinite set of equations called the hierarchy equations for the hierarchy of moments. If we want also higher moments of $u(x)$, we proceed in the same manner multiplying by $u(x_1)$, etc. Thus, the solution involves averaging first (which will later be shown to be wrong) and an arbitrary truncation procedure which has been justified only by its arbitrary assumption.

If, at any level in the hierarchy, the "closure approximation" can be made, then the hierarchy can be terminated. Thus, if

$$\langle n(x_1)n(x)u(x)\rangle = \langle n(x_1)n(x)\rangle\langle u(x)\rangle$$

in (4.2), we can solve for $\langle n(x_1)u(x)\rangle$, evaluate at $x_1 = x$, and substitute into Eq. (4.1) to "solve" for $\langle u(x)\rangle$, assuming L_x is a known operator and the initial conditions are specified. One other possibility exists but appears to be of little physical interest. If there exists a very small correlation length for $n(x)$ compared to u, i.e., if n is varying very rapidly compared to u, the truncation may be justified. However, since a, in a sense, is the solution to be found, *a priori* assumptions on its behavior are undesirable.

The same procedure has been used to get a correlation function for u or its higher moments. The general case gives the infinite set of equations[26,27] (hierarchy equations)

$$\begin{aligned} &L_x\langle u(x)u(x_1)\cdots u(x_j)n(x_{j+1})\cdots n(x_{j+k})\rangle \\ &\quad + \langle n(x)u(x)u(x_1)\cdots u(x_j)n(x_{j+1})\cdots n(x_{j+k})\rangle \\ &\quad = g(x)\langle u(x_1)\cdots u(x_j)n(x_{j+k})\rangle, \quad j, k = 0, 1, 2, \cdots. \end{aligned}$$

The source term $g(x)$ is assumed to be statistically independent of the parameter $n(x)$. We see that an infinite set of equations are needed to find all the moments, i.e., that any moment involves all the moments of higher order and a closure procedure is necessary to get a cutoff. Thus, in the kth member of the hierarchy, we set

$$\begin{aligned} &\langle n(x_1)\cdots n(x_{k-1})n(x)u(x)\rangle \\ &\qquad \cong \langle n(x_1)\cdots n(x_{k-1})n(x)\rangle\langle u(x)\rangle. \end{aligned}$$

B. Connection to Perturbation Theory

We set $\mathfrak{L}_2 = 0$ for simplicity and also assume $\langle \mathfrak{L}_1\rangle = 0$. Taking g as nonrandom, we have

$$\mathfrak{L}u = g,$$
$$[L + \epsilon\mathfrak{L}_1(\alpha)]u(\alpha, \epsilon) = g + O(\epsilon^3).$$

The expectation [dropping $O(\epsilon^3)$] is

$$L\langle u\rangle + \epsilon\langle \mathfrak{L}_1 u\rangle = g. \tag{4.3}$$

But $\langle \mathfrak{L}_1 u\rangle$ is unknown and cannot be separated without perturbation theory. To see if perturbation theory can be avoided, we write, before averaging again,

$$Lu + \epsilon\mathfrak{L}_1 u = g.$$

We multiply by L^{-1}

$$u + \epsilon L^{-1}\mathfrak{L}_1 u = L^{-1}g.$$

Now we multiply by $\mathfrak{L}_1$ and average to get an expression for $\langle \mathfrak{L}_1 u\rangle$. Thus,

$$\langle \mathfrak{L}_1 u\rangle + \epsilon\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1 u\rangle = \langle \mathfrak{L}_1\rangle L^{-1}g = 0.$$

To avoid going on forever, we assume blindly (closure approximation or so-called "local independence") that

$$\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1 u\rangle \cong \langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle.$$

Now,

$$\langle \mathfrak{L}_1 u\rangle = -\epsilon\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle. \tag{4.4}$$

Substituting Eq. (4.4) into Eq. (4.3), we obtain

$$L\langle u\rangle - \epsilon^2\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle = g,$$

which is the same answer as derived by perturbation theory. The significance will be obvious shortly. We could go further in the hierarchy, of course, before invoking "closure." We can return to the expression

for u, multiply by $\mathfrak{L}_1 L^{-1}\mathfrak{L}_1$, and average to get

$$\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1 u\rangle + \epsilon\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1 L^{-1}\mathfrak{L}_1 u\rangle = 0,$$

again separating the second term by assumption into $\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle$.

Now, let g be stochastic:

$$\mathfrak{L}u = g,$$
$$[L + \epsilon\mathfrak{L}_1 + \epsilon^2\mathfrak{L}_2 + O(\epsilon^3)]u = g,$$
$$L\langle u\rangle + \epsilon\langle \mathfrak{L}_1 u\rangle = \langle g\rangle.$$

In the perturbation approach, we have

$$u = u_0 - \epsilon L^{-1}\mathfrak{L}_1 u_0 + O(\epsilon^2).$$

If we use this and multiply by $\mathfrak{L}_1$, we obtain

$$\mathfrak{L}_1 u = \mathfrak{L}_1 u_0 - \epsilon\mathfrak{L}_1 L^{-1}\mathfrak{L}_1 u_0 + O(\epsilon^2),$$
$$\langle \mathfrak{L}_1 u\rangle = \langle \mathfrak{L}_1\rangle\langle u_0\rangle - \epsilon\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u_0\rangle + O(\epsilon^2)$$
$$= -\epsilon\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u_0\rangle$$
$$= -\epsilon\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle.$$

Therefore,

$$L\langle u\rangle - \epsilon^2\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle = \langle g\rangle + O(\epsilon^3),$$

which is the same as the previous result if $\mathfrak{L}_2 = 0$ and $\langle \mathfrak{L}_1\rangle = 0$. Thus, as Keller has pointed out, when the randomness is small, i.e., when the perturbation-theory approach is useful, the average $\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1 u\rangle$ can be written $\langle \mathfrak{L}_1 L^{-1}\mathfrak{L}_1\rangle\langle u\rangle$, and it is not necessary to assume it. It works for perturbation theory, but then it is not needed. If $\mathfrak{L}$ is a differential operator, it is obviously wrong to assume statistical independence of $\mathfrak{L}$ and its operand u, although it is reasonable to assume statistical independence of $\mathfrak{L}$ and g.

The hierarchy method has been widely used in statistical field physics, both for the linear and nonlinear case. In the quantum-mechanical many-body problem or in the theory of turbulence,[*,27] nonlinear stochastic equations arise. The dynamical equations lead to an infinite hierarchy of coupled equations in which given ensemble averages are related to successively higher-order terms. This difficulty occurs, as Kraichnan[27] has pointed out, in the linear case as well. The closure approximation (a truncation and closure of the hierarchy) is always assumed to find an approximation for the desired statistical quantities, but the validity or error involved has not been adequately discussed. It is interesting to note that, if we consider

$$[L + \alpha(t)]u(t) = g,$$

and average it to get

$$L\langle u(t)\rangle + \langle \alpha(t)u(t)\rangle = \langle g\rangle, \tag{4.5}$$

difficulties arise even in truncating at the next level. Thus, if we write (assuming the separation)

$$L\langle \alpha(t_1)u(t)\rangle + \langle \alpha(t_1)\alpha(t)\rangle\langle u(t)\rangle = \langle \alpha(t_1)\rangle\langle g\rangle = 0,$$

it follows that

$$L\langle \alpha(t_1)u(t)\rangle = -\langle \alpha(t_1)\alpha(t)\rangle\langle u(t)\rangle,$$

[it has been assumed that L (or L_t) commutes with $\alpha(t_1)$ which is the reason for not multiplying by $\alpha(t)$]. However, if we then evaluate at $t_1 = t$, a commutator is neglected. (Furthermore, the separation is not valid when $t_1 = t$.) Let us first solve the above for $\langle \alpha(t_1)u(t)\rangle$ by multiplying by L^{-1}, i.e.,

$$\langle \alpha(t_1)u(t)\rangle = -L_t^{-1}\langle \alpha(t_1)\alpha(t)\rangle\langle u(t)\rangle$$
$$= -\langle \alpha(t_1)L_t^{-1}\alpha(t)\rangle\langle u(t)\rangle.$$

Now,

$$\langle \alpha(t)u(t)\rangle = \langle \alpha(t_1)u(t)\rangle_{t_1=t}$$
$$= -\langle \alpha(t)L_t^{-1}\alpha(t)\rangle\langle u(t)\rangle. \tag{4.6}$$

By substituting (4.6), we can rearrange Eq. (4.5) as follows:

$$\langle u(t)\rangle$$
$$= L^{-1}\langle g\rangle[1 + \langle \alpha(t)L_t^{-1}\alpha(t)\rangle]$$
$$= \int_0^t \ell(t-\tau)\langle g(\tau)\rangle\left[1 + \left\langle \alpha(t)\int_0^t \ell(t-\tau)\alpha(\tau)\,d\tau\right\rangle\right]d\tau$$
$$= \int_0^t \ell(t-\tau)\langle g(\tau)\rangle\left[1 + \int_0^t \ell(t-\tau)R_\alpha(t-\tau)\,d\tau\right]d\tau,$$

where R_α is the correlation function for α. We observe that we have $\langle u\rangle$ in terms of $\langle g\rangle$ and an integral kernel or stochastic Green's function[5,12,20]

$$\ell(t-\tau)\left[1 + \int_0^t \ell(t-\tau)R_\alpha(t-\tau)\,d\tau\right],$$

where ℓ and R_α are given.

Suppose that, instead of multiplying by $\alpha(t_1)$, we solve for $u(t)$ and multiply by $\alpha(t)$; thus,

$$Lu(t) + \alpha(t)u(t) = g,$$
$$u(t) = L^{-1}g - L^{-1}\alpha(t)u(t).$$

Multiplying by $\alpha(t)$, we obtain

$$\alpha(t)u(t) = \alpha(t)L^{-1}g - \alpha(t)L^{-1}\alpha(t)u(t),$$

and averaging,

$$\langle \alpha(t)u(t)\rangle = -\langle \alpha(t)L^{-1}\alpha(t)\rangle\langle u(t)\rangle,$$

if we *assume* the separability on the right and a zero mean process for $u(t)$. This obviously gives the same result as the previous procedure if we substitute back into the first averaged equation (4.5). This avoids the commutator problem but makes it still less clear why the separation can be assumed, since $\alpha(t)$ cannot be said to be statistically independent of $u(t)$.

The basic error here is in the averaging. Averaging the equation $\mathfrak{L}u = g$ results in $\langle \mathfrak{L}u \rangle = \langle g \rangle$. If $\mathfrak{L} = L + R$, we have $L\langle u \rangle + \langle Ru \rangle = \langle g \rangle$. If $\mathfrak{L}$ is replaced by L, which clearly means throwing away the R so we have a simple deterministic equation, then $\langle \mathfrak{L}u \rangle$ becomes $\langle Lu \rangle = L\langle u \rangle$, which can be rewritten $\langle \mathfrak{L} \rangle \langle u \rangle$. The misconception has appeared in the literature that the solution of the averaged equation is the expected value of the solution of the random equation, which we see is true if $\mathfrak{L}$ is replaced by L, i.e., if $R = 0$. At the next level of the hierarchy, we write

$$Lu + Ru = g, \tag{4.7}$$

$$u = L^{-1}g - L^{-1}Ru, \tag{4.8}$$

$$\langle u \rangle = L^{-1}\langle g \rangle - L^{-1}\langle Ru \rangle. \tag{4.9}$$

To evaluate $\langle Ru \rangle$ in (4.9), we return to (4.8) and multiply by R to get

$$Ru = RL^{-1}g - RL^{-1}Ru, \tag{4.10}$$

which we average to get

$$\langle Ru \rangle = \langle R \rangle L^{-1}\langle g \rangle - \langle RL^{-1}Ru \rangle. \tag{4.11}$$

If $\langle R \rangle = 0$,

$$\langle Ru \rangle = -\langle RL^{-1}Ru \rangle. \tag{4.12}$$

Now, just as we replaced $\mathfrak{L}$ by L in order to separate $\langle \mathfrak{L}u \rangle$, we now replace $RL^{-1}R$ by $\langle RL^{-1}R \rangle$ in the right-hand side of (4.12). Then $\langle Ru \rangle = -\langle RL^{-1}R \rangle \langle u \rangle$.

Thus,

$$\langle u \rangle = L^{-1}\langle g \rangle + L^{-1}\langle RL^{-1}Ru \rangle$$

becomes

$$\langle u \rangle = L^{-1}\langle g \rangle + L^{-1}\langle RL^{-1}R \rangle \langle u \rangle.$$

$\langle u \rangle$ is the solution of the averaged equation with the operator $RL^{-1}R$ replaced by its average. We, of course, throw away the random part of $RL^{-1}R$, just as we would throw away the random part of $\mathfrak{L}$ if we did this at the first stage of the hierarchy.

It clearly is not true that the averaged equation gives the expectation of the solution of the stochastic equation since, at both levels of the hierarchy discussed, we have replaced a random operator by a deterministic one and, therefore, have lost something.

[In Ref. 28, p. 14, we see that the replacement of the operator by its average should not give the expectation $\langle y \rangle$ since all the terms involving correlation[5] of α would be lost.]

Many treatments of random equations in applications—particularly for waves propagating in a random medium—while differing in detail, are essentially variations of the hierarchy or perturbation methods and, further, involve many questionable and restrictive assumptions. The methods which follow remove many of the restrictions. Applications based on the present discussion will be separately reported.

5. ITERATIVE METHOD

A. Introduction

Suppose we consider[29] the stochastic differential equation $\mathfrak{L}y = x$, where

$$\mathfrak{L} = \sum_{\nu=0}^{n} a_\nu(t) \frac{d^\nu}{dt^\nu}$$

and $x(t)$ and the $a_\nu(t)$ are random functions (or stochastic processes) whose statistics (statistical measures) are known. It is assumed that $x(t; \omega)$ is statistically independent of the $a_\nu(t; \omega)$. The objective is to determine the stochastic Green's function for the desired statistical measure of y, expressed directly in terms of the statistics of the coefficients $a_\nu(t)$. Then, the solution in a statistical sense will have been obtained and we will have avoided asking for a "solution" of the stochastic equation, which is then to be used to find the statistics. The following is based on the earlier work of Samuels[2,3] and on a suggestion of Adomian[12] for an approximation method eliminating the unjustifiable assumptions of *a priori* spectral separation or the closure approximation. The method is elementary in principle but can be extended to partial differential equations and random fields.[7] Suppose that $\mathfrak{L} = L + R$, where L is a deterministic operator and R is a random operator. Then, we can write

$$Ly = x - Ry.$$

To do this, we assume, for example, that

$$a_\nu(t) = \langle a_\nu(t) \rangle + \alpha_\nu(t).$$

Thus,

$$L = \sum_{\nu=0}^{n} \langle a_\nu(t) \rangle \frac{d^\nu}{dt^\nu}$$

and

$$R = \sum_{\nu=0}^{n} \alpha_\nu(t) \frac{d^\nu}{dt^\nu}.$$

Now, writing $\ell(t, \tau)$ as the ordinary Green's function corresponding to the deterministic operator L, we have

$$\begin{aligned} y &= L^{-1}x - L^{-1}Ry \\ &= \int_0^t \ell(t, \tau)x(\tau)\, d\tau - \int_0^t \ell(t, \tau) \sum_{\nu=0}^{n} \alpha_\nu(\tau) \frac{d^\nu y(\tau)}{d\tau^\nu}\, d\tau \\ &\quad + \sum_{\nu=0}^{n} c_\nu \phi_\nu(t). \end{aligned}$$

For simplicity we will take coefficients in L to be constants; $\ell(t, \tau)$ is actually $\ell(t - \tau)$. The limits are a matter of choice. The upper limit can be written

∞ as well since the Green's function is zero if $t - \tau < 0$ for a realizable system. The lower limit can be $-\infty$ or t_0 depending on the memory involved. Using t_0 implies that the earlier values are not significant. The ϕ_ν are a set of independent solutions of the homogeneous equation $Ly = 0$ and the c_ν are arbitrary constants. This can be written as

$$y(t) = F(t) - \int k(t, \tau)y(\tau)\, d\tau,$$

where

$$F(t) = \int \ell(t, \tau)x(\tau)\, d\tau + \sum_{\nu=0}^{n} c_\nu \phi_\nu(t).$$

[It is interesting that the first expression is similar to that for a closed system with feedback. In the first term, $\ell(t, \tau)$ is the impulse function of the forward loop and, in the second term, $k(t, \tau)$ is the feedback impulse function.]

Integration by parts of the remaining integral or use of Green's formula in terms of the adjoint operator yields

$$k(t, \tau) = \sum_{\nu=0}^{n} (-1)^\nu \frac{d^\nu}{d\tau^\nu} [\alpha_\nu(\tau)\ell(t, \tau)].$$

The expectation of y is

$$\langle y(t)\rangle = \langle F(t)\rangle - \int \langle k(t, \tau)y(\tau)\rangle\, d\tau. \qquad (5.1)$$

The difficulties arise in the last term in attempting to separate the ensemble average or expectation into the product of averages over k and over y, since y is not statistically independent of the α involved in k (i.e., in R). It is desirable at this point to avoid the use of *a priori* spectral separation or either explicit or implicit assumptions of small amount of randomness in $\mathfrak{L}$.

If we let the random part be small (e.g., if we let each coefficient $a_i = \gamma_i + \epsilon\alpha_i$ and use the same ϵ in a perturbation series for y), we can then achieve a separation. But without assuming a small random part, slow variations, or white noise, we use in the place of a perturbation series a simple alternating series $y(t) = y_0 - y_1 + y_2 - \cdots$; i.e.,

$$y(t) = \sum_{i=0}^{\infty} (-1)^i y_i(t).$$

This is suggested by the Born expansion and Born approximation used in a most important class of problems involving continuous eigenvalues—that of scattering theory. The state function for an elastic scattering problem satisfies the partial differential equation

$$(\nabla^2 + k^2)\psi(r) = U(r)\psi(r)$$

or, in operator form,

$$F[\psi(r)] = g(r) \equiv U(r)\psi(r).$$

By the Green's function method we write

$$\psi(r) = F^{-1}[g(r)]$$
$$= \psi_0(r) + \int G(r, r')U(r')\psi(r')\, dr',$$

where $\psi_0(r)$ is the solution of the homogeneous equation $F[\psi(r)] = 0$ or $(\nabla^2 + k^2)\psi_0 = 0$. The general problem is a difficult one involving the solution of the integral equation subject to the boundary conditions and the proper asymptotic form for ψ at large distances from the scattering center. One solves by iteration taking ψ_0 as a zeroth approximation and the nth approximation from ψ_{n-1} in the integrand. If the scattering is small, the first approximation is sufficient. We will now proceed in the same manner taking $F(t)$ as y_0. Thus, y_0 is the solution, neglecting R in the original equation. For simplicity, we write $\langle a_\nu(t)\rangle = a_\nu$. If the a_ν are constants, $\ell(t, \tau) = \ell(t - \tau)$. We assume the $a_\nu(t)$ are zero-mean random processes so that $\langle \alpha_\nu(t)\rangle = 0$. If the α_ν are zero, the solution is just $F(t)$. We will neglect the homogeneous solution ϕ_ν for the purposes of this paper. We will not consider stability questions here but only stable solutions. We now have

$$y_0 = F(t) = \int \ell(t, \tau)x(\tau)\, d\tau$$

and each successive y_i is given in terms of the preceding y_{i-1}; thus,

$$y_i = \int k(t, \tau)y_{i-1}(\tau)\, d\tau, \quad i = 1, 2, \cdots.$$

The convergence question is crucial to get more than a formal solution and we will return to it. However, the immediately interesting point is that the identification of $F(t)$ as the first term and the iterative procedure result in the desired separation of ensemble averages of the α's and y. This is because each term of y_i can be worked backward through y_{i-1}, y_{i-2}, to y_0, or $F(t)$, which depends only on $\ell(t, \tau)$ and on $x(t)$ but not on the α's. Thus, the assumption of statistical independence of $\mathfrak{L}$ and x is sufficient, and one does not need to invoke separation of y and the α's by assuming it *a priori* or using a "closure" approximation. Further, we get a *reversed* hierarchy, where each term involves the one before it rather than the one after it. If the randomness *is* small, this gives the same result as perturbation theory. That is, the perturbation expansion accomplishes the same result but is only valid if the expansion parameter is small (small randomness). If the randomness is not small, the number of terms can go to infinity. Before considering the matter further, let us see if the procedure appears useful for higher moments. We can, e.g., find the

correlation function $R_y(t_1, t_2) = \langle y(t_1)y^*(t_2)\rangle$ by multiplying $y(t_1)$ and $y^*(t_2)$ and averaging. Thus,

$$\begin{aligned} R_y(t_1, t_2) &= \langle y(t_1)y^*(t_2)\rangle \\ &= \left\langle \left[F(t_1) - \int k(t_1, \tau_1)y(\tau_1)\,d\tau_1\right] \right. \\ &\quad \left. \times \left[F^*(t_2) - \int k^*(t_2, \tau_2)y^*(\tau_2)\,d\tau_2\right]\right\rangle \\ &= \langle F(t_1)F^*(t_2)\rangle - \int \langle k(t_1, \tau_1)F^*(t_2)y(\tau_1)\rangle\,d\tau_1 \\ &\quad - \int \langle k^*(t_2, \tau_2)F(t_1)y^*(\tau_2)\rangle\,d\tau_2 \\ &\quad + \iint \langle k(t_1, \tau_1)k^*(t_2, \tau_2)y(\tau_1)y^*(\tau_2)\rangle\,d\tau_1\,d\tau_2. \end{aligned}$$

Examination of this expression term by term will show that all the averages will separate in the manner we have discussed when the iteration series is substituted. This is because, each time a y_k is worked backward to y_{k-1}, another k and another integration appear. When we get to y_0, statistical independence of x and the α's allows us to separate the entire average into a product of two averages, one over the product of all the k's and the other involving only the x, which is identified as the autocorrelation of x. Examining the expression for $R_y(t_1, t_2)$ term by term shows that the separation always occurs. Thus, the first term is

$$\begin{aligned} \langle F(t_1)F^*(t_2)\rangle &= \int d\tau \int d\sigma \ell(t_1, \tau)\ell(t_2, \sigma)\langle x(\tau)x(\sigma)\rangle \\ &= \int d\tau \int d\sigma \ell(t_1, \tau)\ell(t_2, \sigma)R_x(\tau, \sigma). \end{aligned}$$

The integrand of the second term R_y with the series substitution is

$$\begin{aligned} &\langle k(t_1, \tau_1)F^*(t_2)y(\tau_1)\rangle \\ &\qquad = \langle k(t_1, \tau_1)F^*(t_2)[y_0(\tau_1) - y_1(\tau_1) + \cdots]\rangle. \end{aligned}$$

The first term of this is zero, since

$$\begin{aligned} \langle k(t_1, \tau_1)F^*(t_2)y_0(\tau_1)\rangle &= \langle k(t_1, \tau_1)F^*(t_2)F(\tau_1)\rangle \\ &= \langle k(t_1, \tau_1)\rangle\langle F^*(t_2)F(\tau_1)\rangle \\ &= 0, \end{aligned}$$

because $\langle \alpha_\nu\rangle = 0$. The second term of the same expression (second term of R_y) is

$$\begin{aligned} &\langle k(t_1, \tau_1)F^*(t_2)y_1(\tau_1)\rangle \\ &\qquad = \left\langle k(t_1, \tau_1)F^*(t_2)\int k(\tau_1, \sigma)F(\sigma)\,d\sigma\right\rangle \\ &\qquad = \int d\sigma \langle k(t_1, \tau_1)k(\tau_1, \sigma)\rangle\langle F^*(t_2)F(\sigma)\rangle. \end{aligned}$$

The third term of the same expression is

$$\begin{aligned} &\langle k(t_2, \tau_1)F^*(t_2)y_2(\tau_1)\rangle \\ &\qquad = \iint \langle k(t_1, \tau_1)k(\tau_2, \beta)k(\beta, \gamma)\rangle\langle F^*(t_2)F(\gamma)\rangle\,d\beta\,d\gamma. \end{aligned}$$

Thus, it is clear that the separations always occur, but successive terms of the kernel will go to higher and higher moments even though only the correlation of the output is required. Remembering that the terms we discussed and the higher moments not explicitly shown were for the second term of R_y, we can write down similar expressions by inspection for the third term of R_y, etc. Finally, we consider the last term of the expression,

$$\begin{aligned} &\iint d\tau_1\,d\tau_2\langle k(t_1, \tau_1)k^*(t_2, \tau_2)y(\tau_1)y^*(\tau_2)\rangle \\ &\qquad = \iint d\tau_1\,d\tau_2\langle k(t_1, \tau_1)k^*(t_2, \tau_2) \\ &\qquad\quad \times [y_0(\tau_1) - y_1(\tau_1) + y_2(\tau_1) - \cdots] \\ &\qquad\quad \times [y_0^*(\tau_2) - y_1^*(\tau_2) + y_2^*(\tau_2) - \cdots]\rangle. \end{aligned}$$

To shorten the writing, let $\int d\tau_1 \int d\tau_2$ be denoted by $\int d\tau$ and $k(t_1, \tau_1)k^*(t_2, \tau_2) = k_1k_2$ which results in

$$\begin{aligned} &\int d\tau\langle k_1k_2y_0(\tau_1)y_0^*(\tau_2)\rangle - \int d\tau\langle k_1k_2y_0(\tau_1)y_1^*(\tau_2)\rangle \\ &\qquad + \int d\tau\langle k_1k_2y_0(\tau_1)y_2^*(\tau_2)\rangle \cdots \\ &\qquad - \int d\tau\langle k_1k_2y_1(\tau_1)y_0^*(\tau_2)\rangle \\ &\qquad + \int dt\langle k_1k_2y_1(\tau_1)y_1^*(\tau_2)\rangle - \cdots, \end{aligned}$$

all of which separate. Thus, in the general expression

$$R_y(t_1, t_2) = \iint G_{\mathcal{K}}(t_1, t_2, \tau_1, \tau_2)R_x(\tau_1, \tau_2)\,d\tau_1\,d\tau_2,$$

we can identify and calculate the stochastic Green's function $G_{\mathcal{K}}$ as a series of terms involving moments of α. Assumption of stationarity simplifies the expressions. If the α_ν are assumed to be Gaussian, products involving an odd number of k's, like $\langle k_1k_2k_3\rangle$, are zero. The even product $\langle k_1k_2k_3k_4\rangle$ involves three averages by pairs; thus,

$$\begin{aligned} &\langle k_1k_2k_3k_4\rangle \\ &\qquad = \langle k_1k_2\rangle\langle k_3k_4\rangle + \langle k_2k_3\rangle\langle k_1k_4\rangle + \langle k_1k_3\rangle\langle k_2k_4\rangle. \end{aligned}$$

The term with $6k$'s involves 15 averages by pairs, etc. However, all of these are given in terms of the autocorrelation of the α, which is known. Thus, the output autocorrelation (statistical measure) is given in terms of the input autocorrelation and a stochastic

Green's function, depending only on the known autocorrelation of $\alpha_v(t)$.

Of course, the input and the operator need not be time dependent. The Green's function can be the response for an impulse in time or frequency or position depending on the quantities of interest in a physical problem. If an explicit expression for $G_{\mathfrak{L}}$ is written for the Gaussian case, the results should correlate with the stochastic Green's function for the first-order stochastic differential equation.[19]

Further results for the iterative approach will deal with the case of a differential operator involving a single random coefficient.

B. Differential Operator with One Random Coefficient

In the stochastic differential operator

$$\mathfrak{L} = \sum_{v=0}^{n} a_v(t) \frac{d^v}{dt^v},$$

let all the $a_v(t)$ be deterministic except one. Because it will be easier to show boundedness, we will suppose the one random coefficient is $a_0(t) = \langle a_0(t)\rangle + \alpha(t)$ and assume that the nonrandom coefficients are actually constants, so that $\ell(t, \tau) = \ell(t - \tau)$. The random coefficient need not be a_0; it can be $a_i(t)$ where $i < n$ and the same treatment will apply since, in $(L + R)y = x$,

$$y = L^{-1}x - L^{-1}Ry \tag{5.2a}$$

$$= \int_0^t \ell(t-\tau)x(\tau)\,d\tau - \int_0^t \ell(t-\tau)R_\tau[y(\tau)]\,d\tau \tag{5.2b}$$

$$= \int_0^t \ell(t-\tau)x(\tau)\,d\tau - \int_0^t R_\tau^\dagger[\ell(t-\tau)]y(\tau)\,d\tau \tag{5.2c}$$

$$= \int_0^t \ell(t-\tau)x(\tau)\,d\tau - \int_0^t (-1)^i \frac{d^i}{d\tau^i}[\ell(t-\tau)\alpha_i(\tau)]y(\tau)\,d\tau \tag{5.2d}$$

$$= y_0(t) - \int_0^t k(t,\tau)y(\tau)\,d\tau, \tag{5.2e}$$

where only the form of k changes if $i \neq 0$. Expression (5.2c) is obtained by use of Green's formula (the bilinear concomitant is zero).

However, we consider the case where a_0 is random as stated. The equation to be considered is, therefore, of the form

$$[L + \alpha(t)]y(t) = x(t),$$

where L is the deterministic operator and $\alpha(t)$ and $x(t)$ are stochastic processes. We write

$$y(t) = F(t) - \int_0^t k(t,\tau)y(\tau)\,d\tau,$$

where

$$F(t) = \int_0^t \ell(t,\tau)x(\tau)\,d\tau.$$

Here $F(t)$ properly includes the solutions of the homogeneous equation, which we have neglected for simplicity, $\ell(t, \tau)$ is the Green's function for L, and $k(t, \tau) = \ell(t, \tau)\alpha(\tau)$. We let $y(t)$ be given by

$$y(t) = \sum_{i=0}^{\infty} (-1)^i y_i(t)$$

and identify $F(t)$ as y_0. Then,

$$\begin{aligned} y(t) &= F(t) - \int_0^t k(t,\tau)y_0(\tau)\,d\tau + \int_0^t k(t,\tau)y_1(\tau)\,d\tau \\ &\quad - \int_0^t k(t,\tau)y_2(\tau)\,d\tau + \cdots \\ &= F(t) - \int_0^t k(t,\tau)F(\tau)\,d\tau \\ &\quad + \int_0^t d\tau \int_0^\tau d\gamma\, k(t,\tau)k(\tau,\gamma)F(\gamma) \\ &\quad - \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\sigma\, k(t,\tau)k(\tau,\gamma)k(\gamma,\sigma)F(\sigma) \\ &\quad + \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\sigma \int_0^\sigma d\xi\, k(t,\tau)k(\tau,\gamma) \\ &\quad \times k(\gamma,\sigma)k(\sigma,\xi)F(\xi) - \cdots. \end{aligned}$$

Taking the expectation, we obtain

$$\begin{aligned} \langle y(t)\rangle &= \langle F(t)\rangle - \int_0^t d\tau \langle k(t,\tau)\rangle\langle F(\tau)\rangle \\ &\quad + \int_0^t d\tau \int_0^\tau d\gamma \langle k(t,\tau)k(\tau,\gamma)\rangle\langle F(\gamma)\rangle \\ &\quad - \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\sigma \langle k(t,\tau)k(\tau,\gamma)k(\gamma,\sigma)\rangle\langle F(\sigma)\rangle \\ &\quad + \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\sigma \int_0^\sigma d\xi \\ &\quad \times \langle k(t,\tau)k(\tau,\gamma)k(\gamma,\sigma)k(\sigma,\xi)\rangle\langle F(\xi)\rangle \cdots, \end{aligned}$$

where the ensemble averages all separate [assuming only the statistical independence of α and x, since each y_i is expressed in terms of the preceding y_{i-1} until we get to y_0 which is $F(t)$ and which depends only on x]. We assume $F(t) \neq 0$ and have implicitly assumed the operator $\mathfrak{L}$ can be written as the sum of L and R and that the Green's function for L exists. Since the central limit theorem is so often involved in physical applications, it is of interest now to investigate the result if $\alpha(t)$ is assumed to be Gaussian. Also, for simplicity, let both $\alpha(t)$ and $x(t)$ be stationary processes.

The assumption that $\alpha(t)$ is a (zero-mean) Gaussian process causes the terms involving the expectation of an odd number of k's (or α's) above to vanish. Thus,

$$\begin{aligned}\langle y(t)\rangle = &\int_0^t d\tau \ell(t,\tau)\langle x(\tau)\rangle \\ &+ \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\delta \ell(t,\tau)\ell(\tau,\gamma)\ell(\gamma,\delta) \\ &\times \langle \alpha(\tau)\alpha(\gamma)\rangle\langle x(\delta)\rangle \\ &+ \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\delta \int_0^\delta d\sigma \int_0^\sigma d\eta \ell(t,\tau)\ell(\tau,\gamma) \\ &\times \ell(\gamma,\delta)\ell(\delta,\sigma)\ell(\sigma,\eta) \\ &\times \langle \alpha(\tau)\alpha(\gamma)\alpha(\delta)\alpha(\sigma)\rangle\langle x(\eta)\rangle + \cdots . \end{aligned} \tag{5.3}$$

Let

$$\langle \alpha(\tau)\alpha(\gamma)\rangle = R_\alpha(\gamma - \tau)$$

and

$$\begin{aligned}&\langle \alpha(\tau)\alpha(\gamma)\alpha(\delta)\alpha(\sigma)\rangle \\ &\quad = R_\alpha(\gamma-\tau)R_\alpha(\sigma-\delta) + R_\alpha(\delta-\tau)R_\sigma(\sigma-\gamma) \\ &\qquad + R_\alpha(\sigma-\tau)R_\alpha(\delta-\gamma)\end{aligned}$$

because of the Gaussian property. Now,

$$\begin{aligned}\langle y(t)\rangle = &\int_0^t d\tau \ell(t,\tau)\langle x(\tau)\rangle \\ &+ \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\delta \ell(t,\tau)\ell(\tau,\gamma) \\ &\times \ell(\gamma,\delta)R_\alpha(\gamma-\tau)\langle x(\delta)\rangle \\ &+ \int_0^t d\tau \int_0^\tau d\gamma \int_0^\gamma d\delta \int_0^\delta d\sigma \int_0^\sigma d\eta \ell(t,\tau)\ell(\tau,\gamma) \\ &\times \ell(\gamma,\delta)\ell(\delta,\sigma)\ell(\sigma,\eta) \\ &\times [R_\alpha(\gamma-\tau)R_\alpha(\sigma-\delta) + R_\alpha(\delta-\tau)R_\alpha(\sigma-\gamma) \\ &+ R_\alpha(\sigma-\tau)R_\alpha(\delta-\gamma)]\langle x(\eta)\rangle + \cdots .\end{aligned}$$

We now make the following changes of variables: In the second term, $\gamma - \tau = \xi$, $d\gamma = d\xi$; in the third term, $\gamma - \tau = \xi$, $d\gamma = d\xi$; also $\sigma - \delta = \rho$; $\delta - \tau$ is unchanged; $\sigma - \gamma = \delta + \rho - \tau - \xi$, $\sigma - \tau = \delta + \rho - \tau$, and $\delta - \gamma = \delta - \tau - \xi$.

With these changes we have

$$\begin{aligned}\langle y(t)\rangle = &\int_0^t d\tau \ell(t,\tau)\langle x(\tau)\rangle \\ &+ \int_0^t d\tau \int_{-\tau}^0 d\xi \int_0^{\xi+\tau} d\delta \ell(t,\tau) \\ &\times \ell(\tau,\tau+\xi)\ell(\tau+\xi,\delta)R(\xi)\langle x(\delta)\rangle \\ &+ \int_0^t d\tau \int_{-\tau}^0 d\xi \int_0^{\tau+\xi} d\delta \int_{-\delta}^0 d\rho \int_0^{\delta+\rho} d\eta \ell(t,\tau) \\ &\times \ell(\tau,\tau+\xi)\ell(\tau+\xi,\delta)\ell(\delta,\delta+\rho)\ell(\delta+\rho,\eta) \\ &\times [R(\xi)R(\rho) + R(\delta-\tau)R(\delta+\rho-\tau-\xi) \\ &+ R(\delta+\rho-\tau)R(\delta-\tau-\xi)]\langle x(\eta)\rangle + \cdots .\end{aligned}$$

We note that the first term

$$\int_0^t d\tau \ell(t,\tau)\langle x(\tau)\rangle = L^{-1}(t)\langle x(t)\rangle$$

by definition of the Green's function $\ell(t,\tau)$.

In the second term, the similar integral

$$\int_0^{\xi+\tau} d\delta \ell(\tau+\xi,\delta)x(\delta)$$

becomes $L^{-1}(\tau+\xi)\langle x(\tau+\xi)\rangle$ or $L^{-1}(\tau+\xi)\langle x(t)\rangle$ since x is stationary. The second term becomes

$$\begin{aligned}&\int_0^t d\tau \int_{-\tau}^0 d\xi L^{-1}(\xi+\tau)\ell(t,\tau)\ell(\tau,\tau+\xi)R(\xi)\langle x(t)\rangle \\ &\qquad = \int_0^t d\tau \ell(t,\tau)f(\tau) = L^{-1}(t)f(t),\end{aligned}$$

if we let

$$f(\tau) = \int_{-\tau}^0 d\xi L^{-1}(\xi+\tau)\ell(\tau,\tau+\xi)R(\xi)\langle x(t)\rangle .$$

Therefore, the second term is

$$\begin{aligned}&L^{-1}(t)\int_{-\tau}^0 d\xi L^{-1}(\xi+t)\ell(t,t+\xi)R(\xi)\langle x(t)\rangle \\ &\qquad = L^{-1}(t)\int_0^t d\tau L^{-1}(\tau)\ell(t,\tau)R(\tau-t)\langle x(t)\rangle \\ &\qquad = L^{-1}(t)\int_0^t d\tau \ell(t,\tau)f(\tau) \\ &\qquad = L^{-1}(t)L^{-1}(t)L^{-1}(t)R(0)\langle x(t)\rangle .\end{aligned}$$

In the third term,

$$\int_0^{\delta+\rho} d\eta \ell(\delta+\rho,\eta)g(\eta) = L^{-1}(\delta+\rho)g(\delta+\rho);$$

thus, the entire term can be written

$$\begin{aligned}&\int_0^t d\tau \int_{-\tau}^0 d\xi \int_0^{\tau+\xi} d\delta \int_{-\delta}^0 d\rho L^{-1}(\delta+\rho) \\ &\qquad \times \ell(t,\tau)\ell(\tau,\tau+\xi)\ell(\tau+\xi,\delta)\ell(\delta,\delta+\rho) \\ &\qquad \times [R(\xi)R(\rho) + R(\delta-\tau)R(\delta+\rho-\tau-\xi) \\ &\qquad + R(\delta+\rho-\tau)R(\delta-\tau-\xi)]\langle x(t)\rangle .\end{aligned}$$

Now, the ρ integration can be written as

$$\begin{aligned}&\int_0^\delta d\rho L^{-1}(\rho)\ell(t,\tau)\ell(\tau,\tau+\xi)\ell(\tau+\xi,\delta)\ell(\delta,\rho) \\ &\qquad \times [R(\xi)R(\rho-\delta) + R(\delta-\tau)R(\rho-\tau-\xi) \\ &\qquad + R(\rho-\tau)R(\delta-\tau-\xi)]\langle x(t)\rangle \\ &\quad = \int_0^\delta d\rho \ell(\delta,\rho)f(\rho) = L^{-1}(\delta)f(\delta) \\ &\quad = L^{-1}(\delta)L^{-1}(\delta)\ell(t,\tau)\ell(\tau,\tau+\xi)\ell(\tau+\xi,\delta) \\ &\qquad \times [R(\xi)R(0) + R(\delta-\tau)R(\delta-\tau-\xi) \\ &\qquad + R(\delta-\tau)R(\delta-\tau-\xi)]\langle x(t)\rangle ,\end{aligned}$$

where we let $\rho' = \rho + \delta$ and $d\rho' = d\rho$ (and then drop primes). The entire term is

$$\int_0^t d\tau \int_{-\tau}^0 d\xi \int_0^{\tau+\xi} d\delta \quad \text{(above expression)}$$
$$= \int_0^t d\tau \int_{-\tau}^0 d\xi \int_0^{\tau+\xi} d\delta \ell(\tau + \xi, \delta) f(\delta)$$
$$= \int_0^t d\tau \int_{-\tau}^0 d\xi L^{-1}(\tau + \xi) L^{-1}(\tau + \xi) L^{-1}(\tau + \xi)$$
$$\times \ell(t, \tau)\ell(\tau, \tau + \xi)$$
$$\times [R(\xi)R(0) + R(\xi)R(0) + R(\xi)R(0)]\langle x(t)\rangle.$$

Let $\tau + \xi = \xi'$ and $d\xi = d\xi'$ (and then drop primes). The above becomes

$$\int_0^t d\tau \int_0^\tau d\xi L^{-1}(\xi) L^{-1}(\xi) L^{-1}(\xi) \ell(t, \tau)\ell(\tau, \xi)$$
$$\times [R(\xi - \tau)R(0) + R(\xi - \tau)R(0)$$
$$+ R(\xi - \tau)R(0)]\langle x(t)\rangle$$
$$= \int_0^t d\tau \ell(t, \tau) f(\tau) = L^{-1}(t) f(t)$$
$$= L^{-1}(t) \int_0^t d\xi L^{-1}(\xi) L^{-1}(\xi) L^{-1}(\xi) \ell(t, t)\ell(t, \xi)$$
$$\times [R(\xi - t)R(0) + R(\xi - t)R(0)$$
$$+ R(\xi - t)R(0)]\langle x(t)\rangle$$
$$= L^{-1}(t) \int_0^t d\xi \ell(t, \xi) f(\xi) = L^{-1}(t) L^{-1}(t) f(t)$$
$$= L^{-1}(t) L^{-1}(t) L^{-1}(t) L^{-1}(t) L^{-1}(t)$$
$$\times [R(0)R(0) + R(0)R(0) + R(0)R(0)]\langle x(t)\rangle$$
$$= 3[L^{-1}(t)]^5 R^2(0)\langle x(t)\rangle.$$

Therefore,

$$\langle y(t)\rangle = L^{-1}(t)\langle x(t)\rangle + [L^{-1}(t)]^3 R(0)\langle x(t)\rangle$$
$$+ 3[L^{-1}(t)]^5 R^2(0)\langle x(t)\rangle + \cdots$$
$$= \int_0^t d\tau G(t, \tau)\langle x(\tau)\rangle,$$

where the "stochastic Green's function" for this particular "statistical measure," i.e., the expectation or mean, is

$$G(t, \tau) = \ell(t, \tau)[1 + [L^{-1}(\tau)]^2 R_\alpha(0)$$
$$+ 3[L^{-1}(\tau)]^4 R_\alpha^2(0) + \cdots].$$

Additional terms can be constructed without difficulty, since in Eq. (5.3) it is easy to write the following terms by inspection and use the expression in products of the correlation function for the expectation of the product of six, eight, or more α's. Then the procedure followed above is repeated over two additional integrations each time. Actually, the general term[30] can be found and the problem is easily adapted to machine computation. We see that the stochastic Green's function is an infinite series beginning with $\ell(t, \tau)$ with the next term involving a twofold integration, the next a fourfold integration, etc. If α is small enough so that $R_\alpha(0)$ is small enough, $G(t, \tau) \to \ell(t, \tau)$. The result here corresponds to the perturbation-theory result for the first-order correction since the Gaussian assumption does not affect the first correction term. Thus, $\langle y\rangle = L^{-1}\langle x\rangle + L^{-1}\langle \alpha L^{-1}\alpha\rangle L^{-1}\langle x\rangle$. However, we are not limited to a perturbation result and no nonvalid assumptions are incorporated in the separation of $\langle Ly\rangle$ into $\langle L\rangle\langle y\rangle$. In operator notation, one can write immediately

$$[L + \alpha]y = x,$$
$$y = L^{-1}x - L^{-1}\alpha y$$
$$= L^{-1}x - L^{-1}\alpha[y_0 - y_1 + y_2 - \cdots]$$
$$= L^{-1}x - L^{-1}\alpha L^{-1}x$$
$$+ L^{-1}\alpha L^{-1}\alpha L^{-1}x - \cdots,$$
$$\langle y\rangle = L^{-1}\langle x\rangle + L^{-1}\langle \alpha L^{-1}\alpha\rangle L^{-1}\langle x\rangle + \cdots,$$

where as before it is assumed that expectation and integration can be interchanged. Thus, the problem has become completely elementary. If we assume stationarity, this will give us the same result as before with the stochastic Green's function $G(t, \tau)$. Also, if L is a constant coefficient operator, $\ell(t, \tau)$ is $\ell(t - \tau)$; however, we will leave it in the above form.

The result for the Gaussian case can now be rewritten in the following convenient form:

$$\langle y(t)\rangle = \langle F(t)\rangle + \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \ell(t_1, \tau_1)\ell(\tau_1, \tau_2)$$
$$\times \langle \alpha(\tau_1)\alpha(\tau_2)\rangle\langle F(\tau_2)\rangle$$
$$+ \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 \ell(t_1, \tau_1)$$
$$\times \ell(\tau_1, \tau_2)\ell(\tau_2, \tau_3)\ell(\tau_3, \tau_4)$$
$$\times \langle \alpha(\tau_1)\alpha(\tau_2)\alpha(\tau_3)\alpha(\tau_4)\rangle\langle F(\tau_4)\rangle + \cdots.$$

The general term is the n-fold integral

$$\int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{n-1}} d\tau_n \ell(t, \tau_1)\ell(\tau_1, \tau_2) \cdots \ell(\tau_{n-1}, \tau_n)$$
$$\times \langle \alpha(\tau_1) \cdots \alpha(\tau_n)\rangle\langle F(\tau_n)\rangle,$$

where $n = 0, 2, 4, 6, \cdots$.

Writing $\alpha(\tau_i) = \alpha_i$ for convenience, we note that

$$\langle \alpha_1\alpha_2\rangle = R_{12},$$
$$\langle \alpha_1\alpha_2\alpha_3\alpha_4\rangle = R_{12}R_{34} + R_{13}R_{24} + R_{14}R_{23}$$

or three terms involving two pairs each of six averages. Similarly, $\langle \alpha_1\alpha_2 \cdots \alpha_6\rangle$ gives five terms of three pairs each or fifteen averages; $\langle \alpha_1 \cdots \alpha_8\rangle$ gives seven terms of four pairs each or twenty-eight averages; $\langle \alpha_1 \cdots \alpha_n\rangle$

for even n gives $n-1$ terms of $\frac{1}{2}n$ pairs each, so that the number of averages involved is always $\binom{n}{2}$.

For the general term, $\langle \alpha_1 \cdots \alpha_n \rangle$ can be represented by $e_{ijk}\cdots R_{ij}R_{kl}\cdots$, where the number of indices involved is n and each index varies from 1 to n. The symbol e is a number defined to be zero, if any of the indices $i, j, k, l, \cdots$ are repeated, and zero, unless the second symbol of each pair is greater than the first; thus, $j > i$, $l > k$, etc., so that they are always in the natural order. The symbol e has the value of one-half if none of the above conditions is violated. Thus, $e_{123\ldots} = \frac{1}{2}$. The summation rule applies if any index on e and an index on R are the same. Equivalently, $e = 1$ and we add the condition that $\ell = 0$ if the R's are reversed from the natural ascending order. Thus, $R_{12}R_{34}$ is nonzero, but $R_{34}R_{12}$ is zero. This amounts then to a commutation pairwise and throwing out of terms which are really the same since the internal order of the α's or the order of the R's does not matter. Now,

$$\langle y(t)\rangle = \langle F(t)\rangle + \cdots + \int_0^t d\tau_1 \cdots \int_0^{\tau_{n-1}} d\tau_n \times \ell(t_1,\tau_1)\ell(\tau_1,\tau_2)\cdots\ell(\tau_{n-1},\tau_n) \times e_{ijk\ldots}R_{ij}R_{kl}\cdots\langle F(\tau_n)\rangle + \cdots,$$

with $n = 2, 4, \cdots, \infty$.

We will assume a bounded input $x(t)$ in an interval $[0, T]$ and a time-limited[31] stationary stochastic process $\alpha(t)$, by which we mean R vanishes for $|t| \geq T$. If $\ell(t,\tau)$ is continuous in the interval, it must be bounded. Similarly, $\alpha(t)$ is assumed continuous with probability one so that almost all sample functions are bounded. The absolute value of the general term is less than $M_1M_2^nM_3^{n+1}t^n/n!$, if M_1 is the bound for $x(t)$, M_2 is the bound for α, and M_3 is the bound for ℓ. Integration of the n-fold integral

$$\int_0^t \cdots \int_0^{\tau_{n-1}} d\tau_1 \cdots d\tau_n$$

gives $t^n/n!$.

For finite observation times, the series converges since the series for e^t converges. By placing the bound on $F(t)$ rather than $x(t)$, we can combine M_2 and M_3 to write not only the general term but also the nth partial sum as less than $Me^{M't}$. Actually, since only even values of n occur here rather than all, we have

$$M[\tfrac{1}{2}(e^{M't} + e^{-M't}) - 1] = M[\cosh(M't) - 1]$$

as the bound. Thus, at $t = 0$, the bound is zero and $\langle y(t)\rangle = 0$. At $t = T$, the bound is $M[\cosh(M'T) - 1]$, where M is the bound for $x(t)$ in $[0, T]$.

y is not necessarily stationary and we observe that, for the non-Gaussian case, the odd terms do not then drop out. However, it is easy to see convergence holds. Of course, the mean value and the correlation do not specify a general random function $x(t)$ uniquely. But random functions encountered in physical problems are often assumed to be Gaussian, so that all their finite-dimensional distribution functions can be assumed to be multidimensional Gaussian. Then the mean value and the correlation completely specify $\alpha(t)$ since they determine all distribution functions. Stationarity in the strict sense and wide sense do not need to be distinguished for Gaussian processes, and both the mean and the correlation are finite. The higher moments $\langle \alpha(\tau_1)\cdots\alpha(\tau_n)\rangle$ depend only on elements of the covariance matrix (and the mean which was assumed zero). That the limit as $\tau \to \infty$ of $R(\tau)$ is zero is usually clear from physical considerations.

In the correlation case we proceed as before:

$$\begin{aligned} y(t) &= L^{-1}x(t) - L^{-1}\alpha(t)y(t) \\ &= F(t) - L^{-1}\alpha(t)y(t), \\ R_y &= \langle y(t_1)y(t_2)\rangle = \langle [F(t_1) - L^{-1}(t_1)\alpha(t_1)y(t_1)][F(t_2) - L^{-1}(t_2)\alpha(t_2)y(t_2)]\rangle \\ &= \langle F(t_1)F(t_2)\rangle - L^{-1}(t_1)\langle \alpha(t_1)F(t_2)y(t_1)\rangle - L^{-1}(t_2)\langle \alpha(t_2)F(t_1)y(t_2)\rangle + L^{-1}(t_1)L^{-1}(t_2)\langle \alpha(t_1)\alpha(t_2)y(t_1)y(t_2)\rangle. \end{aligned}$$

Let $y(t) = y_0(t) - y_1(t) + y_2(t) - \cdots$, where $y_0(t) = F(t) = L^{-1}x(t)$; then,

$$\begin{aligned} R_y &= \langle F(t_1)F(t_2)\rangle - L^{-1}(t_1)\langle \alpha(t_1)F(t_2)[y_0(t_1) - y_1(t_1) + y_2(t_1)\cdots]\rangle - L^{-1}(t_2)\langle \alpha(t_2)F(t_1)[y_0(t_2) - y_1(t_2) \\ &\quad + y_2(t_2)\cdots]\rangle + L^{-1}(t_1)L^{-1}(t_2)\langle \alpha(t_1)\alpha(t_2)[y_0(t_1) - y_1(t_1)\cdots][y_0(t_2) - y_1(t_2) + \cdots]\rangle \\ &= \langle F(t_1)F(t_2)\rangle - L^{-1}(t_1)\langle \alpha(t_1)F(t_2)y_0(t_1)\rangle + L^{-1}(t_1)\langle \alpha(t_1)F(t_2)y_1(t_1)\rangle - L^{-1}(t_1)\langle \alpha(t_1)F(t_2)y_2(t_1)\rangle\cdots \\ &\quad - L^{-1}(t_2)\langle \alpha(t_2)F(t_1)y_0(t_2)\rangle + L^{-1}(t_2)\langle \alpha(t_2)F(t_1)y_1(t_2)\rangle \\ &\quad - L^{-1}(t_2)\langle \alpha(t_2)F(t_1)y_2(t_2)\rangle\cdots \\ &\quad + L^{-1}(t_1)L^{-1}(t_2)\langle \alpha(t_1)\alpha(t_2)y_0(t_1)y_0(t_2)\rangle \\ &\quad - L^{-1}(t_1)L^{-1}(t_2)\langle \alpha(t_1)\alpha(t_2)[y_0(t_1)y_1(t_2) + y_0(t_2)y_1(t_1) + \cdots]\rangle + L^{-1}(t_1)L^{-1}(t_2)\langle \alpha(t_1)\alpha(t_2)y_0(t_1)y_2(t_2) + \cdots\rangle. \end{aligned}$$

The first term is

$$\langle F(t_1)F(t_2)\rangle = \langle L^{-1}x(t_1)L^{-1}x(t_2)\rangle$$
$$= L^{-1}(t_1)L^{-1}(t_2)\langle x(t_1)x(t_2)\rangle.$$

The second term is

$$-L^{-1}(t_1)\langle\alpha(t_1)L^{-1}(t_2)x(t_2)L^{-1}(t_1)x(t_1)\rangle = 0.$$

since $\langle\alpha\rangle$ separates out and is equal to zero. The next term is

$$L^{-1}(t_1)\langle\alpha(t_1)F(t_2)L^{-1}(t_1)\alpha(t_1)F(t_1)\rangle$$

or

$$L^{-1}(t_1)\langle\alpha(t_1)L^{-1}(t_1)\alpha(t_1)\rangle\langle F(t_2)F(t_1)\rangle, \quad \text{etc.}$$

We see the separations occur, but the labor of calculating terms becomes repellent and the convergence question becomes still more complicated. However, the solution can be carried out as before.

6. GENERAL STOCHASTIC OPERATOR

For stochastic differential operators which are completely random rather than a sum of a deterministic and a random operator or for calculating higher moments, it is still desirable to find other methods of proceeding. An attack on this general problem was suggested in 1961.[19] The method attempts to find a stochastic Green's function for the desired statistical measure of the dependent variable and is a generalization of Green's function theory. Consideration of the example[19] of a first-order stochastic differential equation as well as the form of the solutions for the "random sampling" operator[19] suggests that the approach is feasible for the general case. The reasons for this are clear, even though more work is needed. An nth-order differential equation can be given as a system of first-order equations and a first-order equation can be solved in the desired form.[19]

The general approach has led to determination of the power spectral density for a random process which is randomly sampled in time according to a probability law and to the correlation function of the first-order stochastic differential equation (also obtained by Tikhonov[8]). In principle, it should lead to complete statistical determination and can be generalized to the nonlinear (stochastic) equations called "reducible to linear" by Pugachev.

The preliminary work has been discussed previously,[5,12] and its extensions will be reported later.

ACKNOWLEDGMENT

This work has been supported by the National Aeronautics and Space Administration first under contract NAS 8-21044 and presently under Grant NGR 11-003-020.

[1] R. Syski, in *Modern Nonlinear Equations*, T. L. Saaty, Ed. (McGraw-Hill, New York, 1967), Chap. 8.
[2] J. C. Samuels, IRE Trans. PGIT, Suppl. **5**, 248 (1959).
[3] J. C. Samuels and A. C. Eringen, J. Math. & Phys. **38**, 83 (1959).
[4] U. Grenander, *Probability and Statistics* (John Wiley & Sons, New York, 1959), pp. 108–129.
[5] G. Adomian, thesis, University of California, Los Angeles, 1961 [Part I published in Rev. Mod. Phys. **35**, 185 (1963)].
[6] A. T. Bharucha-Reid, *Random Equations* (Academic Press Inc., New York) (to appear).
[7] L. Sibul (private communication).
[8] V. I. Tikhonov, Automation and Remote Control **19**, 705 (1958).
[9] K. J. Aström, Intern. J. Control **1**, 301 (1965).
[10] B. Chelpanov, Appl. Math. Mech. **26**, 1145 (1962).
[11] J. B. Keller, in *Proceedings of Symposia in Applied Mathematics*, Richard Bellman, Ed. (American Mathematical Society, Providence, R.I., 1964), Vol. 16, pp. 145–170.
[12] G. Adomian, in *Proceedings of Symposia in Applied Mathematics*, Richard Bellman, Ed. (American Mathematical Society, Providence, R.I., 1964), Vol. 16, pp. 1–39.
[13] Statistical measures, statistical operators, stochastic operators, and the relationship of stochastic mappings on Cartesian-product spaces and the random operators of space are discussed in a forthcoming book by the author. (*Linear Stochastic Operators*, Academic Press.)
[14] L. L. Foldy, Phys. Rev. **67**, 107 (1945).
[15] W. C. Meecham, Thompson-Ramo-Wooldridge Space Technology Laboratories, Report No. (AD268788), 1961.
[16] V. Twersky, J. Math. Phys. **3**, 700, 724 (1962).
[17] E. T. Jaynes, Am. Scientist **53**, 136A (1965).
[18] U. Grenander, *Probabilities on Algebraic Structures* (John Wiley & Sons, New York, 1963).
[19] See Ref. 5, Chap. 4.
[20] An example in this form appears in a detailed consideration of the random sampling of a random process in the author's dissertation, Ref. 5.
[21] In this section, we use Keller's notation. See Ref. 11.
[22] G. Adomian, Proc. Inst. Radio Eng. **50**, 203 (1962).
[23] Y. M. Chen, J. Math. Phys. **5**, 1541 (1964).
[24] For this method, see Ref. 11.
[25] Tamm–Dancoff method (I. Tamm, Zh. Fiz. **9**, 449 (1945); Green's function method of Martin and Schwinger (P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959); BBGKY hierarchy (N. N. Bogoliubov, in *Studies in Statistical Mechanics*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland, Amsterdam, 1962), pp. 5–116.
[26] J. M. Richardson, in *Proceedings of Symposia in Applied Mathematics*, Richard Bellman, Ed. (American Mathematical Society, Providence, R.I., 1964), Vol. 16, pp. 290–302.
[27] R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[28] G. Adomian, *Theory of Random Systems, Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes*, Prague, Aug 31–Sept 11, 1965 (Publishing House of the Czechoslovak Academy of Sciences, Prague, 1967), pp. 205–222.
[29] See Ref. 28. The iterative method is included here (with minor corrections) for completeness and availability. Detailed application of the method to the propagation of electromagnetic waves in a random continuum or in a random distribution of scatterers will appear in a subsequent paper.
[30] For example, see J. H. Laning and R. H. Battin, *Random Processes in Automatic Control* (McGraw-Hill, New York, 1956), p. 83, or K. S. Miller, *Multidimensional Gaussian Processes* (John Wiley and Sons, New York, 1964), p. 71.
[31] A time-limited stationary stochastic process is physically realizable [see the theorem of A. V. Balakrishnan, IEEE. Trans. Inform. Theory, **11**, 154 (1965)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Crossing Relations for Center-of-Mass- and Breit-System Canonical Amplitudes

A. CHAKRABARTI
Centre de Physique Théorique de l'Ecole Polytechnique, 17, rue Descartes, Paris V^e, France

(Received 27 December 1968)

We derive the crossing relations for the canonical amplitudes, starting with the case of four particles of nonzero rest mass. We give the results for both the center-of-mass- and the Breit-frame amplitudes and compare two different points of view.The definition of the "Breit system" is adapted to a unified treatment of all the three types of momentum transfer. It is shown that the Breit-frame amplitudes need no preliminary adjustments of constraints or the definition of a "generalized" amplitude in the alternative point of view. We also show that the usual convention about the continuation of the covariant spinor amplitudes involve, for the particles to be crossed, a rotation π of the "physical" spin about the direction of the continued momentum. The corresponding results for the helicity and the transversity amplitudes are derived as corollaries. Finally, we discuss the case of amplitudes involving zero-mass particles and some remarks are added concerning the kinematic singularities. The definition of the canonical and the transversity amplitudes are compared and some useful Lorentz-transformation formulas are collected together. The transformations connecting the center of mass and the Breit frames are parametrized in terms of the invariants. The well-known invariant amplitudes for πN scattering are used to verify our canonical formula.

1. INTRODUCTION

As a part of a program of a systematic study of certain properties and possibilities of direct applications of the canonical amplitudes, we will derive in the following sections their crossing relations using both the center of mass (c.m.) and the Breit systems. (We consider only the case of two initial and two final particles.) These results for the canonical amplitudes (Appendix A), where one uses space-fixed projections of the physical spins involved, complement the study of crossing relations for amplitudes involving body-fixed spin projections, such as helicity and transversity amplitudes.[1–5]

Though in this paper we will not carry our study far enough to include a discussion of the L–S-coupled partial-wave development,[6] to which the canonical amplitudes lead in a natural fashion, we would like to mention that one of our principal aims in studying these amplitudes is to display as explicitly as possible the physical significances of the orbital angular momentum contribution in the L–S-coupling mode. Interesting examples occur when one considers such aspects as the behavior of the amplitudes at the thresholds and pseudothresholds[7] and selection rules in some Regge pole models.[8] Apart from the above-mentioned and other possibilities of application, there is also the agreeable fact that the transformation properties of the physical spin acquire an intuitively appealing and, indeed, visualizable form[9] in the canonical representation. However, some authors seem to be under the impression that systematic and convenient crossing properties are prerogatives, among "physical amplitudes" (as contrasted to covariant spinor amplitudes having a less direct physical significance), of the helicity and the transversity ones. Hence, we start by showing, in this paper, that we obtain very simple "diagonalized" crossing matrices for the canonical amplitudes, when the scattering plane is chosen to be the (x, y) plane, the particle spins being projected along the z axis, which is normal to the scattering plane.

Having mentioned this, we should immediately make the necessary distinction between the so-called "transversity" amplitudes for which also the spins are said to be projected along the normal to the scattering plane (particularly since the present author has realized from certain discussions that some confusion is, unfortunately, apt to arise). The similarities and distinctions between these two amplitudes are discussed at length in Appendix A and Sec. 3. Here we may just mention that when one speaks of projections along the normal, a pertinent question is: What is it exactly that is being projected? As shown in Appendix A, a possible source of confusion lies here.

Again, as is shown in Sec. 3, *if, when we transform from one amplitude to the other, we also rotate the momenta from the* (x, y) *to the* (z, x) *plane*, the coefficients regroup themselves, giving a direct one-to-one proportionality,[10] though for the same choice of scattering plane[say, (x,y)plane]a canonical amplitude corresponds to a somewhat complicated superposition of transversity amplitudes (and vice versa). The factor of proportionality is, however, not trivial, since it includes (for the s channel) the scattering angle θ_S which, in general, becomes complex when continued to the t-channel physical region. (See the remarks in Sec. 5 concerning the effect of this term on the kinematic singularities). One interest of the above-mentioned relation would be to give directly an L–S partial-wave expansion of the transversity

amplitudes (in terms of the canonical series), instead of passing via the helicity amplitude and helicity coupling, and then trying to decouple and recouple the various terms in order to bring out the channel spins and the orbital contributions.

Let us now come back to our main interest, namely, the direct derivation of the canonical crossing relations. We will derive the helicity and the transversity crossing matrices (in both the c.m. and Breit systems) as corollaries, indicating also their direct derivations as a check on the various signs and phases involved. The inverse procedure of starting with the helicity (or transversity) amplitudes and then deriving from them the required results for the canonical case would, of course, have been possible. But, in our opinion, this would have tended (as will be seen from the discussion that follows immediately) to obscure certain fundamental features of the problem we wish to emphasize. In our derivation (Sec. 2) we have tried to display step by step, and as explicitly as possible, the consequences of the successive fundamental postulates concerning continuation and crossing, in terms of the physical canonical amplitudes.[11] There are the two following principal aspects of the problem:

(a) Suppose we start with an s-channel c.m. amplitude $\langle p_3\sigma_3, p_4\sigma_4| T |p_1\sigma_1, p_2\sigma_2\rangle$; we want to relate it after continuation and crossing to the t-channel c.m. amplitudes $\langle P_3\sigma_3', P_1\sigma_1'| T |P_4\sigma_4', P_2\sigma_2'\rangle$ (Throughout, we denote the particles to be crossed as particles 1 and 4.) The essential point to note is that the particle momenta satisfy different constraints in the s- and t-channel c.m. amplitudes and that, for the results to be consistent, the process of continuation itself must not violate the constraints one has at the starting point of continuation.

In the technique we have tried to emphasize (Sec. 2A), one expresses (using the invariance of the T matrix under Lorentz transformations), *before starting the continuation and using only real transformations*, the s- and t-channel c.m. amplitudes in terms of amplitudes with particle momenta, say p_1'' and P_1'', respectively, with $i = 1, 2, 3, 4$, such that $-p_1''^{\mu}$, $p_2''^{\mu}$, $p_3''^{\mu}$, $-p_4''^{\mu}$ are, respectively, exactly the same function of the invariants s and t as

$$P_1''^{\mu}, P_2''^{\mu}, P_3''^{\mu}, P_4''^{\mu} \quad \text{(Condition I).}$$

(The s and t values for the two sets belong as yet to the two distinct physical regions.)

As will be shown explicitly (Figs. 5–10 and Appendix D) Condition I is fulfilled when p_i'' and P_i'' refer to the "Breit frame"[12,13] in the s and t channel, respectively.

However, in the rather particular case when $p_3 - p_1 = p_2 - p_4$, which gives the momentum exchange, is a timelike 4-vector with positive energy (which is possible to have for suitable mass values), we can, *using strictly real transformations*, satisfy Condition I with $P_i'' = P_i$. That is, we obtain, transforming only the s-channel amplitudes, the typical "t-channel constraints" in the sense of Condition I. In fact, this is the case we study first in Sec. 2A, since it enables us to derive in a relatively simple fashion all the necessary final formulas (2.6)–(2.13), which [as will be shown in Secs. 2B and 2D, consistency check (iii)] remain essentially valid for arbitrary momentum transfer.[14] In particular, in Sec. 2 we check the consistency of this point of view with the one adopted by TW.[1]

In this connection, one point must be noted: *The adjustment of constraints involves in our formalism pure Lorentz transformations followed by rotations* (2.5) *and* (2.6) *or* (B12) and (B20), *whereas for* TW *it seems that only a pure Lorentz transformation is sufficient to obtain what they call the "generalized amplitude," the one to be continued directly.* The source of this apparent contradiction lies in the result (A23), namely, a pure rotation about the y axis gives no extra phase factor for the (z,x)-plane helicity amplitudes. Indeed, $\cos\alpha_1'$ [Eq. (2.16)] corresponding to their general amplitude becomes complex when the momenta are continued to their final t-channel values. Turning the whole t-channel configuration through any constant angle about the normal is somewhat trivial, but α_1' is a definite function of s and t, involving singularities. This has to be re-absorbed through an extra rotation as indicated. If the continuation process is itself allowed to violate such a constraint, then the logical necessity of the pure Lorentz transformation itself is no longer evident. The necessity of the rotation in question would become apparent even for helicity amplitudes, if we start with a scattering plane different from (z, x).

It is, of course, good to know that a particular simplification occurs in the (z, x) plane and to profit from it, but totally ignoring the logical role of the rotation does not seem to be justified. All these problems do not arise when we consider the Breit-frame crossing relations, since Condition I is already satisfied; hence, no preliminary adjustment or "generalization" is necessary. This fact makes the Breit-frame crossing relations (2.39), (3.18), and (3.19) particularly simple.

(b) Having adjusted the constraints (necessary for the c.m. amplitudes), we have to tackle the problem of continuation. For the c.m. amplitudes we have to continue both the transformed amplitude and the

transformation coefficients [as, for example, the two factors on the rhs of (2.6)].

First, let us take up the central problem of continuation of the amplitude to the t-channel physical region (Sec. 2C).

With regard to proofs, it is now usual to refer to the field theoretical results of Bros, Epstein, and Glaser,[15,5] concerning a certain possible path of continuation for the covariant spinor amplitudes involving only massive particles, and to the works of Hepp and Williams.[15] We will throughout adopt the following attitude: We will *assume* that the covariant spinor amplitudes can be continued (without any change in the spin indices or the phase) from one given point in the s-channel physical region to another given point in the t-channel one—not necessarily restricting ourselves to points corresponding to space-like momentum transfer and eventually including (Sec. 4) amplitudes with zero-mass particles. Without trying to specify the path, we will examine the consequences of the foregoing postulate for the physical canonical amplitudes by expressing them in terms of the spinor amplitudes. Equations (2.25) and (2.26) give the precise content of this statement.

The crucial point is that in the transformation coefficients (2.21)–(2.23) *the sign of the energy is to be taken into account in such a way that such coefficients be invariant under the substitution* $p_\mu \to -p_\mu$. This is essential since [see the discussion following (2.34)] the canonical states $|p, \sigma\rangle$ and $|-p, \sigma\rangle$ transform in the same way and so do the spinor states $|p, A\rangle$ and $|-p, A\rangle$. This last point can be seen readily on writing the generators of the two fundamental spinor representations as

$$\mathbf{M}_S = -i\mathbf{P} \times \frac{\partial}{\partial \mathbf{P}} + \mathbf{S},$$

$$\mathbf{N}_S = -iP^0 \frac{\partial}{\partial \mathbf{P}} \pm i\mathbf{S}. \tag{1.1}$$

Hence, the coefficients that transform one basis into another must evidently preserve this property.

When the sign of the energy is thus taken into account [Eq. (2.23)], the coefficient $[p]_+$ continued to $-P$ becomes $[-P]_+$, which is different from $[-P]_- = [+P]_+$. *As a consequence of Eqs.* (2.27) *and* (2.28), *the physical spins (of the particles to be crossed) undergo a rotation* $\pm\pi$ *(depending on the determination of a square root) about the final direction of the momentum* P.

When the z axis is normal to the scattering plane, this involves [Eq. (2.31)] simply a change of sign along with an extra phase factor for the canonical spin indices.[16] In our formalism, *the continuation of the helicity indices (in any plane) involves a phase factor* (3.13), though a very simple one. This is to be compared with the convention of TW.[1] The simplicity of the Breit-frame helicity and transversity crossing relations (3.18) and (3.19) is noteworthy.

So far we have said nothing about the continuation of the transformation coefficients. There, however, ambiguities may arise only through the different possible continuations of certain square roots involved, such as $\sqrt{s}$ in $\tilde{\theta}_S$ of (2.37). The example considered in (C9)–(C18) will, it is hoped, suffice to illustrate the essential point.[17]

In Sec. 4 we introduce "mixed amplitudes" involving particles of zero mass. The formalism is based on the zero-mass representation studied elsewhere[18] and may be compared with the approach of Ader, Capedeville, and Navelet.[19]

In Appendix C, we compare our general canonical formulas, for the case of "πN scattering," with those obtained through the use of the well-known invariant amplitudes. Again, the simplicity of the Breit-frame case (C21)–(C23) is to be noted.

2. CROSSING RELATIONS FOR CANONICAL AMPLITUDES

A. Transformations Leading to t-Channel Constraints

We start by considering a canonical scattering amplitude (see Appendix A) in the s-channel c.m. frame, with the z axis perpendicular to the scattering plane and with the directions of the momenta $\mathbf{p}_i$, $i = 1, 2, 3, 4$, as shown in Fig. 1. (For the present, we

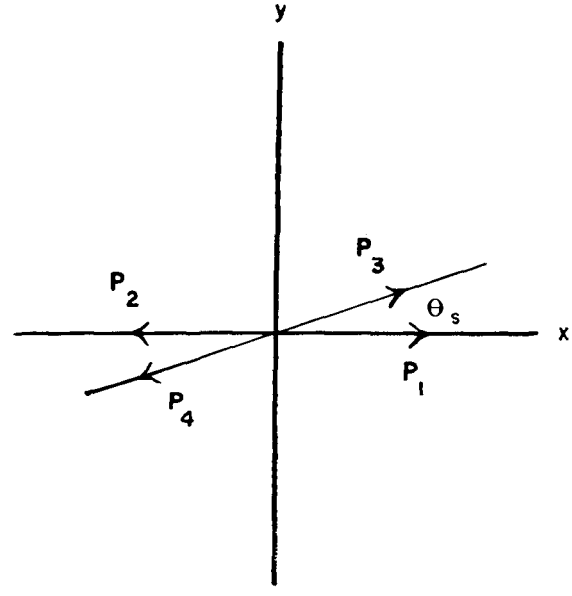


FIG. 1. s-channel (c.m.).

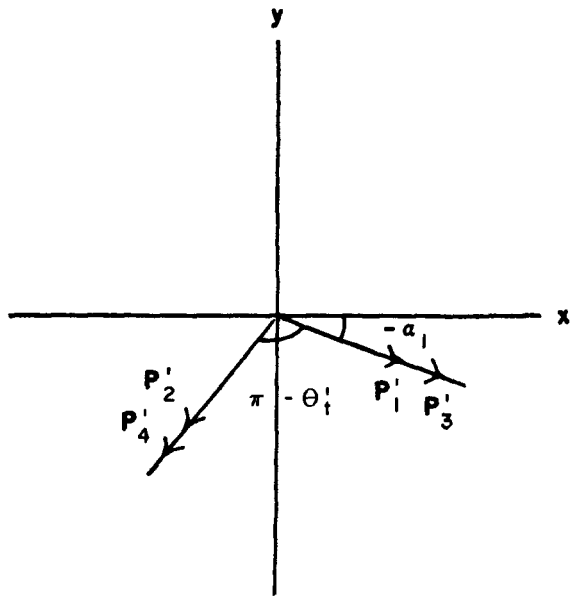


FIG. 2. After $\Lambda_{(t)}$.

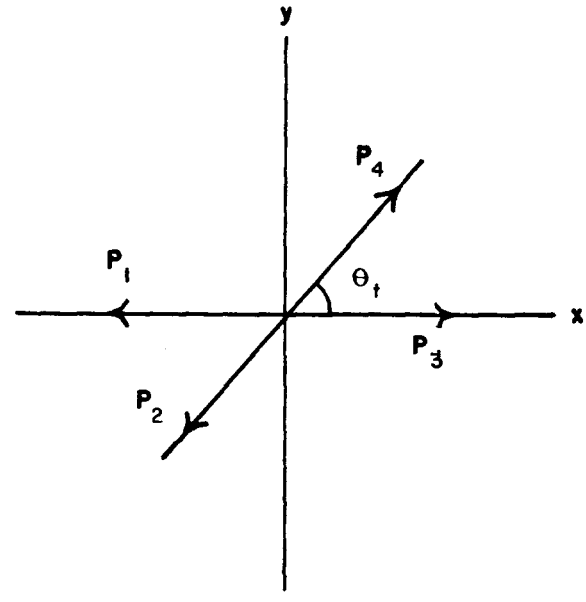


FIG. 4. t-channel (c.m.).

consider only particles of nonzero rest mass. The case of zero-mass particles will be taken up in Sec. 4.)

We propose to express it in terms of an amplitude satisfying the "t-channel constraints." As already explained, we mean by this that the 4-momenta corresponding to the final transformed s-channel amplitude (Fig. 3) must satisfy the following condition: $-p_1''^{\mu}$, $p_2''^{\mu}, p_3''^{\mu}, -p_4''^{\mu}, \mu = 0, 1, 2, 3$, must be the same functions of the invariants s and t and the masses as the respective t-channel momenta (Fig. 3) $P_1^{\mu}, P_2^{\mu}, P_3^{\mu}, P_4^{\mu}$. The values of s and t will, of course, correspond to different regions for the two sets.

Thus, we must have,

$$\mathbf{p}_3'' - \mathbf{p}_2'' = \mathbf{p}_2'' - \mathbf{p}_4'' = 0,$$

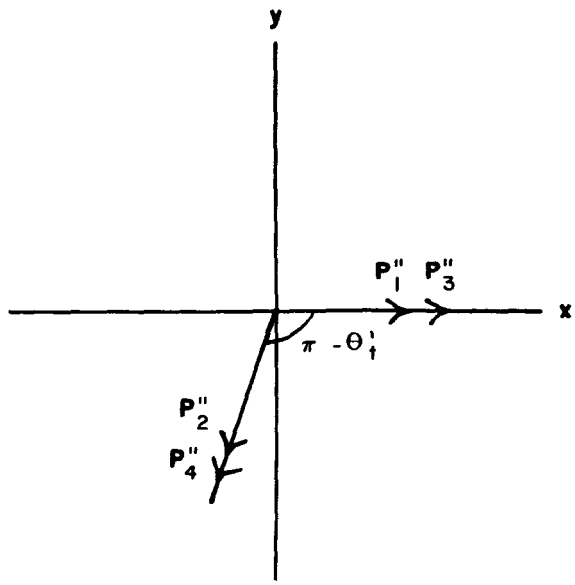


FIG. 3. After $R_s(-\alpha_1)$.

with $\mathbf{p}_3''$ parallel to $\mathbf{P}_3$ as indicated. The angle between $\mathbf{p}_3''$ and $\mathbf{p}_2''$ is to be $(\pi - \theta_t')$, where θ_t' is the same function of s and t as the t-channel scattering angle θ_t.

We shall reach configuration 3 (Fig. 3) from configuration 1 (Fig. 1) in two steps. Let $\Lambda_{(t)}$ be a pure Lorentz transformation such that (Appendix B1)

$$\Lambda_{(t)} \cdot (p_3 - p_1) = (\sqrt{t}, \mathbf{0}) = \Lambda_{(t)} \cdot (p_2 - p_4). \quad (2.1)$$

In general, such a Lorentz transformation is a complex one. But let us start by considering a case where $p_3 - p_1$ is *a timelike 4-vector with positive energy* when (using the positive determination of $\sqrt{t}$) $\Lambda_{(t)}$ is a real Lorentz transformation. We have

$$\begin{aligned}&\langle p_3\sigma_3, p_4\sigma_4| T |p_1\sigma_1, p_2\sigma_2\rangle \\ &\quad = \langle p_3\sigma_3, p_4\sigma_4| \Lambda_{(t)}^{-1} T \Lambda_{(t)} |p_1\sigma_1, p_2\sigma_2\rangle \\ &\quad = \langle p_3'\sigma_3, p_4'\sigma_4| T |p_1'\sigma_1, p_2'\sigma_2\rangle\, e^{-i(\omega_1\sigma_1+\omega_2\sigma_2-\omega_3\sigma_3-\omega_4\sigma_4)},\end{aligned} \quad (2.2)$$

where the ω_i, $i = 1, 2, 3, 4$, are the respective Wigner rotations, undergone by the canonical spins around the z axis, and

$$p_i' = \Lambda_{(t)} \cdot p_i. \quad (2.3)$$

The configuration of the p_i' are shown in Fig. 2. Let α_i be the rotation (in the positive or counterclockwise sense) undergone by the momentum $\mathbf{p}_i$ under $\Lambda_{(t)}$.

As a consequence of (2.1), we have (counting all the rotations in the positive sense)

$$\begin{aligned}\alpha_3 &= (\alpha_1 - \theta_S), \\ \alpha_4 &= (\alpha_2 - \theta_S).\end{aligned} \quad (2.4)$$

(The scattering angle θ_S is given in Appendix B). The

angles α_1 and α_2 may be obtained explicitly from the formulas in Appendices A and B.

Let us now consider the rotation $R_z(-\alpha_1)$ leading to Fig. 3. We have

$$\langle p_3'\sigma_3, p_4'\sigma_4| R_z^{-1}(-\alpha_1)TR_z(-\alpha_1)|p_1'\sigma_1, p_2'\sigma_2\rangle = \langle p_3''\sigma_3, p_4''\sigma_4| T|p_1''\sigma_1, p_2''\sigma_2\rangle e^{i(\sigma_1+\sigma_2-\sigma_3-\sigma_4)\alpha_1}. \quad (2.5)$$

Hence, from (2.2) and (2.5), we obtain

$$\langle p_3\sigma_3, p_4\sigma_4| T|p_1\sigma_1, p_2\sigma_2\rangle = \langle p_3''\sigma_3, p_4''\sigma_4| T|p_1''\sigma_1, p_2''\sigma_2\rangle e^{i\varphi}, \quad (2.6)$$

where

$$\begin{aligned}\varphi &= -(\omega_1\sigma_1 + \omega_2\sigma_2 - \omega_3\sigma_3 - \omega_4\sigma_4) + (\sigma_1 + \sigma_2 - \sigma_3 - \sigma_4)\alpha_1 \\ &= -(\chi_1\sigma_1 - \chi_2\sigma_2 - \chi_3\sigma_3 + \chi_4\sigma_4) - (\alpha_2 - \alpha_1)\sigma_2 \\ &\quad + (\alpha_3 - \alpha_1)\sigma_3 + (\alpha_4 - \alpha_2 + \alpha_2 - \alpha_1)\sigma_4 \quad (2.7)\end{aligned}$$

and

$$\chi_1 = -(\alpha_1 - \omega_1), \quad \chi_3 = -(\alpha_3 - \omega_3),$$
$$\chi_2 = +(\alpha_2 - \omega_2), \quad \chi_4 = +(\alpha_4 - \omega_4). \quad (2.8)$$

The additional negative signs for χ_1 and χ_3 are to compensate for the fact that, under $\Lambda_{(t)}$, $\mathbf{p}_1$ and $\mathbf{p}_3$ and, hence, the respective canonical spins turn in a negative (clockwise) sense, while $\mathbf{p}_2$ and $\mathbf{p}_4$ undergo positive rotations. With this definition we can take $\sin\chi_i$ as positive for all the i's in a uniform fashion. These angles are given by the formulas in Appendices A2 and B2, and expressed in terms of the invariants, they become (in the notation of Appendix B)

$$\cos\chi_1 = -\frac{(s + m_1^2 - m_2^2)(t + m_1^2 - m_3^2) + 2m_1^2\Delta}{s_{12}t_{31}},$$

$$\sin\chi_1 = \frac{2m_1\sqrt{\Phi}}{s_{12}t_{31}}, \quad \Delta \equiv m_2^2 + m_3^2 - m_1^2 - m_4^2,$$

$$\cos\chi_2 = \frac{(s + m_2^2 - m_1^2)(t + m_2^2 - m_4^2) - 2m_2^2\Delta}{s_{12}t_{42}},$$

$$\sin\chi_2 = \frac{2m_2\sqrt{\Phi}}{s_{12}t_{42}},$$

$$\cos\chi_3 = \frac{(s + m_3^2 - m_4^2)(t + m_3^2 - m_1^2) - 2m_3^2\Delta}{s_{34}t_{31}},$$

$$\sin\chi_3 = \frac{2m_3\sqrt{\Phi}}{s_{34}t_{31}},$$

$$\cos\chi_4 = -\frac{(s + m_4^2 - m_3^2)(t + m_4^2 - m_2^2) + 2m_4^2\Delta}{s_{34}t_{42}},$$

$$\sin\chi_4 = \frac{2m_4\sqrt{\Phi}}{s_{34}t_{42}}. \quad (2.9)$$

Further, it may be verified that, just as

$$(\alpha_3 - \alpha_1) = (\alpha_4 - \alpha_2) = -\theta_S, \quad (2.10)$$

so

$$(\alpha_2 - \alpha_1) = \theta_t', \quad (2.11)$$

where θ_t' is obtained from θ_S given in Appendix B by the interchanges

$$s \rightleftarrows t, \quad (m_1, m_2, m_3, m_4) \rightleftarrows (m_4, m_2, m_3, m_1). \quad (2.11')$$

This is consistent with the relation implied by Fig. 4, namely,

$$\alpha_2 + (\pi - \theta_t') + (-\alpha_1) = \pi. \quad (2.12)$$

Thus, finally,

$$\varphi = -[(\chi_1\sigma_1 - \chi_2\sigma_2 - \chi_3\sigma_3 + \chi_4\sigma_4) + \theta_S(\sigma_3 + \sigma_4) + \theta_t'(\sigma_2 - \sigma_4)]. \quad (2.13)$$

So far we have been considering a timelike momentum transfer, and the transformed momenta and the relevant rotation angles have all been real. For spacelike momentum transfer, the preceding transformations and the corresponding momenta and rotations become complex. For lightlike momentum transfer the transformed momenta even become infinite. For the spacelike case, we may try to proceed by invoking the invariance of the T matrix under complex Lorentz transformation and using the corresponding complex parameters in (2.2) and (2.5).

If we want to confine ourselves to real transformations, we may use a different approach, indicated in Figs. 5–10. We start by relating the center of mass to the Breit frame in both channels.

To be more precise, in the s channel we use the pure Lorentz transformation $\Lambda_{(B)}$ (instead of $\Lambda_{(t)}$),

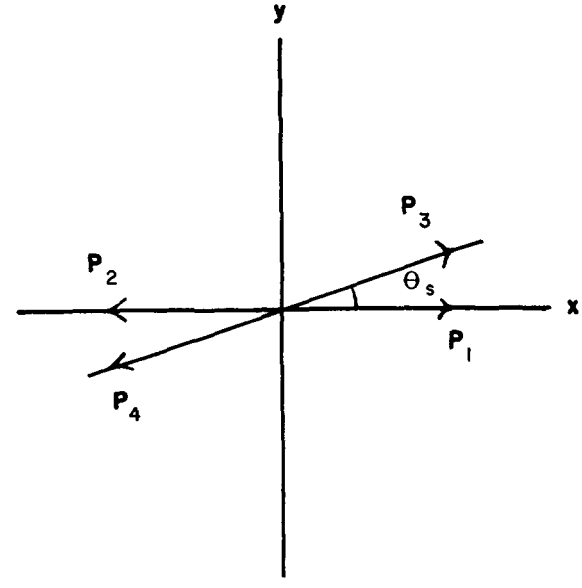


FIG. 5. s-channel (c.m.).

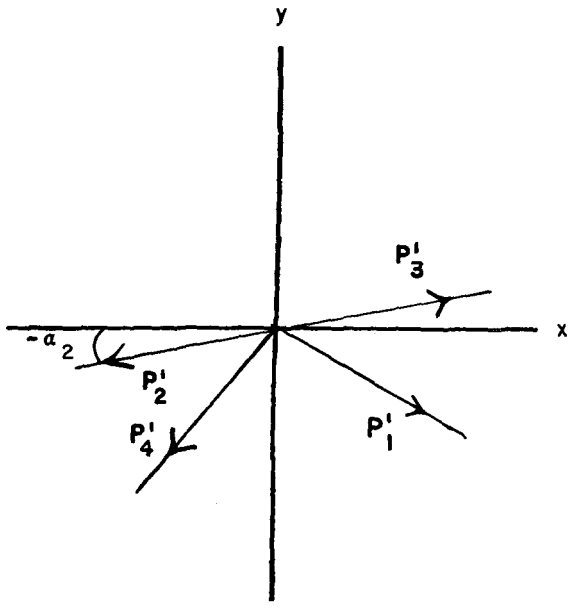


FIG. 6. After $\Lambda_{(B)}$.

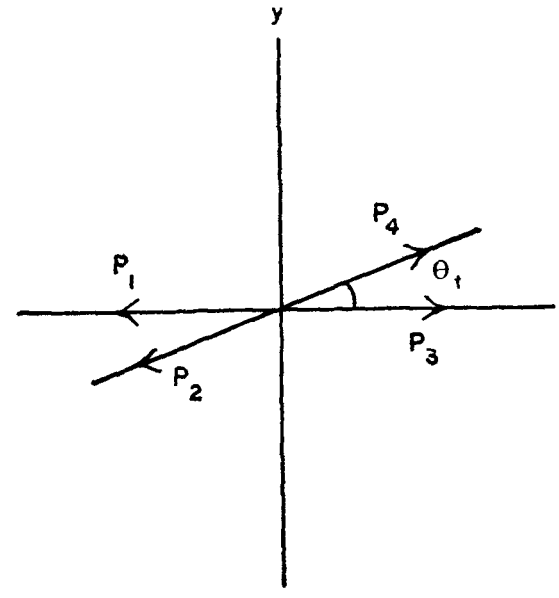


FIG. 8. t-channel (c.m.).

where

$$\Lambda_{(B)} \cdot (p_2 + p_3) = (\sqrt{v}, \mathbf{0}) \tag{2.14}$$

with

$$v = (p_2 + p_3)^2 = s + t + \Delta,$$
$$\Delta = m_2^2 + m_3^2 - m_1^2 - m_1^2. \tag{2.15}$$

Then, instead of $R_z(-\alpha_1)$ of (2.5) we use $R_z(-\alpha_2')$ corresponding to the rotation undergone by $\mathbf{p}_2$ (Fig. 6). Similarly, in the t channel we successively use the pure Lorentz transformation reducing $P_2 + P_3$ to rest and then the rotation $R_z(-\beta_2)$ (Fig. 9) compensating the rotation undergone by $\mathbf{P}_3$.

The explicit expressions for the useful parameters corresponding to these transformations are given in Appendix B. The result immediately essential for us is that, corresponding to the Figs. 7 and 10, $-p_1''^{\mu}$, $p_2''^{\mu}$, $p_3''^{\mu}$, $-p_4''^{\mu}$ and $P_1''^{\mu}$, $P_2''^{\mu}$, $P_3''^{\mu}$, $P_4''^{\mu}$ are the same functions of the invariants s and t for each of the indices 1, 2, 3, and 4, respectively (Condition I). If we leave out, as we shall do, the extremely particular case where p_2 and p_3 are *both lightlike with parallel 3-momenta*, $\Lambda_{(B)}$ and, hence, the rotations that follow are always real.

Thus, we see that, using strictly real transformations, we can reach in both channels frames which

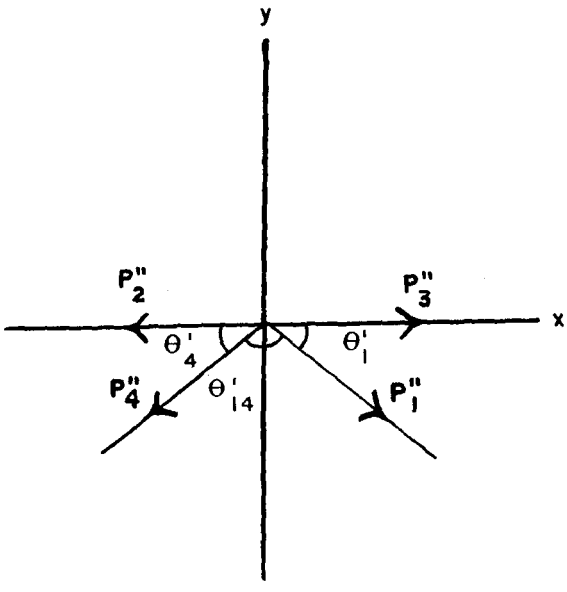


FIG. 7. After $R_z(-\alpha_2)$.

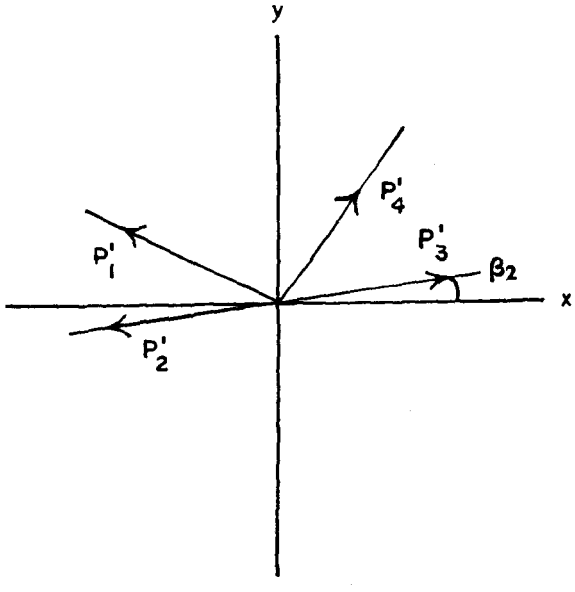


FIG. 9. After $\Lambda_{(B)}$.

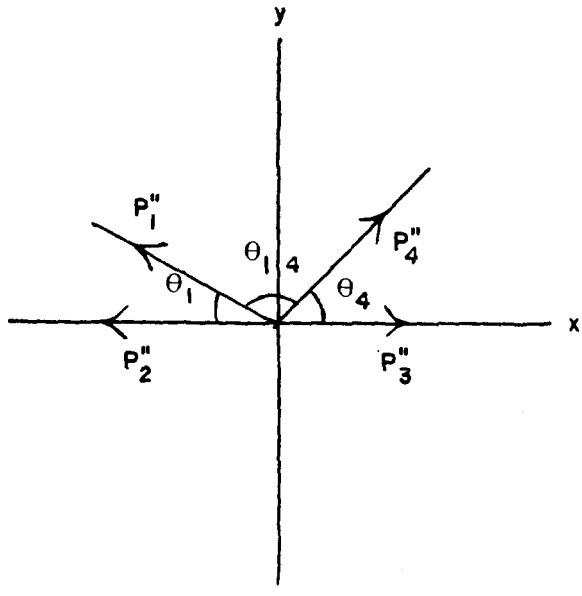


FIG. 10. After $R_z(-\beta_3)$.

possess the remarkable property that Condition I is already built in, no preliminary adjustment being necessary. This fact renders the derivation of the crossing relation between these channels particularly simple in terms of Breit frames for any momentum exchange, *if we assume* the rule for continuation of the amplitude (to be discussed in Sec. 2C) to be valid.

The crossing relations between the s- and t-channel c.m. frames can be obtained by passing via the Breit frames, again for any momentum exchange. As we will verify later on, the result thus obtained for the c.m. frames would be the same as that obtained from a somewhat different and quite convenient point of view, which we will now discuss in the following subsection.

B. Alternative Method: Continuation of Arbitrary Real Transformations

Instead of using $\Lambda_{(t)}$ and $R_z(-\alpha_1)$ [Eqs. (2.2) and (2.5)], we may also start by considering an arbitrary (real) pure Lorentz transformation, Λ, say, followed by the rotation $R_z(-\alpha_1')$, where

$$\cos\alpha_1' = \frac{\mathbf{p}_1 \cdot (\Lambda \cdot \mathbf{p}_1)}{|\mathbf{p}_1|\,|\Lambda \cdot \mathbf{p}_1|}. \tag{2.16}$$

The next step is to continue the parameters of the transformations and, hence, the momenta such that at the end of continuation

$$(-\Lambda \cdot p_1, \Lambda \cdot p_2, \Lambda \cdot p_3, -\Lambda \cdot p_4)_{\text{continued}} = (P_1, P_2, P_3, P_4), \tag{2.17}$$

the t-channel c.m. values.

Our object is to verify that this would give the same final result as that obtained by formally writing the right-hand side of (2.6) *not only for timelike but for any momentum transfer* [with φ given by (2.7) and (2.13)].and then continuing the s and t values to their end-point t-channel ones (assuming, of course, that such a continuation is possible). This involves the continuation of the transformed amplitude and that of the transformation coefficient. The necessary rules regarding the continuation of the amplitude will be given in the next section. But since, concerning this point, the same basic convention is to be used in both cases (or even if we start from Breit instead of c.m. frames), all we need verify is that the transformation coefficients obtained in Secs. 2A and 2B match at the end of continuation. This is practically evident, but still we would like to make it quite explicit.

From the formulas given in Appendix A2 (and utilized in Appendix B) we find the various rotations, undergone by the 3-momenta and the respective canonical spins, to be given (for a Lorentz transformation corresponding to the 4-velocity u'', say) in terms of p_i^0, $(\Lambda \cdot p_i)^0$, and u''^0, $i = 1, 2, 3, 4$, and such expressions as $|\mathbf{p}_i \times \mathbf{u}''|$. Since starting from the same initial energies we arrive (by definition) after continuation at the same final energies, we need only check the last expression.

In the method of Sec. 2A we start [Eq. (B27)] with

$$|\mathbf{p}_i \times \mathbf{u}''| = \left(\frac{p_{in}p_{fi}\sin\theta_S}{t^{\frac{1}{2}}}\right)_{(s)} = \left(\frac{\Phi^{\frac{1}{2}}}{2(st)^{\frac{1}{2}}}\right)_{(s)}, \tag{2.18}$$

the subscript indicating that the values of the invariants are in the s-channel physical region. In the method of Sec. 2B the corresponding expressions can easily be seen to reduce (at the end of continuation) to

$$\left|(\Lambda \cdot \mathbf{p}_i) \times \frac{(\Lambda \cdot \mathbf{p}_1 + \Lambda \cdot \mathbf{p}_2)}{s^{\frac{1}{2}}}\right|_{\text{continued}} = \frac{|\mathbf{P}_1|\,|\mathbf{P}_2|\sin\theta_t}{s^{\frac{1}{2}}} = \frac{\Phi^{\frac{1}{2}}}{2(st)^{\frac{1}{2}}}, \tag{2.19}$$

which is just the continuation of the right-hand side of (2.18). As a final check we note that, since in Sec. 2B the senses of the rotations are to be determined at the end of continuation (Fig. 11), we obtain, exactly as in Fig. 4, negative rotations about the z axis for particles 1 and 3 and positive ones for particles 2 and 4. Hence, our verification is complete.

C. Continuation of the Transformed Amplitude

In terms of the covariant spinor amplitudes (using the conventions of Ref. 5 for positive energy), we can

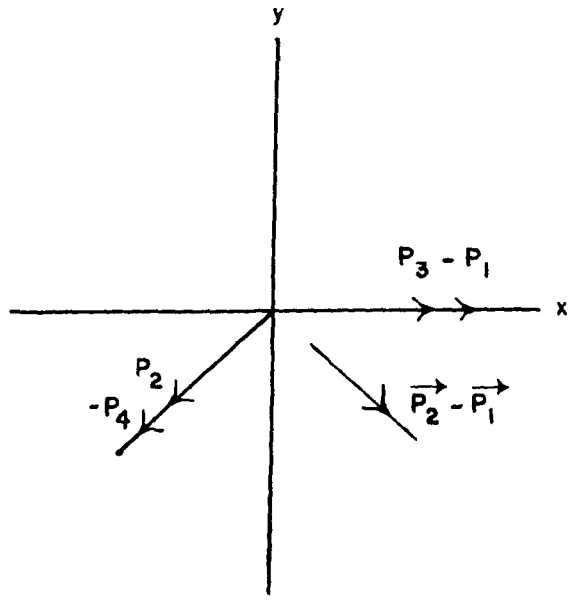


FIG. 11. After continuation.

write the canonical ones as

$$|p, \sigma\rangle_\epsilon = |p, A\rangle_\epsilon \, \mathfrak{D}^S_{A\sigma}([p]_\epsilon), \tag{2.20}$$

$$_\epsilon\langle p, \sigma| = {}_\epsilon\langle p, \dot{A}| \, \mathfrak{D}^S_{\dot{A}\sigma}([p]_\epsilon C), \tag{2.21}$$

where

$$C = \begin{vmatrix} & 1 \\ -1 & \end{vmatrix}, \quad \epsilon = \frac{p_0}{|p_0|} = \pm 1, \tag{2.22}$$

and

$$[p]_\epsilon = \frac{m + \epsilon(p_0 + \mathbf{p}\cdot\boldsymbol{\tau})}{[2m(m + |p_0|)]^{\frac{1}{2}}}. \tag{2.23}$$

(See the remarks, concerning the foregoing use of ϵ, in the Introduction.) Then, factoring out the Wigner rotations, we can write Eq. (2.6) as

$$\begin{aligned}\langle p_3\sigma_3, p_4\sigma_4| T |p_1\sigma_1, p_2\sigma_2\rangle &= \langle p_3''\dot{A}_3, p_4''\dot{A}_4| T |p_1''A_1, p_2''A_2\rangle \\ &\times \mathfrak{D}^{S_1}([p_1'']_+)_{A_1\sigma_1}\mathfrak{D}^{S_2}([p_2'']_+)_{A_2\sigma_2} \\ &\times \mathfrak{D}^{S_3}_{A_3\sigma_3}([p_3'']_+C)\mathfrak{D}^{S_4}_{A_4\sigma_4}([p_4'']_+C)e^{i\varphi}. \end{aligned} \tag{2.24}$$

In the method of Sec. 2B we have an exactly similar expression, where p_i'' and φ are given in terms of arbitrary parameters to be continued. In any case, *our postulate concerning the continuation of the transformed canonical amplitude can be expressed as*

$$\begin{aligned}(\langle p_3''\sigma_3, p_4''\sigma_4| T |p_1''\sigma_1, p_2''\sigma_2\rangle)_{\text{continued}} &= \langle P_3\dot{A}_3, -P_4\dot{A}_4| T |-P_1A_1, P_2A_2\rangle \\ &\times \mathfrak{D}^{S_1}_{A_1\sigma_1}([-P_1]_+)\mathfrak{D}^{S_2}_{A_2\sigma_2}([+P_2]_+) \\ &\times \mathfrak{D}^{S_3}_{A_3\sigma_3}([+P_3]_+C)\mathfrak{D}^{S_4}_{A_4\sigma_4}([-P_4]_+C) \end{aligned} \tag{2.25}$$

$$\begin{aligned}&= \langle P_3\sigma_3, -P_4\sigma_4'| T |-P_1\sigma_1', P_2\sigma_2\rangle \\ &\times \mathfrak{D}^{S_1}_{\sigma_1\sigma_1}([-P_1]_-^{-1}[-P_1]_+) \\ &\times \mathfrak{D}^{S_4}_{\sigma_4\sigma_4}(C^{-1}[-P_4]_-^{-1}[-P_4]_+C), \end{aligned} \tag{2.26}$$

where the P_i are the final t-channel c.m. values. But,

$$\begin{aligned}[-P]_-^{-1}[-P]_+ &= [+P]_+^{-1}[-P]_+ \\ &= \frac{m + P_0 - \mathbf{P}\cdot\boldsymbol{\tau}}{[2m(P_0 + m)]^{\frac{1}{2}}}\cdot\frac{-P_0 + m - \mathbf{P}\cdot\boldsymbol{\tau}}{[2m(-P_0 + m)]^{\frac{1}{2}}} \\ &= \frac{-\mathbf{P}\cdot\boldsymbol{\tau}}{[-\mathbf{P}^2]} = \pm i\hat{\mathbf{P}}\cdot\boldsymbol{\tau} = e^{\pm i\pi\hat{\mathbf{P}}\cdot\frac{1}{2}\boldsymbol{\tau}}, \end{aligned} \tag{2.27}$$

according to the determination of the square root chosen. Here the carat denotes the unit vector. Thus, for particle 1, we have on the right-hand side of (2.26) *a rotation* $(\mp)\pi$ *about the direction* $\hat{\mathbf{P}}_1$. [We will use the + sign in (2.27) unless otherwise and explicitly specified.]

Similarly,

$$C^{-1}[-P]_-^{-1}[-P]_+C = (i\hat{\mathbf{P}}\cdot\boldsymbol{\tau})^*, \tag{2.28}$$

the complex conjugate of the above rotation. Let us note the following alternative forms for eventual uses:

$$\mathfrak{D}^S(i\hat{\mathbf{P}}\cdot\boldsymbol{\tau}) = \mathfrak{D}^S(\varphi, 2\theta, -(\pi + \varphi)) \tag{2.29}$$

$$= \mathfrak{D}^S(\pi + \varphi, \pi - \theta, 0)\mathfrak{D}^S(0, 0, \pi) \times \mathfrak{D}^{S^{-1}}(\pi + \varphi, \pi - \theta, 0) \tag{2.30}$$

$$= \mathfrak{D}^S(\varphi, \pi + \theta, 0)\mathfrak{D}^S(0, 0, \pi)\mathfrak{D}^{S^{-1}}(\varphi, \pi + \theta, 0). \tag{2.30'}$$

In particular, for $P_3 = 0$ or $\theta = \frac{1}{2}\pi$,

$$\mathfrak{D}^S_{\sigma'\sigma}(\varphi, \pi, -(\pi + \varphi)) = (-1)^{S-\sigma}e^{i(\pi+2\varphi)\sigma}\delta_{\sigma,-\sigma'} \tag{2.31}$$

$$= (-1)^S e^{i2\varphi\sigma}\delta_{\sigma,-\sigma'}. \tag{2.31'}$$

Since in the t-channel c.m. system (Fig. 4) $\varphi_1 = \pi$ and $\varphi_1 = \theta_t$, we finally obtain

$$\begin{aligned}(\langle p_3\sigma_3, p_4\sigma_4| T |p_1\sigma_1, p_2\sigma_2\rangle)_{\text{continued}} &= \langle P_3\sigma_3, -P_4 - \sigma_4| T |-P_1 - \sigma_1, P_2\sigma_2\rangle \\ &\times (-1)^{S_1-\sigma_1+S_4-\sigma_4}e^{-i\pi\sigma_1-i(\pi+2\theta_t)\sigma_4}\cdot(e^{i\varphi})_{\text{continued}}. \end{aligned} \tag{2.32}$$

The same argument gives the following result for the Breit frames (Figs. 7 and 10):

$$\begin{aligned}(\langle p_3\sigma_3, p_4\sigma_4| T |p_1\sigma_1, p_2\sigma_2\rangle_{(B)})_{\text{continued}} &= \langle P_3\sigma_3, -P_4 - \sigma_4| T |-P_1 - \sigma_1, P_2\sigma_2\rangle \\ &\times (-1)^{S_1-\sigma_1+S_4-\sigma_4}e^{-i\pi(\sigma_1+\sigma_4)-i2(\theta_1\sigma_1+\theta_4\sigma_4)}. \end{aligned} \tag{2.33}$$

The relevant angles are given in Appendix B.

D. Crossing

So far we have been discussing adjustment of constraints and analytic continuations. Let us now

consider what may really be termed "crossing"—namely, replacing an ingoing by an outgoing state (and vice versa) with the components of the 4-momenta changing sign. (Internal quantum numbers, not exhibited explicitly, are to be charge conjugated.)

When we explicitly take into account both signs of energy, the generators of rotations and pure Lorentz transformations can be written in the canonical representation as (see, for example, Ref. 20, p. 78)

$$\mathbf{M} = -i\mathbf{P} \times \frac{\partial}{\partial \mathbf{P}} + \mathbf{S},$$

$$\mathbf{N} = -iP^0 \frac{\partial}{\partial \mathbf{P}} - \frac{\mathbf{P} \times \mathbf{S}}{P^0 + \epsilon m}, \tag{2.34}$$

where

$$\epsilon = P^0/|P^0|, \quad m = +\sqrt{P^2}.$$

The S are well-known $(2s+1)$-dimensional Hermitian generators of rotation.

As a consequence, under the transformations

$$U(\Lambda) = e^{-i\mathbf{M}\cdot\hat{\mathbf{n}}\theta} \quad \text{or} \quad e^{-i\mathbf{N}\cdot\hat{\mathbf{n}}\chi},$$

the following canonical states (Appendix A) are multiplied by the same Wigner rotation matrices, namely, $|p, \sigma\rangle$ and $|-p, \sigma\rangle$. Hence, noting the relation

$$\mathfrak{D}^S_{\sigma\sigma'}(R) = (-1)^{\sigma-\sigma'}\mathfrak{D}^{S*}_{-\sigma,-\sigma'}(R), \tag{2.35}$$

we find that the states $|-p, -\sigma\rangle$ and $(-1)^{S-\sigma}\langle p, \sigma|$ again transform in the same way (the choice of the constant part of the phase being, of course, arbitrary). Hence, we postulate the crossing relation

$$\langle P_3\sigma_3, -P_4 - \sigma_4| T |-P_1 - \sigma_1, P_2\sigma_2\rangle = (-1)^{\eta}(-1)^{S_1-\sigma_1+S_4-\sigma_4}\langle P_3\sigma_3, P_1\sigma_1| T |P_4\sigma_4, P_2\sigma_2\rangle, \tag{2.36}$$

where η is a constant phase factor.

From (2.32) and (2.36) we finally obtain for the c.m. systems, denoting by $\tilde{\theta}_S$ the value (in general complex) of θ_S continued to the t-channel physical region,

$$\begin{aligned}(\langle p_3\sigma_3, p_4\sigma_4| T |p_1\sigma_1, p_2\sigma_2\rangle)_{\text{continued}} &= (-1)^{\eta}\exp\{-i[\tilde{\theta}_S(\sigma_3+\sigma_4) + \theta_t(\sigma_2+\sigma_4) \\ &\quad + \pi(\sigma_1+\sigma_4) + (\chi_1\sigma_1 - \chi_2\sigma_2 - \chi_3\sigma_3 + \chi_4\sigma_4)]\} \\ &\quad \times \langle P_3\sigma_3, P_1\sigma_1| T |P_4\sigma_4, P_2\sigma_2\rangle.\end{aligned} \tag{2.37}$$

The angles χ_i, $i = 1, 2, 3, 4$, are given by (2.9).

Denoting by $\mathcal{A}^{(s)}_{(s,t)}$ and $\mathcal{A}^{(t)}_{(s,t)}$ respectively, the s- and t-channel c.m. amplitudes expressed in terms of the invariants, we will often write the relation as

$$\mathcal{A}^{(t)}_{\sigma_3\sigma_1,\sigma_4\sigma_2} = (-1)^{-\eta}e^{-i[\tilde{\theta}_S(\sigma_3+\sigma_4)+\theta_t(\sigma_2+\sigma_4)+\pi(\sigma_1+\sigma_4)} \times e^{i(\chi_1\sigma_1-\chi_2\sigma_2-\chi_3\sigma_3+\chi_4\sigma_4)}\tilde{\mathcal{A}}^{(s)}_{\sigma_3\sigma_4,\sigma_1\sigma_2} \tag{2.38}$$

where $\tilde{\mathcal{A}}^{(s)}$ denotes the value continued to the t-channel physical region.

In an analogous way, for the corresponding Breit-frame amplitudes ($\mathcal{B}^{(s)}$, $\mathcal{B}^{(t)}$ say), we obtain, expressing everything in terms of the final t-channel s and t values,

$$\mathcal{B}^{(t)}_{\sigma_3\sigma_1,\sigma_4\sigma_2} = (-1)^{-\eta}e^{i[\pi(\sigma_1+\sigma_4)+2(\theta_1\sigma_1+\theta_4\sigma_4)]}\tilde{\mathcal{B}}^{(s)}_{\sigma_3\sigma_4,\sigma_1\sigma_2}. \tag{2.39}$$

We may note the following checks of consistency:

(i) Instead of transforming back to canonical amplitude, as in (2.26), and then using (2.36), we could first utilize the usually accepted crossing relation for the spinor amplitudes,

$$\langle P_3\dot{A}_3, -P_4\dot{A}_4| T |-P_1A_1, P_2A_2\rangle = (-1)^{\eta'}\langle P_3\dot{A}_3, P_1\dot{A}_1| T |P_4A_4, P_2A_2\rangle. \tag{2.40}$$

The final result would be the same. We prefer to break up the process into two stages [(2.26) and (2.36)] in order to display as explicitly as possible the different assumptions involved in terms of the physical-spin indices themselves.

(ii) As a consequence of (2.27) and (2.28), we get back exactly the same results, (2.38) and (2.39), if we pass via the spinor amplitudes

$$\langle p_3A_3, p_4A_4| T |p_1\dot{A}_1, p_2\dot{A}_2\rangle,$$

instead of via (2.24).

(iii) Passing via the Breit frames, namely, using (2.39), (B16) and (B21), we obtain for the c.m. amplitudes

$$\mathcal{A}^{(t)}_{\sigma_3\sigma_1,\sigma_4\sigma_2} = (-1)^{-\eta}\exp\{i[\pi(\sigma_1+\sigma_4) + 2(\theta_1\sigma_1+\theta_4\sigma_4) - \tilde{\varphi}_s + \varphi_t]\}\tilde{\mathcal{A}}^{(s)}_{\sigma_3\sigma_4,\sigma_1\sigma_2}, \tag{2.41}$$

where $\tilde{\varphi}_s$ denotes φ_s continued to the t-channel physical region. Now, we have to note that

$$\chi_i = (\tilde{\chi}^s_i + \chi^t_i). \tag{2.42}$$

(In fact, χ_s remains real for $s < 0$.) As a consequence, (2.41) reduces exactly to (2.38).

Finally, we note that, if in Fig. 10 we add an extra rotation making P_4'' coincide with the x axis (with a corresponding extra rotation in Fig. 7), then the phase in (2.39) becomes just $(-1)^{\eta}e^{-i(\pi+2\theta_{14})\sigma_1}$. This gives the simplest possible crossing matrix when $s_1 = 0$. Similarly, we can choose to make P_1'' coincide with the x axis.

3. RELATIONS WITH THE HELICITY AND TRANSVERSITY AMPLITUDES IN THE (z, x) PLANE

We will now discuss the relations between the (x, y)-plane canonical amplitudes and the (z, x)-plane helicity and transversity amplitudes. Let us first take up the c.m. amplitudes and, subsequently, the corresponding Breit-frame ones. Let

$$U = R_x(-\tfrac{1}{2}\pi)R_z(-\tfrac{1}{2}\pi). \tag{3.1}$$

Then, starting with the s-channel c.m. canonical amplitude (Fig. 1), we have

$$\begin{aligned}\langle p_3\sigma_3, p_4\sigma_4|\, T\,|p_1\sigma_1, p_2\sigma_2\rangle &= \langle p_3\sigma_3, p_4\sigma_4|\, U^{-1}TU\,|p_1\sigma_1, p_2\sigma_2\rangle \\ &= e^{i\frac{1}{2}\pi(\sigma_1+\sigma_2-\sigma_3-\sigma_4)}\langle p_3'\sigma_3', p_4'\sigma_4'|\, T\,|p_1'\sigma_1', p_2'\sigma_2'\rangle \\ &\quad\times u^{S_1}_{\sigma_1'\sigma_1}u^{S_2}_{\sigma_2'\sigma_2}u^{S_3^*}_{\sigma_3'\sigma_3}u^{S_4^*}_{\sigma_4'\sigma_4},\end{aligned} \tag{3.2}$$

where the momenta $\mathbf{p}_i'$ are now as indicated in Fig. 12, with

$$|\mathbf{p}_i'| = |\mathbf{p}_i|,$$

and the u^{S_i} are (in the notation of Kotansky[4]) the matrices corresponding to $R_x(-\frac{1}{2}\pi)$, namely,

$$\mathfrak{D}^S(\tfrac{1}{2}\pi, \tfrac{1}{2}\pi, -\tfrac{1}{2}\pi).$$

They diagonalize the d matrices corresponding to the rotations about the y axis, giving

$$(u^S \cdot d^S(\theta) \cdot u^{S^*})_{\sigma'\sigma} = e^{i\theta\sigma}\delta_{\sigma\sigma'}. \tag{3.3}$$

(A detailed study of the properties of u^S can be found in the papers of Kotansky.[4])

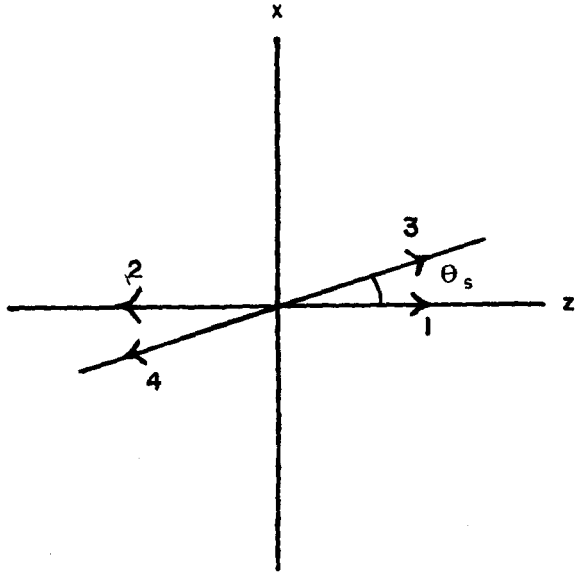


FIG. 12. s-channel c.m. amplitude in the (z, x) plane.

Let us now define the (z, x)-plane helicity amplitudes (Appendix A) as

$$\begin{aligned}&\langle p_3'\sigma_3', p_4'\sigma_4'|\, T\,|p_1'\sigma_1', p_2'\sigma_2'\rangle \\ &= \langle p_3'\lambda_3, p_4'\lambda_4|\, T\,|p_1'\lambda_1, p_2'\lambda_2\rangle \\ &\quad\times d^{S_1}_{\sigma_1'\lambda_1}(0)\, d^{S_2}_{\sigma_2'\lambda_2}(\pi)\, d^{S_3}_{\sigma_3'\lambda_3}(\theta_S)\, d^{S_4}_{\sigma_4'\lambda_4}(\pi+\theta_S).\end{aligned} \tag{3.4}$$

In the above formula we have introduced

$$\langle p_4'\lambda_4| = \langle p_4'\sigma_4'|\, d^{S_4}_{\sigma_4'\lambda_4}(\pi+\theta_S), \tag{3.5}$$

instead of the more habitual convention (which we distinguish here by a $\sim$ sign)

$$\langle p_4', \lambda_4|^{\sim} = \langle p_4', \sigma_4|\, e^{-i\pi\sigma_4'}\, d^{S_4}_{\sigma_4'\lambda_4}(\pi-\theta_S), \tag{3.6}$$

in order to obtain the simplest possible correspondence between the canonical and transversity amplitudes [see the comments following (3.10)]. The relation between the two definitions is given by

$$\langle p_4', \lambda_4| = (-1)^{S_4}e^{-i\pi\lambda_4}\langle p_4', \lambda_4|^{\sim} = (-1)^{S_4-\lambda_4}\langle p_4', \lambda_4|^{\sim}. \tag{3.7}$$

In the language of "tetrads" (see Appendix A) the factor $(-1)^{-\lambda_4}$ corresponds to a rotation π of $n^{(1)}$ and $n^{(2)}$ about the axis $\mathbf{p}_4'$.

The transversity amplitude is defined as

$$\begin{aligned}&\langle p_3'\lambda_3, p_4'\lambda_4|\, T\,|p_1'\lambda_1, p_2', \lambda_2\rangle \\ &= \langle p_3'\tau_3, p_4'\tau_4|\, T\,|p_1'\tau_1, p_2'\tau_2\rangle\, u^{S_1^*}_{\tau_1\lambda_1}u^{S_2^*}_{\tau_2\lambda_2}u^{S_3}_{\tau_3\lambda_3}u^{S_4}_{\tau_4\lambda_4}.\end{aligned} \tag{3.8}$$

Hence from (3.2), (3.4) and (3.8), using (3.3) (and other relevant properties of u^{S_i}),we obtain

$$\begin{aligned}&\langle p_3\sigma_3, p_4\sigma_4|\, T\,|p_1\sigma_1, p_2\sigma_2\rangle \\ &= e^{i\frac{1}{2}\pi(\sigma_1+\sigma_2-\sigma_3-\sigma_4)+i\pi(\sigma_2-\sigma_4)-i\theta_S(\sigma_3+\sigma_4)} \\ &\quad\times \langle p_3'\tau_3, p_4'\tau_4|\, T\,|p_1'\tau_1, p_2'\tau_2\rangle\, \delta_{\tau_1\sigma_1}\delta_{\tau_2\sigma_2}\delta_{\tau_3\sigma_3}\delta_{\tau_4\sigma_4}.\end{aligned} \tag{3.9}$$

Similarly, for the t-channel c.m. systems (Figs. 4 and 13), we obtain,

$$\begin{aligned}&\langle P_3\sigma_3, P_1\sigma_1|\, T\,|P_4\sigma_4, P_2\sigma_2\rangle \\ &= e^{i\frac{1}{2}\pi(\sigma_4+\sigma_2-\sigma_3-\sigma_1)+i\pi(\sigma_2-\sigma_1)+i\theta_t(\sigma_2+\sigma_4)} \\ &\quad\times \langle P_3'\tau_3, P_1'\tau_1|\, T\,|P_4'\tau_4, P_2'\tau_2\rangle\delta_{\tau_1\sigma_1}\delta_{\tau_2\sigma_2}\delta_{\tau_3\sigma_3}\delta_{\tau_4\sigma_4}.\end{aligned} \tag{3.10}$$

Had we chosen the convention (3.6) [instead of (3.5)], we would have obtained in (3.8) a factor $(-1)^{S_4-\sigma_4}e^{i\sigma_4(\pi-\theta_S)}\delta_{\sigma_4,-\tau_4}$ instead of $e^{-i(\pi+\theta_S)\sigma_4}\delta_{\sigma_4\tau_4}$ and, similarly, a factor proportional to $\delta_{\tau_2-\sigma_2}$ in (3.9) instead of $\delta_{\tau_2\sigma_2}$. We have chosen our convention in order to avoid such changes of sign.

Substituting (3.9) and (3.10) in (2.37), we obtain the simple crossing relation for the c.m. (z, x)-plane

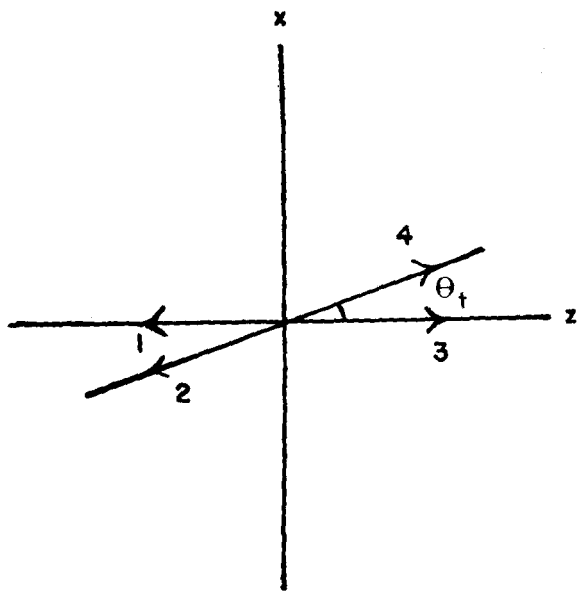


FIG. 13. t-channel c.m. amplitude in the (z, x) plane.

transversity amplitudes as

$$(\langle p_3\tau_3, p_4\tau_4|\ T\ |p_1\tau_1, p_2\tau_2\rangle)_{\text{continued}} = (-1)^{\eta} e^{i\pi(\tau_1+\tau_4)-i(\chi_1\tau_1-\chi_2\tau_2-\chi_3\tau_3+\chi_4\tau_4)} \times \langle P_3\tau_3, P_1\tau_1|\ T\ |P_4\tau_4, P_2\tau_2\rangle. \quad (3.11)$$

For the helicity amplitudes we obtain

$$(\langle p_3\lambda_3, p_4\lambda_4|\ T\ |p_1\lambda_1, p_2\lambda_2\rangle)_{\text{continued}} = (-1)^{\eta}(u^{S_1^*}_{\lambda_1'\tau_1} e^{i(\pi-\chi_1)\tau_1} u^{S_1^*}_{\tau_1\lambda_1})(u^{S_2}_{\lambda_2'\tau_2} e^{i\chi_2\tau_2} u^{S_2^*}_{\tau_2\lambda_2}) \times (u^{S_3^*}_{\lambda_3'\tau_3} e^{i\chi_3\tau_3} u^{S_3}_{\tau_3\lambda_3})(u^{S_4}_{\lambda_4'\tau_4} e^{i(\pi-\chi_4)\tau_4} u^{S_4}_{\tau_4\lambda_4}) \times \langle P_3\lambda_3', P_1\lambda_1'|\ T\ |P_4\lambda_4', P_2\lambda_2'\rangle.$$

Hence,

$$(\langle p_3\lambda_3, p_4\lambda_4|\ T\ |p_1\lambda_1, p_2\lambda_2\rangle)_{\text{continued}} = (-1)^{\eta} e^{i\pi(\lambda_1'-\lambda_4')} \langle P_3\lambda_3', P_1\lambda_1'|\ T\ |P_4\lambda_4', P_2\lambda_2'\rangle \times d^{S_1}_{\lambda_1'\lambda_1}(\chi_1)\, d^{S_2}_{\lambda_2'\lambda_2}(-\chi_2)\, d^{S_3}_{\lambda_3'\lambda_3}(\chi_3)\, d^{S_4}_{\lambda_4'\lambda_4}(-\chi_4). \quad (3.12)$$

[If we prefer to change the signs of χ_2 and χ_4 in (3.12), we have to use the relation $d^S_{\lambda'\lambda}(\chi) = (-1)^{\lambda'-\lambda} d^S_{\lambda'\lambda}(-\chi)$.]

As a consistency check for the different phases and signs involved, it is interesting to note how the above relation (3.12) can be directly obtained for the helicity amplitudes. This may be done in the following steps:

(i) Corresponding to $\Lambda_{(t)}$ of (2.2), the d matrices appearing in (3.12) are directly obtained for the helicity amplitudes by using (A22) and *noting the signs in the definitions* (2.8).

(ii) The rotation $R_y(-\alpha_1)$ gives no extra phase factor, as may be seen from (A23).

(iii) As is particularly evident from (2.30) and (2.30′), the continuation gives the phase factors corresponding to

$$\mathfrak{D}^{S_1}(0, 0, \pi) \quad \text{and} \quad \mathfrak{D}^{S_4^*}(0, 0, \pi);$$

i.e., the factor

$$e^{-i\pi(\lambda_1'-\lambda_4')} = (-1)^{-2(S_1+S_4)} e^{i\pi(\lambda_1'-\lambda_4')}. \quad (3.13)$$

(iv) The consistency of transformation properties leads to the crossing relation [to be compared with (2.36)]

$$\langle P_3\lambda_3', -P_4\lambda_4'|\ T\ |-P_1\lambda_1', P_2\lambda_2'\rangle = (-1)^{\eta}(-1)^{2(S_1+S_4)} \langle P_3\lambda_3', P_1\lambda_1'|\ T\ |P_4\lambda_4', P_2\lambda_2'\rangle. \quad (3.14)$$

The above constant phase factor, rather than (see Ref. 1) such a one as $(-1)^{S_1-\lambda_1'+S_4-\lambda_4'}$, is a direct consequence of our definition (3.5) and may be verified as follows: In the (z, x) plane ($\mathbf{p}$ making an angle θ with the z axis),

$$\begin{aligned} |-p, \lambda\rangle &= |-p, \sigma\rangle\, d_{\sigma\lambda}(\pi+\theta) \\ &\xrightarrow{\text{cross}} \langle p, -\sigma|\,(-1)^{S-\sigma}\, d_{\sigma\lambda}(\pi+\theta) \\ &= \langle p, \lambda'|\, d^S_{\lambda',-\sigma}(-\theta)(-1)^{S-\sigma}\, d_{\sigma\lambda}(\pi+\theta) \\ &= \langle p, \lambda'|\,(-1)^{S+\lambda'}\, d_{-\lambda',\lambda}(\pi) \\ &= \langle p, \lambda'|\,(-1)^{2S}. \end{aligned} \quad (3.15)$$

Hence, (3.12) is exactly verified. Before comparing our results (3.11) and (3.12) with the corresponding ones of Ref. 5, various differences of convention are to be taken into account, such as a change of sign of χ_3 and χ_4, the reversal in Ref. 5 of the projection axis after crossing, and so on.

Let us now consider the Breit-frame amplitudes. Starting with the configuration of s-channel Breit frame (Fig. 7) and using U [Eq. (3.1)] to pass to the corresponding configuration in the (z, x) frame, we obtain [omitting the primes in Fig. 7 and using p_1' for the (z, x) plane momenta as in (3.2)] for the transversity amplitude [instead of (3.9)]

$$\langle p_3\sigma_3, p_4\sigma_4|\ T\ |p_1\sigma_1, p_2\sigma_2\rangle_{(B)} = e^{i\frac{1}{2}\pi(\sigma_1+\sigma_2-\sigma_3-\sigma_4)+i(2\pi-\theta_1)\sigma_1-i(\pi+\theta_4)\sigma_4} \times e^{i\pi\sigma_2} \langle p_3'\tau_3, p_4'\tau_4|\ T\ |p_1'\tau_1, p_2'\tau_2\rangle_{(B)}\, \delta_{\tau_1\sigma_1}\delta_{\tau_2\sigma_2}\delta_{\tau_3\sigma_3}\delta_{\tau_4\sigma_4} \quad (3.16)$$

$$= (-1)^{2S_1} e^{i[\frac{1}{2}\pi(\sigma_1+\sigma_2-\sigma_3-\sigma_4)+\pi(\sigma_2-\sigma_4)-i(\theta_1\sigma_1+\theta_4\sigma_4)]} \times \langle p_3'\tau_3, p_4'\tau_4|\ T\ |p_1'\tau_1, p_2'\tau_2\rangle_{(B)}\, \delta_{\tau_1\sigma_1}\delta_{\tau_2\sigma_2}\delta_{\tau_3\sigma_3}\delta_{\tau_4\sigma_4}. \quad (3.16')$$

[The suffix (B) indicates the Breit frame.] Similarly for the t-channel, starting from Fig. 10, we obtain

$$\langle P_3\sigma_3, P_1\sigma_1|\, T\, |P_4\sigma_4, P_2\sigma_2\rangle_{(B)} = e^{i[\frac{1}{2}\pi(\sigma_4+\sigma_2-\sigma_3-\sigma_1)+\pi(\sigma_2-\sigma_1)+(\theta_1\sigma_1+\theta_4\sigma_4)]} \times \langle P_3'\tau_3, P_1'\tau_1|\, T\, |P_4'\tau_4, P_2'\tau_2\rangle_{(B)}\, \delta_{\tau_1\sigma_1}\delta_{\tau_2\sigma_2}\delta_{\tau_3\sigma_3}\delta_{\tau_4\sigma_4}. \tag{3.17}$$

[It should not be forgotten that θ_1 and θ_4 correspond to different s and t values in (3.16) and (3.17), corresponding to s and t physical regions.]

From (3.16), (3.17), and (2.39) we obtain the crossing relation for the Breit-frame transversity amplitudes. This is extremely simple, namely,

$$\mathcal{B}^{(t)}_{\tau_3\tau_1,\tau_4\tau_2} = (-1)^{-\eta} e^{i\pi(\tau_1-\tau_4)} \tilde{\mathcal{B}}^{(s)}_{\tau_3\tau_4,\tau_1\tau_2}. \tag{3.18}$$

The corresponding relations for the helicity amplitudes are

$$\mathcal{B}^{(t)}_{\lambda_3\lambda_1,\lambda_4\lambda_4} = (-1)^{-\eta} e^{i\pi(\lambda_1-\lambda_4)} \mathcal{B}^{(s)}_{\lambda_3\lambda_4,\lambda_1\lambda_2}. \tag{3.19}$$

Thus, even for helicity amplitudes, we obtain a very simple "diagonalized" crossing matrix. The direct derivation of (3.19) is very easy. We need only note that, of the steps (i)–(iv) needed to derive (3.12), we no longer need (i) and (ii). The operations (iii) and (iv) give the required phase factor.

4. AMPLITUDES INVOLVING PARTICLES OF ZERO REST MASS

In this case we use a "mixed amplitude." The particles of zero mass will be taken, as is natural, in helicity eigenstates, the helicity being restricted to circular polarizations ($\lambda = \pm S$). For the other particles of positive rest mass, we will continue to use the canonical spin projection. The scattering plane will again be taken as the (x, y) plane. We will consider c.m. amplitudes. [When we have four zero-mass particles *and* total initial and final lightlike 4-momenta, the c.m. frame is, of course, no longer defined. There may also be other special features for four zero-mass particles. We will exclude such a case.]

Elsewhere[18] we have discussed how the canonical representations for particles of zero and positive mass can be considered in an entirely unified way. The essential point is that, though the usual factorization of the Wigner rotations corresponding to the passage via the rest frame of massive particles, namely,

$$\mathcal{D}^S(R_W) = \mathcal{D}^S([\Lambda \cdot p]^{-1} \cdot \Lambda \cdot [p]),$$

is no longer well defined when the mass tends to zero, the total expression for R_W exhibits no such problem. Hence, without insisting on passing via the different little groups for the two cases, we can continue to use essentially the same transformation formula (Appendix A) for the zero-mass particles with the additional simplification that, for $m = 0$, the canonical spin rotations become identical with those undergone by the 3-momentum $\mathbf{p}$. That is, we have $\omega = \alpha$ both for pure rotations and pure Lorentz transformations when the mass is zero. This, of course, implies the conservation of helicity. For further details, Ref. 18 may be consulted. As an explicit example showing how the gauge is fixed in our representation, we refer to the Sec. 4 of Ref. 21 which shows how for photons we directly obtain the vector potential in the Coulomb gauge.

For

$$\mathbf{p} = |\mathbf{p}|\,(\cos\theta, \sin\theta, 0),$$

we define the (x, y)-plane helicity states in terms of the above mentioned zero-mass canonical states as

$$|p, \lambda\rangle = |p, \sigma\rangle\, d^S_{\sigma\lambda}(\tfrac{1}{2}\pi) e^{-i\theta\sigma}. \tag{4.1}$$

(This convention differs somewhat from that used in Ref. 18, but is more convenient for our present purpose.)

Let Λ be a pure Lorentz transformation in the (x, y) plane. Then

$$\begin{aligned} U(\Lambda)\,|p, \lambda\rangle &= |\Lambda \cdot p, \sigma\rangle\, e^{-i\omega\sigma}\, d^S_{\sigma\lambda}(\tfrac{1}{2}\pi) e^{-i\theta\sigma} \\ &= |\Lambda \cdot p, \lambda'\rangle\, d^S_{\lambda'\sigma}(\tfrac{1}{2}\pi) e^{i(\alpha-\omega)\sigma}\, d^S_{\sigma\lambda}(\tfrac{1}{2}\pi) \\ &= |\Lambda \cdot p, \lambda\rangle, \quad \text{since} \quad \omega = \alpha. \end{aligned} \tag{4.2}$$

Similarly, for any rotation about the z axis,

$$U(R_z(\varphi))\,|p, \lambda\rangle = |R_z \cdot p, \lambda\rangle. \tag{4.3}$$

Thus, in our formalism there is no contribution to the phase factor φ of Eq. (2.6) from the zero-mass particles.

With regard to the continuation of the transformed amplitude (Sec. 2C), we note that, *since we consider only circular polarization* ($\lambda = \pm S$), the canonical states differ from the corresponding covariant spinor states only by a factor $(p_0)^{\pm S}$. [See Sec. 3 in Ref. 18 and, in particular, the remark following Eq. (3.9).] Hence, exactly as the spinor states, our helicity states should be continued without any change of the helicity index or additional phase factors.

Corresponding to the crossing convention (Sec. 2D)

$$|-p, \sigma\rangle \xrightarrow{\text{cross}} \langle p, -\sigma|\,(-1)^{S-\sigma},$$

we have for our (x, y)-plane helicity states

$$\begin{aligned} |-p, \lambda\rangle &= |-p, \sigma\rangle\, d^S_{\sigma\lambda}(\tfrac{1}{2}\pi) e^{-i(\pi+\theta)\sigma} \qquad (4.4) \\ &\to \langle p, -\sigma|\,(-1)^{S-\sigma}\, d^S_{\lambda\sigma}(\tfrac{1}{2}\pi) e^{-i(\pi+\theta)\sigma} \\ &= (-1)^{-S}\,\langle p, \lambda'|\, d^S_{\lambda',-\sigma}(-\tfrac{1}{2}\pi)\, d^S_{\sigma\lambda}(\tfrac{1}{2}\pi) \\ &= e^{-i\pi\lambda}\,\langle p, \lambda|. \qquad (4.5) \end{aligned}$$

A different convention, if used in writing the rhs of (4.4), may lead to a different over-all constant phase [$(-1)^{2S}$, for example]. But such differences are not essential for our purposes.

The above-mentioned results, when put together, lead to the conclusion that the only difference owing to the presence of zero-mass helicity states in the crossing relation (2.38) is the additional simplification that, when a σ_i is replaced by λ_i, it is simply to be suppressed in the coefficients of the s and t dependent angles $(\chi_i, \theta_s, \theta_t)$. This is, we may write the crossing relation (2.38) as

$$\mathcal{A}^{(t)}_{\sigma_3\sigma_1,\sigma_4\sigma_2} = (-1)^{-\eta} e^{+i\pi(\sigma_1+\sigma_4)} e^{+i\psi} \tilde{\mathcal{A}}^{(s)}_{\sigma_3\sigma_4,\sigma_1\sigma_2}, \quad (4.6)$$

where

$$\psi = \tilde{\theta}_S(\sigma_3 + \sigma_4) + \theta_t(\sigma_4 + \sigma_2) + (\chi_1\sigma_1 - \chi_2\sigma_2 - \chi_3\sigma_3 + \chi_4\sigma_4). \quad (4.7)$$

Then, when one or more of the σ_i are invariant helicity indices λ_i, *they should be replaced by zeros in the rhs of* (4.7). This is all that needs be done. For example,

$$\mathcal{A}^{(t)}_{\sigma_3\sigma_1,\lambda_4\lambda_2} = (-1)^{-\eta} \times \exp[i\pi(\sigma_1 + \lambda_4) + i(\tilde{\theta}_S\sigma_3 + \chi_1\sigma_1 - \chi_3\sigma_3)]\tilde{\mathcal{A}}^{(s)}_{\sigma_3\lambda_4,\sigma_1\lambda_2}. \quad (4.8)$$

This example includes Compton scattering as a particular case.

So far we have discussed only c.m. amplitudes. The corresponding results for Breit (or other particular) frames, as well as those for other types of combinations of spin projections for the different particles involved, may be obtained readily. The above discussion is sufficient to illustrate the basic features of our formalism.

5. REMARKS

Though we do not intend to give a systematic discussion of the kinematic singularities of the canonical amplitudes in this article, let us conclude with a few relevant remarks. (The proofs of these remarks will not be given here, but the verifications are not difficult.)

In comparing the canonical [Eq. (2.38)] and the transversity [Eq. (3.11)] crossing relations we should not consider the extra factors in (2.38) necessarily as sources of additional complications. To take just one example, let us consider (in Fig. 4) the t-channel threshold $t = (m_4 + m_2)^2$. The possible poles and branch points of the transversity amplitudes at this threshold[22,23,5] are exactly cancelled out by the θ_t term making the canonical amplitude regular at this threshold. Elsewhere, however, (such as at $t = 0$) the θ_t term may indeed create an additional singularity.

Similarly the simplicity of the Breit-frame crossing relations (2.39), (3.18), and (3.19) does not necessarily simplify the discussion of the kinematic singularities, since $\tilde{\mathcal{B}}^{(s)}$ (unlike $\tilde{\mathcal{A}}^{(s)}$) cannot be taken to be free of t singularities even when it is not on the physical boundary.

The inevitable conclusion is that each type of amplitude has its particular advantages and difficulties. Since the crossing relations and other related properties of the helicity and the transversity and also those of the spinor[24] amplitudes have fairly often been discussed recently, we have tried to draw attention to the corresponding basic properties of the canonical amplitudes, for reasons already mentioned in the introduction.

APPENDIX A

1. Comparison of the Definitions of Canonical and "Transversity" States

Starting with a state at rest of mass m and spin projection σ along the z axis, we define the canonical state[20] for an arbitrary momentum p as

$$|p, \sigma\rangle = U(\Lambda^{-1}_{(p)})\,|(m, \mathbf{0}), \sigma\rangle, \quad (A1)$$

where $\Lambda_{(p)}$ is a *pure* Lorentz transformation such that

$$\Lambda_{(p)}, p = (m, \mathbf{0}). \quad (A2)$$

(Unless otherwise specified, we will consider particles of positive rest mass. The discussion of the zero-mass case is confined to Sec. 4.)

In this representation the canonical spin operator, defined[25] as

$$\frac{1}{m}\left(\mathbf{W} - \frac{W^0}{P^0 + m}\mathbf{P}\right), \quad (A3)$$

reduces to its simplest form, namely $\mathbf{S}$, the well-known $(2S + 1)$-dimensional generators of rotations (e.g., $\mathbf{S}$ is simply $\frac{1}{2}\boldsymbol{\tau}$ for spin $\frac{1}{2}$ in terms of the Pauli matrices). This simple form of $\mathbf{S}$ is valid for an arbitrary frame. We are not obliged to pass via the rest frame each time in order to define the parameter σ in (A1), this being the eigenvalue of

$$S_z = \mathbf{S} \cdot \hat{\mathbf{k}}, \quad \hat{\mathbf{k}} = (0, 0, 1). \quad (A4)$$

For a state defined as

$$|p, \zeta\rangle = |p, \sigma\rangle\, \mathfrak{D}^S_{\sigma\zeta}(R), \quad (A5)$$

where R is an arbitrary rotation, the parameter ζ may be considered to be the eigenvalue of

$$S_\zeta = \mathbf{S} \cdot \hat{\mathbf{k}}', \quad \text{where} \quad \hat{\mathbf{k}}' = R \cdot \hat{\mathbf{k}}. \quad (A6)$$

In particular, when R is $R(\theta, \varphi)$, corresponding to the direction of the momentum $\mathbf{p}$, we have

$$S_\zeta = S_p = \mathbf{S} \cdot \hat{\mathbf{p}}. \quad (A7)$$

This is called the helicity projection, the helicity states being usually denoted by $|p, \lambda\rangle$.

If

$$R = R_x(\tfrac{1}{2}\pi)R(\theta, \varphi), \tag{A8}$$

then

$$\hat{\mathbf{k}}' = R_x(-\tfrac{1}{2}\pi) \cdot \hat{\mathbf{p}}. \tag{A9}$$

Thus, for example, in Fig. 14

$$\hat{\mathbf{p}} = (\sin\theta, 0, \cos\theta),$$
$$\hat{\mathbf{k}}' = (\sin\theta, \cos\theta, 0). \tag{A10}$$

Such a projection is used when the so-called "transversity" states are defined in terms of the helicity states as

$$|p, \tau\rangle = |p, \lambda\rangle \, \mathfrak{D}^S_{\lambda\tau}(R_x(-\tfrac{1}{2}\pi)).$$

Thus, we see that, *according to the definition* (A3) *of the canonical spin, the "transversity" projection is not, in general, along the normal to the* (z, x) *plane*, though it lies in the (x, y) plane.

However, since for $\hat{\mathbf{k}}'$ given by (A10)

$$\mathbf{S} \cdot \hat{\mathbf{k}}' = [R_z(\theta) \cdot \mathbf{S}] \cdot [R_z(\theta) \cdot \hat{\mathbf{k}}']$$
$$= [R_z(\theta) \cdot \mathbf{S}] \cdot \hat{\mathbf{k}}'', \quad \text{where} \quad \hat{\mathbf{k}}'' = (0, 1, 0), \tag{A11}$$

by modifying the definition of $\mathbf{S}$ to

$$\mathbf{S}' = R_z(\theta) \cdot \mathbf{S} \tag{A12}$$

(an additional rotation about $\hat{\mathbf{k}}''$ may be included), we can consider it to be projected along the normal or y axis. (In fact, it becomes a question of mere terminology, since by suitably modifying the definition of the operator to be projected we can consider it to be projected on any axis whatsoever. This does not affect the transformation properties of the state in question.)

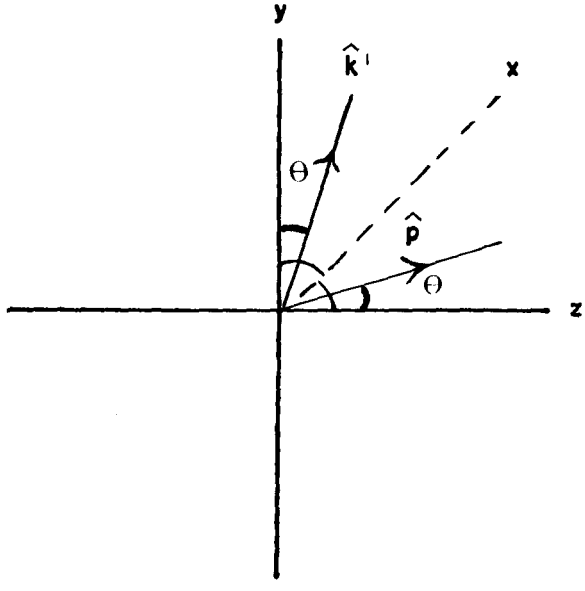


FIG. 14. "Transversity" projection.

In the language of "tetrads,"[5,9] the definition (A3) corresponds to the particular choice

$$S_i = m^{-1}W \cdot n_{(i)}, \quad i = 1, 2, 3, \tag{A13}$$

where

$$n_{(i)} = \Lambda^{-1}_{(p)} \cdot \nu_{(i)}, \tag{A14}$$

with $\Lambda_{(p)}$ as in (A2), and

$$\nu_{(1)} = (0, 1, 0, 0), \quad \nu_{(2)} = (0, 0, 1, 0),$$
$$\nu_{(3)} = (0, 0, 0, 1). \tag{A15}$$

However, we can replace $\Lambda_{(p)}$ by $\Lambda'_{(p)}$, such that

$$\Lambda'^{-1}_{(p)} = R'_{(p)} \cdot \Lambda^{-1}_{(p)} \cdot R \tag{A16}$$

and R is any rotation and R' is one about the axis $\mathbf{p}$, and still have the relations

$$\Lambda'^{-1}_{(p)}(m, \mathbf{0}) = p, \quad p \cdot n'_{(i)} = 0, \quad n'_{(i)} \cdot n'_{(j)} = -\delta ij, \tag{A16'}$$

where

$$n'_{(i)} = \Lambda'^{-1}_{(p)} \cdot \nu_{(i)}.$$

Such arbitrariness is often exploited in the choice of tetrads (see Ref. 5 for some relevant choices). In particular, along with other differences, it makes it possible to introduce [in terms of the canonical spin (A3)] such a redefinition as (A12) for particles whose momenta are not collinear with the z axis. While this is quite legitimate, such arbitrariness must not become a source of confusion when we talk about spin projections normal to the scattering plane for the different cases of canonical and transversity amplitudes.

2. Some Useful Transformation Formulas

Let us consider a pure Lorentz transformation Λ, corresponding to the 4-velocity u'', such that

$$\Lambda \cdot (1, \mathbf{0}) = u'', \tag{A17}$$

and let

$$\Lambda \cdot p = p' \quad \text{or} \quad \Lambda \cdot u = u', \quad \text{with} \quad u = p/m. \tag{A18}$$

The angle α turned through by $\mathbf{p}$ is given by

$$\cos\alpha = \frac{\mathbf{u} \cdot \mathbf{u}'}{|\mathbf{u}|\,|\mathbf{u}'|}$$
$$= \frac{\mathbf{u}^2(u''_0 + 1) - (u_0 u''_0 - u'_0)(u_0 + u'_0)}{(u''_0 + 1)\,|\mathbf{u}|\,|\mathbf{u}'|}, \tag{A19}$$

and

$$\sin\alpha = \frac{\pm|\mathbf{u} \times \mathbf{u}'|\,(u_0 + u'_0)}{(u''_0 + 1)\,|\mathbf{u}|\,|\mathbf{u}'|}, \quad |\mathbf{u}| = (u_0^2 - 1)^{\frac{1}{2}}$$

[for zero mass we just replace u and u' by p and p', respectively, in (A19)]. The positive or negative sign is to be chosen according to the sense of rotation from $\mathbf{u}$ to $\mathbf{u}'$ (or $\mathbf{u}''$).

Under this transformation the state (A1) undergoes a Wigner rotation, the canonical spin turning in the same sense as $\mathbf{u}$ (about the same axis $\mathbf{u} \times \mathbf{u}'$). The angle of rotation is given by

$$\tan \tfrac{1}{2}\omega = \frac{\pm|\mathbf{u} \times \mathbf{u}''|}{(1 + u_0 + u_0' + u_0'')}. \tag{A20}$$

The difference $(\alpha - \omega)$ can be verified to be given by

$$\cos(\alpha - \omega) = \frac{p_0 p_0' - m^2 u_0''}{|\mathbf{p}|\,|\mathbf{p}'|} = \frac{u_0 u_0' - u_0''}{|\mathbf{u}|\,|\mathbf{u}'|} \tag{A21}$$

(for zero mass ω becomes equal to α).

In helicity states, when both $\mathbf{p}$ and $\mathbf{p}'$ are in the (z, x) plane, we have, with

$$\mathbf{p} = |\mathbf{p}|\,(\sin\theta, 0, \cos\theta),$$

$$\begin{aligned} U(\Lambda)\,|p, \lambda\rangle &= |p', \sigma'\rangle\, d^S_{\sigma'\sigma}(\omega)\, d^S_{\sigma\lambda}(\theta) \\ &= |p', \lambda'\rangle\, d^S_{\lambda'\sigma'}(-(\alpha + \theta))\, d^S_{\sigma'\lambda}(\omega + \theta) \\ &= |p', \lambda'\rangle\, d^S_{\lambda'\lambda}(\omega - \alpha). \end{aligned} \tag{A22}$$

[If we use the convention (3.6) for $\mathbf{p}$ or $\mathbf{p}'$ below the z axis, then additional phase factors may arise for those cases.] For the general case the helicity transformation formula is much less simple in terms of α and ω,[26] but (A22) will be sufficient for our purpose.

For a pure rotation about the y axis [with $\mathbf{p}$ in the (z, x) plane], putting ω equal to α in (A22), we obtain

$$R_y(\varphi)\,|p, \lambda\rangle = |p', \lambda\rangle. \tag{A23}$$

The fact that there is no phase factor on the rhs has an important consequence for the crossing relations.

Finally, let us note briefly the consequences of parity conservation. Under the parity operation P the canonical states can be shown to transform as

$$P\,|p, \sigma\rangle = \eta\,|\underline{p}, \sigma\rangle, \tag{A24}$$

where $\underline{p} = (p_0, -\mathbf{p})$ and η is the intrinsic parity.

Hence, for a canonical amplitude with the momenta taken to be in the (x, y) plane, considering the combined operation of parity and a rotation π about the z axis, we have for parity conserving interactions

$$\begin{aligned} &\langle p_3\sigma_3, p_4\sigma_4|\, T\,|p_1\sigma_1, p_2\sigma_2\rangle \\ &\quad = \langle p_3\sigma_3, p_4\sigma_4|\, U^{-1}TU\,|p_1\sigma_1, p_2\sigma_2\rangle, \quad U = R_z(\pi)P, \\ &\quad = \eta e^{-i\pi(\sigma_1+\sigma_2-\sigma_3-\sigma_4)}\,\langle p_3\sigma_3, p_4\sigma_4|\, T\,|p_1\sigma_1, p_2\sigma_2\rangle, \end{aligned} \tag{A25}$$

where $\eta = \eta_1\eta_2\eta_3\eta_4$. Hence, such amplitudes must vanish unless

$$\eta e^{-i\pi(\sigma_1+\sigma_2-\sigma_3-\sigma_4)} = 1. \tag{A26}$$

APPENDIX B

1. Transformations from c.m. to Breit Systems

Our aim is to express the parameters corresponding to the s- and t-channel Breit systems (as indicated in Figs. 7 and 10, respectively) and also the transformation coefficients connecting them to the c.m. systems in terms of the invariants s and t.

Let us start with the well-known formulas for the s-channel c.m. system. We have (corresponding to Fig. 1 or 5)

$$\begin{aligned} s &= (p_1 + p_2)^2 = (p_3 + p_4)^2, \\ t &= (p_3 - p_1)^2 = (p_2 - p_4)^2, \end{aligned} \tag{B1}$$

and

$$s + t + u = \sum_{i=1}^{4} m_i^2.$$

The scattering angle θ_S is given by

$$\cos\theta_S = \frac{s(t - u) + (m_1^2 - m_2^2)(m_3^2 - m_4^2)}{s_{12}s_{34}} \tag{B2}$$

and

$$\sin\theta_S = \frac{2s^{\frac{1}{2}}\Phi^{\frac{1}{2}}}{s_{12}s_{34}},$$

where

$$s_{ij} = s_{ji} = \{[s - (m_i + m_j)^2][s - (m_i - m_j)^2]\}^{\frac{1}{2}} \tag{B3}$$

and

$$\begin{aligned} \Phi = {} & stu - s(m_2^2 - m_4^2)(m_1^2 - m_3^2) \\ & - t(m_1^2 - m_2^2)(m_3^2 - m_4^2) \\ & - (m_1^2 + m_4^2 - m_3^2 - m_2^2)(m_1^2 m_4^2 - m_2^2 m_3^2). \end{aligned} \tag{B4}$$

For the momenta we have

$$p_i^0 = \frac{s + m_i^2 - m_j^2}{2\sqrt{s}}, \quad |\mathbf{p}_i| = \frac{s_{ij}}{2\sqrt{s}}, \tag{B5}$$

where

$$i = 1, 2 \leftrightarrow j = 2, 1,$$

and

$$i = 3, 4 \leftrightarrow j = 4, 3.$$

Thus,

$$p_{in} = |\mathbf{p}_1| = |\mathbf{p}_2| = s_{12}/2\sqrt{s}$$

and

$$p_{fi} = |\mathbf{p}_3| = |\mathbf{p}_4| = s_{34}/2\sqrt{s}. \tag{B5'}$$

[The corresponding t-channel expressions are given by analogous expressions through the substitutions (2.11').]

Now, (corresponding to Fig. 6) let us consider the pure Lorentz transformation

$$\Lambda_{(B)} \cdot (p_2 + p_3) = (v\ , \mathbf{0}), \tag{B6}$$

where

$$\begin{aligned} v &= (p_2 + p_3)^2 = (p_2 + p_1 + p_3 - p_1)^2 \\ &= s + t + 2(p_2^0 + p_1^0)(p_3^0 - p_1^0) \\ &= s + t + \Delta, \quad \Delta = m_2^2 + m_3^2 - m_1^2 - m_4^2. \end{aligned} \tag{B7}$$

We may also note the relation

$$(v - m_2^2 - m_3^2) = -(u - m_2^2 - m_3^2). \tag{B8}$$

If u'' be the 4-velocity corresponding to $\Lambda_{(B)}$ [in the sense of (A17)], then

$$u'' = v^{-\frac{1}{2}}(p_2^0 + p_3^0, -\mathbf{p}_2 - \mathbf{p}_3)$$

and, hence,

$$u_0'' = \frac{2s + \Delta}{2(sv)^{\frac{1}{2}}}. \tag{B9}$$

Let

$$p_i' = \Lambda_{(B)} \cdot p_i.$$

Then,

$$p_2'^0 = \frac{v + m_2^2 - m_3^2}{2\sqrt{v}}, \qquad p_3'^0 = \frac{v + m_3^2 - m_2^2}{2\sqrt{v}}$$

and

$$p_1'^0 = \frac{s - t + m_3^2 - m_2^2}{2\sqrt{v}}, \quad p_4'^0 = \frac{s - t + m_2^2 - m_3^2}{2\sqrt{v}}. \tag{B10}$$

We may also note that

$$|\mathbf{p}_2'| = |\mathbf{p}_3'| = \frac{v_{23}}{2\sqrt{v}},$$

where

$$v_{23} = \{[v - (m_2 + m_3)^2][v - (m_2 - m_3)^2]\}^{\frac{1}{2}}. \tag{B11}$$

Let us consider the (positive) rotation undergone by $\mathbf{p}_2$ under $\Lambda_{(B)}$. Using (A19), (B2), (B5), and (B10), we finally obtain

$$\sin \alpha_2' = \frac{v^{\frac{1}{2}}(s + m_2^2 - m_3^2) + s^{\frac{1}{2}}(v + m_2^2 - m_3^2)}{((s^{\frac{1}{2}} + v^{\frac{1}{2}})^2 - t)s_{12}v_{23}}\Phi^{\frac{1}{2}}. \tag{B12}$$

The angles θ_1', θ_4', and θ_{14}' in Fig. 7 are given below. (We may omit the primes for the angles without causing any confusion. We have only to remember that s- and t-channel expressions, though apparently the same functions of the invariants, correspond to different s and t values.) We obtain

$$\begin{aligned} 4\,|\mathbf{p}_i'|\,|\mathbf{p}_2'| \cos\theta_i' &= \mathbf{p}_i'^2 - \mathbf{p}_j'^2 + 4\mathbf{p}_2'^2, \\ 4\,|\mathbf{p}_i'|\,|\mathbf{p}_2'| \sin\theta_i' &= \{[4\mathbf{p}_2'^2 - (|\mathbf{p}_1'| + |\mathbf{p}_4'|)^2][4\mathbf{p}_2'^2 - (|\mathbf{p}_1'| - |\mathbf{p}_4'|)^2]\}^{\frac{1}{2}}, \end{aligned} \tag{B13}$$

where

$$i = 1, 4 \leftrightarrow j = 4, 1.$$

Thus,

$$\begin{aligned} 4v\,|\mathbf{p}_1'|\,|\mathbf{p}_2'| \cos\theta_1' &= v_{23}^2 + (s - t)(m_3^2 - m_2^2) + v(m_4^2 - m_1^2) \\ 4v\,|\mathbf{p}_4'|\,|\mathbf{p}_2'| \cos\theta_4' &= v_{23}^2 + (s - t)(m_2^2 - m_3^2) + v(m_1^2 - m_4^2). \end{aligned} \tag{B14}$$

Similarly,

$$4v\,|\mathbf{p}_1'|\,|\mathbf{p}_4'| \cos\theta_{14}' = (\Delta^2 - 4st - v_{23}^2). \tag{B15}$$

Denoting the s-channel c.m. and Breit frame amplitudes by $\mathcal{A}^{(s)}$ and $\mathcal{B}^{(s)}$, respectively, we obtain (for z axis, normal to the scattering plane)

$$\mathcal{A}^{(s)}_{\sigma_3\sigma_4,\sigma_1\sigma_2} = \mathcal{B}^{(s)}_{\sigma_3\sigma_4,\sigma_1\sigma_2} e^{i\varphi_s}, \tag{B16}$$

where

$$\begin{aligned} \varphi_s &= -(\omega_1'\sigma_1 + \omega_2'\sigma_2 - \omega_3'\sigma_3 - \omega_4'\sigma_4) \\ &\quad + (\sigma_1 + \sigma_2 - \sigma_3 - \sigma_4)\alpha_2' \\ &= (\theta_1'\sigma_1 + \theta_4'\sigma_4) - \theta_B(\sigma_3 + \sigma_4) \\ &\quad - (\chi_1^s\sigma_1 - \chi_2^s\sigma_2 - \chi_3^s\sigma_3 + \chi_4^s\sigma_4) \end{aligned} \tag{B17}$$

and [in terms of the differences $(\alpha - \omega)$ of (A21)], in an obvious notation,

$$\begin{aligned} \chi_1^s &= -(\alpha_1' - \omega_1'), \quad \chi_3^s = -(\alpha_3' - \omega_3'), \\ \chi_2^s &= (\alpha_2' - \omega_2'), \qquad \chi_4^s = (\alpha_4' - \omega_4'). \end{aligned} \tag{B18}$$

[As in (2.8), we have again introduced extra negative signs for particles 1 and 3 to compensate for the fact that $\mathbf{p}_1$ and $\mathbf{p}_3$ turn in the negative sense. A careful consideration of the senses of the different rotations is essential for all the terms of (B17).]

The corresponding formulas for t channel may be similarly obtained. It may be verified that (Fig. 9) $-P_1'^0, P_2'^0, P_3'^0, -P_4'^0$ are respectively given by the same formulas (B10) as $p_1'^0, p_2'^0, p_3'^0, p_4'^0$. The angles θ_1, θ_4, and θ_{14} (Fig. 10) are again the same function of the invariant as θ_1', θ_4', and θ_{14}', respectively [giving the required correspondence (2.15)]. The initial differences between the s- and t-channel c.m. parameters are compensated by the fact that now, instead of (B9), we have

$$u_0'' = \frac{2t + \Delta}{2(tv)^{\frac{1}{2}}} \tag{B19}$$

and the rotation β_3' undergone by P_3' (Fig. 9) is given [as compared to (B12)] by

$$\sin \beta_3 = \frac{v^{\frac{1}{2}}(t + m_3^2 - m_1^2) + t^{\frac{1}{2}}(v + m_3^2 - m_1^2)}{((v^{\frac{1}{2}} + t^{\frac{1}{2}})^2 - s)t_{31}v_{23}}\Phi^{\frac{1}{2}}. \tag{B20}$$

Corresponding to (B16), we now have

$$\mathcal{A}^{(t)}_{\sigma_3\sigma_1,\sigma_4\sigma_2} = \mathcal{B}^{(t)}_{\sigma_3\sigma_1,\sigma_4\sigma_2} e^{i\varphi_t}, \tag{B21}$$

where

$$-\varphi_t = (\theta_1\sigma_1 + \theta_4\sigma_4) - \theta_t(\sigma_4 + \sigma_2) - (\chi_1^t\sigma_1 - \chi_2^t\sigma_2 - \chi_3^t\sigma_3 + \chi_4^t\sigma_4) \tag{B22}$$

and

$$\chi_1^t = -(\beta_1 - \omega_1), \quad \chi_2^t = -(\beta_2 - \omega_2), \quad \chi_3^t = (\beta_3 - \omega_3), \quad \chi_4^t = (\beta_4 - \omega_4). \tag{B23}$$

(Again it is crucial to note the senses of the different rotations.)

2. The Lorentz Transformation $\Lambda_{(t)}$ of (2.1)

Let us now consider $\Lambda_{(t)}$, the pure Lorentz transformation introduced in (2.1), which is real for the special case of timelike momentum transfer considered there.

Here

$$u'' = t^{-\frac{1}{2}}(p_3^0 - p_1^0, -\mathbf{p}_3 + \mathbf{p}_1). \tag{B24}$$

Hence,

$$u''^0 = \Delta/2(st)^{\frac{1}{2}}, \quad \Delta = m_2^2 + m_3^2 - m_1^2 - m_4^2. \tag{B25}$$

For the particle 1, we have

$$p_1'^0 = \frac{p_1\cdot(p_3 - p_1)}{t^{\frac{1}{2}}} = -\frac{2p_1^2 + (t - p_1^2 - p_3^2)}{2t^{\frac{1}{2}}} = -\frac{t + m_1^2 - m_3^2}{2t^{\frac{1}{2}}} \tag{B26}$$

and, similarly,

$$p_2'^0 = \frac{t + m_2^2 - m_4^2}{2t^{\frac{1}{2}}}, \quad p_3'^0 = \frac{t + m_3^2 - m_1^2}{2t^{\frac{1}{2}}}, \quad p_4'^0 = -\frac{t + m_4^2 - m_2^2}{2t^{\frac{1}{2}}}.$$

Thus, we see that $-p_1'^0$, $p_2'^0$, $p_3'^0$, $-p_4'^0$ are given respectively by the same functions of t as the corresponding t-channel c.m. energies (Fig. 2),

$$P_1^0, P_2^0, P_3^0, P_4^0.$$

We may verify readily that exactly the same correspondence holds, as required, between the space components of p_i' and P_i.

Again,

$$|\mathbf{p}_i \times \mathbf{u}''| = \frac{p_{in}p_{ft}\sin\theta_S}{t^{\frac{1}{2}}} = \frac{\Phi^{\frac{1}{2}}}{2(st)^{\frac{1}{2}}}, \quad i = 1, 2, 3, 4. \tag{B27}$$

Thus, substituting (B26) and (B27) into (A19) and (A20), we directly obtain α_i and ω_i. In fact, we need only calculate [in view of (2.13)] the relatively simple expressions (A21) along with

$$|\mathbf{p}_i|\,|\mathbf{p}_i'|\sin(\alpha_i - \omega_i) = \pm\frac{m_i\Phi^{\frac{1}{2}}}{2(st)^{\frac{1}{2}}}. \tag{B28}$$

Thus, we immediately obtain the expressions (2.9).

APPENDIX C: VERIFICATION OF THE CANONICAL CROSSING RELATIONS FOR "πN SCATTERING"

We now propose to derive the canonical crossing relations for the s-channel elastic scattering of a spin-zero boson and a spin-$\frac{1}{2}$ fermion (πN scattering being a particular case) via decomposition into invariant amplitudes and the use of Dirac formalism. The purpose is to verify that this well-known technique gives a result consistent with the general formulas for canonical amplitudes derived in Sec. 2.

As is well known, we can write the amplitudes (assuming parity conservation) as

$$\mathcal{A}^{(s)}_{\sigma'\sigma} = \bar{u}_{\sigma'}(p_3)[A + \tfrac{1}{2}B\gamma\cdot(p_2 + p_4)]u_\sigma(p_1), \quad \pi N \to \pi N, \tag{C1}$$

and

$$\mathcal{A}^{(t)}_{\sigma'\sigma} = \bar{u}_{\sigma'}(P_3)[A + \tfrac{1}{2}B\gamma\cdot(P_2 - P_4)]v_\sigma(P_1), \quad \pi\pi \to N\bar{N}. \tag{C2}$$

In our formalism the solutions u and v of the Dirac equation are

$$u_\pm = 2^{-\frac{1}{2}}\begin{vmatrix} [p]_+\cdot\chi_\pm \\ [p]_+^{-1}\cdot\chi_\pm \end{vmatrix}, \quad v_\pm = 2^{-\frac{1}{2}}\begin{vmatrix} [p]_+C^{-1}\cdot\chi_\pm \\ [p]_+^{-1}C\cdot\chi_\pm \end{vmatrix}, \tag{C3}$$

where

$$[p]_+ = \frac{m + p_0 + \mathbf{p}\cdot\boldsymbol{\tau}}{[2m(m + p_0)]^{\frac{1}{2}}}, \quad C = \begin{vmatrix} & 1 \\ -1 & \end{vmatrix}, \tag{C3'}$$

and

$$\chi_+ = \begin{vmatrix} 1 \\ 0 \end{vmatrix}, \quad \chi_- = \begin{vmatrix} 0 \\ 1 \end{vmatrix}.$$

The form (C3) of the solutions implies the following representation of the γ matrices:

$$\gamma_0 = \begin{vmatrix} & I_2 \\ I_2 & \end{vmatrix}, \quad \gamma^5 = \begin{vmatrix} I_2 & \\ & -I_2 \end{vmatrix},$$

$$\boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma} = \begin{vmatrix} \boldsymbol{\tau} & \\ & -\boldsymbol{\tau} \end{vmatrix}, \quad \boldsymbol{\sigma} = \begin{vmatrix} \boldsymbol{\tau} & \\ & \boldsymbol{\tau} \end{vmatrix} = \gamma^5\boldsymbol{\alpha}, \tag{C4}$$

the solutions being eigenfunctions of the canonical spin projection Σ_z, where $\boldsymbol{\Sigma}$ [see (A3) and the following comments] is given in the Dirac representation[25] by

$$\boldsymbol{\Sigma} = \frac{P^0}{m}\boldsymbol{\sigma} + \frac{i\boldsymbol{\alpha}\times\mathbf{P}}{m} - \frac{(\boldsymbol{\sigma}\cdot\mathbf{P})\mathbf{P}}{m(P_0 + m)}. \tag{C5}$$

Let us first consider the *c.m. amplitudes*. Let

$$p^0 = p_1^0 = p_3^0, \quad q^0 = p_2^0 = p_4^0. \tag{C6}$$

Then, with the momenta in the (x, y) plane (Fig. 1),

$$\mathcal{A}_{\pm\pm}^{(s)} = (2m)^{-1}\{(p_0 + m)[A + B(p_0 + q_0 - m)] - (p_0 - m)[A - B(p_0 + q_0 + m)]e^{\mp i\theta_S}\} \tag{C7}$$

and

$$\mathcal{A}_{\mp\pm}^{(s)} = 0, \tag{C7'}$$

where θ_S is given by (B2) with

$$m_1 = m_3\,(= m_N) = m, \tag{C8}$$

and

$$m_2 = m_4\,(= m_\pi) = \mu.$$

In continuing the rhs of (C7) to the t-channel physical region where $s < 0$ and $t > 0$, with A and B supposed holomorphic, the only ambiguity arises from the two possibilities

$$s^{\frac{1}{2}} \to \pm i(-s)^{\frac{1}{2}}. \tag{C9}$$

(Here we assume $\Phi^{\frac{1}{2}}$ to be positive at the end of continuation, as in Ref. 5.) Denoting all the continued parameters by adding a "$\sim$" sign, we obtain, corresponding to the two signs of (C9),

$$[\exp(-i\tilde{\theta}_S)]_{(\pm)} = (\exp i\tilde{\theta}_S)_{(\mp)} = -\frac{\{(-st)^{\frac{1}{2}} \mp [su - (m^2 - \mu^2)^2]^{\frac{1}{2}}\}^2}{S^2} \tag{C10}$$

$$= -\frac{\{(-st)^{\frac{1}{2}} \mp [-(st + S^2)]^{\frac{1}{2}}\}^2}{S^2}, \tag{C10'}$$

where

$$S^2 = [s - (m + \mu)^2][s - (m - \mu)^2]. \tag{C10''}$$

Similarly,

$$(\tilde{p}_0)_{(\pm)} = \mp i\frac{s + m^2 - \mu^2}{2(-s)^{\frac{1}{2}}}, \qquad (\tilde{q}_0)_{(\pm)} = \mp i\frac{s + \mu^2 - m^2}{2(-s)^{\frac{1}{2}}}. \tag{C11}$$

Substituting (C10) and (C11) into (C7) and carefully rearranging the terms, we finally obtain

$$(\tilde{\mathcal{A}}_{++}^{(s)})_{(\pm)} = \mp i\frac{(-st)^{\frac{1}{2}} \mp [-(st + S^2)]^{\frac{1}{2}}}{S}e^{-i\chi}\mathcal{A}_{++}^{(t)} \tag{C12}$$

and

$$(\tilde{\mathcal{A}}_{--}^{(s)})_{(\pm)} = \pm i\frac{(-st)^{\frac{1}{2}} \pm [-(st + S^2)]^{\frac{1}{2}}}{S}e^{i\chi}\mathcal{A}_{--}^{(t)}, \tag{C13}$$

where

$$e^{\mp i\chi} = \frac{-(s + m^2 - \mu^2)t^{\frac{1}{2}} \mp i2m[-(st + S^2)]^{\frac{1}{2}}}{S(t - 4m^2)^{\frac{1}{2}}}. \tag{C14}$$

The t-channel amplitudes may be written in terms of the invariants as

$$\mathcal{A}_{\pm\pm}^{(t)} = \mp[(Pm^{-1}A - Q\cos\theta_t B) \pm iP_0 m^{-1}Q\sin\theta_t B], \tag{C15}$$

where

$$P_0 = \tfrac{1}{2}(t)^{\frac{1}{2}}, \quad P = \tfrac{1}{2}(t - 4m^2)^{\frac{1}{2}}, \quad Q = \tfrac{1}{2}(t - 4\mu^2)^{\frac{1}{2}} \tag{C16}$$

and

$$e^{i\theta_t} = \frac{(s - u) + i2[-(st + S^2)]^{\frac{1}{2}}}{(t - 4m^2)^{\frac{1}{2}}(t - 4\mu^2)^{\frac{1}{2}}}. \tag{C17}$$

Also, consistently with the continuation of (C7'), we have

$$\mathcal{A}_{\pm\mp}^{(t)} = 0. \tag{C15'}$$

In order to compare (C12) and (C13) with (2.37) or (2.38), we have to note that from (2.9) and (C8)

$$e^{\mp\frac{1}{2}i(\chi_1 - \chi_3)} = e^{\mp i(\frac{1}{2}\pi + \chi)} = \mp ie^{\mp i\chi}. \tag{C18}$$

Hence, putting $\sigma_2 = \sigma_4 = 0$ in (2.37) or (2.38), we obtain

$$\tilde{\mathcal{A}}_{\pm\pm}^{(s)} = (-1)^{\eta+1}e^{\mp i\frac{1}{2}\tilde{\theta}_S}e^{\mp i\chi}\mathcal{A}_{\pm\pm}^{(t)}. \tag{C19}$$

If, for example, by starting from (C10') we utilize the determination

$$(e^{-i\frac{1}{2}\theta_S})_{(\pm)} = (e^{+i\frac{1}{2}\theta_S})_{(\mp)} = \mp i\frac{(-st)^{\frac{1}{2}} \pm [-(st + S^2)]^{\frac{1}{2}}}{S}, \tag{C20}$$

we must then have

$$(-1)^{\eta+1} = 1, \tag{C21}$$

for (C12) and (C13) to agree with (C17). In any case, we have the required consistency to within a constant over-all phase factor, which we make no attempt to determine explicitly. It is possible to choose other determinations of the square root leading to a different value of η. It depends also on the relative sign of u and v.

It is amusing to note that, for the parity conserving amplitudes discussed above, the coefficient from the rotation $R_z(-\alpha_1)$ given by (2.5) becomes unity and the role of this operation is hidden just as for the helicity amplitudes in the (z, x) plane. Its necessity becomes evident if we add into (C1) the parity changing terms $\gamma_5(C + \frac{1}{2}D\gamma \cdot (p_2 + p_4))$ with a corresponding modification of (C2).

Let us now consider the *canonical Breit-frame amplitudes*. We start again with (C1) and (C2), replacing $\mathcal{A}^{(s)}$ and $\mathcal{A}^{(t)}$ by the symbols $\mathcal{B}^{(s)}$ and $\mathcal{B}^{(t)}$, respectively, to indicate that the momenta now satisfy the Breit-frame constraints, namely those indicated in Figs. 7 and 10, respectively. Of crucial importance is the fact that the momenta now satisfy Condition I. The explicit forms are given in Appendix B, but are not even necessary for our present purpose. All we need to note is that, as a direct consequence of Condition I,

$$\tilde{\mathcal{B}}^{(s)}_{\sigma'\sigma} = \bar{u}_{\sigma'}(P_3)[A + \tfrac{1}{2}B\gamma\cdot(P_2 - P_4)]u_{\sigma}(-P_1). \quad \text{(C22)}$$

Since in our case the momenta lie in the (x, y) plane, it is quite easy to verify that

$$u_{\pm}(-P) = \left(\pm i\frac{P_x \pm iP_y}{|\mathbf{P}|}\right)v_{\pm}(P), \quad \text{(C23)}$$

where we have used for the sake of definiteness the first determination of the square root indicated in (2.27). Thus, referring to the angle θ_1 in Fig. 10, we have

$$\tilde{\mathcal{B}}^{(s)}_{\sigma'\sigma} = \mp ie^{\mp i\theta_1}\mathcal{B}^{(t)}_{\sigma'\sigma}. \quad \text{(C24)}$$

This agrees with (2.39) to within a constant phase factor. Thus, we see that for Breit-frame amplitude the verification involves very little calculation.

We would like to mention one final point. If we write the positive and negative energy solutions of the Dirac equation

$$(\gamma\cdot P - m)\psi = 0$$

as

$$u^{(\epsilon)}_{\pm}(p) = (Q^{(\epsilon)})^{-1}(p)\begin{vmatrix}\chi_{\pm}\\ \epsilon\chi_{\pm}\end{vmatrix}, \quad \text{(C25)}$$

where

$$Q^{(\epsilon)}(p) = \frac{\epsilon m + \gamma^0(\gamma\cdot p)}{[2\epsilon m(\epsilon m + p_0)]^{\frac{1}{2}}} = \frac{m + \epsilon\gamma^0(\gamma\cdot p)}{[2m(m + |p_0|)]^{\frac{1}{2}}} \quad \text{(C26)}$$

is the canonical transformation introduced in Ref. 25 with the role of the sign of the energy mode explicit, we find that

$$\begin{aligned} v_{\sigma}(+\mathbf{p}) &= (-1)^{\frac{1}{2}-\sigma}\gamma^5 u^{(\mp)}_{-\sigma}(+\mathbf{p}) \\ &= (-1)^{\frac{1}{2}-\sigma}u^{(-)}_{-\sigma}(-\mathbf{p}). \end{aligned} \quad \text{(C27)}$$

Hence, (C2) can be written as

$$\begin{aligned} \mathcal{A}^{(t)}_{\sigma'\sigma} = \bar{u}^{(+)}_{\sigma'}(\mathbf{P}_3)[A + \tfrac{1}{2}\gamma\cdot(P_2 - P_4)] \\ \times ((-1)^{\frac{1}{2}-\sigma}u^{(-)}_{-\sigma}(-\mathbf{P}_1)). \end{aligned} \quad \text{(C28)}$$

Thus, in terms of $u^{(\epsilon)}_{\sigma}$, the language becomes quite consistent with that of the general canonical case [compare (2.36)], where the rotation and Lorentz-transformation properties are sufficient to determine the crossing behavior, without it being necessary to invoke other mutually distinct aspects of a particle and its antiparticle.

[1] T. L. Trueman and G.-C. Wick, Ann. Phys. (N.Y.) **26**, 322 (1964).

[2] I. J. Muzinich, J. Math. Phys. **5**, 1481 (1964).

[3] E. L. Surkov, Yad. Fiz. **1**, 1113 (1965) [Sov. J. Nucl. Phys. **1**, 792 (1965)].

[4] A. Kotansky, Acta Phys. Polon. **29**, 699; **30**, 629 (1966).

[5] G. Cohen-Tannoudji, A. Morel, and H. Navelet, Ann. Phys. (N.Y.) **46**, 239 (1968).

[6] A. Chakrabarti, J. Math. Phys. **5**, 922, 1747 (1964); thesis, University of Paris, 1965 (other important sources are cited in the above articles).

[7] S. Frautschi and L. Jones, Phys. Rev. **167**, 1335 (1968); J. D. Jackson and G. E. Hite, *ibid.* **169**, 1248 (1968); J. Franklin, *ibid.* **170**, 1606 (1968).

[8] M. Barmawi, Phys. Rev. **166**, 1846 (1968).

[9] A. Chakrabarti, Nuovo Cimento **43**, 576 (1966) (Sec. 1.1 and Appendix A).

[10] In Sec. 3 we choose our conventions so as to make this correspondence as simple as possible, avoiding any change of sign of the spin indices. This is also reflected in the crossing relations.

[11] Again, it is for this very reason that we do not perform the continuation and crossing, in terms of the spinor amplitudes, at one stroke [Sec. 2D(i)].

[12] By "Breit frame" we mean the one in which the uncrossed particles 2 and 3 have equal and opposite momenta, chosen, moreover, for the canonical amplitudes, to be along the x axis. Thus, for example, for πN scattering (Appendix C) the nucleon and the meson would not appear to be reflected separately at the "brick wall," which is the y axis in our case. Instead, the incoming meson (nucleon) would be "reflected" as the outgoing nucleon (meson). This is not the usual choice. But it is an essential one for our purposes and need not cause any confusion.

[13] J. Bjørneboe and Z. Koba, Nucl. Phys. **B7**, 53 (1968); Z. Koba, *ibid.* **B8**, 351 (1968). A different definition of the "Breit frame" is used by these authors, which is related to the reduction of the 4-vector momentum exchange corresponding to the u channel of the form $(0, 0, 0, \sqrt{-u})$ for $u < 0$. (They replace u by t, considering crossing between s and u channels.) This leads to expressions for the momenta with quite interesting properties. As will be evident from the argument in the following section, for these frames [satisfying also (A3)] our formalism for continuation (Sec. 2C) would give again the canonical crossing relation (2.39), where now θ_1 and θ_4 will refer to angles in these frames. For the transversity and helicity amplitudes we would get back (3.18) and (3.19), respectively.

For $u > 0$, Eq. (A3) may be attained by first reducing $p_2 - p_3$ to $(\sqrt{u}, \mathbf{0})$ in both the s and t channels. For $u^2 = 0$ we have to adopt yet another prescription. Our transformation (B6) is applicable uniformly to all the above-mentioned cases.

[14] This exercise also serves to indicate without ambiguity the final t-channel configuration (Fig. 4) at which we must arrive after continuation and crossing. Turning the entire t-channel configuration around the z axis through any arbitrary angle (not a function s, t) is rather trivial. But let us compare the configuration (Fig. 4) to the one obtained from it through a rotation π around the x axis. The corresponding amplitude may be shown to be related through the reversal of sign of all the canonical spin indices, as well as through a phase factor. Thus, if the crossing relation between the amplitudes, corresponding to Figs. 1 and 4, conserve, as it will turn out, all the spin indices including their signs, then those between the amplitudes for Fig. 1 and Fig. 4, rotated as above, must reverse all the spins *including those of the uncrossed particles*. This last possibility never arises in our formalism and, as should be particularly clear from the discussion in Appendic C, is certainly not related to the ambiguities because of the different possible continuations of the square roots involved. This feature should be compared to the comments referring to Fig. 5 of Ref. 3.

[15] J. Bros, H. Epstein, and V. Glaser, Commun. Math. Phys. **1**,

240 (1965); K. Hepp, Helv. Phys. Acta **36**, 355 (1963); D. N. Williams, Lawrence Radiation Laboratory, University of California, Berkeley, Preprint, UCRL 11113.

[16] Since there is again another change of sign [Eq. (2.36)] involved in the crossing over of the two particles, the final crossing formula (2.37) conserves all the spin indices with their signs, allowing us to really "diagonalize" the crossing matrix.

[17] The simultaneous consideration of the alternative possibilities of continuation becomes important when one studies the continuations of cross sections and density matrices. We do not, however, intend to discuss these aspects in this article.

[18] A. Chakrabarti, J. Math. Phys. **7**, 949 (1966).

[19] J. P. Ader, M. Capedeville, and H. Navelet, (private communication).

[20] H. Joos, Fortschr. Physik **10**, 65 (1962).

[21] A. Chakrabarti, J. Math. Phys. **8**, 1367 (1967).

[22] A. Kotansky, Nuovo Cimento **56A**, 737 (1968).

[23] A. McKerrel, J. Math. Phys. **9**, 1824 (1968).

[24] R. Hwa, Phys. Rev. **180**, 1588 (1968).

[25] A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963).

[26] V. I. Ritus, Zh. Eksp. Teor. Fiz. **40**, 352 (1961) [Sov. Phys.—JETP **13**, 240 (1961)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Radiation from a Dipole Imbedded in an Inhomogeneous Media with Critical Coupling Regions

E. BAHAR
Electrical Engineering Department, University of Nebraska, Lincoln, Nebraska 68508

(Received 23 June 1969)

Maxwell's equations for inhomogeneous media are transformed into loosely coupled first-order differential equations by taking into account the local features of the refractive index profile. A set of generalized WKB type solutions are derived for the electromagnetic fields which are valid for regions of critical coupling. These solutions also merge with the familiar WKB solutions for vertically polarized waves propagating in slowly varying media.

1. INTRODUCTION

It was first pointed out by Galejs[1,2] that, for vertically polarized waves propagating in an inhomogeneous medium with an exponential profile, there exist rigorous solutions in terms of standard tabulated functions. Until recently this was apparently the only case for which standard, full-wave solutions were known for the propagation of vertically polarized waves in inhomogeneous media. It is well known that, for horizontally polarized waves, there exists an extensive family of dielectric profiles for which the solutions may be written in terms of standard functions.[2-4] Several analytical solutions have been developed in the past to extend the solutions for the propagation of horizontally polarized waves in a medium with an arbitrarily varying dielectric profile. These include iterative solutions, infinite series expansions, and uniform asymptotic expansions.[2-8] In addition, extensive numerical methods have been developed.[2-4,9] However, the problem of propagation of vertically polarized waves in inhomogeneous media has not been treated as extensively in the literature. A principal factor that contributes to the special difficulty associated with vertically polarized waves is that, for this case, rigorous solutions to the problem of propagation in linearly stratified media do not exist. The existence of such rigorous full-wave solutions for horizontally polarized waves (the Airy integral functions) has contributed immensely to the understanding of the behavior of these waves in regions of critical coupling (where regular WKB-type solutions fail) and to the development of uniform asymptotic expansions.[7]

Since the propagation from an electric dipole in an inhomogeneous dielectric medium involves differential equations similar to those governing vertically polarized waves, we shall, in this paper, consider this problem in detail. In an attempt to obtain a better understanding of the behavior of vertically polarized waves in regions of critical coupling, an appropriate substitute is found for the Airy integral functions. The method used here to solve this problem is based on the development of a set of loosely coupled first-order differential equations by taking into account the local features of the refractive index profile. This is referred to as a generalized WKB solution.[10] It is shown that these solutions merge with the regular WKB-type solutions for regions in which the refractive index profile is slowly varying.

2. STATEMENT OF THE PROBLEM

The electromagnetic fields radiated by an electric dipole immersed in an inhomogeneous dielectric material are derived. The dielectric coefficient of the

240 (1965); K. Hepp, Helv. Phys. Acta **36**, 355 (1963); D. N. Williams, Lawrence Radiation Laboratory, University of California, Berkeley, Preprint, UCRL 11113.

[16] Since there is again another change of sign [Eq. (2.36)] involved in the crossing over of the two particles, the final crossing formula (2.37) conserves all the spin indices with their signs, allowing us to really "diagonalize" the crossing matrix.

[17] The simultaneous consideration of the alternative possibilities of continuation becomes important when one studies the continuations of cross sections and density matrices. We do not, however, intend to discuss these aspects in this article.

[18] A. Chakrabarti, J. Math. Phys. **7**, 949 (1966).

[19] J. P. Ader, M. Capedeville, and H. Navelet, (private communication).

[20] H. Joos, Fortschr. Physik **10**, 65 (1962).

[21] A. Chakrabarti, J. Math. Phys. **8**, 1367 (1967).

[22] A. Kotansky, Nuovo Cimento **56A**, 737 (1968).

[23] A. McKerrel, J. Math. Phys. **9**, 1824 (1968).

[24] R. Hwa, Phys. Rev. **180**, 1588 (1968).

[25] A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963).

[26] V. I. Ritus, Zh. Eksp. Teor. Fiz. **40**, 352 (1961) [Sov. Phys.—JETP **13**, 240 (1961)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Radiation from a Dipole Imbedded in an Inhomogeneous Media with Critical Coupling Regions

E. BAHAR
Electrical Engineering Department, University of Nebraska, Lincoln, Nebraska 68508

(Received 23 June 1969)

Maxwell's equations for inhomogeneous media are transformed into loosely coupled first-order differential equations by taking into account the local features of the refractive index profile. A set of generalized WKB type solutions are derived for the electromagnetic fields which are valid for regions of critical coupling. These solutions also merge with the familiar WKB solutions for vertically polarized waves propagating in slowly varying media.

1. INTRODUCTION

It was first pointed out by Galejs[1,2] that, for vertically polarized waves propagating in an inhomogeneous medium with an exponential profile, there exist rigorous solutions in terms of standard tabulated functions. Until recently this was apparently the only case for which standard, full-wave solutions were known for the propagation of vertically polarized waves in inhomogeneous media. It is well known that, for horizontally polarized waves, there exists an extensive family of dielectric profiles for which the solutions may be written in terms of standard functions.[2-4] Several analytical solutions have been developed in the past to extend the solutions for the propagation of horizontally polarized waves in a medium with an arbitrarily varying dielectric profile. These include iterative solutions, infinite series expansions, and uniform asymptotic expansions.[2-8] In addition, extensive numerical methods have been developed.[2-4,9] However, the problem of propagation of vertically polarized waves in inhomogeneous media has not been treated as extensively in the literature. A principal factor that contributes to the special difficulty associated with vertically polarized waves is that, for this case, rigorous solutions to the problem of propagation in linearly stratified media do not exist. The existence of such rigorous full-wave solutions for horizontally polarized waves (the Airy integral functions) has contributed immensely to the understanding of the behavior of these waves in regions of critical coupling (where regular WKB-type solutions fail) and to the development of uniform asymptotic expansions.[7]

Since the propagation from an electric dipole in an inhomogeneous dielectric medium involves differential equations similar to those governing vertically polarized waves, we shall, in this paper, consider this problem in detail. In an attempt to obtain a better understanding of the behavior of vertically polarized waves in regions of critical coupling, an appropriate substitute is found for the Airy integral functions. The method used here to solve this problem is based on the development of a set of loosely coupled first-order differential equations by taking into account the local features of the refractive index profile. This is referred to as a generalized WKB solution.[10] It is shown that these solutions merge with the regular WKB-type solutions for regions in which the refractive index profile is slowly varying.

2. STATEMENT OF THE PROBLEM

The electromagnetic fields radiated by an electric dipole immersed in an inhomogeneous dielectric material are derived. The dielectric coefficient of the

medium of propagation $\epsilon(z)$ is assumed to be an arbitrary function of the z coordinate. An electric dipole of strength $I\Delta s$ is assumed to be oriented along the z axis at the point $z = z_0$. (See Fig. 1.) The electric-current density vector $\mathbf{J}$ in a circular, cylindrical coordinate system is

$$\mathbf{J} = J_z\mathbf{a}_z = (I\Delta s/2\pi\rho)\delta(z - z_0)\delta(\rho)\mathbf{a}_z, \quad (2.1)$$

in which $\delta(z - z_0)$ and $\delta(\rho)$ are Dirac δ functions of z and ρ, respectively, and $\mathbf{a}_z$ is the z-directed unit vector. An $e^{i\omega t}$ time harmonic dependence is assumed throughout this paper. From the symmetry of the problem, we note that the magnetic field has only an H_ϕ component. Thus, using Maxwell's curl equations, it follows that

$$-\frac{\partial H_\phi}{\partial z} = i\omega\epsilon E_\rho, \quad (2.2)$$

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho H_\phi) = i\omega\epsilon E_z + J_z, \quad (2.3)$$

and

$$\frac{\partial E_\rho}{\partial z} - \frac{\partial E_z}{\partial\rho} = -i\omega\mu_0 H_\phi, \quad (2.4)$$

in which E_ρ and E_z are the ρ and z components of the electric field, respectively, and the permeability μ_0 is assumed to be the same as that for free space. Eliminating E_z from (2.4) by using (2.3), one obtains

$$i\omega\epsilon\frac{\partial E_\rho}{\partial z} - L_\rho[H_\phi] - k^2n^2H_\phi = -\frac{\partial J_z}{\partial\rho}, \quad (2.5a)$$

in which $k = \omega(\mu_0\epsilon_0)^{\frac{1}{2}}$ is the free-space wave number, n is the refractive index (a function of z), and the operator L_ρ is defined as

$$L_\rho[H_\phi] = \left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{1}{\rho^2}\right)H_\phi. \quad (2.5b)$$

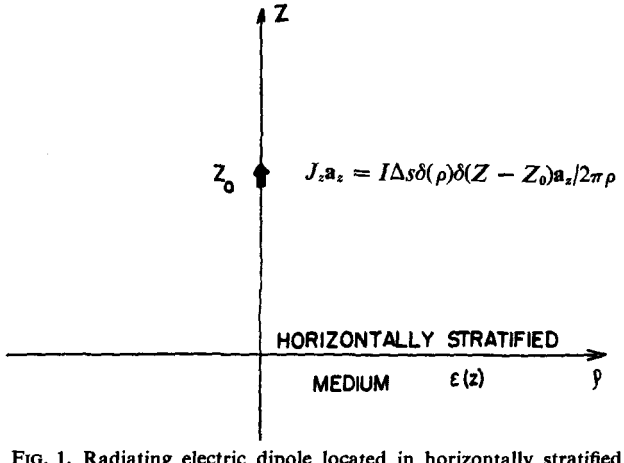


FIG. 1. Radiating electric dipole located in horizontally stratified medium.

Inspection of the operator L_ρ indicates the following separable form for H_ϕ:

$$H_\phi = \int_0^\infty h(z, \lambda)J_1(\lambda\rho)\,d\lambda, \quad (2.6)$$

in which $J_1(\lambda\rho)$ is the Bessel's function of order 1 and argument $\lambda\rho$. The problem then reduces to determining the expression for $h(z, \lambda)$, which is a function of z and the variable of integration λ. In view of the expression (2.6) for H_ϕ, we express the Dirac δ function $\delta(\rho)$ in terms of its Fourier–Bessel transform.[11] Thus,

$$\frac{\delta(\rho)}{\rho} = \int_0^\infty J_0(\lambda\rho)\lambda\,d\lambda \quad (2.7a)$$

and

$$\frac{\partial J_z}{\partial\rho} = -\frac{I\Delta s}{2\pi}\delta(z - z_0)\int_0^\infty \lambda^2J_1(\lambda\rho)\,d\lambda. \quad (2.7b)$$

Hence, (2.5a) reduces to the following differential equation:

$$i\omega\epsilon\frac{\partial E_\rho}{\partial z} - k^2q^2H_\rho = \frac{I\Delta s}{2\pi}\delta(z - z_0)\int_0^\infty \lambda^2J_1(\lambda\rho)\,d\lambda, \quad (2.8a)$$

in which

$$kq = (k^2n^2 - \lambda^2)^{\frac{1}{2}}. \quad (2.8b)$$

In the above expression, the square root is chosen such that q lies in the fourth quadrant. In view of (2.2) and (2.8), it is obvious that E_ρ has the same ρ dependence as H_ϕ. Hence, we write E_ρ as

$$E_\rho = \int_0^\infty e(z, \lambda)J_1(\lambda\rho)\,d\lambda, \quad (2.9)$$

in which $e(z, \lambda)$ is as yet an unknown function of z and λ. Eliminating E_ρ from (2.2) and (2.8), we obtain the following differential equation for H_ϕ:

$$L_z[H_\phi] \equiv \frac{\partial^2H_\phi}{\partial z^2} - \frac{d\ln(n^2)}{dz}\frac{\partial H_\phi}{\partial z} + k^2q^2H_\phi = -\frac{I\Delta s}{2\pi}\delta(z - z_0)\int_0^\infty \lambda^2J_1(\lambda\rho)\,d\lambda, \quad (2.10a)$$

in which

$$\frac{d\ln(n^2)}{dz} = \frac{1}{n^2}\frac{\partial}{\partial z}(n^2).$$

Thus, using (2.6), we derive the following ordinary differential equation from (2.10a):

$$L_z[h(z, \lambda)] = -(I\Delta s\lambda^2/2\pi)\delta(z - z_0). \quad (2.10b)$$

For a homogeneous dielectric, the Green's function solution for $h(z, \lambda)$ is known.[12] Thus, for $n = 1$, we obtain,

$$h(z, \lambda) = (I\Delta s\lambda^2/4\pi ikq)\exp(-ikq\,|z - z_0|) \quad (2.11a)$$

and

$$H_\phi = \frac{I\Delta s}{4\pi}\int_0^\infty \frac{J_1(\lambda\rho)}{ikq}\exp(-ikq\,|z - z_0|)\lambda^2\,d\lambda. \tag{2.11b}$$

Thus, noting that[2]

$$\frac{e^{-ikR}}{R} = \int_0^\infty \frac{J_0(\lambda\rho)}{ikq}\exp(-ikq\,|z - z_0|)\lambda\,d\lambda, \tag{2.11c}$$

we obtain

$$H_\phi = -\frac{I\Delta s}{4\pi}\frac{\partial}{\partial\rho}\left(\frac{e^{-ikR}}{R}\right), \tag{2.11d}$$

in which $R = [\rho^2 + (z - z_0)^2]^{\frac{1}{2}}$. Using (2.3) and (2.4), it is now possible to solve for E_ρ and E_z. In the general case, however, for an arbitrarily varying function $\epsilon(z)$, it is not possible to obtain a solution for $h(z, \lambda)$ in terms of standard tabulated functions. The problem of radiation from the electric dipole considered here is related to the problem of propagation of vertically polarized waves incident upon an inhomogeneous dielectric medium. Until recently, there was apparently only one nonuniform refractive-index profile (the exponential profile) for which the solutions could be expressed in terms of standard tabulated functions.[1,2]

In the next section, we shall study this special case and then proceed to consider the general case in which $\epsilon(z)$ is an arbitrary function of z. This, of course, is the main purpose of this paper. In this latter case, we shall find it more convenient to directly use primary equations (2.2) and (2.8) for E_z and H_ϕ rather than the secondary equations such as (2.9).

Using (2.6) and (2.9), we can factor the common ρ dependence of E_ρ and H_ϕ in (2.2) and (2.8). Thus, we obtain the ordinary first-order coupled differential equations for e and h,

$$\frac{dh}{dz} = -i\omega\epsilon e \tag{2.12a}$$

and

$$i\omega\epsilon\frac{de}{dz} = k^2q^2h + \frac{I\Delta s}{2\pi}\lambda^2\delta(z - z_0). \tag{2.12b}$$

In the following sections, solutions will be derived for the above equations (2.12). These solutions for $h(z, \lambda)$ can then be substituted into the expressions for H_ϕ[(2.6)]. If one only seeks far field solutions, it is possible to apply the familiar saddle-point methods to solve (2.6) since the principal contributions to the field at the receiver come from a narrow range of the values of λ (corresponding to a thin beam centered around the ray linking the receiver to the source). In general, however, it may be necessary to compute the integral (2.6) numerically.

3. RADIATION FROM AN ELECTRIC DIPOLE IMMERSED IN A MEDIUM WITH AN EXPONENTIALLY VARYING DIELECTRIC COEFFICIENT

We consider here the special case in which the refractive-index profile is given by

$$n^2(z) = n_0^2e^{\beta z}, \tag{3.1}$$

in which n_0 is the complex value of n at the reference level $z = 0$. The constant parameter β is assumed to be a positive real number in this section.

For the case $\beta > 0$, the Green's function solution to (2.10b) is

$$\begin{aligned} h(z, \lambda) &= h_0vJ_\nu(v)H_\nu^{(2)}(v_0), \quad z < z_0, \\ &= h_0vH_\nu^{(2)}(v)J_\nu(v_0), \quad z > z_0, \end{aligned} \tag{3.2a}$$

in which

$$v = (2k/\beta)n \tag{3.2b}$$

and

$$\nu = [1 + (2\lambda/\beta)^2]^{\frac{1}{2}}, \tag{3.2c}$$

where the square root is chosen such that $\mathrm{Re}\,\nu > 0$. The constant h_0 is determined by the discontinuity of the derivative of $h(z, \lambda)$ at $z = z_0$. Thus, it follows that[12]

$$h_0 = -\frac{I\Delta s\lambda^2/2\pi}{v_0W_0\,(dv_0/dz)} = \frac{I\Delta s\lambda^2}{4ikn}, \tag{3.2d}$$

in which W_0 is the value of the Wronskian $W[J_\nu(v_0), H_\nu^{(2)}(v_0)]$ at $v_0 = (2k/\beta)n(z_0)$. The Debye–Watson or WKB expansion for $H_\nu^{(2)}(v)$ (*assuming n_0 is real*) is[2]

$$H_\nu^{(2)}(v) \approx \left(\frac{2}{\pi vC}\right)^{\frac{1}{2}}\exp(\tfrac{1}{4}i\pi)\exp\left(-i\int_\nu^v C\,dv\right), \tag{3.3a}$$

for $\nu \gg 1$ $v \gg 1$ and $\nu/v \ll 1$, where

$$C = [1 - (\nu/v)^2]^{\frac{1}{2}} \approx [1 - (\lambda/kn)^2]^{\frac{1}{2}} = q/n. \tag{3.3b}$$

Similarly,

$$J_\nu(v) = \left(\frac{2}{\pi vC}\right)^{\frac{1}{2}}\cos\left(\int_\nu^v C\,dv - \tfrac{1}{4}\pi\right). \tag{3.4a}$$

The above asymptotic solutions are appropriate for the region above the critical coupling layer $z = z_r$, with $v(z_r) \equiv \nu$ and $C(z_r) \equiv 0$, for a particular value of λ, and for which the location of the dipole z_0 is above this critical layer $z = z_r$. Note that for $z > z_0$ the solution exhibits the characteristics of an upgoing wave and that for $z < z_0$ it exhibits a standing wave behavior. The ratio of the amplitudes of the upgoing to the downgoing waves R at a reference level z_R can be calculated using (3.4a):

$$R = i\exp\left(-2i\int_\nu^{v_R} C\,dv\right), \quad \nu \ll v_R = v(z_R) < v_0. \tag{3.4b}$$

Furthermore, using the series expansion for $J_\nu(v)$ for large negative values of z ($v \to 0$), we get for $\nu \gg 1$

$$J_\nu(z \to -\infty) = \frac{(\frac{1}{2}v)^\nu}{\nu!} \to \frac{kn_0}{\beta}\frac{e^{\lambda z}}{\nu!} \to 0. \quad (3.5)$$

Thus, below the critical layer ($z \to -\infty$), the solution exhibits the behavior of an evanescent wave. We now investigate the behavior of the solutions in the vicinity of the critical coupling layer. For the case $z_0 > z_r$ and $z \approx z_r$, the uniform asymptotic expansion for $J_\nu(v)$ may be written as[13]

$$J_\nu(v) \approx \left(\frac{4\zeta}{1-u^2}\right)^{\frac{1}{4}} \frac{\mathrm{Ai}\,(\nu^{\frac{2}{3}}\zeta)}{\nu^{\frac{1}{3}}}, \quad (3.6a)$$

in which Ai is the Airy integral function

$$\tfrac{2}{3}\zeta^{\frac{3}{2}} = \ln\left(\frac{1+(1-u^2)^{\frac{1}{2}}}{u}\right) - (1-u^2)^{\frac{1}{2}}, \quad (3.6b)$$

and

$$u = v/\nu = 2kn_0 e^{\frac{1}{2}\beta z}/\beta\nu \approx e^{\frac{1}{2}\beta(z-z_r)}. \quad (3.6c)$$

Thus, for

$$\nu \gg 1, \quad z \approx z_r, \quad u \approx 1 + \tfrac{1}{2}\beta(z - z_r), \quad \text{and} \quad \zeta \approx -(\tfrac{1}{2})^{\frac{2}{3}}\beta(z - z_r), \quad (3.7a)$$

we get

$$J_\nu(v) \approx (2/\nu)^{\frac{1}{3}}\,\mathrm{Ai}\,[-\nu^{\frac{2}{3}}(\tfrac{1}{2})^{\frac{2}{3}}\beta(z - z_r)] \approx (2/\nu)^{\frac{1}{3}}\,\mathrm{Ai}\,[-\lambda^{\frac{2}{3}}\beta^{\frac{1}{3}}(z - z_r)]. \quad (3.7b)$$

The above expression for $J_\nu(v)$ clearly indicates the transition from a standing wave solution for $z > z_r$ to an evanescent wave for $z < z_r$. Similarly, for the case when the source is located below the critical level $z_0 < z_r$, the uniform asymptotic expansion for $H_\nu^{(2)}(v)$ in the vicinity of the critical level $z = z_r$ is

$$H_\nu^{(2)}(v) = (2/\nu)^{\frac{1}{3}}\,\mathrm{Ai}\,[-\lambda^{\frac{2}{3}}\beta^{\frac{1}{3}}(z - z_r)e^{-\frac{2}{3}\pi i}]. \quad (3.8)$$

It can be shown that[3,14]

$$\mathrm{Ai}\,[\zeta e^{-\frac{2}{3}\pi i}] = -\tfrac{1}{2}e^{\frac{2}{3}\pi i}[\mathrm{Ai}\,(\zeta) + i\,\mathrm{Bi}\,(\zeta)]. \quad (3.9)$$

Now, comparing (3.9) with (3.8), $\zeta = -\lambda^{\frac{2}{3}}\beta^{\frac{1}{3}}(z - z_r)$. Thus, the uniform asymptotic expansion for $H_\nu^{(2)}(v)$ demonstrates the transition from upward and downward evanescent waves for $z < z_r$ to an upward traveling wave for $z > z_r$. Comprehensive presentations of the properties of the Airy integral functions are given by Budden[3] and Wait.[2]

Since the parameter λ determines the critical level $z = z_r$ and λ is the variable of integration in the expression (2.6) for H_ϕ, we see how the field of the dipole, imbedded in a homogeneous or inhomogeneous medium (2.11a), (3.2a), may be regarded as a superposition of vertically polarized waves that are either propagating or evanescent. Indeed, λ/kn (2.8b) or ν/v (3.3b) may be regarded as the local sine of the angle of incidence of these elementary waves at any level z in the horizontally stratified medium.[2] However, for the case of the exponential profile, it can also be seen how the classification of the elementary waves as propagating or evanescent is dependent on the variable z and on z_0. With the aid of the uniform asymptotic expansions, we are able to follow the transition of a particular elementary wavelet ($\lambda =$ const) from a traveling to an evanescent wave as a function of z.

The discussion in this section is essential to the analysis presented in the following section, in which the problem of radiation from a dipole imbedded in a medium with an arbitrarily varying refractive-index profile is studied by regarding the medium of propagation as consisting of infinitesimally thin layers with exponentially varying profiles such as (3.1). Of course, in the general case the parameter β is also a function of z.

4. TRANSFORMATION OF THE EQUATIONS FOR $e(z, \lambda)$ AND $h(z, \lambda)$ INTO LOOSELY COUPLED FIRST-ORDER DIFFERENTIAL EQUATIONS

The refractive-index profile, which is an arbitrary function of z, may be expressed as

$$n^2(z) = n_0^2 \exp\int_0^z \beta\, dz, \quad (4.1a)$$

in which β and n_0 are defined as

$$\beta = \frac{1}{n^2}\frac{d}{dz}(n^2) = \frac{d\ln(n^2)}{dz} = \frac{2}{n}\frac{dn}{dz} \quad (4.1b)$$

and

$$n_0^2 = n^2(0). \quad (4.1c)$$

Thus, β is proportional to the relative slope of the refractive-index profile. On substituting (4.1) in (2.10b), we cannot obtain a general solution for $h(z, \lambda)$ in terms of standard functions unless, for instance, $\beta =$ const (Sec. 3). Therefore, the primary first-order differential equations (2.12) for $h(z, \lambda)$ and $e(z, \lambda)$ are solved instead of (2.10b). Noting that both h and e may be expressed in terms of two linearly independent functions, we define h and e as

$$h = \phi_1 + \phi_2 \quad (4.2a)$$

and

$$-i\omega\epsilon e = G_1\phi_1 + G_2\phi_2, \quad (4.2b)$$

in which ϕ_1 and ϕ_2 are independent functions corresponding, for example, to upward and downward propagating waves in an isotropic homogeneous medium.

The functions G_1 and G_2 may be chosen arbitrarily, with the restriction $G_1 \neq G_2$. Substituting (4.2) into (2.12), we get

$$\frac{d}{dz}(\phi_1 + \phi_2) = G_1\phi_1 + G_2\phi_2 \tag{4.3a}$$

and

$$n^2 \frac{d}{dz}\left(\frac{G_1\phi_1 + G_2\phi_2}{n^2}\right) = -k^2q^2(\phi_1 + \phi_2) - \frac{I\Delta s\lambda^2}{2\pi}\delta(z - z_0). \tag{4.3b}$$

Multiplying (4.3a) by G_2 and subtracting it from (4.3b) yields

$$\begin{aligned}\phi_1'(G_1 - G_2) + \phi_1[G_1G_2 + k^2q^2 + n^2(G_1/n^2)'] \\ + \phi_2[G_2^2 + k^2q^2 + n^2(G_2/n^2)'] \\ = -(I\Delta s\lambda^2/2\pi)\delta(z - z_0).\end{aligned} \tag{4.4}$$

A similar expression for ϕ_2' can be written directly by inspection. It is obvious that the set of equations for ϕ_1 and ϕ_2 would be greatly simplified if the coefficient of ϕ_2 in (4.4) were made to vanish by an appropriate choice of the functions G_2. However, on setting this coupling coefficient equal to zero, we obtain a nonlinear differential equation for G_2. In order to transform it into a linear differential equation we make the following substitutions[10]:

$$G_1 = \frac{d \ln g_1}{dz} = \frac{g_1'}{g_1} \quad \text{and} \quad G_2 = \frac{d \ln g_2}{dz} = \frac{g_2'}{g_2}. \tag{4.5}$$

Substituting (4.5) into (4.4), we get

$$\begin{aligned}\phi_1' - \frac{g_1'}{g_1}\phi_1 + \frac{g_2}{W(g_2, g_1)}L_z(g_1)\phi_1 \\ = \frac{g_1}{W(g_1, g_2)}L_z(g_2)\phi_2 + \frac{g_1g_2}{W(g_1, g_2)}\frac{I\Delta s\lambda^2}{2\pi}\delta(z - z_0),\end{aligned} \tag{4.6a}$$

in which the operator L_z is defined in (2.10a) and the Wronskian W is defined as

$$W(g_1, g_2) = -W(g_2, g_1) = g_1g_2' - g_2g_1'. \tag{4.6b}$$

A similar expression involving ϕ_2' is obtained by interchanging subscripts 1 and 2 in (4.6a). From the expression (4.6a), it is obvious that the differential equations for $\phi_{1,2}$ decouple only if we can choose $g_{1,2}$ such that $L_z(g_{1,2}) = 0$.

Thus, for the case of the homogeneous dielectric, we choose $g_{1,2} = e^{\mp ikq}$ and, for the case of the exponential refractive index profile, we may choose

$$g_{1,2} = vH_\nu^{(2)}(v)$$

as defined in (3.2). In these cases, of course, we would obtain solutions for $h(z, \lambda)$ and $e(z, \lambda)$ no different from those derived in the earlier sections. While in Secs. 2 and 3 we determine the Green's function solution for second-order differential operators, here we seek the Green's function for a first-order differential operator. In this paper, we are considering radiation in a medium with an arbitrarily varying refractive index for which no standard solution to $L(g_{1,2}) = 0$ is known. The special solutions considered earlier will only assist us in making a judicious choice for the auxiliary functions g_1 and g_2. This aspect of the problem is now considered in some detail since it is of primary importance.

Equation (4.6a) and a similar one involving ϕ_2' can be written in matrix form as follows:

$$\phi' - G\phi = S\phi + D\delta(z - z_0), \tag{4.7a}$$

where

$$\phi = \begin{bmatrix}\phi_1 \\ \phi_2\end{bmatrix}, \quad G = \begin{bmatrix}G_1 & 0 \\ 0 & G_2\end{bmatrix}, \quad S = \begin{bmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}, \quad D = \begin{bmatrix}D_1 \\ D_2\end{bmatrix} \tag{4.7b}$$

and the elements of the matrix S are

$$S_{12} = -S_{22} = \frac{g_1L(g_2)}{W(g_1, g_2)}, \quad S_{21} = -S_{11} = \frac{g_1L(g_1)}{W(g_2, g_1)}, \tag{4.7c}$$

and

$$D_1 = -D_2 = \frac{g_1g_2I\Delta s\lambda^2}{2\pi W(g_1, g_2)}. \tag{4.7d}$$

Thus, in order to make the equations for $\phi_{1,2}$ loosely coupled, we must generate $|S_{12}| \ll |G_1|$ and $|S_{21}| \ll |G_2|$. If this is achieved, we note that the functions $G_{1,2}$ (the logarithmic derivatives of $g_{1,2}$) can be interpreted as local propagation coefficients. Thus, for slowly varying media, we refer to solution (2.11a) and set $G_{1,2} = \mp ikq$; thus, (4.5) yields

$$g_{1,2} = \exp\int_{z_0}^{z} \mp ikq\, dz. \tag{4.8}$$

The above solutions constitute the well-known phase memory concept. In this case, it can be shown that

$$S_{11} = S_{22} = -S_{12} = -S_{21} = \frac{1}{2}\frac{d}{dz}\ln\left(\frac{n^2}{q}\right). \tag{4.9}$$

Thus, if we neglect the cross coupling terms (S_{12}, S_{21}) only and assume outward going waves for $|z| \to \infty$, we obtain for $\phi_{1,2}$ expressions in terms of the familiar WKB solutions

$$\phi_1 = \frac{An}{q^{\frac{1}{2}}} \exp\left(-ik\int_{z_0}^{z} q\, dz\right), \quad z > z_0,$$
$$= 0, \quad z < z_0,$$

and

$$\phi_2 = 0, \quad z > z_0,$$
$$= \frac{An}{q^{\frac{1}{2}}} \exp\left(+ik\int_{z_0}^{z} q\, dz\right), \quad z < z_0. \quad (4.10a)$$

The constant A is determined by the discontinuity in the value of $\phi_{1,2}$ at $z = z_0$ which is equal to $D_{1,2}$, the coefficient of $\delta(z - z_0)$ [(4.7d)]. Thus, we obtain

$$A = \frac{I\Delta s\lambda^2}{4\pi ik}\left(\frac{1}{q^{\frac{1}{2}}n}\right)_{z=z_0}. \quad (4.10b)$$

Obviously, these WKB solutions are not appropriate if the refractive-index profile is rapidly varying or if it consists of regions of critical coupling ($q \to 0$). To overcome this problem, we shall seek expressions for $g_{1,2}$ from the solutions in Sec. 3. Thus, noting that the functions $G_{1,2}$ are equal to the logarithmic derivatives of the "local" solutions, we assume that

$$g_1 = \exp\int_{z_0}^{z} \frac{\partial}{\partial z} \ln\,[vH_\nu^{(2)}(v)]\, dz$$
$$= \exp\int_{z_0}^{z} \frac{d}{dv}[\ln vH^{(2)}(v)]kn\, dz$$

and

$$g_2 = \exp\int_{z_0}^{z} \frac{\partial}{\partial z} \ln\,[vH_\nu^{(1)}(v)]\, dz$$
$$= \exp\int_{z_0}^{z} \frac{d}{dv}[\ln vH_\nu^{(1)}(v)]kn\, dz, \quad (4.11a)$$

in which the argument v and the order ν of the Bessel's functions are given by (3.2b) and (3.2c), respectively. Note, however, that in this case β is a function of z [(4.1)]. The symbol for the partial derivative $\partial/\partial z$ is used whenever we wish to imply that β is to be regarded as an independent variable. Thus, partial and total differentiation with respect to z may be expressed in terms of d/dv as follows:

$$\frac{\partial}{\partial z} = \frac{\partial v}{\partial z}\frac{d}{dv} = kn\frac{d}{dv}, \quad \frac{d}{dz} = \frac{dv}{dz}\frac{d}{dv} = \left(1 - \frac{2\beta'}{\beta^2}\right)\frac{d}{dv}. \quad (4.11b)$$

In regions for which $\beta = \text{const}$, $g_{1,2}$ reduce to $vH_\nu^{(2)}(v)$ and $vH_\nu^{(1)}(v)$, respectively, and all the coupling coefficients vanish. In general, however, the above choice [(4.11)] for $g_{1,2}$ results in the following expressions for S_{11} and $W(g_1, g_2)$:

$$S_{11} = \frac{\pi vH_\nu^{(2)}(v)H_\nu^{(1)}(v)}{2ikn}\left(\left\{\frac{\partial}{\partial z}\ln\,[vH_\nu^{(2)}(v)]\right\}^2 - \beta\frac{\partial}{\partial z}\ln\,[vH_\nu^{(2)}(v)] + k^2q^2\right)\frac{1}{\beta^2}\frac{d\beta}{dz},$$

$$\frac{g_1g_2}{W(g_1, g_2)} = \frac{\pi vH_\nu^{(2)}(v)H_\nu^{(1)}(v)}{4ikn}. \quad (4.12)$$

To obtain the expression for S_{22}, we simply interchange the orders of the Hankel functions. These expressions for the coupling coefficients (4.12) are not singular for regions of critical coupling ($q \to 0$). Also, for regions in which n^2 is rapidly varying, β is large and the coupling is small. Furthermore, for slowly varying media ($v \gg 1$), the Debye–Watson expansion for the Hankel functions (4.11) indicate that the expressions for $g_{1,2}$ reduce to the familiar WKB solutions

$$g_{1,2} \approx \frac{n}{q^{\frac{1}{2}}} \exp\left(\mp ik\int_{z_0}^{z} q\, dz\right). \quad (4.13)$$

It is interesting to note that (4.13) is an appropriate approximation for (4.11), for both positive and negative values of β, provided that $|\beta/kn| \ll 1$.

We shall now derive generalized WKB solutions for the fields of the dipole immersed in a medium with a single critical coupling layer (otherwise, n^2 is an arbitrary function; see Fig. 2). Thus, we require that, for large positive z, the solution behaves like an upward going wave and, for large negative z, the solution behaves as an evanescent wave (the net power flow is in the positive z direction). We first solve (4.7), assuming (4.11) for $g_{1,2}$ and neglecting all the coupling terms:

$$\phi_1 = g_1A_1, \qquad \phi_2 = g_2A_2, \qquad z > z_0,$$
$$= g_1(A_1 - D_1), \qquad = g_2(A_2 - D_2), \qquad z < z_0. \quad (4.14a)$$

The discontinuity in $\phi_{1,2}$ at $z = z_0$ is

$$D_{1,2} = \frac{\pm I\Delta s\lambda^2 v_0 H_\nu^{(1)}(v_0)H_\nu^{(2)}(v_0)}{8ikn}. \quad (4.14b)$$

Now, for $z \to \infty$, $h(z, \lambda)$ must represent an upward traveling wave; thus, $A_2 = 0$. We have seen that, for the solution to represent an evanescent wave, $h(z, \lambda)$ must be proportional to $J_\nu(v)$; thus, for $z \to \infty$,

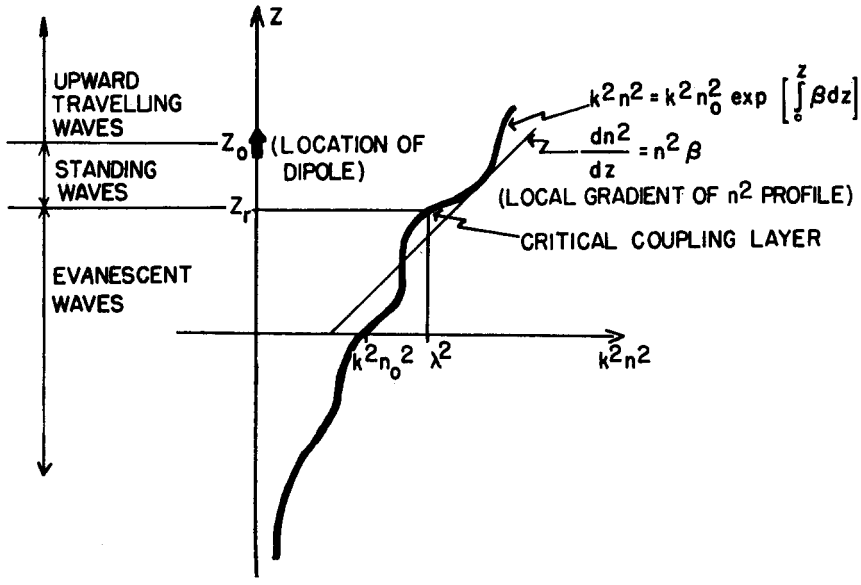


FIG. 2. Inhomogeneous refractive-index profile, regarded locally as an exponentially varying profile.

$\phi_1 + \phi_2 \sim J_\nu(v)$. Summing up, we have

$$h(z,\lambda) \approx A_1 \exp\int_{z_0}^{z} \frac{\partial}{\partial z}[\ln vH_\nu^{(2)}(v)]\,dz, \quad z > z_0,$$

$$h(z,\lambda) \approx (A_1 - D_1)\exp\int_{z_0}^{z} \frac{\partial}{\partial z}[\ln vH_\nu^{(2)}(v)]\,dz + D_1 \exp\int_{z_0}^{z} \frac{\partial}{\partial z}[\ln vH_\nu^{(1)}(v)], \quad z < z_0. \tag{4.15a}$$

For the case of the exponential profile, we obtain

$$A_1 = \frac{2D_1 J_\nu(v_0)}{H_\nu^{(1)}(v_0)}. \tag{4.15b}$$

Thus, the solution (4.15) reduces to (3.2) for $\beta =$ const. In a similar way, it can be shown that, if the boundary conditions are such that the solutions behave like outward going waves for large $|z|$, the generalized WKB solution is

$$h(z,\lambda) = D_1 \exp\int_{z_0}^{z} \frac{\partial}{\partial z}[\ln vH_\nu^{(2)}(v)]\,dz, \tag{4.16}$$

where the orders of the Hankel functions are 2 and 1 for $z > z_0$ and $z < z_0$, respectively.

It is obvious that (4.17) reduces to (4.10a) for slowly varying media with no regions of critical coupling and to (2.11) for homogeneous media ($\beta =$ 0). We may now proceed to derive higher-order solutions for $h(z, \lambda)$ in which the coupling coefficients are not neglected. Two distinct iterative methods to derive higher-order approximations to the coupled differential equations have been discussed earlier in some detail.[10]

5. CONCLUDING REMARKS

It is shown that Maxwell's equations for inhomogeneous dielectric media may be solved by transforming them into a set of loosely coupled first-order differential equations, if one chooses a set of auxiliary functions $g_{1,2}$ that take into account the local features of the refractive index profile. These auxiliary functions $g_{1,2}$ are chosen such that at every level they are linearly independent wave solutions for a medium with the same relative gradient of the refractive-index profile β as that of the medium under consideration. The logarithmic derivatives of these local solutions $g_{1,2}$ are identified as the local propagation coefficients $G_{1,2}$.

It is interesting to point out that the coupling between the new dependent variables $\phi_{1,2}$ can be derived directly by considering the reflection and transmission coefficients at the interface between two media with exponentially varying refractive-index profiles, characterized by constant parameters β and $\beta + d\beta$, respectively. The self- and cross-coupling terms S_{11} and S_{12}, respectively, have therefore also been identified as differential reflection and transmission coefficients.[10]

The generalized WKB solutions are valid for regions of critical coupling as well as for slowly varying regions in which they have been shown to merge with the familiar WKB solutions. The Green's function solutions for the first-order differential equations are readily solved by imposing a discontinuity in the value of the function at the source point z_0 [equal to D (4.7d)] and the appropriate boundary conditions.

In this paper it has been assumed for simplicity of presentation that n_0^2 is real. However, these solutions

may be generalized in a straightforward manner for the case in which n_0^2 is complex. Thus, for example, if the conductivity σ of the medium is large ($n_0^2 \approx i\sigma_0/\omega\epsilon_0$), it is convenient to express the local solutions $g_{1,2}$ in terms of the modified Bessel's functions.[1,2] The generalization of the above solutions for anisotropic, inhomogeneous media can be approached by transforming Maxwell's equations into four coupled first-order differential equations for the upward and downward ordinary and extraordinary waves.[15]

Recently, Westcott[16] derived solutions for the propagation of vertically polarized waves for several new refractive-index profiles in terms of standard transcendental functions. Among these profiles is

$$n = n_0 \tanh \alpha(z - z_0), \tag{5.1}$$

in which $n_0, \alpha,$ and z_0 are constants. Following the technique developed in the previous section, we may readily generalize these solutions for refractive-index profiles with two layers of critical coupling. In this case, we express $n(z)$ as follows:

$$n = n_0 \tanh \int_{z_0}^{z} \alpha(z)\, dz, \tag{5.2a}$$

in which α can be shown to be defined as

$$\alpha = \frac{dm}{dz}\frac{1}{1 - m^2}, \quad m = \frac{n}{n_0}. \tag{5.2b}$$

In this case, coupling between the two local solutions is proportional to $d\alpha/dz$ rather than $d\beta/dz$ [(4.12)]. The local solutions for this representation of the refractive-index profile (5.2) are the hypergeometric functions rather than the Bessel functions (3.2). In order to study the behavior of these local wave solutions as a function of the coordinate z, it is necessary to develop the expressions for the circuit relations for these hypergeometric functions.[17]

ACKNOWLEDGMENTS

We would like to thank S. W. Maley, F. G. Ullman, and J. R. Wait for their encouragement and comments. The manuscript was typed by Mrs. M. Alles. This work has been partially sponsored by a grant from the University of Nebraska Engineering Research Center.

[1] J. Galejs, IRE Trans. Antennas and Propagation **9**, 554 (1961)
[2] J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon Press, New York, 1962).
[3] K. G. Budden, *Radio Waves in the Ionosphere* (Cambridge University Press, London, 1961).
[4] J. Galejs, *Antennas in Inhomogeneous Media* (Pergamon Press, New York, 1968).
[5] H. Bremmer, Physica **15**, 593 (1949).
[6] D. S. Jones, *The Theory of Electromagnetism* (Pergamon Press, New York, 1964).
[7] A. Erdélyi, *et al.*, *Higher Transcendental Functions* (McGraw-Hill Co., New York, 1953), Vol. 2.
[8] L. B. Felsen and N. Marcuvitz, Brooklyn Polytechnic Institute, Brooklyn, N.Y., Research Report No. PIBMPI-1225-64, 1964.
[9] Y. Inoue and S. Horowitz, Radio Sci. **1**, 957 (1966).
[10] E. Bahar, J. Math. Phys. **8**, 1735 (1967).
[11] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Co., New York, 1953).
[12] B. Friedman, *Principles and Techniques of Applied Mathematics* (John Wiley and Sons, New York, 1964).
[13] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Appl. Math. Series No. 55 (Dept. of Commerce, National Bureau of Standards, Washington, D.C., 1964).
[14] J. C. P. Miller, *The Air-Integral*, British Assoc. Math. Tables, Vol. B (Cambridge University Press, Cambridge, 1946).
[15] O. E. H. Rydbeck and A. Hjalmarson, Research Laboratory of Electronics, Chalmers University of Technology, gebraltargatan 5G, Gothenburg S, Sweden, Research Report No. 61, 1966.
[16] B. S. Westcott, Proc. Cambridge Phil. Soc. **66**, 675 (1970).
[17] J. Heading, Radio Sci. **4**, 441 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

$O(5)$ Bases for Nuclear Seniority Model*

K. AHMED AND R. T. SHARP
Physics Department, McGill University, Montreal, Canada

(Received 25 April 1969)

Two complete sets of bases are given for the general representation of $O(5)$, utilizing the $SU(2)$ subgroup which classifies nuclear seniority model states. The first set, which are analogous to the $SU(3) \supset O(3)$ states of Elliott and Harvey, are formed by projecting good $SU(2)$ states out of certain "intrinsic" states. The second set are analogous to the $SU(3) \supset O(3)$ states defined by Moshinsky and Bargmann; each state is characterized by a particular simple term which appears in the expansion of that state and no other. Transformation matrices between the two sets of states are given. The utility of the states for the calculation of generator matrix elements and Clebsch–Gordan coefficients is discussed. The seniority model states of Hecht and Parikh are discussed briefly.

I. INTRODUCTION

The group $O(5)$ finds application in the treatment (seniority model) of the pairing force between particles in the same nuclear shell.[1–3] As often happens when group theory is used in nuclear physics, the physically significant states are not mathematically canonical ones, i.e., are not bases of subgroups which provide the maximum number of state labels (other examples are the Wigner supermultiplet scheme[4] and Elliott's $SU(3)$ classification of levels[5]). There are then not enough naturally occurring quantum numbers to identify the states completely, a circumstance referred to as the internal labeling problem.

In this paper we present two complete sets of $O(5)$ basis states for the noncanonical but physically significant chain $O(5) \supset SU(2)$. The first set are analogous to the $SU(3) \supset O(3)$ states of Elliott and Harvey[5] and are formed by projecting good $SU(2)$ states from certain "intrinsic" states. These "Elliott" states are then utilized to derive a second set of "symmetric" states, which are analogous to the $SU(3) \supset O(3)$ states defined by Moshinsky and Bargmann.[6] A feature of this second set is that, although each state may be a sum of many terms in a certain expansion, it is characterized uniquely by a single one of those terms.

The basis states are derived in the next section. The simple bases $(p, 0)$ and $(0, q)$ are considered first and the general basis (p, q) is then constructed by combining $(p, 0)$ and $(0, q)$. An appendix contains a brief interpretation of $O(5)$ in terms of the seniority model and discusses the Hecht–Parikh states.[2,3]

II. POLYNOMIAL BASES

Mathematically canonical bases, belonging to the chain $O(5) \supset SU(2) \times SU(2)$ have been discussed by, among others, Hecht[2] and Kemmer, Pursey, and Williams[7]; the notation of Sharp and Pieper[8] is followed here. Two sets of three generators, $S_\pm$, S_3 and $T_\pm$, T_3, generate two independent $SU(2)$ subgroups, while the other four, $U_\pm$, $V_\pm$, transform by the $(\frac{1}{2}, \frac{1}{2})$ representation of $SU(2) \times SU(2)$. The four quantum numbers S, S_3, T, and T_3 label the basis states completely.

The $O(5) \supset SU(2)$ scheme, which concerns us, utilizes basis states of the $SU(2)$ subgroup generated by $U_\pm$, U_3 where $U_3 = S_3 + T_3$ (the subgroup generated by $V_\pm$, V_3, where $V_3 = S_3 - T_3$, is not independent since the $U_\pm$ do not commute with the $V_\pm$). V_3 is a U-scalar and its eigenvalue serves as a state label, along with U and U_3. The other generators form two U-vectors with respective components $-2S_+$, V_+, $2T_-$ and $-2T_+$, V_-, $2S_-$. A different $SU(2)$ subgroup of $O(5)$, of interest in connection with nuclear surface quadrupole vibrations,[2,9] is not considered here. We denote the general basis state by

$$\left| \begin{matrix} p & q \\ \alpha\ V_3\ U\ U_3 \end{matrix} \right\rangle.$$

The additional label α will be discussed later; for the simple representations $(p, 0)$ and $(0, q)$, it is not required and may be suppressed.

We use nine variables,

$$\begin{aligned}
\alpha &= \left| \begin{matrix} 1\ 0 \\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2} \end{matrix} \right\rangle, & \delta &= \left| \begin{matrix} 1\ 0 \\ \frac{1}{2}\ \frac{1}{2}\ -\frac{1}{2} \end{matrix} \right\rangle, \\
\gamma &= \left| \begin{matrix} 1\ 0 \\ -\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2} \end{matrix} \right\rangle, & \beta &= \left| \begin{matrix} 1\ 0 \\ -\frac{1}{2}\ \frac{1}{2}\ -\frac{1}{2} \end{matrix} \right\rangle, \\
\theta &= \left| \begin{matrix} 0\ 1 \\ 1\ 0\ 0 \end{matrix} \right\rangle, & \xi &= \left| \begin{matrix} 0\ 1 \\ -1\ 0\ 0 \end{matrix} \right\rangle, \\
\eta &= \left| \begin{matrix} 0\ 1 \\ 0\ 1\ 1 \end{matrix} \right\rangle, & \lambda &= \left| \begin{matrix} 0\ 1 \\ 0\ 1\ 0 \end{matrix} \right\rangle, \\
\zeta &= \left| \begin{matrix} 0\ 1 \\ 0\ 1\ -1 \end{matrix} \right\rangle,
\end{aligned} \tag{1}$$

to represent the basis states of the two fundamental

representations (1, 0) and (0, 1). The variables may be identified with the same ones in Ref. 8, except η whose sign is reversed.

$(p, 0)$ States

The general $(p, 0)$ state, a polynomial of degree p in the $(1, 0)$ variables, may be written down immediately by coupling the states of U-spin $\frac{1}{2}p + \frac{1}{2}V_3$ in the variables α, δ to the states of U-spin $\frac{1}{2}p - \frac{1}{2}V_3$ in the variables γ, β:

$$\left|\begin{matrix} p & 0 \\ V_3 & U \quad U_3 \end{matrix}\right\rangle = \sum_u \alpha^{\frac{1}{2}p+\frac{1}{2}V_3+u}\delta^{\frac{1}{2}p+\frac{1}{2}V_3-u}\gamma^{\frac{1}{2}p-\frac{1}{2}V_3+U_3-u}\beta^{\frac{1}{2}p-\frac{1}{2}V_3-U_3+u}\,[(\tfrac{1}{2}p + \tfrac{1}{2}V_3 + u)!\,(\tfrac{1}{2}p + \tfrac{1}{2}V_3 - u)!$$
$$\times\,(\tfrac{1}{2}p - \tfrac{1}{2}V_3 + U_3 - u)!\,(\tfrac{1}{2}p - \tfrac{1}{2}V_3 - U_3 + u)!]^{-\frac{1}{2}}\left\langle \begin{matrix} \frac{1}{2}p + \frac{1}{2}V_3 & \frac{1}{2}p - \frac{1}{2}V_3 \\ u & U_3 - u \end{matrix}\middle|\, \begin{matrix} U \\ U_3 \end{matrix}\right\rangle. \tag{2}$$

V_3 goes by integer steps from $-\frac{1}{2}p$ to $\frac{1}{2}p$ and U by integer steps from $|V_3|$ to $\frac{1}{2}p$. The last factor in Eq. (2) is an ordinary $SU(2)$ Clebsch–Gordan coefficient; we use this notation throughout.

$(0, q)$ States

The states of $(0, q)$ are polynomials of degree q in the (0, 1) variables. We find

$$\left|\begin{matrix} 0 & q \\ V_3 & U \quad U \end{matrix}\right\rangle = \left(\frac{(q + U + V_3 + 1)!!\,(q + U - V_3 + 1)!!\,[\frac{1}{2}(q - U + V_3)]!\,[\frac{1}{2}(q - U - V_3)]!\,(2U + 1)!}{(2q + 1)!!\,2^U}\right)$$
$$\times \frac{\eta^U}{U!}\sum_x \frac{\xi^{\frac{1}{2}(q-V_3-U)-x}\theta^{\frac{1}{2}(q+V_3-U)-x}(\lambda^2 - 2\eta\zeta)^x}{x!\,[\frac{1}{2}(q - U - V_3) - x]!\,[\frac{1}{2}(q - U + V_3) - x]!\,(2U + 2x + 1)!!}. \tag{3}$$

The state (3) is derived by means of the condition

$$(\partial_\eta\partial_\zeta + \partial_\xi\partial_\theta - \tfrac{1}{2}\partial_\lambda^2)\left|\begin{matrix} 0 & q \\ V_3 & U \quad U \end{matrix}\right\rangle = 0$$

which implies orthogonality to "unwanted" states containing powers of the $O(5)$ scalar $\eta\zeta + \xi\theta - \frac{1}{2}\lambda^2$ [they belong to representations lower than (p, q)]. Normalization is effected with the help of the equality

$$\left\langle \begin{matrix} 0 & q \\ V_3 + 1 & U + 1 \quad U + 1 \end{matrix}\middle|\, S_+ \,\middle|\, \begin{matrix} 0 & q \\ V_3 & U \quad U \end{matrix}\right\rangle = \left\langle \begin{matrix} 0 & q \\ V_3 & U \quad U \end{matrix}\middle|\, S_- \,\middle|\, \begin{matrix} 0 & q \\ V_3 + 1 & U + 1 \quad U + 1 \end{matrix}\right\rangle.$$

A state with $U_3 < U$ is found by cranking with U_- and may be obtained from (3) by the replacement

$$\eta^U \to \left(\frac{2^{U+U_3}(U + U_3)!\,(U - U_3)!}{(2U)!}\right)^{\frac{1}{2}} U!$$
$$\times \sum_y \frac{\eta^y\lambda^{U+U_3-2y}\zeta^{\,-U_3}}{2^y y!\,(y - U_3)!\,(U + U_3 - 2y)!}. \tag{4}$$

V_3 goes by integer steps from $-q$ to q and U by steps of 2 from $q - |V_3|$ down to 1 or 0.

(p, q) States

Before deriving general (p, q) states, we outline the procedure to be followed and describe the different types of state which arise. The method follows closely one used recently for $SU(3) \supset O(3)$ states.[10]

It is basic to our approach that states of (p, q) can be constructed out of products of the states of the simple representations $(p, 0)$ and $(0, q)$,[11] i.e., that they lie in the space spanned by the simple product states

$$|(pV_{3p}U_p,\, q\,V_3 - V_{3p}\,U_q)V_3UU_3\rangle$$
$$= \sum_{U_{3p}} \left|\begin{matrix} p & 0 \\ V_{3p} & U_p \quad U_{3p} \end{matrix}\right\rangle \left|\begin{matrix} 0 & q \\ V_3 - V_{3p} & U_q \quad U_3 - U_{3p} \end{matrix}\right\rangle$$
$$\times \left\langle \begin{matrix} U_p & U_q \\ U_{3p} & U_3 - U_{3p} \end{matrix}\middle|\, \begin{matrix} U \\ U_3 \end{matrix}\right\rangle. \tag{5}$$

These simple product states are orthonormal and have definite values of V_3, U, U_3. The space which they span includes all representations in the Clebsch–Gordan series for $(p, 0) \otimes (0, q)$.

Next we introduce certain states $|pqKV_3\rangle$ which belong to the representation (p, q) and which are eigenstates of V, V_3, U_3 with $V = \frac{1}{2}p + q$, $U_3 = K$; we call them intrinsic states since they play a role similar to the intrinsic states of Elliott[5] in $SU(3)$ theory [however, they do not appear to have a direct physical interpretation as do their $SU(3)$ analogs]. They may be obtained by cranking with V_- from states on the boundary of the weight diagram and present no multiplicity problem.

To obtain states of good U, we expand the intrinsic states in simple product states and retain only those terms which belong to the desired value of U; in this way, the projected states

$$\left|\begin{matrix} p & q \\ K & V_3 \quad U \quad K \end{matrix}\right\rangle$$

are obtained. The second K in the state symbol is actually the U_3 value and may be changed by cranking with $U_\pm$ to obtain

$$\left| \begin{matrix} p & & q & \\ K & V_3 & U & U_3 \end{matrix} \right\rangle.$$

The projected states are not linearly independent in general, but by suitably restricting the range of K a complete linearly independent set of bases is obtained. These nonredundant projected states we call Elliott states by analogy with those defined by Elliott[5] for $SU(3)$.

Although the Elliott states solve the degeneracy problem, they are not the most convenient for applications. Hence, we introduce the symmetric basis which treats $(p, 0)$ and $(0, q)$ on much the same footing. Each symmetric state is characterized by the presence of a particular term in its simple-product-state expansion. The characteristic term is stretched, i.e., has $U_p = \underline{U}_p$, $U_q = \underline{U}_q$ with $\underline{U}_p + \underline{U}_q = U$ and has the largest consistent value of V_{3p}. The term which characterizes one symmetric state is absent from the expansion of all others. Hence the symmetric state may be labeled by $\underline{U}_p$ and denoted by

$$\left| \begin{matrix} p & & q & \\ \underline{U}_p & V_3 & U & U_3 \end{matrix} \right\rangle.$$

It is easy to express Elliott states in terms of symmetric states and, by inverting the equations, to express symmetric states in terms of Elliott states, thereby obtaining their expansion in simple product states.

The intrinsic states are

$$\begin{aligned} &|pqKV_3\rangle \\ &\quad = [(\tfrac{1}{2}p + q - V_3)!\,(\tfrac{1}{2}p + K)!\,(\tfrac{1}{2}p - K)!\,q!]^{-1} \\ &\qquad \times [2^{\frac{1}{2}}(\lambda\partial_\theta + \xi\partial_\lambda) + \gamma\partial_\alpha - \beta\partial_\delta]^{\frac{1}{2}p+q-V_3} \\ &\qquad \times \alpha^{\frac{1}{2}p+K}\delta^{\frac{1}{2}p-K}\theta^q; \end{aligned} \tag{6}$$

to avoid unimportant factors later, they are normalized according to

$$\begin{aligned} &\langle pqKV_3 \,|\, pqKV_3\rangle \\ &\quad = (p + 2q)!/(\tfrac{1}{2}p + q + V_3)!\,(\tfrac{1}{2}p + q - V_3)! \\ &\qquad \times (\tfrac{1}{2}p + K)!\,(\tfrac{1}{2}p - K)!\,q!\,. \end{aligned}$$

Now expand $|pqKV_3\rangle$ in simple product states:

$$\begin{aligned} |pqKV_3\rangle = \sum_{U, V_{3p}, U_p, U_q} &|(pV_{3p}U_p, q\,V_3 - V_{3p}\,U_q)V_3UK\rangle \\ &\times B^{pq}_{KV_3U}(V_{3p}U_pU_q). \end{aligned} \tag{7}$$

The coefficient $B^{pq}_{KV_3U}(V_{3p}U_pU_q)$ is just the scalar product $\langle (pV_{3p}U_p, q\,V_3 - V_{3p}\,U_q)V_3UK \,|\, pqKV_3\rangle$. To evaluate it, one substitutes the expanded form of Eq. (6) in the right side. In the left, the sum (5) is inserted, but only one term contributes, that with $U_{3p} = K$; the expressions (2) and (3) with (4) are then substituted. The x-sum is evaluated with the help of the formula

$$\begin{aligned} &\sum_x \frac{2^x}{x!\,[\frac{1}{2}(q - U - V) - x]!\,[\frac{1}{2}(q - U + V) - x]!\,(2U + 2x + 1)!!} \\ &\qquad = \frac{(2q + 1)!!}{(q + U + V + 1)!!\,(q + U - V + 1)!!\,[\frac{1}{2}(q - U + V)]!\,[\frac{1}{2}(q - U - V)]!}, \end{aligned} \tag{8}$$

which follows from the normalization of the state (3). The result is

$$\begin{aligned} B^{pq}_{KV_3U}(V_{3p}U_pU_q) = (-1)^{\frac{1}{2}p-V_{3p}} &\left[\frac{(2q + 1)!!\,(2U_q + 1)(U_p - V_{3p})!\,(U_p + V_{3p})!}{(q + U_q + V_3 - V_{3p} + 1)!!\,(q + U_q - V_3 + V_{3p} + 1)!!}\right]^{\frac{1}{2}} \\ &\times \left[\frac{(U_p + K)!\,(U_p - K)!\,2^{U_q - U_p + \frac{1}{2}p}}{[\frac{1}{2}(q - U_q + V_3 - V_{3p})]!\,[\frac{1}{2}(q - U_q - V_3 + V_{3p})]!\,(\frac{1}{2}p - U_p)!\,(\frac{1}{2}p + U_p + 1)!}\right]^{\frac{1}{2}} \\ &\times \left\langle \begin{matrix} U_p & U_q & \\ K & 0 & K \end{matrix} \right| \left. \begin{matrix} U \\ K \end{matrix} \right\rangle \sum_x \frac{(-1)^x}{(U_p - V_{3p} - x)!\,(U_p + K - x)!\,x!\,(V_{3p} - K + x)!}. \end{aligned} \tag{9}$$

The projected states are obtained by dropping the U-sum in Eq. (7):

$$\begin{aligned} &\left| \begin{matrix} p & & q & \\ K & V_3 & U & U_3 \end{matrix} \right\rangle \\ &\quad = \sum_{V_{3p}, U_p, U_q} |(pV_{3p}U_p, q\,V_3 - V_{3p}\,U_q)V_3UU_3\rangle \\ &\qquad \times B^{pq}_{KV_3U}(V_{3p}U_pU_q); \end{aligned} \tag{10}$$

K, V_3, and U vary by integer steps in the ranges

$$\begin{aligned} &-\tfrac{1}{2}p \le K \le \tfrac{1}{2}p, \\ &-\tfrac{1}{2}p - q \le V_3 \le \tfrac{1}{2}p + q, \\ &\max(\epsilon_p, |V_3| - q) \le U \\ &\qquad \le \min(\tfrac{1}{2}p + q, p + q - |V_3|), \end{aligned} \tag{11}$$

where ϵ_p is 0 or $\frac{1}{2}$, depending on whether p is even or odd; if $V_3 + q$ is odd, $U = 0$ is excluded.

To obtain the nonredundant set of Elliott states, K is further restricted to the range

$$\max(\epsilon_p, U - q, \tfrac{1}{2}(U + |V_3| - q)) \leq K \leq \min(\tfrac{1}{2}p, U), \quad K \neq 0, \quad \text{for} \quad U + V_3 + q \text{ odd}. \tag{12}$$

Examination of the expansion (10) shows that the Elliott states are independent; for if K is in the range (12), the sum (10) contains certain stretched terms, those with $U_p = \underline{U}_p = K$, and $U_q = \underline{U}_q = U - K$ which are not included in states of any higher K (and the same V_3, U). That they are complete follows from counting them, for their total number is $\frac{1}{6}(p+1) \times (q+1)(p+q+2)(p+2q+3)$, the dimensionality of the representation (p, q). The Elliott states therefore form a basis. They are nonorthonormal with metric matrix given, according to Eq. (10), by

$$\left\langle \begin{matrix} p & & q & \\ K' & V_3' & U' & U_3' \end{matrix} \middle| \begin{matrix} p & & q & \\ K & V_3 & U & U_3 \end{matrix} \right\rangle = \delta_{V_3'V_3}\delta_{U'U}\delta_{U_3'U_3} \sum_{V_{3p}, U_p, U_q} B^{pq}_{K'V_3U}(V_{3p}U_pU_q) \times B^{pq}_{KV_3U}(V_{3p}U_pU_q). \tag{13}$$

The symmetric states

$$\left| \begin{matrix} p & & q & \\ \underline{U}_p & V_3 & U & U_3 \end{matrix} \right\rangle = \sum_{V_{3p}, U_p, U_q} |(pV_{3p}U_p, q\,V_3 - V_{3p}\,U_q)V_3UU_3\rangle \times A^{pq}_{\underline{U}_pV_3U}(V_{3p}U_pU_q) \tag{14}$$

are those states of (p, q) with eigenvalues V_3, U, U_3, with the coefficients A chosen so that

$$A^{pq}_{\underline{U}_pV_3U}(\underline{V}_{3p}(\underline{U}_p'), \underline{U}_p', \underline{U}_q(\underline{U}_p')) = \delta_{\underline{U}_p\underline{U}_p'}. \tag{15}$$

$\underline{U}_p$ has the same range as K in Eq. (12), i.e.,

$$\max(\epsilon_p, U - q, \tfrac{1}{2}(U + |V_3| - q)) \leq \underline{U}_p \leq \min(\tfrac{1}{2}p, U), \quad \underline{U}_p \neq 0, \quad \text{for} \quad U + V_3 + q \text{ odd}. \tag{16}$$

$\underline{U}_q(\underline{U}_p)$, $\underline{V}_{3p}(\underline{U}_p)$ have the meanings

$$\underline{U}_q = U - \underline{U}_p, \tag{17}$$

$$\underline{V}_{3p}(\underline{U}_p) = \min(V_3 - \underline{U}_q + q, \underline{U}_p - \epsilon'), \tag{18}$$

where $\epsilon' = 0$ or 1, depending on whether $V_3 - \underline{U}_p + \underline{U}_q + q$ is even or odd. The significance of Eqs. (15), (17), and (18) is that each stretched term with maximum V_{3p} appears (with unit coefficient) in only one symmetric state; the state is labeled by the $\underline{U}_p$ of the term in question. From the discussion in the preceding paragraph of the independence of the Elliott states, it is apparent that such symmetric states exist and are independent; since their number is equal to that of the Elliott states, they are complete and form a basis.

Elliott states can be expanded in terms of symmetric states simply by examining the coefficients of the stretched terms in Eq. (10):

$$\left| \begin{matrix} p & & q & \\ K & V_3 & U & U_3 \end{matrix} \right\rangle = \sum_{\underline{U}_p} \left| \begin{matrix} p & & q & \\ \underline{U}_p & V_3 & U & U_3 \end{matrix} \right\rangle D^{pqV_3U}_{\underline{U}_pK}. \tag{19}$$

Moreover, Eq. (19) can be solved for the symmetric states in terms of the Elliott states:

$$\left| \begin{matrix} p & & q & \\ \underline{U}_p & V_3 & U & U_3 \end{matrix} \right\rangle = \sum_{K} \left| \begin{matrix} p & & q & \\ K & V_3 & U & U_3 \end{matrix} \right\rangle \times (D^{-1})^{pqV_3U}_{K\underline{U}_p}. \tag{20}$$

Explicit formulas for D and D^{-1} are given in Appendix B. Combining Eqs. (10) and (20) gives the coefficients A of Eq. (14) in the form

$$A^{pq}_{\underline{U}_pV_3U}(V_{3p}U_pU_q) = \sum_K B^{pq}_{KV_3U}(V_{3p}U_pU_q)(D^{-1})^{pqV_3U}_{K\underline{U}_p}. \tag{21}$$

The metric matrix for the symmetric states is

$$\left\langle \begin{matrix} p & & q & \\ \underline{U}_p' & V_3' & U' & U_3' \end{matrix} \middle| \begin{matrix} p & & q & \\ \underline{U}_p & V_3 & U & U_3 \end{matrix} \right\rangle = \delta_{V_3'V_3}\delta_{U'U}\delta_{U_3'U_3} \sum_{V_{3p}, U_p, U_q} A^{pq}_{\underline{U}_p'V_3U}(V_{3p}U_pU_q) \times A^{pq}_{\underline{U}_p,V_3U}(V_{3p}U_pU_q). \tag{22}$$

This completes our results. The utility of the symmetric states is that each is characterized by a unique stretched term in its simple-product-state expansion; this facilitates the expansion of a compound state in terms of symmetric states—it is only necessary to examine the coefficients of the characteristic terms. In this way generator matrix elements, transformation matrix elements, Clebsch–Gordan coefficients can all be evaluated directly in the noncanonical scheme with comparative ease; we leave such applications for the future. Sebe's[12] elegant method of obtaining $SU(3) \supset O(3)$ Clebsch–Gordan coefficients in the Elliott basis can be adapted to obtain $O(5) \supset SU(2)$ Clebsch–Gordan coefficients in the Elliott basis of this paper.

ACKNOWLEDGMENTS

We thank Professors H. C. von Baeyer and C. S. Lam for helpful discussions.

APPENDIX A

Details of the application of $O(5)$ to nuclear states are given in the classic papers of Flowers, Edmonds and Flowers, Helmers,[1] Hecht,[2] and Parikh[3]; see also Talmi and de Shalit.[13] A brief interpretation is given here chiefly to relate notations.

The seniority model of pairing forces describes a system of neutrons and protons in the same j shell. A pair means two nucleons whose angular momenta are coupled to 0 and whose isospins are coupled to 1; the pairing contribution to the binding is effective between the particles of such pairs.

The states of s unpaired nucleons with isospin t together with an arbitrary number of pairs (consistent with the Pauli principle) form a basis for the $O(5)$ representation (p,q) with $p = 2t$ and $q = j - \frac{1}{2}s - t + \frac{1}{2}$. Then s is called the seniority, t the reduced isospin. The eigenvalues U, U_3 specify the total isospin and V_3 is related to the number of particles N by $N = 2j + 1 + 2V_3$. The states with the smallest value of V_3, i.e., $V_3 = -\frac{1}{2}p - q = \frac{1}{2}s - j - \frac{1}{2}$ have no pairs; those with the largest value, $V_3 = \frac{1}{2}p + q = j + \frac{1}{2} - \frac{1}{2}s$, have as many pairs as are consistent with the Pauli principle. The $U_\pm$ are isospin ladder operators; S_+, V_+, T_- create a pair with 3-component of isospin 1, 0, -1, respectively; S_-, V_-, T_+ are the corresponding annihilation operators.

Hecht[2] suggests using T_p, the pair isospin, as the degeneracy label. Parikh's suggestion[3] of n_α, the α-particle number, amounts to the same thing; for there is just one operator of degree $\frac{1}{2}(N - s) = q + \frac{1}{2}p + V_3$ in S_+, V_+, T_- with isospin T_p, T_{3p}, namely,

$$\Omega^{V_3}_{T_p,T_{3p}} = [(T_p + T_{3p})!\,(T_p - T_{3p})!]^{\frac{1}{2}} \times (V_+^2 + 4S_+T_-)^{\frac{1}{4}(2q+p+2V_3-2T_p)} \times \sum_x \frac{(-1)^x S_+^x V_+^{T_p+T_{3p}-2x} T_-^{x-T_{3p}}}{x!\,(x - T_{3p})!\,(T_p + T_{3p} - 2x)!}. \tag{A1}$$

As pointed out by Parikh, $V_+^2 + 4S_+T_-$ creates an "α particle," four nucleons with zero angular momentum and zero isospin; so Parikh's n_α is related to Hecht's T_p by $n_\alpha = \frac{1}{2}(2q + p + 2V_3 - 2T_p)$. All states can be reached by operating with

$$\Omega^{V_3}_{T_p,T_{3p}} \text{ on } \left| \begin{matrix} p & q \\ -\frac{1}{2}p - q & \frac{1}{2}p \end{matrix} \; U_3' \right\rangle,$$

the no-pair states with reduced isospin $t = \frac{1}{2}p$. In fact, the Hecht–Parikh state of definite U_p, U_{3p} is

$$\left| \begin{matrix} p & q \\ T_p & V_3 \quad U_p \end{matrix} \; U_{3p} \right\rangle = \sum_{T_{3p}} \Omega^{V_3}_{T_p,T_{3p}} \left| \begin{matrix} p & q \\ -\frac{1}{2}p - q & \frac{1}{2}p \end{matrix} \; U_3 - T_{3p} \right\rangle \times \left\langle \begin{matrix} T_p & \frac{1}{2}p \\ T_{3p} & U_3 - T_{3p} \end{matrix} \middle| \begin{matrix} U \\ U_3 \end{matrix} \right\rangle. \tag{A2}$$

The Hecht–Parikh states present difficulties, however. The label T_p is not unique; in general, many values of T_p yield the same nondegenerate state and the number of T_p values corresponding to a degenerate V_3, U exceeds the multiplicity in question (see Ref. 13, pp. 415 and 425). This redundancy could possibly be overcome by suitably restricting the range of T_p to give the correct multiplicity in each case. In general, each T_p state is a linear combination of all the $\underline{U}_p$ (symmetric) states in the degenerate subspace (in this, they contrast with the K states, each of which contains just one U_p state not contained in states of higher K). This nontriangularity of the $T_p - \underline{U}_p$ transformation matrix complicates the proof of the independence of T_p states and makes inherently more difficult the task of solving for $\underline{U}_p$ states in terms of T_p states. Moreover, the coefficients which arise when a T_p state is expanded in simple product states contain many sums (seven). For these reasons, we abandoned the Hecht–Parikh states and followed Elliott's idea of projecting from intrinsic states.

APPENDIX B

The expansion coefficients D in Eq. (19) for the projected, and, in particular, the Elliott states in terms of symmetric states are just the coefficients of the characteristic stretched terms in Eq. (10):

$$D^{pqV_3U}_{\underline{U}_pK} = B^{pq}_{KV_3U}(\underline{V}_{3p}\underline{U}_p\underline{U}_q), \tag{B1}$$

where $\underline{U}_q$ and $\underline{V}_{3p} \equiv \underline{V}_{3p}(\underline{U}_p)$ are given by Eqs. (17) and (18).

Equation (19) may be written

$$|K\rangle' = \sum_{\underline{U}_p} |\underline{U}_p\rangle' \sum_u \frac{(-1)^{K-u+\frac{1}{2}a}}{(K - u + \frac{1}{2}a)!\,(\frac{1}{2}a - K + u)!\,(U - \frac{1}{2}a + u)!\,(U - \frac{1}{2}a - u)!}, \tag{B2}$$

where for brevity we have put

$$|K\rangle' = \left| \begin{matrix} p & & q & \\ K & V_3 & U & U_3 \end{matrix} \right\rangle \left[\frac{(2U)!}{(2q+1)!!\,(U+K)!\,(U-K)!} \right]^{\frac{1}{2}}, \tag{B3}$$

$$|\underline{U}_p\rangle' = \left| \begin{matrix} p & & q & \\ \underline{U}_p & V_3 & U & U_3 \end{matrix} \right\rangle \frac{(-1)^{\frac{1}{2}p-\underline{U}_p+a}}{\underline{U}_q!} \left[\frac{(2\underline{U}_q+1)a!\,(2\underline{U}_p-a)!\,2^{\underline{U}_q-\underline{U}_p+\frac{1}{2}p}}{(q+V_3+a+1)!!\,(q+U-V_3-a+1)!!} \right]^{\frac{1}{2}}$$
$$\times \left[\frac{(2\underline{U}_p)!\,(2\underline{U}_q)!}{[\frac{1}{2}(q-U+V_3+a)]!\,[\frac{1}{2}(q+2\underline{U}_p-U-V_3-a)]!\,(\frac{1}{2}p-\underline{U}_p)!\,(\frac{1}{2}p+\underline{U}_p+1)!} \right]^{\frac{1}{2}}, \tag{B4}$$

$$a = \underline{U}_p - \underline{V}_p = \max(U - V_3 - q, \epsilon') \tag{B5}$$

and substituted $x = K + \frac{1}{2}a - u$.

We rewrite Eq. (B2) as

$$|K\rangle' = \sum_u |u\rangle \frac{(-1)^{K-u+\frac{1}{2}a}}{(K-u+\frac{1}{2}a)!\,(\frac{1}{2}a-K+u)!}, \tag{B6}$$

with

$$|u\rangle = \sum_{\underline{U}_p} |\underline{U}_p\rangle' / (\underline{U}_p - \tfrac{1}{2}a + u)!\,(\underline{U}_p - \tfrac{1}{2}a - u)!. \tag{B7}$$

Then (B6) can be solved for $|u\rangle$ by noting that the expansion of $|K\rangle'$ in terms of $|u\rangle$ is equivalent to that of $z^{K-\frac{1}{2}a}(z-1)^a/a!$ in terms of z^u; and (B7) can be solved for $|\underline{U}_p\rangle'$ by noting that the expansion of $|u\rangle$ in terms of $|\underline{U}_p\rangle'$ is equivalent to that of $(-1)^u J_{2u}(z)$ in terms of $(-1)^{\underline{U}_p-\frac{1}{2}a}(\frac{1}{2}z)^{2\underline{U}_p-a}$. The results are

$$|u\rangle = \sum_K |K\rangle' \frac{(-1)^a a(K-u+\frac{1}{2}a-1)!}{(K-u-\frac{1}{2}a)!}, \tag{B8}$$

$$|\underline{U}_p\rangle' = 2\sum_u |u\rangle \frac{(-1)^{\underline{U}_p-\frac{1}{2}a-u} u(u+\underline{U}_p-\frac{1}{2}a-1)!}{(u-\underline{U}_p+\frac{1}{2}a)!}. \tag{B9}$$

Combining Eqs. (B3), (B4), (B8), and (B9) shows that D^{-1} of Eq. (20) is given by

$$(D^{-1})^{pqV_3U}_{K\underline{U}_p} = 2\underline{U}_q! \left(\frac{(q+V_3+a+1)!!\,(q+U-V_3-a+1)!!\,[\frac{1}{2}(q-U+V_3+a)]!}{(2\underline{U}_q+1)!\,(a-1)!\,(2\underline{U}_p-a)!\,2^{\underline{U}_q-\underline{U}_p+\frac{1}{2}p}} \right)^{\frac{1}{2}}$$
$$\times \left(\frac{[\frac{1}{2}(q+2\underline{U}_p-U-V_3-a)]!\,(\frac{1}{2}p-\underline{U}_p)!\,(\frac{1}{2}p+\underline{U}_p+1)!\,(2U)!}{(2\underline{U}_p)!\,(2q+1)!!\,(U+K)!\,(U-K)!} \right)^{\frac{1}{2}}$$
$$\times \sum_u \frac{(-1)^{\frac{1}{2}p-\frac{1}{2}a-u} u(u+\underline{U}_p-\frac{1}{2}a-1)!\,(K-u+\frac{1}{2}a-1)!}{(u-\underline{U}_p+\frac{1}{2}a)!\,(K-u-\frac{1}{2}a)!}. \tag{B10}$$

If $a = 0$ in Eq. (B10), the factors $(K-u+\frac{1}{2}a-1)!/(a-1)!\,(K-u-\frac{1}{2}a)!$ should be replaced by δ_{Ku} and the u-sum carried out; if, in addition, $K = \underline{U}_p = 0$, the factors $K(K+\underline{U}_p-\frac{1}{2}a-1)!$ should be replaced by unity.

* Research supported by the National Research Council of Canada.

[1] B. H. Flowers, Proc. Roy. Soc. (London) **A212**, 248 (1952); A. R. Edmonds and B. H. Flowers, *ibid.* **A214**, 515 (1952); K. Helmers, Nucl. Phys. **23**, 594 (1961); K. T. Hecht, *ibid.* **A102**, 11 (1967); R. P. Hemenger, University of Michigan, Ann Arbor, Mich., Preprint 07591-3-T, 1968.

[2] K. T. Hecht, Nucl. Phys. **63**, 177 (1965).

[3] J. C. Parikh, Nucl. Phys. **63**, 214 (1965).

[4] E. P. Wigner, Phys. Rev. **51**, 106 (1937).

[5] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128 and 562 (1958); J. P. Elliott and M. Harvey, *ibid.* **A272**, 557 (1963). See also M. Harvey, *Advances in Nuclear Physics* (Plenum Press, New York, 1968), Vol. I.

[6] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 77 (1961).

[7] N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968); see also W. J. Holman, University of Miami, Miami, Fla., Preprint CTS-M-68-1, 1968.

[8] R. T. Sharp and S. C. Pieper, J. Math. Phys. **9**, 663 (1968).

[9] S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).

[10] R. T. Sharp and H. C. von Baeyer, Nucl. Phys. **A140**, 118 (1970).

[11] See, for example, R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962), Sec. E, Theorem IV.

[12] T. Sebe, Nucl. Phys. **A109**, 65 (1968).

[13] A. de Shalit and I. Talmi, *Nuclear Shell Theory* (Academic Press, New York, 1963).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Electromagnetic Form Factors in $SL(n, C)$-Calculation of Ultraspherical Polynomials

N. S. THORNBER*
Physics Department, University of Bristol, Bristol, England

(Received 29 April 1969)

We derive a simple expression for the functions encountered in electromagnetic form factor calculations done with infinite-component fields. Our expression holds for the general case of $SL(n, C)$. We relate our functions to Gegenbauer polynomials.

INTRODUCTION

One of the unsolved problems in particle physics is predicting the electromagnetic form factors of particles. A theoretical understanding of the electromagnetic structure of strongly interacting particles would be quite desirable. In this paper we shall deal with one approach to this subject, namely, infinite-component field theory.

Infinite-component fields have been used to derive the electromagnetic form factors of particles.[1] In particular, Fronsdal *et al.*[2,3] have computed the electromagnetic form factors for the case of spin 0 and for the proton. For the proton the calculation was done in an $SL(6, C)$ framework, and one had to deal with the ultraspherical functions appropriate to $SL(6, C)$. The analogous functions that arise in an $SL(n, C)$ framework are dealt with in the present paper. Our results should prove useful to anyone doing calculations with infinite-component fields in the framework of $SL(n, C)$.

Our main result is an expression for the $SL(n, C)$ functions in terms of the mth derivative of a rather simple term. This result was conjectured by Fronsdal and White[3] and proved by them[2,3] for $n = 2, 4$, and 6; it is proved for general n (n even) in the present paper. We also relate our expression to the well-known Gegenbauer polynomials. One may notice that our results may be directly used to illustrate *why* particles with integral spin and nonzero baryon number *should not* exist in nature.[4]

In Sec. I, we explain the problem and review the infinite-component-field formalism. The application of this formalism to the physical problem of electromagnetic structure of particles is discussed, and the symbols are defined. In Sec. II, we define and discuss the functions to be used in the actual calculation. Section III contains the derivation of our results. In Sec. IV, we summarize our results and remark on a simple application of these results illustrating why certain particles should not exist in nature.

I. INFINITE-COMPONENT FIELDS AND ELECTROMAGNETIC FORM FACTORS

The problem we are dealing with is the calculation of an electromagnetic vertex, shown in Fig. 1. A strongly interacting particle interacts with a photon; the "blob" represents the interaction. The strongly interacting particle has initial momentum p and final momentum p'. We are interested in the electromagnetic form factor associated with the transition of the particle from momentum p to momentum p'.

To set up the problem, we represent the strongly interacting particle by a wavefunction in momentum space. Ordinarily, one would expect this function to be a scalar (under Lorentz transformations) for a spin-0 particle, a function ψ_μ with one Lorentz index, satisfying $p^\mu\psi_\mu = 0$, for a spin-1 particle, and so on. In the infinite-component-field formalism this idea is generalized. Instead of having a wavefunction that contains only one spin (i.e., spin-0 or spin-1, etc., one now takes a generalized wavefunction having, say, N indices: $\psi_{\mu_1,\cdots,\mu_N}$. This generalized wavefunction contains many spins; i.e., it can be written as a sum of pieces, where each piece corresponds to a particle with a definite spin. For a physical particle with a given spin, one simply projects out the piece of $\psi_{\mu_1,\cdots,\mu_N}$ that corresponds to this spin.

To describe the physical process in Fig. 1, namely the electromagnetic vertex of a strongly interacting particle with a given spin, we thus write down the generalized wavefunctions $\psi_{\mu_1,\cdots,\mu_N}(p)$ and $\bar{\psi}^{\mu_1,\cdots,\mu_N}(p')$ and take their projections. [Here p and p' are the initial and final momenta of the strongly interacting particles, and $\bar{\psi}^{\mu_1,\cdots,\mu_N}$ transforms like $(\Psi_{\mu_1,\cdots,\mu_N})^*$.] The charge part of the vertex in Fig. 1 is then given, in first Born approximation, by

$$[(p + p')_\mu/2m][P\bar{\psi}^{\mu_1,\cdots,\mu_N}(p')]Q(P\psi_{\mu_1,\cdots,\mu_N}(p)),$$

where m is the mass of the strongly interacting particle, Q is the charge operator, and P is the

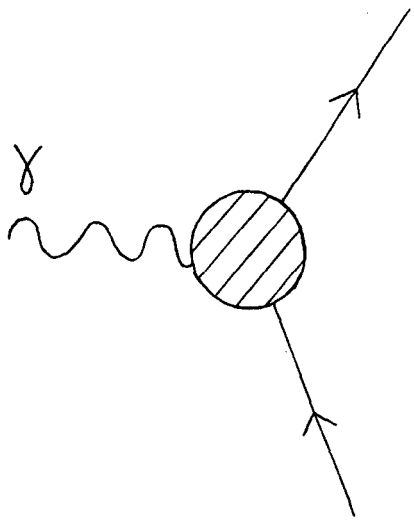


FIG. 1. Vertex for electromagnetic form factor. The solid lines represent the strongly interacting particle.

projection operator that projects out the desired spin of the particle.

We now direct our attention to the more general case, where the strongly interacting particle belongs to a specific representation of $SU(6)$. In this case, one uses a slightly more general wavefunction,

$$\psi^{\dot{B}_1,\cdots,\dot{B}_N}_{A_1,\cdots,A_{N+k}}(p),$$

to replace the wavefunction $\psi_{\mu_1,\cdots,\mu_N}$. [For $SU(6)$ one sets $k=3$; N is as yet unspecified.] The indices $A_1,\cdots,A_{N+k}$, $\dot{B}_1,\cdots,\dot{B}_N$, are spinor indices; the A's are undotted, while the B's are dotted spinor indices. If the strongly interacting particle belongs to the 56-dimensional representation of $SU(6)$, then the appropriate projection of the general wavefunction turns out to be[2]:

$$\psi^{\dot{B}_1,\cdots,\dot{B}_N}_{A_1,\cdots,A_{N+k}}(p) \to m^{-N} S p^{\dot{B}_1}_{A_1}\cdots p^{\dot{B}_N}_{A_N}\psi_{A_{N+1},\cdots,A_{N+k}}(p). \tag{1}$$

Here m is the mass of the strongly interacting particle, p is its momentum, and $\psi_{A_{N+1},\cdots,A_{N+k}}(p)$ is a wavefunction with three indices and corresponding to momentum p. $p^{\dot{B}}_A$ is a 6×6 matrix defined as

$$p^{\dot{B}}_A = (p_0 - \mathbf{p}\cdot\boldsymbol{\sigma})^{\beta}_{\alpha}\delta^{b}_{a},$$

where $(p_0, \mathbf{p})$ is the four-momentum of the strongly interacting particle, $\boldsymbol{\sigma}$ is the set of 2×2 Pauli spin matrices, and a and b run from 1 to 3. The symbol S stands for symmetrization in the indices $A_1,\cdots A_{N+k}$.

Thus, for the case of a particle scattering from momentum p to momentum p', assuming the particle belongs to the completely symmetric (56) representation of $SU(6)$, one obtains for the charge part of the vertex of Fig. 1 (to lowest Born approximation):

$$\psi^{\dot{B}_1,\cdots,\dot{B}_N}_{A_1,\cdots,A_{N+k}}(p) \to m^{-N} S p^{\dot{B}_1}_{A_1}\cdots p^{\dot{B}_N}_{A_N}\psi_{A_{N+1},\cdots,A_{N+k}}(p),$$
$$\bar{\psi}^{A_1,\cdots,A_{N+k}}_{\dot{B}_1,\cdots,\dot{B}_N}(p') \to m^{-N} S\bar{\psi}^{A_1,\cdots,A_k}(p')p'^{A_{k+1}}_{\dot{B}_1}\cdots p'^{A_{k+N}}_{\dot{B}_N}, \tag{2}$$

Vertex

$$= \frac{(p+p')_\mu}{2m}\,\bar{\psi}^{A_1,\cdots,A_{N+k}}_{\dot{B}_1,\cdots,\dot{B}_N}(p')Q\psi^{\dot{B}_1,\cdots,\dot{B}_N}_{A_1,\cdots,A_{N+k}}(p)$$
$$\to \frac{(p+p')_\mu}{2m}\,m^{-2N}(S\bar{\psi}^{A_1,\cdots,A_k}(p')p'^{A_{k+1}}_{\dot{B}_1}\cdots p'^{A_{k+N}}_{\dot{B}_N})$$
$$\times Q(Sp^{\dot{B}_1}_{A_1}\cdots p^{\dot{B}_N}_{A_N}\psi_{A_{N+1},\cdots,A_{N+k}}(p)). \tag{3}$$

Here

$$p^{\dot{B}}_A = (p_0 - \mathbf{p}\cdot\boldsymbol{\sigma})^{\beta}_{\alpha}\delta^{b}_{a}, \quad \text{and} \quad p'^{B}_{A} = (p'_0 + \mathbf{p}\cdot\boldsymbol{\sigma})^{\beta}_{\alpha}\delta^{b}_{a},$$

where the initial momentum of the strongly interacting particle is $(p_0, \mathbf{p})$, and the final momentum is $(p'_0, \mathbf{p}')$.

For a strongly interacting particle belonging to the completely symmetric representation of $SU(n)$ [instead of $SU(6)$, as discussed above], exactly the same form of the vertex holds true. The only difference is that, whereas in the $SU(6)$ case we put the index $k=3$, in the case of $SU(n)$ we now put $k=\frac{1}{2}n$ in Eq. (3).

The vertex obtained for the case of a proton-$SU(6)$ has in fact been explicitly evaluated by Cocho *et al.*[2] We shall not attempt an evaluation of the vertex for $SU(n)$ here,[5] but rather derive a simple expression for some of the complicated functions arising in this context. These functions are defined in the following section.

II. DEFINITION OF FUNCTIONS

In the previous section we discussed and wrote down the expression for the charge part of electromagnetic vertex of Fig. 1, namely Eq. (3). This expression is rather complicated, but it can be written in a simpler form,[2] namely,

$$\text{Vertex} = \frac{(p+p')_\mu}{2m}\,\bar{\psi}^{A_1,\cdots,A_k}(p')(f_0\delta^{A'_1}_{A_1}\cdots\delta^{A'_k}_{A_k}$$
$$+ f_1\delta^{A'_1}_{A_1}\cdots\delta^{A'_{k-1}}_{A_{k-1}}T^{A'_k}_{A_k}$$
$$+\cdots+ f_k T^{A'_1}_{A_1}\cdots T^{A'_k}_{A_k})\psi_{A'_1,\cdots,A'_k}(p), \tag{4}$$

where $T^{A'}_{A} = m^{-2}p^{\dot{C}}_{A}p'^{A'}_{\dot{C}}$ and f_j are functions of $p\cdot p'$. Cocho *et al.*[2] are then able to show (the proof is rather long) that the functions f_j are given by

$$f_j = \binom{N+k}{k}^{-1}\sum_{i=j}^{k}\binom{k}{i}\binom{i}{j}(-1)^{i-j}Q^{(i)}_{N-i+j}, \tag{5}$$

where the Q functions are defined by

$$
\begin{aligned}
{}^{k}Q_N^{(i)} &= (y - y^{-1})^{-1} \sum_{j=0}^{N-1} (y^{N-j} - y^{-N+j})\, {}^{k}Q_j^{(i-1)}, \\
{}^{k}Q_N^{(0)} &\equiv {}^{k}Q_N \\
&= m^{-2N}[S(pp')_{A_1}^{B_1} \cdots (pp')_{A_N}^{B_N}]\delta_{B_1}^{A_1} \cdots \delta_{B_N}^{A_N}. \quad (6)
\end{aligned}
$$

Here

$$y = m^{-2}\{p \cdot p' + [(p \cdot p')^2 - m^4]^{\frac{1}{2}}\}.$$

It is these functions ${}^{k}Q_N^{(i)}$ that we wish to investigate in the present paper. It turns out that they can be expressed in a rather simple form:

$${}^{k}Q_N^{(i)} = [(i + k - 1)!]^{-1}\left(\frac{d}{d\omega}\right)^{i+k-1}\left(\frac{y^{N+k} - y^{-N-k}}{y - y^{-1}}\right), \quad i \geq 0, \quad (7)$$

where

$$\omega = y + y^{-1} + 2, \quad \text{and} \quad k = \tfrac{1}{2}n,$$

This form has been derived by Fronsdal *et al.*[2,3] for $k = 1, 2$, and 3; it is derived in the present paper for general k.

We wish to remark that the result (7) (to be derived in the following section) is quite useful: using Eq. (7) it is easy to show[4] that particles with integral spin and nonzero baryon number have no charge coupling (i.e., the charge vertex considered in this paper *vanishes* for such particles).

III. DERIVATION FOR ${}^{k}Q_N^{(i)}$

In this section we derive Eq. (7) for $SL(n, C)$. Our first step is to obtain a recursion relation relating the $Q_j^{(0)}$ to each other. (For ease in writing, the index k will be temporarily suppressed.) To do this we look at the definition of Q_j [Eq. (6)] to obtain

$$
\begin{aligned}
Q_N = (1/N!)[&(N-1)!\, Q_0 \operatorname{Tr}(pp')^N m^{-2N} \\
&+ (N-1)!\, Q_1 \operatorname{Tr}(pp')^{N-1} m^{-2(N-1)} \\
&+ \cdots + (N-1)!\, Q_{N-1} \operatorname{Tr}(pp')^1 m^{-2}]. \quad (8)
\end{aligned}
$$

But $\operatorname{Tr}(pp')^j = km^{2j}(y^j + y^{-j})$, as can be easily shown by induction. Thus,

$$
\begin{aligned}
Q_N = (k/N)[&Q_0(y^N + y^{-N}) + Q_1(y^{N-1} + y^{-N+1}) \\
&+ \cdots + Q_{N-1}(y + y^{-1})];
\end{aligned}
$$

i.e., we have obtained the recursion relation[6]

$$Q_N = (k/N) \sum_{j=0}^{N-1} (y^{N-j} + y^{-N+j}) Q_j. \quad (9)$$

Our next step is to show that[7]

$$Q_N^{(i+1)} = (i + k)^{-1} \frac{d}{d\omega} Q_N^{(i)}. \quad (10)$$

This is shown by induction on N: The case $N = 1$ is easy to verify, using the definition of $Q_N^{(i)}$ [Eq. (6)]. Assuming Eq. (10) to be true for $N = 2, 3, \cdots, \mathcal{N}$, we then prove Eq. (10) is also true for $N = \mathcal{N} + 1$. Here we obtain an equation containing the index (i); the equation is verified by induction on (i) [the case $i = 0$ is verified separately, using Eq. (6) and the induction formula Eq. (9)]. The actual details are fairly straightforward, but very messy.

We now proceed with the proof of our general formula Eq. (7), using induction on k. For $k = 1$, we need to show that

$${}^{k=1}Q_N^{(i)} = (i!)^{-1}\left(\frac{d}{d\omega}\right)^{i}\left(\frac{y^{N+1} - y^{-N-1}}{y - y^{-1}}\right). \quad (11)$$

But it is easy to show, using Eq. (9) with $k = 1$, that

$${}^{k=1}Q_N^{(0)} = \frac{y^{N+1} - y^{-N-1}}{y - y^{-1}}. \quad (12)$$

(This expression is also normalized correctly: $Q_0^{(0)} = 1$.) Taking the ith derivative of both sides and using Eq. (10), we easily verify Eq. (11).

Proceeding with the proof by induction, we next assume Eq. (7) for $k = 2, 3, \cdots, K$, and proceed to prove Eq. (7) for $k = K + 1$. First we show that

$${}^{k=K+1}Q_N^{(0)} = (K!)^{-1}\left(\frac{d}{d\omega}\right)^{K}\left(\frac{y^{N+K+1} - y^{-N-K-1}}{y - y^{-1}}\right). \quad (13)$$

[If we can show this, then Eq. (7) for $k = K + 1$ follows immediately from Eq. (10).] Now the right-hand side of Eq. (13) is equal to ${}^{k=K}Q_{N+1}^{(1)}$ by induction hypothesis; i.e., we must show that

$${}^{k=K+1}Q_N^{(0)} = {}^{k=K}Q_{N+1}^{(1)}. \quad (14)$$

This equation is easily seen to hold for $N = 0$. Now assume Eq. (14) for $N = 1, 2, \cdots, \mathcal{N} - 1$, and prove Eq. (14) for $N = \mathcal{N}$.

Using Eq. (9) for ${}^{k=K+1}Q_{\mathcal{N}}^{(0)}$ and Eq. (6) for ${}^{k=K}Q_{\mathcal{N}+1}^{(1)}$, we thus need to show that

$$
\begin{aligned}
&[(K+1)/\mathcal{N}] \sum_{j=0}^{\mathcal{N}-1} (y^{\mathcal{N}-j} + y^{-\mathcal{N}+j})({}^{k=K+1}Q_j^{(0)}) \\
&= (y - y^{-1})^{-1} \sum_{j=0}^{\mathcal{N}} (y^{\mathcal{N}+1-j} - y^{-\mathcal{N}-1+j})({}^{k=K}Q_j^{(0)}). \quad (15)
\end{aligned}
$$

But by induction hypothesis on N,

$${}^{k=K+1}Q_j^{(0)} = {}^{k=K}Q_{j+1}^{(1)}, \quad j \leq \mathcal{N} - 1.$$

Putting this into Eq. (15) and expanding ${}^{k=K}Q_{j+1}^{(1)}$ by use of Eq. (6), we obtain a double sum on the left-hand side of Eq. (15). Interchanging the order of

summation, we do the sum over j explicitly. We obtain the following equation that needs to be shown true:

$$[(K+1)/\mathcal{N}]\sum_{l=0}^{\mathcal{N}} {}^{k=K}Q_l^{(0)}(\mathcal{N}-l)(y^{\mathcal{N}+1-l}-y^{-\mathcal{N}-1+l}) = \sum_{j=0}^{\mathcal{N}} {}^{k=K}Q_j^{(0)}(y^{\mathcal{N}+1-j}-y^{-\mathcal{N}-1+j}). \quad (16)$$

But

$$\sum_{l=0}^{\mathcal{N}} l({}^{k=K}Q_l^{(0)})(y^{\mathcal{N}+1-l}-y^{-\mathcal{N}-1+l}) = K\sum_{l=0}^{\mathcal{N}} {}^{k=K}Q_l^{(0)}(\mathcal{N}-l)(y^{\mathcal{N}+1-l}-y^{-\mathcal{N}-1+l}), \quad (17)$$

as can be seen by using Eq. (9) for ${}^{l}Q_l^{(0)}$ on the left-hand side, inverting the order of summation on the left-hand side, and then doing the sum over l explicitly. Thus, [from Eq. (17)]

$$\sum_{l=0}^{\mathcal{N}} l({}^{k=K}Q_l^{(0)})(y^{\mathcal{N}+1-l}-y^{-\mathcal{N}-1+l}) = [K/(1+K)]\sum_{l=0}^{\mathcal{N}} \mathcal{N}({}^{k=K}Q_l^{(0)})(y^{\mathcal{N}+1-l}-y^{-\mathcal{N}-1+l})$$

and, hence, Eq. (16) is true.

Thus, we have finished the proof of Eq. (7). This is our main result; we, thus, have obtained a rather compact expression for the functions ${}^{k}Q_N^{(i)}$ encountered in an infinite-component field formalism. For the special case of $i = 0$, our general result reduces to

$${}^{k}Q_N^{(0)} = [(k-1)!]^{-1}\left(\frac{d}{d\omega}\right)^{k-1}\left(\frac{y^{N+k}-y^{-N-k}}{y-y^{-1}}\right). \quad (18)$$

The functions ${}^{k}Q_N^{(i)}$ can also be related to Gegenbauer polynomials. In fact, in the notation of Ref. 8,

$${}^{k=1}Q_N^{(0)} = U_N(x) = C_N^{(1)}(x) = (y^{N+1}-y^{-N-1})/(y-y^{-1}),$$

$$x \leftrightarrow \tfrac{1}{2}(\omega-2) = \tfrac{1}{2}(y+y^{-1}).$$

Here U_N is a Tschebycheff polynomial of the second kind and $C_N^{(1)}$ is a Gegenbauer polynomial (also called ultraspherical polynomial). We then immediately have

$${}^{k}Q_N^{(i)} = [(i+k-1)!]^{-1}\left(\frac{d}{d\omega}\right)^{i+k-1} C_{N+k-1}^{1}(x).$$

Thus,

$${}^{k}Q_N^{(i)} = C_{N-i}^{k+i}(x), \quad x \leftrightarrow \tfrac{1}{2}(\omega-2), \quad (19)$$

and C_{N-i}^{k+i} is a Gegenbauer polynomial.

IV. SUMMARY

We have thus derived an expression for the functions ${}^{k}Q_N^{(i)}$ that arise in calculations of form factors using infinite-component fields. This expression [Eq. (7)] has been conjectured by Fronsdal and White, but had been proved only for $k = 1$, 2, and 3.[2,3] It is proved in the present paper for general k. In addition, we show that

$${}^{k}Q_N^{(i)} = C_{N-i}^{k+i}(x), \quad x \leftrightarrow \tfrac{1}{2}(\omega-2),$$

where $C_{N-i}^{k+i}(x)$ is a Gegenbauer polynomial. Using the main result [Eq. (7)], one can directly show that particles with integral sign and nonzero baryon number have no electromagnetic charge coupling.[4]

ACKNOWLEDGMENTS

The author wishes to thank the University of Bristol for its hospitality and the Science Research Council for financial support. Thanks are also extended to the Stanford University Physics Department and to SLAC, where this work was begun.

* Present address: Department of Mathematics, Newark College of Engineering, Newark, N.J. 07102

[1] A summary of work using infinite-component fields in connection with form factors is listed by C. Fronsdal, Phys. Rev. **182**, 1564 (1969).

[2] G. Cocho, C. Fronsdal, H. Ar-Rashid, and R. White, Phys. Rev. Letters **17**, 275 (1966).

[3] C. Fronsdal and R. White, Phys. Rev. **163**, 1835 (1967).

[4] N. S. Thornber, Phys. Rev. **D1**, 431 (1970). By "particle" we mean an "elementary" particle with a symmetric spin-isospin wavefunction.

[5] The result of Cocho *et al.* in Ref. 2 [$SU(6)$] does not agree with experiment unless additional pole terms are included. It has been shown by Barut *et al.* [A. O. Barut, D. Corrigan, and H. Kleinert, Phys. Rev. Letters **20**, 167 (1968)] that the group $O(4, 2)$ does give the experimentally correct form factors.

For $n = 2k$ with k even, the vertex considered in the present paper turns out to vanish (see Ref. 4). For $n = 2k$ with k odd (the only other case), we conjecture that the form factor has the form $(1 - t/4m^2)^{N+1}$, where $N = -\frac{3}{4}n$, in conventional Dirac theory. For $SL(n = 2, C)$ with Pauli 2-spinors, this reduces to $(1 - t/4m^2)^{-\frac{3}{4}}$, whereas for $SL(n = 6, C)$ with Dirac spinors, this reduces to $(1 - t/4m^2)^{-\frac{3}{2}}$, in agreement with Ref. 3 and 2, respectively.

[6] This recursion relation was obtained by Cocho *et al.* (Ref. 2) for the case of $k = 3$.

[7] This relation was found in Refs. 2 and 3 for $k = 1$, 2, and 3.

[8] *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill, 1953), Vol. 2.

Remarks on the Equations of Motion in Odd-Dimensional Spaces and *TCP* Invariance

T. S. SANTHANAM*
International Atomic Energy Agency, International Centre for Theoretical Physics
Trieste, Italy

(Received 20 June 1969)

It is shown that the problem of *TCP* noninvariance of the equations of motion in odd-dimensional spaces is connected with the irreducibility of the representations of the Clifford algebra C_{2n+1}. It is further shown that the above difficulty can be overcome by using representations of C_{2i+1}, $i > n$,whether the particle is massive or massless.

Recently there have been attempts to look into the problem of *TCP* invariance of the equations of motion in odd-dimensional spaces. In particular it has been shown[1] that the *TCP* theorem may break down in an odd-dimensional space–time. In this paper we show that this problem is connected with the irreducibility of the representations of the Clifford algebra C_{2n+1}, and is independent of the question of the mass of the particle. It is further argued that the above difficulty can be overcome by using representations of C_{2i+1}, $i > n$, which become reducible representations of C_{2n+1}.

It is well known[2] that only $2n + 1$ matrices Γ_μ ($\mu = 1, \cdots, 2n + 1$) exist of dimension $2^n \times 2^n$ satisfying the Clifford algebra

$$\{\Gamma_\mu, \Gamma_\nu\}_+ = 2\delta_{\mu\nu}, \quad \mu, \nu = 1, \cdots, 2n + 1, \quad (1a)$$

$$\Gamma^2 = I, \quad (1b)$$

and

$$\Gamma_1 \cdots \Gamma_{2n+1} = (i)^n I. \quad (1c)$$

These $(2n + 1)$ matrices of dimension $2^n \times 2^n$ form the irreducible base elements of C_{2n+1} which is of dimension 2^n. The complete set of 2^n elements of C_{2n+1} can be obtained through the product

$$X_{(\rho_1 \cdots \rho_{2n})} = (\Gamma_1)^{\rho_1}(\Gamma_2)^{\rho_2} \cdots (\Gamma_{2n})^{\rho_{2n}}, \quad (2a)$$

where the integers ρ satisfy

$$0 \leq \rho_\nu \leq 1, \quad \nu = 1, \cdots, 2n. \quad (2b)$$

The explicit construction of the Γ_μ's in terms of Pauli matrices is known[3] and, more recently, an elegant algorithm has been given[4] for this.

The equation of motion in $(2n + 1)$-dimensional space–time can be written as[5]

$$(\Gamma_\mu \partial_\mu + m)\psi(r_1, \cdots, r_{2n}, t) = 0, \quad \mu = 1, \cdots, 2n + 1. \quad (3)$$

The *TCP* operator can be defined[1] as the one which takes the differential operator ∂_μ into $-\partial_\mu$. If one uses all the $(2n + 1)\Gamma_\mu$'s in Eq. (3), in view of Eq. (1), no operator S exists which satisfies

$$S\Gamma_\mu S^{-1} = -\Gamma_\mu. \quad (4)$$

This is the conclusion of Ref. 1. So it looks as if one cannot write an equation in odd space–time which is *TCP* invariant. It may be easily seen that this difficulty can be traced to the simple fact that Γ_μ's form a *complete set* satisfying Eq. (1) and if they are irreducible no operator S exists which will anticommute with all Γ_μ's, independent of whether or not the particle has mass.

On the other hand, if one uses reducible representations for the Γ_μ's one can find an S which satisfies Eq. (4). To fix the point, if the Γ_μ's are of dimension $2^{n+1} \times 2^{n+1}$, $\mu = 1, \cdots, 2n + 1$; then any of the other two matrices of dimension $2^{n+1} \times 2^{n+1}$ from the complete set of $2n + 3$ matrices satisfying Eq. (1) can be used as the *TCP* operator. For example, if the Dirac equation is written in $(4 + 1)$-dimensional space–time, $\mu = 1, \cdots, 5$, using not the irreducible 4×4 matrices but the reducible 8×8 matrices, then any one of the other two anticommuting matrices of dimension 8×8 from the complete set of seven matrices can be used as the *TCP* operator. Thus there is no problem of *TCP* invariance in odd space–time if we use representations of C_{2i+1}, $i > n$,which become the reducible representations of C_{2n+1}.[6]

ACKNOWLEDGMENTS

The author is grateful to Professor Abdus Salam and Professor P. Budini, and the International Atomic Energy Agency for hospitality at the International Centre for Theoretical Physics, Trieste.

* On leave of absence from Matscience, Institute of Mathematical Sciences, Madras, India.

[1] S. P. Rosen, J. Math. Phys. **9**, 1593 (1968).

[2] See, for example, H. Boerner, *Representation of groups* (North-Holland Publ. Co., Amsterdam, 1963), Chap. 8.

[3] R. Brauer and H. Weyl, Am. J. Math. **47**, 447 (1935).

[4] A. Ramakrishnan, J. Math. Anal. Appl. **20**, 9 (1967).

[5] T. S. Santhanam and P. S. Chandrasekaran, Progr. Theoret. Phys. (Kyoto) **41**, 264 (1969).

[6] Similar conclusions have recently been advanced by more indirect arguments by V. I. Fushchich (Academy of Sciences of Ukrainian SSR, Institute for Theoretical Physics, Kiev, Preprint 69-17, 1969). Actually, an equation in 5-dimensional space–time has been given by M. Bakri, J. Math. Phys. **10**, 298 (1969), which is *CPT* invariant. See also A. Ramakrishnan, J. Math. Anal. Appl. **22**, 39 (1968).