Crossing Matrices for Helicity Amplitudes, Application to Crossed Channel Partial-Wave Analysis, and Reggeization*

IVAN J. MUZINICH

University of Washington, Seattle, Washington
(Received 11 February 1964)

The crossing relations of two-body scattering amplitudes for reactions involving particles with spin are derived for the helicity amplitudes of Jacob and Wick. From the crossing relations, the differential cross section and polarization of the direct channel are related to the analytic continuation of the crossed channel helicity amplitude. The differential cross section is then expressed in terms of the crossed channel partial waves, and rules are given for treating the exchange of fixed angular momentum poles and Regge poles for two-body processes involving particles with higher spin.

I. INTRODUCTION

THE crossing relations relate the scattering amplitude for the direct process of some two-particle scattering process to the analytic continuation of the scattering amplitude for the crossed process. It is the concern of this communication to relate the c.m. helicity amplitude for processes with spin to the analytic continuation of the c.m. helicity[1] amplitude for the "crossed process." The statement of crossing for the helicity amplitudes is useful for practical applications of the dispersion theory of strong interactions and has been worked out for processes such as πN and NN scattering[2] in detail.

The crossing relations were known in quantum field theory as the substitution rule and follow from the principle of analytic continuation in the linear momenta, which is assumed in the S matrix theory.[3,4]

It is the latter point of view which we consider here. In Sec. II, the crossing relations are given for the invariant spinor functions (M functions) introduced by Stapp.[3] And from the crossing relations for the M functions, the crossing relations for the c.m. helicity amplitudes are derived. The crossing relations are given in terms of the crossing matrices, which are rotations (unitary matrices in spin space). The details concerning the angles of rotation are given in an Appendix.

In Sec. III, two theorems are proven from the crossing relations relating the polarization and differential cross section to the c.m. helicity amplitude for the "crossed process."

In Sec. IV, the cross section for a two-body reaction is related to the partial wave helicity amplitudes of the "crossed reaction," and simple rules are given for the calculation of the cross section due to the exchange of a fixed angular momentum pole or a Regge pole for processes involving particles with spin.

II. CROSSING MATRICES

We begin with a discussion of the kinematic preliminaries. Let the mass, spin, and momentum of the particles in the two-particle process be denoted by m_i, s_i, and k_i, where $i = 1, 2, 3, 4$; we will denote

[1] This problem has also been studied by G. C. Wick and T. L. Treuman of the Brookhaven National Laboratory. They have reached conclusions similar to ours [Ann. Phys. (N. Y.) **26**, 322 (1964)].

[2] M. L. Goldberger, M. T. Grisaru, S. W. MacDowell, and D. Y. Wong, Phys. Rev. **120**, 2250 (1960) for the NN problem, and G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, *ibid.* **106**, 1377 (1957) for the πN problem.

[3] H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

[4] G. F. Chew, *S Matrix Theory of Strong Interactions* (W. A. Benjamin, Inc., New York, 1962).

Fig. 1. Two-body scattering diagram.

the helicities of the particles by lower-case Greek letters when the need arises. There are three physical processes related by analytic continuation of the linear momenta.

$$\text{(i)} \qquad 1 + 2 \rightarrow 3 + 4,$$
$$\text{(ii)} \qquad 1 + \bar{3} \rightarrow \bar{2} + 4,$$

and

$$\text{(iii)} \qquad 1 + \bar{4} \rightarrow 2 + \bar{3}.$$

All of the momenta are taken to be into the scattering diagram Fig. 1, and conservation of 4-momentum reads

$$k_1 + k_2 + k_3 + k_4 = 0. \qquad \text{(II.1)}$$

We define the usual invariants by

$$s = (k_1 + k_2)^2 = (k_3 + k_4)^2,$$
$$t = (k_1 + k_3)^2 = (k_2 + k_4)^2, \qquad \text{(II.2)}$$

and

$$u = (k_1 + k_4)^2 = (k_2 + k_3)^2.$$

Each of the momenta has the property[5]

$$k_i^2 = m_i^2 \quad \text{and} \quad s + t + u = \sum_{i=1}^{4} m_i^2.$$

In the s channel, particles 1 and 2 with momenta k_1 and k_2 are incoming, particles 3 and 4 with momenta $-k_3$ and $-k_4$ are outgoing (Fig. 2). The physical region for the invariant variables is

$$s \geq \max\ [(m_1 + m_2)^2, (m_3 + m_4)^2],$$
$$t \leq t_{\min}(s), \qquad \text{(II.3)}$$
$$u \leq u_{\min}(s),$$

since the masses are unequal the minimum momentum transfers $t_{\min}$ and $u_{\min}$ are not zero.[6]

Fig. 2. Direct channel (s-channel process).

---

[5] The metric here is such that $A \cdot B = A^\circ B^\circ - \mathbf{A} \cdot \mathbf{B}$, where $A = (A^\circ, \mathbf{A})$ and $B = (B^\circ, \mathbf{B})$ are 4-vectors.
[6] We will have no need to know the precise value of the minimum momentum transfers. The minimum momentum transfers $t_{\min}$ and $u_{\min}$ are defined by forward and backward scattering.
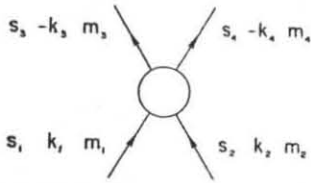

Fig. 3. Crossed channel (t-channel process).

In the t channel, particle 1 and antiparticle $\bar{3}$ with momenta k_1 and k_3 are incoming, antiparticle $\bar{2}$ and particle 4 with momenta $-k_2$ and $-k_4$ are outgoing (Fig. 3). The physical region for the invariant variables is

$$s \leq s_{\min}(t),$$
$$u \leq u_{\min}(t), \qquad \text{(II.4)}$$
$$t \geq \max\ [(m_1 + m_3)^2, (m_2 + m_4)^2].$$

There is another channel, the u channel, where the physical region for the invariants is defined by

$$s \leq s_{\min}(u),$$
$$t \leq t_{\min}(u), \qquad \text{(II.5)}$$
$$u \geq \max\ [(m_1 + m_4)^2, (m_2 + m_3)^2].$$

In the following, the s channel will be referred to as the direct channel and the t and u channels will be referred to as crossed channels. The S matrix for the s channel is written as

$$\langle \lambda' - k_3; \mu' - k_4|\ (S - 1)\ |\lambda, k_1; \mu k_2\rangle$$
$$= -i(2\pi)^4 \delta(k_1 + k_2 + k_3 + k_4)H', \qquad \text{(II.6)}$$

where λ, μ, λ', and μ' are the helicities of particles 1, 2, 3, and 4, respectively. The c.m. differential cross section per unit c.m. solid angle is related to the H amplitude by

$$\frac{d\sigma}{d\Omega} = \frac{k'}{k}\ |\Phi_{\lambda'\mu'.\lambda\mu}(\mathbf{k}', \mathbf{k})|^2, \qquad \text{(II.7)}$$

where

$$\Phi = H/4\pi s^{\frac{1}{2}}. \qquad \text{(II.8)}$$

$(-\mathbf{k}', \mathbf{k}')$ and $(-\mathbf{k}, \mathbf{k})$ are the final and initial c.m. momenta (Fig. 4).
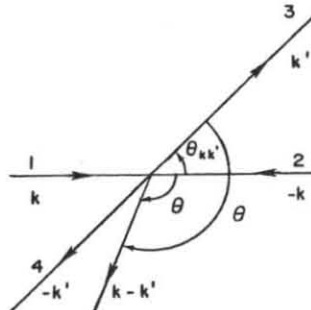
Under the homogeneous proper Lorentz Group,

Fig. 4. Scattering process in s-channel c. m.

the helicity amplitude transforms according to

$$H[K] = \mathfrak{D}^{**}[A(-k_4)] \otimes \mathfrak{D}^{**}[A(-k_3)]H[\Lambda^{-1}(\mathbf{A})K]$$
$$\times \mathfrak{D}^{**}[A^\dagger(k_2)] \otimes \mathfrak{D}^{**}[A^\dagger(k_1)]; \quad (II.9)$$

K represents collectively the incoming and outgoing momenta.[7] Elements of the proper homogeneous Lorentz group are denoted $\Lambda(\mathbf{A})$, where $\pm\mathbf{A}$ are the corresponding elements of the two-by-two unimodular group. The helicity indices are suppressed and matrix notation is used. The unitary matrices $\mathfrak{D}^*$ are the well-known representations of the unitary unimodular group. The unitary matrices $A(k)$ are known as the Wigner rotations in the literature and are defined by

$$A(k) = B^{-1}(k)\mathbf{A}B(\Lambda^{-1}k), \quad (II.10)$$

and the matrix $B(k)$ is defined by

$$\Lambda[B(k)]k_r = k,$$

where $k_r = (m, 0)$.

In the notation of Ref. 7,

$$B(k) = \left[\frac{k^\mu \sigma_\mu}{m}\right]^{\frac{1}{2}} U = U\left[\frac{k^0 \sigma_0 + |\mathbf{k}| \sigma_3}{m}\right]^{\frac{1}{2}}, \quad (II.11)$$

$$U = \exp\left(-\tfrac{1}{2}i\psi\sigma_3\right)\exp\left(-\tfrac{1}{2}i\theta\sigma_2\right)\exp\left(\tfrac{1}{2}i\psi\sigma_3\right), \quad (II.12)$$

the angles ψ, θ are the azimuthal and polar angles of the vector $\mathbf{k}$, and σ_μ are the usual Pauli matrices $\sigma_\mu = (\sigma_0, \mathbf{\sigma})$.

Besides invariance under proper Lorentz transformations, invariance under space reflection and time reversal are assumed throughout. Isotopic spin crossing will not be considered in this paper.[8]

It is useful to define another amplitude the invariant spinor function (M function[3,7]) by

$$M[K] = \mathfrak{D}^{**}[B(-k_4)] \otimes \mathfrak{D}^{**}[B(-k_3)]H[K]$$
$$\times \mathfrak{D}^{**}[B^\dagger(k_2)] \otimes \mathfrak{D}^{**}[B^\dagger(k_1)] \quad (II.13)$$

with the transformation property

$$M[K] = \mathfrak{D}^{**}[\mathbf{A}] \otimes \mathfrak{D}^{**}[\mathbf{A}]M[\Lambda^{-1}(\mathbf{A})K]$$
$$\times \mathfrak{D}^{**}[\mathbf{A}^\dagger] \otimes \mathfrak{D}^{**}[\mathbf{A}^\dagger]. \quad (II.14)$$

In the spin-$\frac{1}{2}$ case for example, Eq. (II.14) becomes[9]

---

[7] A. O. Barut, I. J. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963). Equation (II.9) can also be written as in Eq. (2.1) of Barut *et al.*, where an index transforming cording to $\mathfrak{D}^*$ corresponds to an outgoing particle or an incoming antiparticle. An index transforming according to $\mathfrak{D}^{**}$ corresponds to an incoming particle or an outgoing antiparticle.

[8] For a detailed treatment of isotopic spin crossing. See, for example, L. L. Foldy and R. F. Peierls, Phys. Rev. **130**, 1585 (1963).

[9] The conventions on the spinor indices of M are the same as in Ref. 7. The dotted index transforms according to $A\dagger$ (incoming particle or outgoing antiparticle) and the undotted index transforms according to A (incoming antiparticle or outgoing particle).

$$M_{\alpha\dot\beta}[K] = A_\alpha^{\alpha'} A_{\dot\beta}^{\dot\beta'} M_{\alpha'\dot\beta'}[\Lambda^{-1}(\mathbf{A})K], \quad (II.15)$$

when the spinor indices are displayed explicitly.[9]

Aside from having simple transformation properties, the M functions are expected to be free of kinematical singularities and are the natural objects to consider for analytic properties. The matrix $[k^\mu\sigma_\mu/m]^{\frac{1}{2}}$ in Eq. (II.11) can be written as

$$\left[\frac{k^\mu \sigma_\mu}{m}\right]^{\frac{1}{2}} = \left[\frac{k^0 + m}{2m}\right]^{\frac{1}{2}} + \hat{\mathbf{k}}\cdot\mathbf{\sigma}\left[\frac{k^0 - m}{2m}\right]^{\frac{1}{2}}, \quad (II.16)$$

and we see that the above matrix is analytical except at the kinematical branch points of the square root factors, $k^0 = \pm m$. The factors $[(k^0 \pm m)/2m]^{\frac{1}{2}}$ are precisely the factors that enter from the use of Dirac helicity spinors in the evaluation of the matrix element $H[K]$, and multiplication by $B(k)$ Eq. (II.11) removes these factors.

Note added in proof: Once the square roots in Eq. (II.16) are defined, the analytic continuation of $[k^\mu\sigma_\mu/m]^{\frac{1}{2}}$ and $H[K]$ is defined.

In particular, for spin-$\frac{1}{2}$ spin-0 elastic scattering, the M function is

$$M[K] = [k_1\cdot\sigma/m - k_3\cdot\sigma/m]A$$
$$- [k_3\cdot\sigma/m\bar{\sigma}\cdot(k_2 - k_4)k_1\sigma/m$$
$$- \sigma\cdot(k_2 - k_4)]B, \quad (II.17)$$

where A and B are the usual scalar amplitudes,[2] and the spin-$\frac{1}{2}$ particle has initial momenta k_1 and final momentum $-k_3$, the spin-0 particle has initial momentum k_2 and final momentum $-k_4$. The matrices $\bar{\sigma}_\mu$ are $(\sigma_0, -\mathbf{\sigma})$. The M function appears in the unitarity condition without kinematical singularities or projections and is the natural object to consider for analytic properties.[3]

We now assume that the M function can be analytically continued from one set of real energy momenta describing a physical process to another set of real energy momenta, but with some different signs.[9a] We assume that the physical sheet is such that are no natural boundaries and that this continuation is possible. The analytically continued four momentum with its sign reversed contributes oppositely to the energy–momentum conservation law, and if the object associated with this energy–momentum 4-vector was formerly a particle in the final (initial) state of a physical process, we interpret the analytically continued M function as

---

[9a] The assumption of the crossing properties for the M functions is not unique. One could make the same assumption for the helicity amplitudes. However, if one assumes that the M functions are completely free from singularities except those required by the unitarity condition, then it is useful to assume that they have crossing properties also.

describing a physical process in which there is an antiparticle in the initial (final) state. This is the only statement of analyticity we will need in the following.

The crossing relations for the M functions follow from the assumption about analyticity. In particular let $M^s_{\alpha_4\alpha_3;\dot\alpha_2\dot\alpha_1}[-k_4, -k_3; k_2, k_1]$ be the M function describing the physical process associated with the s channel (direct channel), and let $M^t_{\alpha_4\alpha_3;\dot\alpha_2\dot\alpha_1}[-k_4, k_3; -k_2, k_1]$ be the M function describing the physical process associated with the t channel (crossed channel) (Fig. 3). The crossing relation is simply

$$M^s_{\alpha_4\alpha_3;\dot\alpha_2\dot\alpha_1}[-k_4, -k_3; k_2, k_1]$$
$$= M^t_{\alpha_4\alpha_3;\dot\alpha_2\dot\alpha_1}[-k_4, -k_3; k_2, k_1] \qquad (II.18)$$

We desire the crossing relations for the c.m. helicity amplitudes; therefore, we use Eq. (II.13) to relate the M function to the helicity amplitude and Eq. (II.9) (Lorentz invariance) to relate the general frame helicity amplitude to the c.m. helicity amplitude.

The Lorentz transformation $\mathbf{A}$ that takes one from the c.m. in the t channel to a general frame is parameterized in the following way.[10]

$$\mathbf{A} = \mathbf{A}(\hat\beta, t) = \cosh \chi/2 + \hat\beta \cdot \sigma \sinh \tfrac{1}{2}\chi,$$
$$\mathbf{k}_1 + \mathbf{k}_3 = -(\mathbf{k}_2 + \mathbf{k}_4) = \hat\beta t^{\frac{1}{2}} \sinh \chi,$$
$$k_1^0 + k_3^0 = -(k_2^0 + k_4^0) = t^{\frac{1}{2}} \cosh \chi. \qquad (II.19)$$

For the helicity convention we have

$$\mathbf{A} = \mathbf{A}(\Psi, \Theta, -\Psi) = R(\Psi, \Theta, -\Psi)$$
$$\times [\cosh \tfrac{1}{2}\chi + \sigma_3 \sinh \tfrac{1}{2}\chi], \qquad (II.20)$$
where

$$R(\Psi, \Theta, -\Psi) = \exp(-i)\Psi\sigma_3/2$$
$$\times \exp(-i)\Theta\sigma_2/2 \exp i\Psi\sigma_3/2, \qquad (II.20')$$

and (Ψ, Θ) are the azimuthal and polar angles of the vector $\mathbf{k}_1 + \mathbf{k}_3$. From Eqs. (II.9) and (II.13) with A given by Eq. (II.20) we obtain

$$M^t[K''] = \mathfrak{D}^{s*}[AB(k_1')] \otimes \mathfrak{D}^{s*}[AB(k_3')]$$
$$\times H^t[\Lambda^{-1}(\mathbf{A})K'']\mathfrak{D}^{s*}[B^\dagger(k_2')\mathbf{A}^\dagger]$$
$$\otimes \mathfrak{D}^{s*}[B^\dagger(k_1')\mathbf{A}^\dagger], \qquad (II.21)$$

where K'' represents the set of momenta $\{-k_4, k_3; -k_2, k_1\}$, and $\Lambda^{-1}(\mathbf{A})K'' = K'$, which represents the set of momenta $\{k_4', k_3'; k_2', k_1'\}$. Since $\mathbf{A}$ is the transformation from the general frame to the c.m.,

the set K' is merely the c.m. momenta, and H^t is the helicity amplitude evaluated in the c.m. of the t channel. Next we analytically continue in the set K'' the expression (II.21) to the physical region for the s channel, Eq. (II.18) (direct channel). (Note $\mathbf{A}$ is also continued since it is a function of the momenta.) Furthermore, we use Eq. (II.13) to relate M to the helicity amplitude for the s channel and we obtain:

$$H^s[K] = \mathfrak{D}^{s*}[A(-k_4)] \otimes \mathfrak{D}^{s*}[A(-k_3)]$$
$$\times H^t[\Lambda^{-1}(\mathbf{A})K]\mathfrak{D}^{s*}[A^\dagger(k_2)] \otimes \mathfrak{D}^{s*}[A^\dagger(k_1)]. \qquad (II.22)$$

Equation (II.22) requires some explanations. The set of momenta $\{-k_4, -k_3; k_2, k_1\}$ are represented by K and are the physical momenta for the s-channel scattering, and

$$A(k) = B^{-1}(k)\mathbf{A}B(\Lambda^{-1}k), \qquad (II.23)$$

where $k = -k_4, -k_3, k_2$, or k_1. Hence, the matrices $A(k)$ are merely Wigner rotations for the Lorentz transformation $\mathbf{A}$. However, when the continuation to the direct channel is made, i.e., going from the real set K'' to the real set K, the variable t which was timelike before the continuation now becomes spacelike, and the vector $\mathbf{k}_1 + \mathbf{k}_3$ which specified the Lorentz transformation $\mathbf{A}$ becomes the momentum transfer $\mathbf{k}_1 - \mathbf{k}_3$ for the s channel. We are dealing with a complex Lorentz transformation in the crossing relation Eq. (II.22), and $A(k)$ Eq. (II.23) is the Wigner rotation for the complex Lorentz transformation. The Lorentz transformation $\mathbf{A}$ becomes complex since t can be negative in the physical region for the s channel, and the quantity $(t)^{\frac{1}{2}}$, which plays the role of the mass in the Lorentz transformation $\mathbf{A}$, is now complex. We obtain this complex Lorentz transformation solely from the continuation of Eq. (II.21) to obtain M^s and H^s.

To obtain our final result we will take the result for $H^s(K)$ and evaluate in the c.m. of the s channel $-(\mathbf{k}_4 + \mathbf{k}_3) = (\mathbf{k}_1 + \mathbf{k}_2) = 0$. Without loss of generality we can take the direction of the Lorentz transformation $\mathbf{A}$ [Eqs. (II.19) and (II.20)] to be along the 3 direction [we will consider scattering in the (1, 3) plane]. In the Appendix it will be shown that the Wigner rotations $A(k)$ corresponding to the complex Lorentz transformation are unitary and are rotations about the 2 direction (transverse to the scattering plane) through real angles. The final result for the crossing matrices is

$$H^s[K] = d^{s*}(\omega_4) \otimes d^{s*}(\omega_3)$$
$$\times H^t[\Lambda^{-1}(\mathbf{A})K]d^{s*}(-\omega_2) \otimes d^{s*}(-\omega_1), \qquad (II.24)$$

---

[10] The formulas on the Lorentz transformation from the c. m. to a general frame which are included here for completeness are contained in G. C. Wick, Ann. Phys. (N. Y.) **18**, 65 (1962).

where the matrices $d^s(\omega)$ are the familiar reduced rotation matrices and ω_i $(i = 1, 2, 3, 4)$ are the real angles of rotation for each of the particles. The angles of rotation will be constructed explicitly in the Appendix. The fact that the angles of rotation are real and the $d^s(\omega)$ are unitary is at first glance surprising and requires detailed derivation. The matrices $d^s(\omega)$ can be inverted easily $[d^s(\omega)]^{-1} = d^s(-w)$ and the inverse crossing relations can be found.

III. DIFFERENTIAL CROSS SECTION AND POLARIZATION

Next we prove two theorems from the crossing relation Eq. (II.24) regarding the unpolarized differential cross section and the polarization of one of the final particles in a two-body process.

The unpolarized differential cross section per unit momentum transfer is for the s channel

$$\frac{d\sigma}{dt} = \frac{1}{16\pi k^2 s} \frac{1}{(2s_1 + 1)(2s_2 + 1)} \text{Tr} \{H^s[K]H^{s\dagger}[K]\},$$
(III.1)

where t is the momentum transfer which is evaluated in the c.m. for the s channel, and $t = m_1^2 + m_3^2 - 2k_1^0 k_3^0 + 2\mathbf{k}'\cdot\mathbf{k}$.

Using the crossing relation Eq. (II.24) and (III.1) we obtain

$$\frac{d\sigma}{dt} = \frac{1}{16\pi k^2 s} \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$
$$\times \text{Tr} \{H^t[\Lambda^{-1}(A)K]H^t[\Lambda^{-1}(A)K]^\dagger\}.$$
(III.1')

The unitary crossing matrices $d^s(\omega)$ disappear when the spins are averaged and summed. Therefore, we have

Theorem (i). The unpolarized differential cross section of the direct channel is the spin average and sum of the product $H^t H^{t\dagger}$, where H^t is the analytic continuation of the c.m. helicity amplitude of the crossed channel.

The polarization of one of the final particles can also be related to the helicity amplitudes of the crossed channel. The polarization[11] P is defined as the expectation value of the total angular momentum of the particle in its rest frame. Following Jacob and Wick we can relate the polarization[12] to the direct channel helicity amplitudes in the following

[11] For the general theory of polarization in scattering processes see, for example, H. P. Stapp, Phys. Rev. **103**, 425 (1956); R. Spitzer and H. P. Stapp, *ibid.* **109**, 540 (1958); M. Jacob and G. C. Wick, Ann. Phys. (N. Y.) **7**, 404 (1959).
[12] One can also use in place of H in Eqs. (III.2) and (III.4) the amplitude Φ defined by Eq. (II.7).

way by standard methods

$$16\pi^2 s \frac{d\sigma}{d\Omega} \mathbf{P} = \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$
$$\times \text{Tr} \{(\mathfrak{D}^{s_f}(U)H^s[K])^\dagger \mathbf{\Sigma} \mathfrak{D}^{s_f}(U)H^s[K]\},$$
(III.2)

where s_f is the spin of the final particle whose polarization is being studied; $U = U(\psi, \theta, -\psi)$ [Eq. (II.12)], where θ and ψ are the polar and azimuthal angles of this final particle. $\mathbf{\Sigma}$ is the spin operator (angular momentum of particle in the rest frame). In terms of the rest frame state vectors we have

$$(\mathbf{\Sigma})^2 |s_f, \lambda\rangle = s_f(s_f + 1) |s_f, \lambda\rangle,$$
$$\Sigma_3 |s_f, \lambda\rangle = \lambda |s_f, \lambda\rangle.$$
(III.3)

In writing the indices explicitly, Eq. (III.2) takes the form

$$16\pi^2 s \frac{d\sigma}{d\Omega} \mathbf{P} = \frac{1}{(2s_1 + 1)(2s_2 + 1)} \sum (\mathfrak{D}^{s_f}_{\lambda\mu}(U)H^s_{\mu\sigma}[K])^*$$
$$\times \langle s_f, \lambda| \mathbf{\Sigma} |s_f, \lambda'\rangle \mathfrak{D}^{s_f}_{\lambda'\mu'}(U)H^s_{\mu'\sigma}[K],$$
(III.4)

only the relevant indices are displayed explicitly, all other indices are summed. Without loss of generality, we can take $\psi = 0$ and consider the scattering in the (1, 3) plane (Fig. 4.) In a parity-conserving theory we need only study the transverse component of the polarization which is in the 2 direction or $(\mathbf{k} \times \mathbf{k}')$ direction; $\mathbf{k}$ and $\mathbf{k}'$ are the initial and final c.m. momenta. In this case, U reduces to $\exp(-i\theta\Sigma_2)$ which commutes with Σ_2, and Eq. (III.2) reduces to

$$16\pi^2 s \frac{d\sigma}{d\Omega} \mathbf{P} = \hat{\mathbf{n}}_\perp \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$
$$\times \text{Tr} \{H^{s\dagger}[K]\Sigma_\perp H^s[K]\},$$
(III.5)

where $\hat{\mathbf{n}}_\perp$ is a unit vector in the transverse direction (2 direction), and $\Sigma_\perp$ is the transverse component of the spin.

Next we use the crossing matrices to relate the helicity amplitude $H^s[K]$ to the helicity amplitude of the cross channel Eq. (II.24), and since the crossing matrices are rotations about the transverse direction they dissappear in the trace Eq. (III.5) (rotations about the transverse direction commute with $\Sigma_\perp$); therefore, we obtain

$$16\pi^2 s \frac{d\sigma}{d\Omega} \mathbf{P} = \hat{\mathbf{n}}_\perp \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$
$$\times \text{Tr} \{H^{t\dagger}[\Lambda^{-1}(A)K]\Sigma_\perp H^t[\Lambda^{-1}(A)K]\}.$$
(III.6)

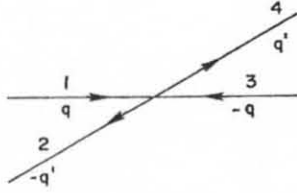We have the *Theorem (ii).* The transverse polarization of a final particle of the direct channel is trace

FIG. 5. Scattering process in t channel c. m.

of the product $H'^\dagger \Sigma_\perp H'$, where H' is the analytic continuation of the c.m. helicity amplitude of the cross channel.

IV. CROSS CHANNEL PARTIAL-WAVE ANALYSIS

In this section, the differential cross section is related to the partial waves of the crossed channel. The analysis of the differential cross section in terms of crossed channel partial waves is not a new idea and has been considered elsewhere,[13] and the discussion here is included for completeness. In particular, only a finite number of partial waves and Regge poles is retained in the crossed channel partial wave amplitude.

The c.m. helicity amplitude in terms of the t channel partial waves $T^J(t)$ is written for unequal mass kinematics following Jacob and Wick:

$$\Phi_{\bar\lambda\mu,\lambda\bar\mu}(\mathbf{q}',\mathbf{q}) = \frac{1}{2(q'q)^{\frac{1}{2}}} \Sigma_J (2J+1) \mathfrak{D}^J_{\lambda-\bar\lambda,\,\mu-\bar\mu}(R') T^J_{\bar\lambda\mu,\lambda\bar\mu},$$

$$= \frac{1}{4\pi(t)^{\frac{1}{2}}} H^t_{\bar\lambda\mu,\lambda\bar\mu}, \qquad (\text{IV.1})$$

where q' and q are the final and initial c.m. momenta for the t channel, λ and μ are the helicities of particles 1 and 4, and $\bar\lambda$ and $\bar\mu$ are the helicities of antiparticles $\bar3$ and $\bar2$ (Fig. 5) and

$$R' = R'(\alpha, \beta, -\alpha)$$

$$= \exp(-i\alpha J_3) \exp(-i\beta J_2) \exp(i\alpha J_3).$$

The quantities β and α are the polar angles of $\mathbf{q}'$ and the direction of the incident beam $\mathbf{q}$ is taken along the 3 direction.

Of course, we cannot continue the partial wave expansion to the physical region of the direct channel Eq. (II.3). The partial wave expansion converges only in the Lehmann elipse for the crossed channel. However, we will keep only a finite number of partial waves in Eq. (IV.1) and approximate the

amplitude with a finite number of poles; the question of convergence does not arise in this approximation although the approximation may not be too good. This approximation amounts to considering the exchange of a finite number of quanta, and Eq. (IV.1) becomes a sum over a finite number of partial waves.

The simplest situation we can encounter is the one quantum exchange approximation. If the particles 1 and $\bar3$ (2 and 4) have the same quantum numbers as some system (elementary particle or resonance) with angular momentum J, M $(-J \le M \le J)$ we can approximate the Jth partial wave amplitude by a single pole[14]

$$T^J_{\bar\mu\mu,\lambda\bar\lambda} = \frac{\Gamma_{\lambda\bar\mu}\Gamma_{\mu\bar\lambda}}{t - t_r + it_r^{\frac{1}{2}}\Gamma}, \qquad (\text{IV.2})$$

where t_r is the mass of the system, Γ its width, the quantities $\Gamma_{\lambda\bar\lambda}$ and $\Gamma_{\mu\bar\mu}$ are the partial widths, which are related to the coupling of the system to particles $\bar2$ and 4 and 1 and $\bar3$. We have assumed that the residue of the pole can be factored.

In determining the number of independent matrix elements and partial widths, we must consider the restrictions implied by angular momentum, space reflection invariance P, and time-reversal invariance T:

$$P: T^J_{-\bar\lambda-\mu,\,-\lambda-\bar\mu}(t) = \eta_2\eta_4\eta_1\eta_3(-1)^{s_2+s_4-s_1-s_3} T^J_{\bar\lambda\mu,\lambda\bar\mu}(t) \qquad (\text{IV.3})$$

$$T: \qquad T^J_{\lambda\bar\mu,\bar\lambda\mu}(t) = T^J_{\bar\lambda\mu,\lambda\bar\mu}(t), \qquad (\text{IV.4})$$

where the η, s are the phase factors denoting the intrinsic parities of the objects involved. For the partial widths which are proportional to the matrix element of the system $X(J, M)$ coupled to the initial and final particles we have, under space-reflection invariance,

$$\Gamma_{\mu\bar\mu} = \langle (J, M), X| \, T \, |(J, M); \mu\bar\mu \rangle$$

$$= \langle (J, M); X| \, P^{-1}TP \, |(J, M); \mu\bar\mu \rangle,$$

$$= (-1)^{J-s_2-s_4}\Gamma_{-\mu-\bar\mu}\eta_x\eta_2\eta_{\bar4}, \qquad (\text{IV.5})$$

and

$$\Gamma_{\lambda\bar\lambda} = (-1)^{J-s_1-s_3}\Gamma_{-\lambda-\bar\lambda}\eta_x\eta_1\eta_3;$$

the quantities $J - s_1 - s_3$ and $J - s_2 - s_4$ are always integers. Consideration of the total angular momentum J of the system leads us to the further restrictions

$$-J \le \lambda - \bar\lambda \le J \quad \text{and} \quad -J \le \mu - \bar\mu \le J; \quad (\text{IV.6})$$

of course, $|\lambda| \le s_1$, $|\bar\lambda| \le s_3$, $|\bar\mu| \le s_2$, and $|\mu| \le s_4$.

13 A. Martin and M. Gourdin (unpublished), have considered the analysis of the cross section in terms of the partial wave amplitudes of the crossed channel from the point of view of assuming that the spin sum can be analytically continued. Also P. K. DeCelles, L. Durand, and R. B. Marr, Phys. Rev. 126, 1882 (1962) have considered the cross section in the single quantum exchange process. They consider the problem from the point of view of the analysis of the vertex functions in the brick wall frame and give the multiple decomposition of the vertex functions.

14 Equation (IV.2) is the Breit-Wigner form of the partial-wave amplitude. This approximation is probably not good for t much different from t_r.

Time-reversal invariance which implies a symmetric S matrix does not give us any further restrictions on the partial widths.

Once J^η (spin parity) is given for the stable or unstable system, the number of independent partial widths will be restricted by Eqs. (IV.5) and (IV.6). Of course, the detailed dependence of the partial widths on t is a question that can only be answered by dynamics. However, the differential cross section is an easy matter to calculate in the one quantum exchange approximation, once the partial widths are given. Using Eq. (IV.2) in Eq. (IV.1) we have

$$\Phi_{\bar{\lambda}\mu,\lambda\bar{\mu}} = \frac{1}{2(q'q)^{\frac{1}{2}}} \frac{\mathfrak{D}^J_{\lambda-\bar{\lambda},\mu-\bar{\mu}}(R')\Gamma_{\lambda\bar{\lambda}}\Gamma_{\mu\bar{\mu}}}{t - t_r + it_r^{\frac{1}{2}}\Gamma}, \quad (IV.7)$$

where $t > \max[(m_1 + m_3)^2, (m_2 + m_4)^2,$

$$s \leq s_{\min}, \quad u \leq u_{\min}, \quad (IV.8)$$

and

$$s = m_1^2 + m_2^2 = 2[(q^2 + m_1^2)^{\frac{1}{2}}(q'^2 + m_2^2)^{\frac{1}{2}} + qq' \cos\beta] \quad (IV.9)$$

in the t-channel c.m. The angle α can be set equal to zero without loss of generality.

We now analytically continue Eq. (IV.7) to the physical region of the direct channel Eq. (II.3) and use Eq. (III.1) [Theorem (i)] to obtain for the direct-channel differential cross section:

$$\frac{d\sigma}{dt} = \frac{\pi}{16sk^2} \frac{F(s, t)}{|t - t_r + it_r^{\frac{1}{2}}\Gamma|^2}, \quad (IV.10)$$

where

$$F(s, t) = \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$
$$\times \sum_{\lambda\bar{\lambda}\mu\bar{\mu}} |\gamma_{\lambda\bar{\lambda}}\gamma_{\mu\bar{\mu}}|^2 |d^J_{\lambda-\bar{\lambda},\mu-\bar{\mu}}(\beta)|^2.$$

Note that the crossing matrices do not complicate the spin sums in view of Eq. (III.1) [Theorem (i)] the s dependence is displayed explicitly in the reduced rotation matrix $d^J(\beta)$. The widths have been redefined $\gamma_{\lambda\bar{\lambda}} = (t^{\frac{1}{2}}/q)\Gamma_{\lambda\bar{\lambda}}$, $\gamma_{\mu\bar{\mu}} = (t^{\frac{1}{2}}/q')\Gamma_{\mu\bar{\mu}}$.

We conclude this section with the calculation of the contribution of Regge poles of the t channel to the differential cross section of the direct channel. The problem of calculating the contribution of Regge poles of the crossed channel has been considered elsewhere.[15] Here the full (not asymptotic) contribution will be obtained.

The reggeization of two particle amplitudes has been considered in detail by the recent work of Gell-Mann et al.[16] They have introduced partial wave amplitudes of well-defined parity.

$$T^{J\pm}_{\bar{\lambda}\mu,\lambda\bar{\mu}} = T^J_{\bar{\lambda}\mu,\lambda\bar{\mu}} \pm \eta_{\bar{3}}\eta_4(-1)^{s_3+s_4-v}T^J_{\bar{\lambda}-\mu,\lambda-\bar{\mu}}. \quad (IV.11)$$

Here $v = 0$ for integral J and $\frac{1}{2}$ for half integral J. A given trajectory will belong to either $+$ or $-$ in a parity conserving theory. The concepts of sense and nonsense channels and compensating trajectories has been considered in detail by these authors. The concept of nonsense channels refers to channels in which J can become less than the difference of the helicities $\lambda - \bar{\lambda}$ or $\mu - \bar{\mu}$.

In order to find the contribution of a Regge pole of the t channel to the cross section of the direct channel we need the contribution to the helicity amplitude of the t channel to use in Eq. (III.1). The helicity amplitude for one Regge pole is written for the azimuthal angle equal to zero:

$$\Phi^t_{\bar{\lambda}\mu,\lambda\bar{\mu}} = -\frac{\pi}{4(qq')^{\frac{1}{2}}} \frac{2\alpha_\pm + 1}{\sin \pi\alpha_\pm} \xi^\pm_{\lambda\bar{\lambda}}(t)\xi^\pm_{\mu\bar{\mu}}(t)$$
$$\times [d^{\alpha_\pm}_{\lambda-\bar{\lambda},\bar{\mu}-\mu}(\pi - \beta)(-1)^{\bar{\lambda}-\lambda} + \epsilon d^{\alpha_\pm}_{\lambda-\bar{\lambda},\mu-\bar{\mu}}(\beta)], \quad (IV.12)$$

where the residue of the Regge pole has been factored[17] into the coupling parameters $\xi_{\lambda\bar{\lambda}}$ $\xi_{\mu\bar{\mu}}$, $\alpha_\pm$ is the position of the Regge pole, the subscript $\pm$ on α and ξ indicates that the pole is associated with either T^{J+} or T^{J-}, and ϵ is the signature of the trajectory.

The d functions in Eq. (IV.12) are not the ordinary reduced rotation matrices. In the evaluation of the d functions in terms of Legendere functions P_α of Appendix A of Ref. 11 (Jacob and Wick) the P_α functions should be replaced by $\mathcal{P}_\alpha = -(\tan \alpha\pi/\pi) \cdot Q_{-\alpha-1}$. The Q functions are the familiar Legendre functions of the second kind.

Using Eq. (III.1) we obtain for the unpolarized differential cross section

$$\frac{d\sigma}{dt} = \frac{\pi^3}{16sk^2}\left(\frac{2\alpha_\pm + 1}{\sin \pi\alpha_\pm}\right)^2 F(\alpha_\pm, s, t), \quad (IV.13)$$

Gribov, Zh. Eksperim. i Teor. Fiz. **41**, 667, 1962 (1961) [English transl.: Soviet Phys.—JETP **14**, 478, 1395 (1962)]. For the NN problem: I. J. Muzinich Phys. Rev. **130**, 1571 (1962); W. Wagner, Phys. Rev. Letters **10**, 202 (1963); D. Sharp and W. Wagner, Phys. Rev. (to be published). M. Gell-Mann in *The proceedings of the* 1962 *International Conference on High-Energy Physics CERN*, (Scientific Informative Service, Geneva 23, Switzerland). For the πN problem: V. Singh, Phys. Rev. **129**, 1889 (1962); M. Gell-Mann, F. Zacharaisen, and S. C. Frautschi, *ibid.* **126**, 2204 (1962).

[16] M. Gell-Mann, M. L. Goldberger, F. E. Low, E. Marx, and F. Zachariasen, Elementary Particles of Conventional Field Theory as Regge Poles III, especially Appendix B. Massachusetts Institute of Technology 1963 (unpublished).

[17] M. Gell-Mann, Phys. Rev. Letters **8**, 263 (1962). See also the work Gribov and Pomeranchuk, Ref. 15.

[15] G. F. Chew and S. C. Frautschi, Phys. Rev. Letters **7**, 394 (1961); G. F. Chew, S. C. Frautschi, and S. Mandelstam, Phys. Rev. **126**, 1202 (1962); V. N. Gribov and I. Ya Pomeranchuk, Phys. Rev. Letters **8**, 343, 412 (1962). V. N.

where

$$F(\alpha_\pm, s, t) = \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$

$$\times \sum_{\lambda\bar\lambda\mu\bar\mu} |d^{\alpha\pm}_{\lambda-\bar\lambda, \bar\mu-\mu}(\pi - \beta)(-1)^{\bar\lambda-\lambda} + \epsilon d^{\alpha\pm}_{\lambda-\bar\lambda, \mu-\bar\mu}(\beta)|^2$$

$$\times |\zeta^\pm_{\lambda\bar\lambda}\zeta^\pm_{\mu\bar\mu}|^2, \qquad\qquad (IV.14)$$

and

$$\zeta_{\lambda\bar\lambda} = (t^{\frac{1}{2}}/q)\xi_{\lambda\bar\lambda}, \qquad \zeta_{\mu\bar\mu} = (t^{\frac{1}{2}}/q')\xi_{\mu\bar\mu}.$$

If we have a finite number N of Regge poles, we obtain

$$\frac{d\sigma}{dt} = \frac{\pi^3}{16sk^2}\left\{\sum_i \frac{2\alpha_i + 1}{\sin \pi\alpha_i} F(\alpha_i, s, t)\right.$$

$$\left. + 2 \operatorname{Re} \sum_{i>j} \frac{2\alpha_i + 1}{\sin \pi\alpha_i}\frac{2\alpha_j + 1}{\sin \pi\alpha_j} G(\alpha_i, \alpha_j; s, t)\right\}, \quad (IV.15)$$

where

$$G(\alpha_i, \alpha_j; s, t) = \frac{1}{(2s_1 + 1)(2s_2 + 1)}$$

$$\times \sum_{\lambda\bar\lambda\mu\bar\mu} [d^{\alpha_i}_{\lambda-\bar\lambda, \bar\mu-\mu}(\pi - \beta)(-1)^{\bar\lambda-\lambda} + \epsilon_i d^{\alpha_i}_{\lambda-\bar\lambda, \mu-\bar\mu}]$$

$$\times [d^{\alpha_j}_{\lambda-\bar\lambda, \mu-\bar\mu}(\pi - \beta)(-1)^{\bar\lambda-\lambda} + \epsilon_j d^{\alpha_j}_{\lambda-\bar\lambda, \mu-\bar\mu}(\beta)]^*$$

$$\times \zeta^i_{\lambda\bar\lambda}\zeta^i_{\mu\bar\mu}(\zeta^j_{\lambda\bar\lambda}\zeta^j_{\mu\bar\mu})^*, \qquad\qquad (IV.16)$$

the indices i and j ($i, j = 1, 2, \cdots, N$) label the Regge pole parameters the $\pm$ subscript for the parity is included in the labels i and j.

Another important question is the number of independent coupling paramaters $\zeta_{\lambda\bar\lambda}$ for a given channel that has the quantum numbers of some given trajectory. This problem has been dealt with for the πN and $N\bar N$ channels[15] and is being studied for the $N_{33}\bar N$ channel (N_{33} is the $J = \frac{3}{2}$ isospin $\frac{3}{2}$ pion nucleon isobar).

The discussion in this section which is applicable for the boson-type Regge trajectories can easily be extended to the fermion trajectories.

V. CONCLUSIONS

From the crossing relations for the helicity amplitudes of two-body processes with higher spin, the cross section and polarization of the direct channel were related in a simple manner to the analytic continuation of the c.m. helicity amplitude for the crossed reaction. The crossing relations are simple in the helicity language and one does not have to go to the intermediate step of relating the helicity amplitudes to scalar amplitudes to complete the crossing. The scalar amplitudes are useful for discussing analytic properties in the invariants which

was not our concern here. It might be added that the discussion in the last section concerning the Regge poles will be greatly complicated if there are cuts in the angular-momentum plane as suggested by Mandelstam.[18]

APPENDIX: THE ANGLES OF ROTATION

The angle of rotation for the Wigner rotation Eqs. (II.10) and (II.23) can be computed by direct evaluation of the three matrices in Eqs. (II.10) and (II.23). This is most easily accomplished by using the representation of the matrices $A(p)$, $\mathbf{A}$, and $B(p)$ on 4-vectors $\Lambda[A(p)]$, $\Lambda[\mathbf{A}]$, and $\Lambda[B(p)]$. We take the Lorentz transformation along the 3 direction and Eq. (II.23) becomes

$$\Lambda[A(p)] = \Lambda[B^{-1}(p')]\Lambda[\mathbf{A}]\Lambda[B(p)],$$

$$= \Lambda^{-1}[B(p')]\Lambda[\mathbf{A}]\Lambda[B(p)], \qquad (A1)$$

where

$$p' = \Lambda[\mathbf{A}]\rho.$$

$$\Lambda[\mathbf{A}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh\chi & \sinh\chi \\ 0 & 0 & \sinh\chi & \cosh\chi \end{bmatrix},$$

$$\Lambda[B(p)] = \begin{bmatrix} \cos\rho & 0 & \sin\rho & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\rho & 0 & \cos\rho & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh\sigma & \sinh\sigma \\ 0 & 0 & \sinh\sigma & \cosh\sigma \end{bmatrix}, \qquad (A1')$$

and $p_0 = m \cosh\sigma$, $|\mathbf{p}| = m \sinh\sigma$. Here χ is the usual Lorentz transformation parameter and ρ and ρ' are the polar angles of the vectors $\mathbf{p}$ and $\mathbf{p}'$. After combining the three matrices in Eq. (A.1) we find

[18] S. Mandelstam, University of Birmingham (unpublished).

$$\Lambda[A(p)] = \begin{bmatrix} \cos\omega & 0 & \sin\omega & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\omega & 0 & \cos\omega & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (A2)$$

where

$$\cos\omega = \cos\rho\cos\rho' + \cosh\chi\sin\rho\sin\rho' \quad (A3)$$

or alternatively

$$\cosh\chi = \cosh\sigma\cosh\sigma' + \cos\omega\sinh\sigma\sinh\sigma'. \quad (A.3')$$

The axis of the Wigner rotation is transverse to the direction of the Lorentz transformation and the plane of the vectors $\mathbf{p}$ and $\mathbf{p}'$, the $(1, 3)$ plane. The angle of rotation is the angle between $\mathbf{p}$ and $\mathbf{p}'$ as seen from the rest system of the particle. The paper of Wick[11] contains other references and the geometrical interpretation of the angle.

If we give the vectors $\mathbf{p}$ and $\mathbf{p}'$ some azimuthal angles ψ and ψ', a Lorentz transformation along the three direction does not change the transverse components therefore $\psi = \psi'$, and the Wigner rotation Eq. (A.2) will become

$$\Lambda[A(p)] = \Lambda[\exp(-i)\psi J_3 \exp(-i)\omega J_2 \exp i\psi J_3]$$

$$= \Lambda[\exp(-i)\hat{n}\cdot J\omega] \quad (A4)$$

instead of $\Lambda[A(p)] = \Lambda[\exp(-i)J_2\omega]$ in Eq. (A.2). Here $\hat{n}$ is a unit vector transverse to the plane of $\mathbf{p}$ and $\mathbf{p}'$.

$$\hat{n} = -\hat{1}\sin\psi + \hat{2}\cos\psi, \quad (A4')$$

where $\hat{1}$ and $\hat{2}$ are unit vectors in the 1 and 2 directions. Thus, the only change that nonzero azimuthal angle can produce in the crossing relation Eqs. (II.23) and (II.24) and polarization etc., is a redefinition of transverse from the 2 direction to the $\hat{n}$ direction. And the crossing matrices in Eq. (II.24) would become $\mathfrak{D}'(\psi, \omega, -\psi)$ instead of $d'(\omega)$. We will set $\psi = 0$.

We now wish to find the angle ω for each of the particles in the scattering diagram Fig. 1 for the complex Lorentz transformation Eq. (II.20), when we continue to the physical region of the direct channel Eq. (II.3). We will take the Lorentz transformation along the direction

$$\hat{\beta} = (\mathbf{k} - \mathbf{k}')/t^{\frac{1}{2}}\sinh\chi, \quad (A5)$$

which we take as the three direction, and

$$k_1^0 - k_3^0 = t^{\frac{1}{2}}\cosh\chi, \quad (A6)$$

of course, $\mathbf{k}$ and $\mathbf{k}'$ will have some angles θ and θ' with respect to $\hat{\beta}$ and the cosine of the c.m. scat-

tering angle for the direct process is

$$\cos\theta_{kk'} = \cos(\theta' - \theta).$$

The vector $\mathbf{k} - \mathbf{k}'$ is the space part of the momentum transfer and $k_1^0 - k_3^0$ is the time component of the momentum transfer for the s channel. Using Eq. (A.3) the angle of rotation can be written

$$m^2\cosh\chi = p^0 p'^0 - |\mathbf{p}|\,|\mathbf{p}'|\cos\omega, \quad (A7)$$

and similarly for each of the particles we have for the transformation A in Eq. (II.20) from Eq. (A7):

$$m_1^2\cosh\chi = k_1'^0 k_1^0 - qk\cos\omega_1,$$
$$m_2^2\cosh\chi = k_2'^0 k_2^0 - q'k\cos\omega_2,$$
$$m_3^2\cosh\chi = k_3'^0 k_3^0 + qk\cos\omega_3, \quad (A8)$$
$$m_4^2\cosh\chi = k_4'^0 k_4^0 + q'k'\cos\omega_4,$$

where the k_i^0 $i = 1, 2, 3, 4$ are evaluated in the c.m. of the s channel and $k_i'^0$ are evaluated in the c.m. of the t channel. Similarly, q, q', and k, k' are the c.m. momenta for the t and s channels. A paper by Kibble[19] contains useful formulas for the relativistic kinematics of two-body reactions; however we list also the following formulas.

$$k_1^0 = \frac{s + m_1^2 - m_2^2}{2s^{\frac{1}{2}}}, \quad k_1'^0 = \frac{t + m_1^2 - m_3^2}{2t^{\frac{1}{2}}},$$

$$k_2^0 = \frac{s + m_2^2 - m_1^2}{2s^{\frac{1}{2}}}, \quad k_2'^0 = \frac{t + m_2^2 - m_4^2}{2t^{\frac{1}{2}}},$$

$$k_3^0 = \frac{s + m_3^2 - m_4^2}{2s^{\frac{1}{2}}}, \quad k_3'^0 = \frac{t + m_3^2 - m_1^2}{2t^{\frac{1}{2}}}, \quad (A9)$$

$$k_4^0 = \frac{s + m_4^2 - m_3^2}{2s^{\frac{1}{2}}}, \quad k_4'^0 = \frac{t + m_4^2 - m_2^2}{2t^{\frac{1}{2}}},$$

$$k = [(s - (m_1 + m_2)^2)(s - (m_1 - m_2)^2)]^{\frac{1}{2}}/2s^{\frac{1}{2}},$$
$$k' = [(s - (m_3 + m_4)^2)(s - (m_3 - m_4)^2)]^{\frac{1}{2}}/2s^{\frac{1}{2}},$$
$$q = [(t - (m_1 - m_3)^2)(t - (m_1 - m_3)^2)]^{\frac{1}{2}}/2t^{\frac{1}{2}},$$
$$q' = [(t - (m_2 + m_4)^2)(t - (m_2 - m_4)^2)]^{\frac{1}{2}}/2t^{\frac{1}{2}}.$$

For the complex Lorentz transformation $\mathbf{A}$ Eq. (II.20), we will show that $\cos\dot{\omega}$ is real and $|\cos\omega| < 1$ in the physical region for the direct channel. Note that k_1^0 and k_3^0 are given by Eq. (A.9) in the formula for $\cosh\chi$ in Eq. (A.6).

Using Eqs. (A8), (A6), and (A9) we obtain for $\cos\omega_1$, for example,

$$\cos\omega_1 = \frac{k_1^0[t - m_1^2 - m_3^2] + 2m_1^2 k_3^0}{k([t - (m_1 + m_3)^2][t - (m_1 - m_3)^2])^{\frac{1}{2}}}. \quad (A10)$$

[19] T. W. B. Kibble, Phys. Rev. 117, 1159 (1960).

Evaluating t in the physical region for the direct channel Eq. (II.3) $t \leq t_{\min}(s) < (m_1 - m_3)^2$ we see that the numerator is real and the argument of the square root is real and positive. Expressing t in terms of c.m. quantities for the direct channel we obtain

$$\cos \omega_1 = \frac{k'k_1^0 z - k_3^0 k}{[(k_1^0 k_3^0 - kk'z)^2 - m_1^2 m_3^2]^{\frac{1}{2}}}, \qquad (A11)$$

where

$$z = \cos \theta_{kk'} \qquad (-1 \leq z \leq 1)$$

and

$$4[(k_1^0 k_3^0 - kk'z)^2 - m_1^2 m_3^2]$$
$$= [t - (m_1 - m_3)^2][t - (m_1 - m_3)^2] > 0.$$

It is easier at this stage to work with the square of Eq. (A11)

$$\cos^2 \omega_1 = \frac{(k_1^0 k'z)^2 + (k_3^0 k)^2 - 2k_1^0 k_3^0 kk'z}{(k_1^0 k_3^0)^2 + k^2 k'^2 z^2 - 2k_1^0 k_3^0 kk'z - m_1^2 m_3^2}. \qquad (A12)$$

Using $(k_1^0)^2 = m_1^2 + k^2$ and $(k_3^0)^2 = m_3^2 + k'^2$, we see that

$$\cos^2 \omega_1 = \frac{k^2 k'^2 z^2 + k^2 k'^2 + m_1^2 k'^2 z^2 + m_3^2 k^2 - 2k_1^0 k_3^0 kk'z}{k^2 k'^2 z^2 + k^2 k'^2 + m_1^2 k'^2 + m_3^2 k^2 - 2k_1^0 k_3^0 kk'z}, \qquad (A13)$$

which clearly satisfies

$$1 \geq \cos^2 \omega_1 \geq 0, \qquad (A14)$$

and thus $\cos \omega_1$ in Eq. (A11) satisfies

$$-1 \leq \cos \omega_1 \leq 1. \qquad (A14')$$

A similar demonstration can be carried out for the other angles. Thus $\cos \omega$ is the cosine of a real angle. The sine of the angle is given by $\sin \omega = (1 - \cos^2 \omega)^{\frac{1}{2}}$ where the sine of the square root is taken to be positive for $-1 \leq \cos \omega \leq 1$. The angle is given by $\exp i\omega = \cos \omega + i \sin \omega$. There is still an arbitrary multiple of 2π that can be added to the angle with the above definition. However, the most that this multiple of 2π can contribute to the amplitude is a factor of a minus sign with no observable effect.

Let us apply the crossing relation Eq. (II.24) with angles given by Eqs. (A8) and (A9) to express the πN helicity amplitudes in terms of the analytic continuation of the Frazer–Fulco amplitudes.[20]

[20] W. R. Frazer and J. R. Fulco, Phys. Rev. **117**, 1609 (1960).

Applying Eq. (II.24) and Eq. (A8) we obtain

$$H_{\lambda'\lambda}^s = \sum_{\mu\mu'} d_{\lambda'\mu'}^{\frac{1}{2}}(\omega_3) H_{\mu'\mu}^t d_{\mu\lambda}^{\frac{1}{2}}(-\omega_1), \qquad (A15)$$

where

$$\cos \omega_3 = -(2E \sin \tfrac{1}{2}\theta_{kk'})(4m^2 - t)^{-\frac{1}{2}},$$

and

$$\cos \omega_1 = (2E \sin \tfrac{1}{2}\theta_{kk'})(4m^2 - t)^{-\frac{1}{2}}. \qquad (A16)$$

Particles 1 and 3 are the fermions. E is the final and initial energy of the nucleon in the πN c.m. t is the momentum transfer for the πN channel; $t = -2k^2(1 - \cos \theta_{kk'})$. In this case the angles ω_1 and ω_3 are supplementary $\omega_1 + \omega_3 = \pi$. The amplitude $H_{\lambda'\lambda}^s$ is the πN helicity amplitude, and $H_{\mu'\mu}^t$ is the analytic continuation of the helicity amplitude (Frazer–Fulco amplitude) here $\mu = $ helicity of antinucleon, $\mu' = $ helicity of nucleon.

After expansion of Eq. (A15) and collection of terms we find

$$H_{++}^s = \sin \omega_1 H_{++}^t - \cos \omega_1 H_{+-}^t, \qquad (A17)$$
$$H_{+-}^s = \cos \omega_1 H_{++}^t + \sin \omega_1 H_{+-}^t.$$

If we express the $H_{\lambda'\lambda}^s$ and $H_{\mu'\mu}^t$ in terms of the usual scaler amplitudes A and B for N scattering and eliminate A and B we find the same relation as above, Eq. (A17) where

$$\cos \omega_1 = (2E \sin \tfrac{1}{2}\theta_{kk'})(4m^2 - t)^{-\frac{1}{2}}$$

and

$$\sin \omega_1 = (1 - \cos^2 \omega_1)^{\frac{1}{2}}$$
$$= (2m \cos \tfrac{1}{2}\theta_{kk'})(4m^2 - t)^{-\frac{1}{2}}.$$

We can see the rotational character of the crossing matrices by using vector notation in the following manner

$$\begin{bmatrix} H_{++}^s \\ H_{+-}^s \end{bmatrix} = \begin{bmatrix} \sin \omega_1 & -\cos \omega_1 \\ \cos \omega_1 & \sin \omega_1 \end{bmatrix} \begin{bmatrix} H_{++}^t \\ H_{+-}^t \end{bmatrix}. \qquad (A18')$$

The crossing relations have also been considered by Barut[21]; however, the crossing relations for the special frame (c.m.) helicity amplitudes are not worked out.

[21] A. O. Barut, Phys. Rev. **130**, 436 (1962).

Asymptotic Behavior of Feynman Integrals with Spin*

J. C. POLKINGHORNE

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England
(Received 21 April 1964)

Some general features are investigated of the dependence of the asymptotic behavior of Feynman integrals upon factors in the numerator of the integrand resulting from particles with spin. These results are used to analyze the high-energy behavior of ladder diagrams for spin-½ nucleons interacting with neutral vector mesons. The leading contribution is shown to consist of terms corresponding to a reggeised nucleon together with certain other terms. The expected cancellation of these other terms by terms associated with a well-defined class of crossed diagrams is verified in detail for the sixth-order case. Finally, other significant diagrams, different from the ladders and their associated crossed diagrams, are investigated and it is shown that they only provide higher-order corrections to the trajectory of the reggeised nucleon.

1. INTRODUCTION

THE possibility that an interaction with neutral vector mesons has the effect of turning the elementary particle pole associated with a spin-½ nucleon into a Regge pole has been discussed by Gell-Mann, Goldberger, Low, Zachariasen, and collaborators[1]. A great part of their analysis was concerned with the behavior of perturbation theory integrals, a method of investigation which, though not rigorous, has had a fruitful heuristic influence on relativistic theories and which they refer to as a "laboratory" in which to test the plausibility of such notions.

The literature developing the theory of the high-energy behavior of Feynman integrals[2-6] has so far been concerned with the simpler case of spinless particles. In Sec. 2 we investigate the effect of extra factors in the numerator of the Feynman integrand such as occur in the case of particles with spin. Some simple examples illustrate typical effects which are given the names of singularity and displacement contributions. This section provides some of the basic "apparatus" for the Feynman integral "laboratory."

These ideas are applied in Sec. 3 to analyze the asymptotic behavior of ladder diagrams for spin-½ nucleons in interaction with neutral spin-1 mesons. The behavior turns out to be somewhat more complicated than was indicated in C but finally it reduces to the term required for the reggeization of the nucleon plus other terms, in general involving higher powers of ln t, which are associated with an effect called a cancellation contribution.

These unwanted cancellation contributions will also occur, but with opposite sign, in a series of diagrams obtained from the ladders by crossing meson lines. It is also necessary to verify that these crossed diagrams do not provide any further unwanted and uncanceled terms. This is done in Sec. 4 for the sixth-order diagrams, both for ln$^3 t$ and ln$^2 t$ terms, and this completes the Reggeization program to order g^6. It is highly plausible that this continues to hold in higher orders but a notation sufficiently succinct to deal successfully with the complications of the general case has not yet been devised.

N-particle intermediate states lead in a spinless theory to Regge poles tending to $l = -N + 1$. If in a theory $N - 1$ of these particles can have spin 1 it is natural to suppose that the well-known translation effect of spin[7] will produce an effect associated with a Regge pole tending to $l = 0$. Thus, it is clear that not only the ladders and the corresponding crossed diagrams must be considered in order to investigate reggeization. In Sec. 5 other significant diagrams are considered. The simple case of $N = 3$ is discussed, although similar considerations would hold for higher values of N. It is shown that these diagrams just correspond to a g^4 term in the trajectory function of the reggeized nucleon. It is important to verify that this is the case for

[1] M. Gell-Mann and M. L. Goldberger, Phys. Rev. Letters 9, 275 (1962); M. Gell-Mann, M. L. Goldberger, F. E. Low, and F. Zachariasen, Phys. Letters 4, 265 (1963); M. Gell-Mann, M. L. Goldberger, F. E. Low, E. Marx, and F. Zachariasen, Phys. Rev. 133, B 145 (1964), referred to as C; M. Gell-Mann, M. L. Goldberger, F. E. Low, V. Singh, and F. Zachariasen, *ibid.* 133, B 161 (1964).
[2] J. C. Polkinghorne, J. Math. Phys. 4, 503, 1393 (1963).
[3] P. G. Federbush and M. T. Grisaru, Ann. Phys. (N. Y.) 22, 263, 299 (1963).
[4] I. G. Halliday, Nuovo Cimento 30, 177 (1963).
[5] G. Tiktopoulos, Phys. Rev. 131, 480, 2373 (1963).
[6] J. D. Bjorken and T. T. Wu, Phys. Rev. 130, 2566 (1963); T. L. Trueman and T. Yao, *ibid.* 132, 2741 (1963); J. C. Polkinghorne, J. Math. Phys. 5, 431 (1964).

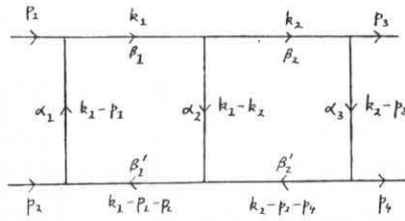[7] Ya. I. Azimov, Phys. Letters 3, 195 (1963).

FIG. 1. The basic diagram, labeling Feynman parameters and momenta.

if these terms corresponded to a *separate* Regge pole, then the corresponding zero power of $\ln t$ in its expansion would be absent and this would have the effect of restoring the Born approximation and spoiling the reggeization result.

2. ASYMPTOTIC BEHAVIOR OF FEYNMAN INTEGRALS

The presence of factors in the numerator of a Feynman integral may affect the asymptotic behavior in three ways.

(i) Explicit powers of t, the asymptotic variable, may appear. They may arise from external momenta present in the original numerator or they may arise from displacements of the origin of the internal momenta which are necessary for symmetric integration.[8]

This effect is crucial for the possibility of Reggeization. A Feynman integral with numerator unity can at most produce asymptotic behavior of the form $t^{-1} \ln^n t$. In order to get a Regge pole associated with $l = 0$ we need a factor of t from the numerator to convert this asymptotic form into $\ln^n t$. The occurrence of such factors in theories of particles with spin is just the translation into Feynman integral terms of the familiar effect of spin producing a shift to the right in the angular momentum plane.[7]

(ii) The presence of internal momentum factors in the numerator produces terms with a decreased power of the denominator after symmetric integration has been performed. If this power is equal to, or less than, the length of the minimal d-lines[4,5] then enhanced asymptotic behavior is obtained.

This effect proves unimportant for reggeization for it only occurs for terms lacking the crucial t factor.

(iii) The presence of internal momentum factors in the numerator may enlarge the class of singular configurations which exist. These singular configurations were first discussed by Tiktopoulos[5] for the

case of a ϕ^3 interaction. He pointed out that if two lines could be added to a d line to form a triangular loop then scaling this enlarged set of lines would also enhance the asymptotic behavior. Thus is because the power of C in the numerator after symmetric integration has been performed is 2 less than the power of D in the denominator. Thus, although the addition of two extra lines adds a factor ρ^2 in the numerator (where ρ is the scaling parameter), this is canceled by an extra factor of ρ^2 in the denominator, since C and D both vanish like ρ because the α's round the closed triangular loop are proportonal to ρ.

The presence of a factor $k_i \cdot k_j$ in the numerator, where k_i and k_j are the momenta around the ith and jth loops respectively, produces after symmetric integration a term with an additional factor

$$[AdjA]_{ij}/C, \tag{2.1}$$

where A is the matrix of the quadratic form in the loop momenta. If the α's round the kth loop are scaled by ρ then both numerator and denominator in (2.1) vanish like ρ unless $i = j = k$ in which case only C vanishes. In this latter case, therefore, a singular configuration would occur if *three* lines could be added to a d line to close the kth loop.

It will be useful to illustrate the operation of these effects by some simple examples. We shall consider integrals whose denominator corresponds to Fig. 1 with all the particles spinless and see the effect of various factors in the numerator:

(a) $(k_1^2 + m^2)$. This cancels the propagator of the line whole parameter is β_1, making the first loop triangular and giving a singular configuration. Then the leading asymptotic behavior is $t^{-1} \ln^3 t$.

If we had not noticed this cancellation, but gone straight ahead with symmetric integration we should have obtained the leading result in a different way as the sum of two terms. One of these terms corresponds precisely to the effect (iii). The other arises from the effect (i) due to displacement terms in k_1^2. These latter include a term with the additional factor

$$2p_1 \cdot p_3 \cdot \alpha_1\alpha_2\alpha_3 C_2/C^2, \tag{2.2}$$

where C_2 is the C function for the second loop. The presence of $\alpha_1\alpha_2\alpha_3$ in (2.2) means that the natural asymptotic behavior $t^{-1} \ln^2 t$ associated with the denominator is depressed to t^{-2} times some power of $\ln t$. However, the lost power of t is recouped by the presence in (2.2) of $2p_1 \cdot p_3 \sim t$. The net result is a term asymptotic like $t^{-1} \ln^3 t$, the extra power of the logarithm being due to the $C_2 C^{-2}$ factor

---

[8] For an account of symmetric integration see Appendix A5 of J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1956).

in (2.2). The precise calculation of terms of this type is greatly facilitated by the Mellin transform method.[6] A brief account of this method is given in the Appendix together with its application to (2.2).

These two ways of producing the $t^{-1} \ln^3 t$ behavior in this case are illustrative of general effects which we shall encounter many times and which will therefore require names. A term produced by the effect (i) will be called a *displacement contribution* and a term produced by (iii) will be called a *singularity contribution*.

(b) $(k_1 - k_2)^2 + m^2$. This cancels the propagator of the line corresponding to α_2, reducing the diagram to the product of two vertex parts whose exact asymptotic behavior is t^0.

If symmetric integration is performed without noticing the cancellation the t^0 term is obtained as a displacement contribution. The $-2k_1 \cdot k_2$ term gives a contribution to the numerator which is

$$-2p_1 \cdot p_2 \cdot \alpha_1 \alpha_3 C_1 C_2 / C^2. \qquad (2.3)$$

The α-dependent factors in (2.3) give an asymptotic behavior of t^{-1} which is converted by the $2p_1 \cdot p_2$ factor into t^0. It is also possible to verify that the $t^{-1} \ln^3 t$, etc., terms cancel among themselves.

These examples show that it is often an economical way of calculating to group together numerator factors in such a way that they just cancel certain propagators in the denominator. The resulting effect upon asymptotic behavior can then often be read off immediately without the need for detailed calculation. This trick proves particularly useful in Reggeization problems and we shall call the terms with enhanced asymptotic behavior obtained in this way *cancellation contributions*. They contain, of course, sums of displacement and singularity contributions.

3. LADDER DIAGRAMS

In this section we consider the ladder diagrams, of the form of Fig. 2, for the interaction of spin-$\frac{1}{2}$ nucleons with neutral spin-1 mesons with coupling constant g. The external mesons may also be spin 1 or they may, for example, be pseudoscalar mesons. We denote their interaction vertex by Γ. If they are spin 1, the special gauge for Γ given in C must be chosen. If they are pseudoscalar mesons then Γ is just γ_5.

The numerator of the Feynman integral is of the form

$$\gamma_\alpha[-i\gamma(p_2 - k_n) + m]\Gamma[-i\gamma(p - k_n) + m]$$
$$\times \gamma_\beta[-i\gamma(p - k_{n-1} - k_n) + m]$$
$$\times \gamma_\alpha \cdots \Gamma[-i\gamma(p_1 - k) + m]\gamma_\omega \qquad (3.1)$$
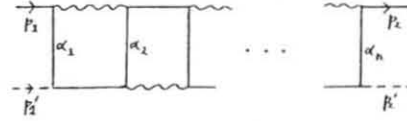

FIG. 2. The $2n$th-order ladder diagram.

evaluated between free-field nucleon spinors, with

$$p = p_1 + p_1' = p_2 + p_2'. \qquad (3.2)$$

The k_i are the loop momenta which are to be integrated over. In order to find the high-energy behavior associated with such a diagram it is necessary to perform certain manipulations to turn the effect of the γ matrices into the formation of scalar products of momenta. It is convenient to have a succinct notation to denote the momentum in a given line of the diagram. We shall use the capital letter of the Feynman parameter associated with the line.

The procedure for manipulating (3.1) is suggested by that discussed in C, although we must keep many more terms. We first move to the right the term $i\gamma A_1$ so that the $i\gamma p_1$ which it contains, both explicitly and also implicitly in the $-i\gamma k_1$ term by displacement, acts on $u(p_1)$ to give $-m$. The $i\gamma p_2$ also implicitly present in $i\gamma k_1$ is easily seen to be negligible since it appears multiplied by $\alpha_2 \cdots \alpha_{n+1}$. At the same time we move $i\gamma A_{n+1}$ to the left. Each anticommutator with a γ factor corresponding to a meson vertex (other than a Γ) has the effect of pulling $i\gamma A_1$ further back to the left, or $i\gamma A_{n+1}$ further back to the right, so that this manipulation is lengthy. When it is completed we have a sum of many terms. These terms fall into four groups:

(i) Those containing no scalar products of A_1 or A_{n+1} with other momenta. Neither p_1 nor p_2 appear explicitly in these terms.

(ii) Terms involving the scalar product of A_1 with another momentum, X, but not involving a scalar product of A_{n+1}. Only p_1 appears explicitly in these terms. Similarly there are terms with the role of A_1 and A_{n+1} interchanged which only involve p_2 explicitly.

(iii) A term with $A_1 \cdot A_{n+1}$. This is the only term considered in C.

(iv) Terms involving the scalar product of A_1 with a momentum X and the scalar product of A_{n+1} with a momentum X'. The line carrying the momentum X must lie to the left of the line carrying the momentum X' reading along the nucleon line in Fig. 2.

In addition to the scalar products explicitly men-

tioned in (i)–(iv), there are in each term $(\pm i\gamma X''+m)$ factors associated with each nucleon line other than A_1, A_{n+1}, X, and X'. The sign is determined by the anticommutations necessary to form the term and is readily determined by inspection. There are also the two external meson vertex factors Γ which are always manipulated to the extreme right and left, respectively. Finally, there may be pairs of internal meson vertex factors $\gamma_\nu \cdots \gamma_\nu$. They must occur to the left (right) of $X(X')$ in (ii) and (iv) and cannot occur in (iii). If these latter factors are present it is necessary to reduce the terms further by moving the implicit factors of $i\gamma p_1$ ($i\gamma p_2$), arising in internal lines by displacement, to the right (left) of these γ matrices.

We must now consider how the scalar products formed in this way affect the asymptotic behavior:

(a) There may be an explicit factor $p_1 \cdot p_2 \sim \frac{1}{2}t$. This comes solely from (iii).

(b) A factor $p_1 \cdot k_i$ can produce a factor of t by displacement. In considering the effect of a factor we write down just the extra terms which would appear in the Mellin transform due to the presence of the factor (see the Appendix). In this case they are

$$(-)^{n+1} \frac{C_{1\ldots j-1}\alpha_{j+1} \cdots \alpha_{n+1}}{C} p_1 \cdot p_2, \quad (3.3)$$

where $C_{l_1,\ldots,l_2}$ is the C function of the loops $l_1 \cdots l_2$. We are only concerned with the $\ln^n t$ and higher terms in $2(n+1)$th order. The presence of the α's in (3.3) depresses the natural asymptotic contribution and means that the term is only important if associated with other factors which have a compensatingly enhancing effect.

Factors of $p_2 \cdot k_i$ produce similar effects but the combination of both types of factor leads to a different result.

(c) A factor of $(p_1 \cdot k_i)(p_2 \cdot k_i)$ can produce significant terms in two ways. It yields a singularity contribution

$$\tfrac{1}{2}(p_1 \cdot p_2)[C_{1\ldots i-1} \cdot \alpha_{i+1} \cdots \alpha_i \cdot C_{i+1\ldots n}]C^{-1}, \quad (3.4)$$

if $i < j$, and a similar expression if $j < i$. The presence of the α's in (3.4) means that the term is unimportant (unless enhanced by other factors) if $i \neq j$. If $i = j$ there are no α's and we obtain

$$\tfrac{1}{2}(p_1 \cdot p_2)[C_{1\ldots i-1} \cdot C_{i+1\ldots n}]C^{-1}, \quad (3.5)$$

which gives $\ln^{n+1} t$ asymptotic behavior.

The displacement contribution gives

$$(-)^{n+1}(p_1 \cdot p_2)^2 \cdot [C_{1\ldots i-1} \cdot \alpha_{i+1} \cdots \alpha_{n+1}]$$
$$\times [\alpha_i \cdots \alpha_j C_{j+1\ldots n}] \cdot C^{-2}. \quad (3.6)$$

The terms we are considering arise only from (iv), in which case $i \leq j$. If $i = j$ then the term is

$$(-)^{n+1}(p_1 \cdot p_2)^2\alpha_1 \cdots \alpha_n \cdot [C_{1\ldots i-1} \cdot C_{i+1\ldots n}]C^{-2}, \quad (3.7)$$

whose α-dependent factors give a multiple pole of the Mellin transform at $\beta = -2$ of identical structure to that given at $\beta = -1$ by the term (3.5). If n is even the asymptotic variable is u, while if n is odd the asymptotic variable is t. Also $p_1 \cdot p_2 \sim \frac{1}{2}t \sim -\frac{1}{2}u$. We thus see that the contributions of (3.5) and (3.7) exactly cancel for all n, since the Mellin transform is with respect to $-t(-u)$.

If $i < j$ then higher powers of some α's appear in (3.6) and the contribution is only significant if it is enhanced by other factors.

(d) A factor of $k_i \cdot k_j$ ($i < j$) yields a singularity contribution

$$2 \cdot [C_{1\ldots i-1} \cdot \alpha_{i+1} \cdots \alpha_i C_{j+1\ldots n}]C^{-1}, \quad (3.8)$$

which to be significant needs both a t and an enhancement from other factors. The corresponding displacement contribution is

$$(-)^{n+1}(p_1 \cdot p_2)\{[C_{1\ldots i-1} \cdot \alpha_{i+1} \cdots \alpha_{n+1}]$$
$$\times [\alpha_1 \cdots \alpha_i \cdot C_{i+1\ldots n}]C^{-2} + i \leftrightarrow j\}. \quad (3.9)$$

The first of the two terms in (3.9) yields by itself an asymptotic behavior $(\ln t)^{i-i-1}$. If such a term could be combined with, say, a $p_1 \cdot p_2$ factor it would lead to a $t(\ln t)^{i-i-1}$ behavior. However, it is not difficult to see that this can never happen owing to the structure of the numerator of ladder diagrams. The terms we are considering arises from the implicit presence of p_1 in k_i and p_2 in k_j due to displacement. In the original numerator $i\gamma \cdot k_i$ terms appear always to the right of $i\gamma \cdot k_j$ terms if $i < j$. In our prescription for reducing the denominator, the necessary scalar product can only result from the 'pulling back' of the implicit $i\gamma \cdot p_1$ and $i\gamma \cdot p_2$ terms by anticommutators with the meson vertices. The only scalar products that can be formed in this way can in fact never be combined with factors that do any thing more than enhance the power of the logarithm. The first significant term of this type arises from the $k_1 \cdot k_5$ factor in (iii) for the tenth-order diagram which can in fact be enhanced to give a $\ln^4 t$ behavior.

The second term in (3.9) can be combined with a further $p_1 \cdot p_2$ factor because the implicit $i\gamma p_1$ appears to the left of the implicit $i\gamma p_2$ factor in the numerator. However, the additional α's present in the second term mean that it can at most give a power of $\ln t$ when so combined and this is such

that it would only be significant if there were enhancement due to further factors.

A factor of k_1^2 yields a singularity contribution

$$2 \cdot [C_1 \ldots_{i-1} \cdot C_{i+1} \ldots_n] C^{-1} \qquad (3.10)$$

and a displacement contribution

$$2(-)^{n+1} (p_1 \cdot p_2) \alpha_1 \cdots \alpha_n [C_1 \ldots_{i-1} \cdot C_{i+1} \ldots_n] C^{-2}. \qquad (3.11)$$

Unlike (3.5) and (3.7) these do not cancel.

In (3.10) and (3.11) the extra significant scaling round the ith loop makes them capable of combining with other factors to give an enhancement. For example with $p_1 \cdot p_2$ they give a behavior of $\ln^{n+1} t$ and with $p_1 \cdot k_n$ they give a behavior of $\ln^n t$ provided $i \neq n$.

The results obtained in (a)–(d) may be summarised as follows. Only the $p_1 \cdot p_2$ factor produces a contribution which is significant by itself. All other factors need further enhancement before they become significant. The only factors capable of enhancing them are k_i^2 factors.

The next step in the analysis of ladder diagrams is to note that all the enhancement due to k_i^2 factors can be expressed in the form of cancellation contributions. This then leads to the expectation that these contributions are cancelled by similar contributions from diagrams with crossed lines.

The argument leading to the cancellation contributions is sufficiently illustrated by considering the example of the sixth-order ladder diagram ($n = 3$). This also provides a specific example of how the general discussion of (a)–(d) works out in practice.

The numerator can be manipulated into the form

$$-8(p_1 - k_1) \cdot (p_2 + k_2) \Gamma[-i\gamma(p + k_2) + m]$$
$$\times [i\gamma(p - k_1 + k_2) + m][-i\gamma(p - k_1) + m]\Gamma$$
$$+ 8(p_1 \cdot k_1)(p_2 \cdot k_1) \Gamma[-i\gamma(p + k_2) + m]\Gamma$$
$$+ 8(p_1 \cdot k_2)(p_2 \cdot k_2) \Gamma[-i\gamma(p - k_1) + m]\Gamma + \cdots, \qquad (3.12)$$

where all the omitted terms are trivially not significant. By the argument of (c), the second and third terms of (3.12) are not significant. In the first term of (3.12) we may use the manipulation given in C to rewrite the expression between Γ factors as

$$[(p + k_2)^2 + m^2][-i\gamma(p - k_1) + m]$$
$$+ [-i\gamma(p + k_2) + m][(p - k_1)^2 + m^2]$$
$$+ [-i\gamma(p + k_2) + m][-i\gamma p - m]$$
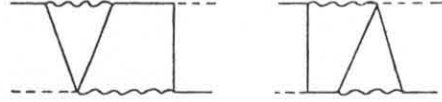$$\times [-i\gamma(p - k_1) + m]. \qquad (3.13)$$

FIG. 3. The reduced diagrams associated with the sixth-order ladder diagram.

The k_1^2 and k_2^2 terms which are capable of giving enhancement are now isolated in a form where they are transparently associated with cancellation contributions.

Thus, apart from the desired term

$$+8(p_1 \cdot p_2) \cdot \Gamma[-i\gamma(p + k_2) + m]$$
$$\times [i\gamma p + m][-i\gamma(p - k_1) + m]\Gamma, \qquad (3.14)$$

all significant contributions are given by cancellation contributions which can be associated with the reduced diagrams of Fig. 3. In each of these diagrams there is a triangular loop which has two meson lines attached to one of its vertices. Similar terms would occur in the reduction of the diagram Fig. 4 and its reflection, in which the corresponding two meson lines are attached to the canceled nucleon lines in the reversed order. However, the occurrence of an extra anticommutation gives a change of sign which produces a cancellation between the sixth-order ladder and these crossed diagrams. This is discussed in detail in the next section.

4. CROSSED DIAGRAMS

The analysis of crossed diagrams is considerably more complicated than that of ladder diagrams. In this paper we only attempt to give a complete discussion of the sixth-order diagram of Fig. 4. The coefficient of t in the denominator associated with this diagram is

$$\alpha_1 G, \qquad (4.1)$$

where

$$G = \alpha_2 \alpha_3 - \beta_2 \beta_2'. \qquad (4.2)$$

If there were no numerator factors this would give a leading asymptotic behavior of t^{-1} with a correction term of order $t^{-2} \ln^3 t$. We shall find that the significant contributions in our problem arise from factors
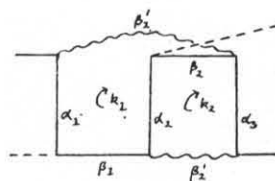

FIG. 4. One of the crossed diagrams associated with the sixth-order ladder diagram, labeling Feynman parameters and loop momenta.

in the numerator which contain an α_1, which depresses the leading contribution, and a t^2, which makes the correction term significant.

In the reduction of the Feynman integral corresponding to Fig. 4 we follow the same procedure used in Sec. 3, that is to say we move $i\gamma A_1$ to the right and $i\gamma A_3$ to the left. This is not, however, quite the most appropriate procedure. The displacement contributions of the two loop momenta are

$$k_1 \sim \alpha_1 C_2 C^{-1} p_1 - G C^{-1} p_2, \qquad (4.3)$$

$$k_2 \sim \alpha_1 (\alpha_2 + \beta_2) C^{-1} p_1 - H C^{-1} p_2, \qquad (4.4)$$

where C_1 and C_2 are the C functions of the two loops and

$$H = (\alpha_3 + \beta_2) C_1 - \beta_2 (\alpha_2 + \beta_2). \qquad (4.5)$$

The G in (4.3) makes the $i\gamma p_2$ implicit in $i\gamma k_1$ negligible but the $i\gamma p_1$ implicit in $i\gamma k_2$ cannot be neglected and must be moved over to the right. However, we find it convenient to use the same reduction procedure as in Sec. 3, to facilitate comparison, and then consider the extra terms arising from moving this $i\gamma p_1$ factor separately.

No significant terms can arise from singularity contributions. The α_1 in (4.1) is associated with a five-line loop; scaling G already involves putting all the parameters round the second loop equal to zero and is in any case only relevant to the correction term, which needs a t^2 to be significant whilst singularity contributions can only be associated with a t.

When displacement contributions are considered the following terms lead to $\ln^3 t$ contributions:

$$(p_1 \cdot p_2) \cdot (p_2 \cdot k_1), \quad (p_1 \cdot p_2) \cdot (p_2 \cdot k_2),$$

$$(p_1 \cdot p_2) \cdot (k_1 \cdot k_2), \quad (p_1 \cdot k_2) \cdot (p_2 \cdot k_1), \qquad (4.6)$$

$$(p_1 \cdot p_2) \cdot k_2^2, \quad (p_1 \cdot k_2) \cdot (p_2 \cdot k_2);$$

while the following terms lead to $\ln^2 t$ contributions:

$$(p_2 \cdot k_1) \cdot (p_2 \cdot k_2), \quad (p_2 \cdot k_1) \cdot (p_2 \cdot k_1),$$

$$(p_2 \cdot k_2) \cdot (k_1 \cdot k_2), \quad (p_2 \cdot k_1) \cdot k_2^2, \qquad (4.7)$$

$$(p_2 \cdot k_1) \cdot (k_1 \cdot k_2).$$

Equations (4.6) and (4.7) list the only significant terms which actually arise in the reduction. It is important to notice that the appearance of G in (4.3) means that terms involving $(p_1 \cdot k_1)$ or k_1^2 are never significant.

The terms that arise in the reduction may be classified as follows:

(i) A term

$$8(p_1 - k_1) \cdot (p_2 + k_2) \Gamma [-i\gamma (-k_1 + p) + m]$$

$$\times [i\gamma (-k_1 + k_2 + p) + m]$$

$$\times [-i\gamma (-k_1 + k_2 + p_2) + m] \Gamma, \qquad (4.8)$$

corresponding to (iii) in Sec. 3. This can be manipulated into the form

$$8(p_1 - k_1) \cdot (p_2 + k_2) \Gamma [-i\gamma (-k_1 + p) + m] \Gamma$$

$$\times \{ [(-k_1 + k_2 + p_2) + m^2]$$

$$- 2p_2 \cdot (-k_1 + k_2 + p_2) \} + \cdots, \qquad (4.9)$$

where all the omitted terms are not significant. The first term in the curly brackets cancels one of the cancellation contributions in (3.13); the other is canceled by a similar term in the reflected diagram. The second term in the curly brackets must be cancelled by other terms associated with Fig. 4. These arise from the second class of contributions we consider.

(ii) Two of the terms corresponding to (iv) of Sec. 3 are

$$16[(p_1 - k_1) \cdot (-k_1 + k_2 + p)] \cdot [(p_2 + k_2)$$

$$\times (-k_1 + k_2 + p)] \Gamma [-i\gamma (-k_1 + p) + m] \Gamma, \qquad (4.10)$$

and

$$-16[(p_1 - k_1) \cdot (-k_1 + k_2 + p)] \cdot [(p_2 + k_2)$$

$$\times (-k_1 + k_2 + p_2)] \Gamma [-i\gamma (-k_1 + p) + m] \Gamma. \qquad (4.11)$$

These two terms together cancel the significant contributions from the second term of (4.7). To see this it is necessary to note that the significant contribution from $(p_1 \cdot p_2) \cdot k_2^2$ is equal to twice that from $(p_1 \cdot k_2) \cdot (p_2 \cdot k_2)$; the significant contribution from $(p_1 \cdot k_2) \cdot (p_2 \cdot k_1)$ equals that from $(p_1 \cdot p_2) \cdot (k_1 \cdot k_2)$; and the significant contribution from $(p_2 \cdot k_1) \cdot k_2^2$ is twice that from $(p_2 \cdot k_2) \cdot (k_1 \cdot k_2)$.

(iii) There are two other terms corresponding to (iv) of Sec. 3 but they both have a factor

$$(p_1 - k_1) \cdot (-k_1 + p), \qquad (4.12)$$

which makes them negligible.

Finally, there are the terms corresponding to (i) and (ii) in Sec. 3. The only significant contributions are those associated with the $i\gamma p_1$ implicit in $i\gamma k_2$. When these are collected together they are found to cancel identically.

This completes the verification of reggeization in sixth order. The same pattern is expected to repeat itself in higher order. For example, in eighth
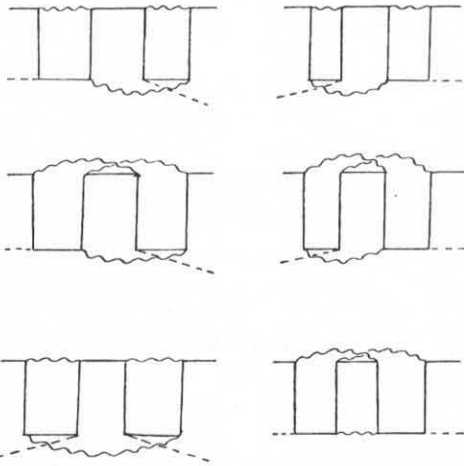
FIG. 5. The crossed diagrams associated with eighth-order ladder diagram.

order the crossed diagrams of Fig. 5 must be considered. However, the number of terms to be taken account of rises sharply with the order and a succinct and powerful notation has not yet been found to handle them.

5. OTHER DIAGRAMS

The diagram of Fig. 6 has a leading asymptotic contribution

$$\Gamma[\alpha_2(s) \ln (-t)/(i\gamma p + m)]\Gamma, \qquad (5.1)$$

with

$$\alpha_2(s) = \frac{g^4(i\gamma p + m)}{2^5 \pi^3} \int_0^1 d\alpha_1 \, d\alpha_2 \, d\alpha_3$$

$$\times \frac{\delta(\alpha_1 + \alpha_2 + \alpha_3 - 1) \cdot [-i\gamma \cdot p\alpha_1 \alpha_2 + mC(\alpha)]}{D(\alpha; s)}, \quad (5.2)$$

where C and D are associated with the contracted diagram Fig. 7. Although a diagram of the type of Fig. 6 would for spinless particles be associated with a Regge pole tending to $l = -2$ this effect is translated to $l = 0$ by the presence of two spin-1 particles in the intermediate state.[7]

If (5.1) is regarded as a term in an expansion of a set of exponentials in powers of $\ln t$ and $\ln (-t)$ then the corresponding zero-order term is just the Born approximation. This Born-approximation term has already been used as the zero-order term in the reggeization program described in C. Thus, if this program is to succeed it is essential that $\alpha_2(s)$ may be considered as a g^4 term in the expansion
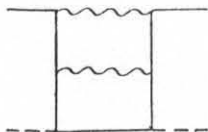

FIG. 7. A contracted diagram associated with Fig. 6.

of the trajectory function associated with the Reggeized nucleon, the g^2 term being the $\alpha(s)$ given in C and considered here in Sec. 3 and 4, and that these terms should not correspond to a separate Regge pole. In order that this should be so it is necessary to exhibit the correct $\alpha(s)^{n_1}\alpha_2(s)^{n_2}$ terms in the expansion of $[\alpha(s) + \alpha_2(s)]^{n_1+n_2}$, which itself arises from the expansion of $(\pm t)^{\pm[\alpha(s)+\alpha_2(s)]}$. The purpose of this section is to outline a proof that this is so for the cases (a) $n_1 = 0$, $n_2 = 2$; (b) $n_1 = 1$, $n_2 = 1$, respectively.

The correct term for case (a) is obtained from the diagram Fig. 8. The manipulation follows that used for the ladder diagrams but there are now two factors of $i\gamma p_1$ to move to the right and two factors of $i\gamma p_2$ to move to the left. A term involving $(p_1 \cdot p_2)^2$ may be formed, which is what is needed to cancel the t^{-2} factor associated with the de-

FIG. 8. The diagram which gives the iteration of the significant contribution of Fig. 6.
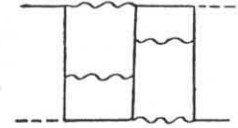
nominator. It is important to realise that the diagram of Fig. 9 is not significant because in its contribution the two factors of $i\gamma p_1$ combine to give p_1^2 without forming a $(p_1 \cdot p_2)^2$ term.

The coefficient of $(p_1 \cdot p_2)^2$ has a numerator containing the terms

$$\Gamma[-i\gamma(p + k_4) + m][i\gamma(p - k_1 + k_4) + m]$$

$$\times [-i\gamma(p - k_1 + k_3) + m][i\gamma(p - k_2 + k_3) + m]$$

$$\times [-i\gamma(p - k_2) + m]\Gamma. \qquad (5.3)$$

If we write

$$p - k_1 + k_4 = (p - k_1 + k_3) - k_3 + k_4, \qquad (5.4)$$
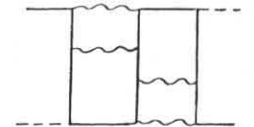$$p - k_2 + k_3 = (p - k_1 + k_3) + k_1 - k_2,$$

then it is possible to extract from (5.3) a term

$$\Gamma[-i\gamma(p + k_4) + m][i\gamma(p - k_2 + k_4) + m]$$

$$\times [-i\gamma(p - k_2) + m][(p - k_1 + k_3)^2 + m^2]\Gamma. \quad (5.5)$$

This is the significant part of (5.3) since it cor-

FIG. 6. Another significant diagram.

FIG. 9. A diagram which does not give a significant contribution.
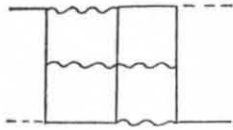
FIG. 10. The effective reduced diagram associated with the term (5.5).

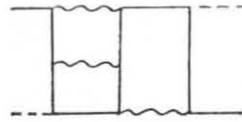responds to a cancellation contribution associated with the effective reduced diagram of Fig. 10 in which there are now *three* 2-lines, giving an asymptotic behavior from the denominator alone of $t^{-2} \ln^2 t$. This cancellation contribution is not canceled by the corresponding crossed diagram since this is just Fig. 9 which is already known not to be significant. The first three factors of (5.5) are now amenable to the type of manipulation given in C and Sec. 3 by writing

$$p - k_2 + k_4 = (p - k_2) + (p + k_4) - p. \qquad (5.6)$$

The first two terms of (5.6) will give cancellation contributions. These are expected to be cancelled by the diagrams of Fig. 11, though this has not been checked in detail. The third term then gives the correct contribution to correspond to (a).

The terms corresponding to (b) are obtained from Fig. 12 and its reflection. There are two factors of $i\gamma p_1$ associated with Fig. 11 and only one factor of $i\gamma p_2$. The desired contribution comes from the displacement contribution associated with the scalar products $(p_1 \cdot p_2) \cdot (p_1 \cdot k_3)$. This gives a Mellin transform proportional to

$$C_{12} C^{-1} \alpha_3 (p_1 \cdot p_2)^2. \qquad (5.7)$$

The α_3 reduces the asymptotic behavior of the denominator to $t^{-2} \ln^2 t$ and the t^{-2} factor is canceled by $(p_1 \cdot p_2)^2$. The remaining spinor factors may be manipulated exactly as in C and Sec. 3 and after cancellation by contributions from the appropriate crossed diagrams the correct coefficient of $\ln^2 t$ remains.
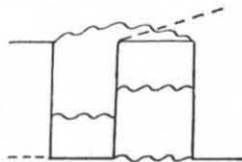
FIG. 11. One of the crossed diagrams associated with Fig. 8.

FIG. 12. The diagram giving the cross-term contribution.

APPENDIX

In this Appendix we summarize some useful results in the Mellin transform method of evaluating asymptotic behavior.[6]

If $\tilde{f}(s, \beta)$ is the Mellin transform of $f(s, -\tau)$ with respect to $\tau = -t$, then

$$\tilde{f}(s, \beta) = N \cdot \Gamma(-\alpha)$$

$$\times \int_0^\infty dx_i \, g(x)^\beta C(x)^{-2-\beta} M(x) e^{-J(s;x)}, \qquad (A1)$$

where x_i are the set of Feynman parameters,

$$DC^{-1} = -[\tau g C^{-1} + J(s, x)], \qquad (A2)$$

C and D are the appropriate Feynman functions, M is the additional numerator term due to spin, N is a constant.

The singularities of $\tilde{f}(s, \beta)$ in β give the asymptotic behavior. A pole of order $m + 1$ at $\beta = -n$ corresponds to a behavior $\tau^{-n} \ln^m \tau$. These singularities arise from the divergence of (A1) at the edge of the region of integration corresponding to a set of x_i equal to zero. The effect of this may be exhibited by introducing a scaling parameter ρ for the set of x_i concerned and integrating by parts with respect to ρ. The singularity for a given value of β may be due to divergencies associated with several distinct sets of x_i. The number of such independent sets gives the order of the pole.

The integrand associated with (2.2) is

$$(\alpha_1 \alpha_2 \alpha_3)^{\beta+1} C_2 C^{-4-\beta} e^{-J}, \qquad (A3)$$

with C and J corresponding to Fig. 1. This is first divergent at $\beta = -2$ and the independent sets are four in number: $\alpha_1, \alpha_2, \beta_1, \beta_1'; \alpha_1; \alpha_2; \alpha_3$. This gives the $t^{-2} \ln^3 t$ behavior discussed in Sec. 3.

Partial Differential Equations with Periodic Coefficients and Bloch Waves in Crystals*

Farouk Odeh

IBM Watson Research Center, Yorktown Heights, New York

AND

Joseph B. Keller

Courant Institute of Mathematical Sciences, New York University, New York, N. Y.

(Received 7 April 1964)

Bloch waves are special solutions of Schrödinger's equation with a periodic real potential. They are plane waves multiplied by periodic functions. In this paper we prove the existence and completeness of Bloch waves and of the related Kohn–Luttinger waves in unbounded domains for a class of partial differential equations which includes the Schrödinger equation. In addition, we discuss the dependence of these waves and the corresponding eigenvalues on the wave vector of the associated plane wave. The results may be interpreted as the analogs for certain partial differential equations of Floquet's theory for ordinary differential equations or as the determination of the spectral representation of certain periodic Hamiltonian operators.

INTRODUCTION

LINEAR differential equations with periodic coefficients often arise in the analysis of periodic structures. For example, the Schrödinger equation for an electron in a crystal is of this type with the spatially periodic potential occurring as a coefficient. In the case of ordinary linear differential equations with periodic coefficients, Floquet's theorem shows that every solution is a linear combination of special solutions, each of which is an exponential function multiplied by a periodic function.[1,2] For partial differential equations a certain corresponding result has been proved by Bloch.[3] It pertains to Bloch waves, which are plane waves multiplied by periodic functions, and they have formed the basis of the theory of electrons in crystals—i.e., of the theory of solids. It is our purpose to prove the existence and completeness of Bloch waves and of the related Kohn–Luttinger waves in unbounded domains for a class of partial differential equations which includes the Schrödinger equation. In addition, we deduce some properties of these waves and the corresponding eigenvalues. In the theory of solids, these results are usually assumed to be true in three dimensions because they have been proved in one dimension with the aid of Floquet's theorem.

Mathematically, our results may be interpreted as the analogs for certain partial differential equations of Floquet's results for ordinary differential equations. They may also be viewed as the determination of the spectral representation associated with certain periodic Hamiltonian operators. The "crystal-momentum" representation of these Hamiltonians, which we consider, is the analog of Fourier analysis for partial differential operators with constant coefficients, and it serves similar purposes.

In Sec. 1, after some preliminary remarks, we define the eigenvalue problem. Section 2 is concerned with proving the existence and completeness of Bloch and Kohn–Luttinger waves and with some of their properties. In Sec. 3, we discuss the dependence of the energy on the wave vector and prove the convergence of what is usually known as the $k \cdot v$ method. We also show that the spectrum is the union of a countable number of intervals, which demonstrates its well-known band structure. In Sec. 4, we conclude with some remarks on Wannier functions and on the representation of the position operator.

1. PRELIMINARIES

The Hamiltonian H governing the motion of an electron in an infinite periodic lattice in three dimensions may be taken to have the form

$$H = -\Delta + V(x). \tag{1.1a}$$

Here, Δ denotes the Laplacian operator, $V(x)$ is the real potential energy, and x is a vector (x_1, x_2, x_3) in the three-dimensional space R_3. If the lattice is invariant under the group generated by some three primitive translations t_1, t_2, t_3, then both V and H will be invariant under the same group. For the sake of clarity we assume that the vectors t_i are mutually orthogonal and directed along the coordinate axes.

[1] E. C. Titchmarch, *Eigenfunction Expansions Associated with Second-Order Differential Equations* (Oxford University Press, London, 1958), Part II.
[2] W. Kohn, Phys. Rev. **115**, 809 (1959).
[3] F. Bloch, Z. Physik **52**, 555 (1928).

The modifications in our analysis necessary to treat a general Bravais lattice are rather obvious.

We consider a more general operator H, invariant under the same group, and defined by

$$H\psi = -\sum_{i,j=1}^{3} \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial \psi}{\partial x_j} \right) + V(x)\psi. \quad (1.1b)$$

We assume that a_{ij} is a real symmetric differentiable matrix which renders H uniformly elliptic and that V is real and continuous, although weaker smoothness conditions would suffice. Under these conditions H, regarded as an operator in $L_2(R_3)$, has a unique self-adjoint extension,[4] which we also denote by H. Its spectral analysis is based on the eigenvalue problem

$$H\psi = \lambda\psi. \quad (1.2)$$

To (1.2) we must add appropriate restrictions on ψ in order to have a well-defined problem. In the one-dimensional case there are no L_2 solutions of (1.2) since all bounded solutions are sums of Bloch waves. Therefore, we seek solutions of (1.2) in the space of bounded continuous functions.

A first step toward the analysis of solutions of (1.2) is provided by the following theorem of Bloch.[3] Consider (1.2) in a rectangular box whose sides have lengths K, L, M which are integer multiples of the primitive translations. Assume that ψ satisfies periodic boundary conditions and that the eigenspace associated with a fixed eigenvalue $\lambda = \lambda_0$ is finite dimensional. Then the eigenfunctions corresponding to λ_0 may be taken to have the form of Bloch waves,

$$\psi_{k,l,m}(x, y, z)$$
$$= \exp\left[2\pi i\left(\frac{kx}{K} + \frac{ly}{L} + \frac{mz}{M} \right) \right] u_{k,l,m}(x, y, z), \quad (1.3)$$

where k, l, m are integers and u is periodic. The proof depends upon a simple group theoretic argument. The corresponding theorem for the whole space does not seem to have been proved, although several authors[5,6] indicate that it is valid. It is often incorrectly stated that Bloch's theorem proves that all bounded solutions of (1.2) are sums of Bloch waves[7]. We prove the existence and completeness of Bloch waves in the whole space.

[4] N. Dunford and J. Schwartz, *Linear Operators, Part II. Spectral Theory* (Interscience Publishers, Inc., New York, 1963).
[5] J. S. Lomont, *Applications of Finite Groups* (Academic Press Inc., New York, 1959).
[6] G. Lyubarskii, *The Application of Group Theory in Physics* (Pergamon Press, Inc., New York, 1960).
[7] N. F. Mott and H. Jones, *The Theory of Metals and Alloys* (Oxford University Press, New York, 1936).

2. EXISTENCE, COMPLETENESS, AND PROPERTIES OF BLOCH WAVES

Let us consider the eigenvalue problem

$$H\psi = \lambda\psi, \quad (2.1)$$

where H is defined by (1.1b) and ψ is bounded in R_3. We suppose that the whole space is decomposed into a countable number of rectangular unit cells with edges equal to the primitive translations, and place the origin of coordinates at a vertex of the "first" unit cell Ω. By using these cells we prove the following lemma, which asserts the existence of Bloch waves.

Lemma 1. For any real vector k there exist a countable number of solutions of (2.1) of the form

$$\psi_n(x, k) = e^{2\pi i k \cdot x} \phi_n(x, k). \quad (2.2)$$

Here, $\phi_n(x, k)$ is a smooth function of x which has the same periodicity as the lattice and $k \cdot x$ denotes the scalar product of the vectors k and x.

Proof: By substituting (2.2) into (2.1), we find that ϕ must satisfy

$$\tilde{H}_k\phi = \sum_{l,m=1}^{3} \left\{ -\frac{\partial}{\partial x_l}\left(a_{lm}\frac{\partial\phi}{\partial x_m} \right) - 4i\pi a_{lm}k_l \frac{\partial\phi}{\partial x_m} \right.$$
$$\left. - 2i\pi\phi k_l \frac{\partial a_{lm}}{\partial x_m} + 4\pi^2\phi a_{lm}k_l k_m \right\} + V\phi = \lambda\phi. \quad (2.3)$$

We now consider the eigenvalue problem (2.3) in the first cell only and impose periodic boundary conditions on ϕ. It is then easy to check that $\tilde{H}_k$, with these boundary conditions, defines a symmetric operator in the space of continuously differentiable functions whose first partial derivatives are absolutely continuous. The smoothness conditions which we have imposed on the matrix a_{ij} and on the potential V are sufficient to assure that there is a unique self-adjoint extension of $\tilde{H}_k$ in $L_2(\Omega)$, which we also denote by $\tilde{H}_k$. This operator $\tilde{H}_k$, being a regular uniformly elliptic self-adjoint operator defined in a bounded domain, possesses a discrete set of eigenvalues $\lambda = \lambda_n$, each of finite multiplicity and corresponding eigenfunctions $\phi_n(x, k)$. Moreover, the $\phi_n(x, k)$ are smooth functions of x.[8] By extending each $\phi_n(x, k)$ to the whole space by periodicity, we obtain a solution of the form (2.2), which proves the lemma.

By varying the vector k over the whole k space, we obtain a set $S = \bigcup_k S_k$ of the eigenvalues $S_k = \{\lambda_n(k)\}$ and a corresponding set of Bloch waves. We prove in Sec. 3 that the set S is exactly the spectrum of H, as an operator in L_2, and in this

[8] S. Agmon, Commun. Pure Appl. Math. 15, 119 (1962).

section we show that the Bloch waves are complete in $L_2(R_3)$. These results remain true even when k is restricted to a unit cell, Ω^*, in the reciprocal lattice defined by the three primitive translations k_i which satisfy the relations $t_i \cdot k_j = \delta_{ij}$.

To show that it suffices to restrict k to the cell Ω^*, we note that the set S_k is the totality of the eigenvalues of the problem (2.1) under the "boundary" condition $\psi(k; r + t) = e^{2\pi i k \cdot t} \psi(k; r)$. But this latter problem does not change if k is replaced by $k + K$, where K is a reciprocal lattice vector, i.e., $K = m_i k_i$ where the m_i are integers, Hence, the set S_k and the corresponding set of solutions are invariant under such a translation. Therefore, the eigenvalues and eigenfunctions corresponding to all k in Ω^* are the same as those corresponding to all k in k space.

In order to prove the completeness of these eigenfunctions, which is the main result of this section, we first prove two lemmas. They define a certain transform of $L_2(R_3)$ into $L_2(\Omega)$ and its inverse, which pave the way for the completeness theorem. For the sake of clarity, but without loss of generality, we assume Ω to be a unit cube and then Ω^* is also a unit cube.

Lemma 2. Let $g(x, k) \in L_2(\Omega \times \Omega^*)$ where Ω, Ω^* are the closed unit cells and let

$$f(x, n) = \int_{\Omega^*} g(x, k) e^{-2\pi i k \cdot n} \, dk, \qquad (2.4)$$

where n is a lattice vector. Then the function $\hat{f}(z)$ defined, a.e., by

$$\hat{f}(z) \equiv \hat{f}(x + n) \equiv f(x, n), \qquad x \in \Omega$$

belongs to $L_2(R_3)$. ("a.e." is used throughout this paper as an abbreviation for "almost every" or "almost everywhere.")

Proof: By Tonelli's theorem, $g(x, k) \in L_2(\Omega^*)$ for almost every x in Ω. Hence, the integral in (2.4) exists and by Parseval's equality, it follows that for x, a.e.,

$$\int_{\Omega^*} |g(x, k)|^2 \, dk = \sum_n |f(x, n)|^2. \qquad (2.5)$$

Upon integrating (2.5) over Ω, we get, since $g \in L_2(\Omega \times \Omega^*)$,

$$\infty > \iint_{\Omega \, \Omega^*} |g(x, k)|^2 \, dx \, dk = \int_\Omega \sum_n |f(x, n)|^2 \, dx$$

$$= \sum_n \int_\Omega |f(x, n)|^2 \, dx$$

$$= \sum_n \int_\Omega |\hat{f}(x + n)|^2 \, dx$$

$$= \int_{R_3} |\hat{f}(z)|^2 \, dz.$$

Here we have interchanged summations and integrations using a special version of Fubini's theorem and the fact that the sums and integrals are convergent in L_2. The last inequality proves the lemma.

Lemma 3. If $h(x) \in L_2(R_3)$, n is a lattice vector, $k \in \Omega^*$, and

$$g(x, k) = \sum_n h(x + n) e^{2\pi i k \cdot n}, \qquad x \in \Omega, \qquad (2.6)$$

then g is defined for a.e. x in Ω and belongs to $L_2(\Omega \times \Omega^*)$.

Proof: Let $h_n(x) = h(x + n)$, $x \in \Omega$; then we have

$$\infty > \int_{R_3} |h(x)|^2 \, dx = \sum_n \int_\Omega |h_n(x)|^2 \, dx$$

$$= \int_\Omega \sum_n |h_n(x)|^2 \, dx. \qquad (2.7)$$

Hence, $\sum_n |h_n(x)|^2 \in L_2$ for a.e. $x \in \Omega$ and therefore $g(x, k)$ is defined for a.e. $x \in \Omega$. Parseval's equality, together with (2.7), now proves that $g \in L_2(\Omega \times \Omega^*)$, and this completes the proof of the lemma.

From (2.6) and (2.4) it follows that

$$h(x + n) = \int_{\Omega^*} g(x, k) e^{-2\pi i k \cdot n} \, dk \equiv f(x, n).$$

Thus, h of Lemma 3 and f of Lemma 2 may be identified. Therefore, for any function f in $L_2(R_3)$ we obtain the representation theorem

$$f(x + n) = \int_{\Omega^*} \sum_m f(x + m) e^{2\pi i k \cdot (m-n)} \, dk.$$

This theorem can also be derived as a special case of a representation theorem due to McGarvey.[9]

We are now in a position to prove the completeness theorem. As in Lemma 1, let $\phi_n(x, k)$, $\lambda_n(k)$ denote, respectively, the orthonormal eigenfunctions and eigenvalues of the auxiliary problem $\tilde{H}_k \phi_n = \lambda_n \phi_n$. Let $\psi_n(x, k) = e^{2\pi i k \cdot x} \phi_n(x, k)$. Then we prove

Theorem 1. The set of Bloch waves $B = \{\psi_n(x, k)\}$, where k varies over the whole closed unit cell Ω^* and n ranges over the positive integers, is complete in $L_2(R_3)$.

Proof: We give the proof in one dimension in which case both Ω and Ω^* are the closed interval $[0, 1]$. The proof applies to higher dimensional cases merely with a change in notation.

We first prove that if $f \in L_2(-\infty, +\infty)$, then the scalar product $\langle f, \psi \rangle$ where $\psi \in B$, exists in the sense of mean convergence. Let $\psi = e^{2\pi i k x} \phi$ and consider

[9] D. McGarvey, J. Math. Anal. Appl. **4**, 366 (1962).

the inner product $\langle f, \psi \rangle$ defined by

$$\langle f, \psi \rangle = \lim_{N \to \infty} \int_{-N}^{+N} f^*(x) e^{2\pi i k x} \phi(x, k) \, dx$$

$$= \lim_{N \to \infty} \sum_{j=-N}^{N-1} \int_{j}^{j+1} f^*(x) e^{2\pi i k x} \phi(x, k) \, dx$$

$$= \lim_{N \to \infty} \int_{0}^{1} \sum_{j=-N}^{N-1} f^*(\xi + j) e^{2\pi i k j} e^{2\pi i k \xi} \phi(\xi, k) \, d\xi$$

$$= \int_{0}^{1} \lim_{N \to \infty} \sum_{j=-N}^{N-1} [f^*(\xi + j) e^{2\pi i k j}] e^{2\pi i k \xi} \phi(\xi, k) \, d\xi.$$

Here we have interchanged the limit and the integration, which can be justified with the aid of Lemma 3. The last integral exists since both ϕ and the sum $\sum_{j=-\infty}^{\infty} f^*(\xi + j) e^{2\pi i k j}$ belong to $L_2(\xi)$, for a.e. k. Thus, the scalar product exists.

To prove the completeness of B we must prove that if $\langle f, \psi \rangle = 0$ for all $\psi \in B$, then $f \equiv 0$. From the above equations we have

$$\langle f, \psi_n \rangle$$
$$= \int_{0}^{1} \left\{ \lim_{N \to \infty} \sum_{j=-N}^{N-1} f^*(\xi + j) e^{2\pi i k j} \right\} \psi_n(\xi, k) \, d\xi. \quad (2.8)$$

Let us define $g(\xi, k)$ by

$$g(\xi, k) = \lim_{N \to \infty} \sum_{j=-N}^{N-1} f(\xi + j) e^{2\pi i k j}. \quad (2.9)$$

Then, by Lemma 3, $g \in L_2[0, 1]$ for a.e. k. Since the functions $\psi_n(\xi, k)$ are complete in $L_2(0, 1)$ for every k, the vanishing of $\langle f, \psi_n \rangle$ for fixed k implies, from (2.8), that $g(\xi, k) = 0$ for a.e. ξ. By varying k in $[0, 1]$, we find that $g(\xi, k) = 0$ for a.e. ξ and k. But since

$$f(x + n) = \int_{0}^{1} g(x, k) e^{-2\pi i k n} \, dk,$$

f vanishes identically. Thus the theorem is proved.

Corollary 1. "Momentum representation of L_2-functions". If $f \in L_2$ and

$$f_N(x) = \int_{k} \sum_{n=0}^{N} \psi_n(x, k) \langle \psi_n(x', k), f(x') \rangle \, dk,$$

then $f_N(x)$ converges in the mean to a function $F(x) = f(x)$ a.e.

Proof: It is sufficient to consider x in the interval $I = [0, 1]$ since $F = f$ a.e. in I implies $F = f$ a.e. in $(-\infty, +\infty)$ because of the quasiperiodic property of the basis functions $\psi_n(x, k)$. Let $f_n(k)$ denote the scalar product $\langle \psi_n, f \rangle$. Then using (2.8), and (2.9) and integrating over k, we have

$$\int_{k} \sum_{n} \psi_n(x, k) f_n(k) \, dk = \int_{k} \sum_{n} \psi_n \langle \psi_n, g(x, k) \rangle \, dk.$$

But, since $g(x, k) \in L_2[0, 1]$ for a.e. k and $\{\psi_n(x, k)\}$ is an orthonormal set which is complete in $[0, 1]$ for every k, we have

$$\sum_{n} \psi_n \langle \psi_n, g \rangle \to g(x, k) \quad \text{in} \quad L_2(0, 1),$$

$$\text{for a.e. } k. \quad (2.10)$$

Integrating (2.10), we get

$$\int_{k} \sum_{n} \psi_n f_n(k) \, dk = \int_{k} \sum_{n} \psi_n \langle \psi_n, g \rangle = \int_{k} g(x, k) \, dk.$$

The last integral is equal to $f(x)$ by Lemmas 2 and 3 above, which proves the corollary.

Corollary 2. "Parseval's equality." The Bloch representation is an isometry in L_2, i.e.,

$$||f|| = \int_{k} \sum_{n} |f_n(k)|^2 \, dk.$$

Here $||f||$ denotes the norm in $L_2(R_3)$. The proof proceeds along lines similar to the two preceding ones, so we omit it.

Corollary 3. "The Kohn–Luttinger (K–L) representation." If k_0 is a fixed vector in Ω^*, then the functions

$$\psi_n(x, k) = e^{2\pi i k \cdot x} \phi_n(x, k_0)$$

are known as the K–L functions[10]. They are complete in $L_2(R_3)$. The proof consists in repeating the proof of Theorem 1 with the new definition of ψ_n.

3. NATURE OF THE SPECTRUM

In Sec. 1 we introduced the auxiliary eigenvalue problem $\tilde{H}_k \phi_n = \lambda_n \phi_n$ and showed that it possesses a discrete set of eigenvalues $S_k = \{\lambda_n(k)\}$. In this section we prove that the spectrum σ of the operator H in L_2 is equal to $\bar{S} = \bigcup_{k \in \Omega^*} S_k$ (in fact, $S = \bar{S} = $ closure of S) and discuss the dependence of the eigenvalues λ on k.

Lemma 4. The set $S = \bigcup_{k} \{\lambda_n(k)\}$ is contained in the spectrum of H.

Proof: Let λ be a point in S. Without loss of generality we may assume that $\lambda = 0$. Then there exists a Bloch wave

$$\psi(x) = e^{2\pi i k \cdot x} \phi(x, k),$$

where $\phi(x, k)$ is periodic and $H\psi = 0$.

To prove that $\lambda \in \sigma$ we construct a singular sequence of functions ψ_N in the domain of H such that $||H\psi_N||/||\psi_N|| \to 0$ as $N \to \infty$. We first choose a mollifying function $\eta(t) \in C^\infty$ such that

[10] W. Kohn and J. M. Luttinger, Phys. Rev. 97, 869 (1955).

$$\eta(t) = \begin{cases} 1, & t \in (-1, 1), \\ 0, & |t| \geq 2, \end{cases}$$

and $|\eta'|$, $|\eta''| \leq 2$. Consider the sequence $\psi_N(x) = \eta(|x|/N)\psi(x)$ where N is an integer. Then

$$\|H\psi_N\| = \left\| H\left[\eta\left(\frac{|x|}{N}\right)\right]\psi \right\|$$

$$\leq \frac{c}{N} \left\| \nabla\eta\left(\frac{|x|}{N}\right)\nabla\psi \right\| + \frac{c}{N^2} \left\| \Delta\eta\left(\frac{|x|}{N}\right)\psi \right\|$$

$$\leq \frac{c}{N} \|\nabla\psi\|_{(N,2N)} + \frac{c}{N^2} \|\psi\|_{(N,2N)}, \qquad (3.1)$$

where c is a generic constant, independent of N, and $\| \; \|_{N,2N}$ denotes the norm in the region between the two spheres of radii N and $2N$, respectively. From (3.1) we have

$$\|H\psi_N\|/\|\psi_N\| = O[\{N^{-1} \|\nabla\psi\|_{N,2N}$$

$$+ N^{-2} \|\psi\|_{N,2N}\}/\|\psi\|_{0,N}]$$

$$= O(1/N) \quad \text{as} \quad N \to \infty.$$

Hence, $\lambda \in \sigma_e$ where σ_e is the essential spectrum[4] of H, so $\lambda \in \sigma$ and the lemma is proved.

Lemma 5. The spectrum of H is contained in $\bar{S}$.

Proof: Let $\mu \neq \infty$ be a real number which is not in $\bar{S}$ and consider the equation $(H - \mu)\phi = f$ where $f \in L_2$. Then from Corollary 1, namely the "Bloch representation" of f, we have

$$\phi = \int_k \sum_i \langle f, \psi_i(x, k)\rangle/(\mu - \lambda_i(k)) \, dk, \qquad (3.2)$$

where the right-hand side of (3.2) is a bounded operator on f since $\mu \notin \bar{S}$. Hence μ is in the resolvent set of H and the lemma follows.

Lemmas 4 and 5 show that the spectrum of H, being a closed set, is equal to $\bar{S}$. It is clear, however, that the spectrum is identical with its essential part since, if there were a point eigenvalue of finite multiplicity, one could easily construct an L_2 eigenfunction common to both H and the group of translations, which is impossible.

We show now that S is simply a countable union of closed intervals. This depends on the following theorem which describes the dependence of an eigenvalue $\lambda = \lambda_n$ on k.

Theorem 2. Let $\lambda_n(k)$ denote the nth isolated eigenvalue of the auxiliary problem (2.3). Then λ_n is an analytic, though not necessarily single-valued, function of k.

Proof: Let $k = k_0 + \epsilon d$, where d is a fixed unit

vector. Upon inserting this value of k into (2.3), the operator $\tilde{H}_k$ can be written in the form

$$\tilde{H}_k \equiv H(\epsilon) = \tilde{H}_{k_0} + \epsilon B = H_0 + \epsilon B.$$

Here $\tilde{H}_{k_0} = H_0$ is the operator $\tilde{H}_k$ with $k = k_0$ and B is defined by

$$B = \sum_{l,m=1}^{3} \left[-4i\pi a_{lm}d_l \frac{\partial}{\partial x_m} - 2i\pi d_l \frac{\partial a_{lm}}{\partial x_m} \right.$$

$$\left. + 4\pi^2 a_{lm}(k_{0l}d_m + k_{0m}d_l) + \epsilon 4\pi^2 a_{lm}d_l d_m \right].$$

To prove the analytic dependence of λ on ϵ amounts to proving the convergence, in ϵ, of the formal perturbation series for $\lambda = \lambda(\epsilon)$. Although this can be deduced from general theorems about regular perturbations of a self-adjoint operator in the sense of Rellich,[11] the situation above is simple enough to give another proof. Let z denote a real number; then for large enough z, the resolvent operator

$$R_z(H_\epsilon) \equiv (H(\epsilon) + z)^{-1} = (H_0 + \epsilon B - z)^{-1}$$

exists as a bounded operator in L_2. This follows simply from the semiboundedness and self-adjointness of $H(\epsilon)$. But

$$R_z(H_\epsilon) = R_z(H_0)[1 + \epsilon B R_z(H_0)]^{-1}. \qquad (3.3)$$

Now, since B is relatively bounded with respect to H_0, i.e., since $\|\nabla\phi\| \leq C[\|\phi\| + \|H_0\phi\|]$, for all ϕ in the domain of H_ϵ, the operator BR_z is bounded, and hence the geometric series for $[1 + \epsilon B R_z]^{-1}$ converges uniformly for small ϵ. But then the eigenvalues $\mu(\epsilon)$ of $R_z(H_\epsilon)$ depend analytically on ϵ.[11] The same statement is then true for $\lambda(\epsilon)$, which proves the theorem. It is clear now that the set $S_n = \{\lambda_n(k)\}$, being the image under a continuous function of the closed unit cell, is itself a closed simply connected set.

We end this section with a few remarks.

(i) We have proved that the spectrum is equal to the closure of $\bigcup_{k \in \Omega^*} \{\lambda_n(k)\} = \bigcup_n S_n$. The sets S_n are closed "intervals" by the above theorem. It is possible to show that the countable union of these sets is again closed because of the asymptotic behavior of the eigenvalues λ_n. Then one has $\sigma = \bigcup_n S_n$.

(ii) In the special case when the eigenvalues $\lambda_n(k) \in S_n$ are all simple—i.e., the case of nondegeneracy of the nth band—the proof of the lemma shows that $\lambda(k)$ is a single-valued holomorphic function of the three complex variables k_1, k_2, k_3 in the domain $\text{Re } k \in \Omega^*$, $|\text{Im } k| \leq \epsilon$ for small enough ϵ.

[11] F. Rellich, "Perturbation Theory of Eigenvalue Problems," New York University Notes, (Courant Institute of Mathematical Sciences, New York, 1953).

(iii) The dependence of the eigenfunctions $\psi_n(x, k)$, regarded as elements in $L_2(\Omega)$, is also governed by Theorem 2. This follows from results of Rellich[11] and Brownell.[12] If the eigenvalues are simple, then we again have holomorphy in the complex sense.

(iv) The perturbation method described in Theorem 2 is usually called the $\mathbf{k} \cdot \mathbf{v}$ method because it was first applied to the case $k = 0$; then $\epsilon = k$ and $i\nabla$ is proportional to v. It is widely used, but no proof of its convergence seems to have been given previously.

4. MISCELLANEOUS REMARKS

A. Operator Representation

The Bloch representation discussed above is simply the "Fourier" analysis of the operator H, and hence it makes H diagonal. As in the theory of Fourier transforms, it is interesting to discuss the form which other operators take in the Bloch representation. One important operator is the position operator x, concerning which the following lemma is useful.

Lemma 6.[13] Let $f(x) \in L_2$, $xf(x) \in L_2$, and let $f_n(k)$ and $g_n(k)$ be the Bloch components of $f(x)$ and $xf(x)$, respectively. Then

$$2\pi g_n(k) = if'_n(k) - i \sum_m A_{n,m}(k)f_m(k) \quad (4.1)$$

where

$$A_{n,m} = \left(\frac{\partial \phi_n}{\partial k}, \phi_m\right) = \int_0^1 \frac{\partial \phi_n^*}{\partial k} \phi_m \, dx$$

and ϕ_n is defined by (2.3).

Proof: We assume that the phase of ϕ is chosen to be analytic in k. The definition of $f_n(k)$ is

$$f_n(k) = \langle \psi_n, f \rangle = \int_{-\infty}^{+\infty} f(\xi)e^{-2\pi i k\xi}\phi_n^*(\xi, k) \, d\xi. \quad (4.2)$$

By Theorem 2 we can differentiate (4.2) with respect to k to get

$$2\pi g_n(k) = i\frac{\partial f_n}{\partial k} - i \int_{-\infty}^{+\infty} f(\xi)e^{-2\pi i k\xi} \frac{\partial \phi_n^*}{\partial k} \, d\xi. \quad (4.3)$$

By substituting for $f(\xi)$ its Bloch representation, we get

$$2\pi g_n = \frac{i\partial f_n}{\partial k} - i \sum_m \int_{k'} f_m(k')$$

$$\times \int_{-\infty}^{+\infty} e^{-2\pi i\xi(k-k')}\phi_m(\xi, k') \frac{\partial \phi_n^*(k)}{dk} \, d\xi \, dk'. \quad (4.4)$$

We now expand the periodic function

$$A_{n,m}(k, k') \equiv (\partial \phi_n^*(\xi, k)/\partial k)\phi_m(\xi, k')$$

in a Fourier series in ξ and substitute into (4.4), and the result follows.

In the Kohn–Luttinger representation, the second term in (4.1) is not present and the x operator corresponds simply to differentiation with respect to k as in the theory of Fourier transforms.

The representations of many other operators are given by Blount.[14] Our Theorems 1 and 2 may be used to make the derivations of those representations completely rigorous.

B. Wannier Functions

Wannier functions are band functions which, by definition, are proportional to

$$a_\nu(x) = \int_{\Omega^*} \psi_\nu(x, k) \, dk. \quad (4.5)$$

Equivalently, if n is a lattice vector, we have

$$a_\nu(x - n) = \int_{\Omega^*} e^{-2\pi ik \cdot n}\psi_\nu(x, k) \, dk. \quad (4.5a)$$

A most important property of Wannier functions is that they are localized, i.e., that $a_\nu(x - n)$ is concentrated around the lattice point $x = n$. This has been proved in the one-dimensional case by Kohn.[2] We now indicate a proof of their localized nature in three dimensions, although in a rather special case.

The definition of the bands, i.e., of the set of eigenvalues $\bigcup_k \{\lambda_\nu(k)\} = S_\nu$ of Theorem 2, depends essentially on the analytic continuation of $\lambda_\nu(0)$ as a function of k. We mentioned in Sec. 2 that the eigenvalue problem $H\psi = \lambda\psi$, where $\psi(x + t) = e^{2\pi ik \cdot t}\psi(x)$, is unchanged when k is increased by a reciprocal lattice vector K. Therefore, for every ν there exists an m such that $\lambda_\nu(k) = \lambda_m(k + K)$. Suppose now (a) that no two bands overlap, and (b) that the eigenvalues in each band are simple. Then one can label the bands in order of increasing energy, and the eigenvalues will be periodic in k. Furthermore, the eigenfunctions can be chosen to be periodic and complex analytic in k, by remark (ii), Theorem 2. Then, application of the Riemann–Lebesgue lemma to the integral (4.5a) proves that the Wannier function decays exponentially with $|x - n|$. This demonstrates that these functions are localized in the case when Conditions (a) and (b) above are satisfied.

[12] F. H. Brownell, J. Math. Anal. Appl. **6**, 190 (1963).
[13] The lemma is proved by Blount,[14] but our proof is somewhat different.

[14] E. I. Blount, in *Solid-State Physics*, edited by F. Seitz and D. Turnbull (Academic Press Inc., New York, 1961), Vol. 13.

A Modified WKB Approximation for Phase Shifts*

MERVINE ROSEN

*School of Physics, University of Minnesota, Minneapolis, Minnesota
and U. S. Naval Research Laboratory, Washington, D. C.†*

AND

DONALD R. YENNIE‡

School of Physics, University of Minnesota, Minneapolis, Minnesota

(Received 15 May 1964)

Extending an idea of Good, a modified WKB approximation using radial wavefunctions having the form of free-particle solutions to the radial wave equation rather than an exponential form is developed. The lowest-order phase shifts are the same as those of the usual WKB approximation, but are improved by the contribution of the next order. The method is applied to two examples: the radial Dirac equation in the high-energy limit and the radial Schrödinger equation.

I. INTRODUCTION

IT is our purpose, in this paper, to develop a modified WKB approximation for partial-wave phase shifts by extending an idea first proposed by Good.[1]

Partial-wave phase shifts are obtained, of course, from the asymptotic behavior of the radial wavefunction. Now, in the customary WKB approximation, one writes the radial wavefunction in the form

$$u(r) = A(r) \exp [iS(r)/\hbar], \qquad (1.1)$$

and assumes a solution to the wave equation can be found by developing $A(r)$ and $S(r)$ in a power series in $\hbar$. As is well known, the resulting wavefunctions are singular at the classical turning point. The question was raised by Good as to whether the functional form used in Eq. (1.1) is the most appropriate one when dealing with radial wave equations. He suggested that rather than the exponential form, the form of the free-particle solutions to the radial equation be used. He then made the customary development in powers of $\hbar$ and was able to define the phase function $S(r)$ in such a way that it is real everywhere except perhaps for a range of r in the vicinity of the turning point and also such that the wavefunction is everywhere finite.[2] The latter makes unnecessary the use of connection formulas across the turning point.

The key point in our generalization of Good's method lies in the recognition of the fact that the higher derivatives of the zero-order functions—the free-particle solutions—can be reduced to zeroth and first derivatives (for the second-order equations we are interested in). Hence, we need use only two functions of $S(r)$. For purposes of illustrating the method, we give two practical examples in Sec. II and III, applying the method to the Dirac radial equation in the high-energy limit and to the Schrödinger equation, respectively. We show that in both cases $S(r)$ may be defined such that it is finite and real everywhere, including the vicinity of the turning point. The lowest-order phase shifts, which are the same as those obtained in the usual WKB approximation, are improved by the next higher order. In the last section we indicate several further possible applications.

II. ELECTRON-SCATTERING PHASE SHIFTS

A. Lowest-Order Approximation to Radial Wavefunctions.

We are interested in scattering at energies high enough so that we may neglect the mass term in the Dirac equation. We begin, therefore, by considering the coupled Dirac radial equations, written in dimensionless form

$$(dF/dx) - [(l + 1)/x]F - (1 - v)G = 0, \qquad (2.1)$$

$$dG/dx + [(l + 1)/x]G + (1 - v)F = 0, \qquad (2.2)$$

where the radial variable is in units of $k^{-1} = \lambda$,

$$k = E/\hbar c, \qquad v(x) = V(r)/E$$

and E, of course, is the total energy. Note that we describe the interaction between the electron and the nuclear charge distribution by a static central potential $V(r)$. For reasons that later become clear, we put Eqs. (2.1) and (2.2) into a more symmetrical

[1] R. H. Good, Jr., Phys. Rev. 90, 131 (1953).
[2] For somewhat related work on modifications of the WKB method which also result in wavefunctions which are non-singular at the classical turning point, see C. E. Hecht and J. E. Mayer, Phys. Rev. 106, 1156 (1957), and J. S. Nodvik, UCLA Tech. Rept. No. 3–1–58.

form by defining the two functions

$$M(x) = F(x) + G(x), \quad N(x) = F(x) - G(x). \quad (2.3)$$

We find that

$$dM/dx = f(x)N, \quad dN/dx = g(x)M, \quad (2.4)$$

where

$$f(x) = (l + 1)/x - (1 - v), \quad (2.5)$$
$$g(x) = (l + 1)/x + (1 - v).$$

In accordance with the ideas expressed in the previous section, we want to find a function $S(r)$ which will enable us to approximate the radial wavefunctions by functions of the form of the unperturbed solutions of Eqs. (2.4). We denote the unperturbed solutions by M_0 and N_0, respectively; they satisfy the equations

$$dM_0/dS = [(l + 1)/S - 1]N_0 \equiv f_0(S)N_0, \quad (2.6)$$
$$dN_0/dS = [(l + 1)/S + 1]M_0 \equiv g_0(S)M_0.$$

The functional behavior of M_0 and N_0 can be inferred from the fact that

$$F_0 = \tfrac{1}{2}(M_0 + N_0) \sim Sj_l(S), \quad (2.7)$$
$$G_0 = \tfrac{1}{2}(M_0 - N_0) \sim Sj_{l+1}(S),$$

where j_l is the spherical Bessel function of order l. The solution regular at the origin is chosen in anticipation of the fact that the modified WKB wavefunctions are well defined everywhere.

In lowest order we, therefore, write

$$M(x) = a_0(x)M_0(S), \quad N(x) = b_0(x)N_0(S), \quad (2.8)$$

and neglect the derivatives of a_0 and b_0. Substituting Eq. (2.8) into Eq. (2.4) we obtain

$$M' = a_0 S' M_0' = a_0 S' f_0 N_0$$
$$= f(x)b_0 N_0,$$

or

$$a_0 f_0 S' = b_0 f. \quad (2.9)$$

Similarly,

$$b_0 g_0 S' = a_0 g, \quad (2.10)$$

where a prime denotes the derivative of a function with respect to its argument. Note that the comparison functions M_0 and N_0 do not appear in Eqs. (2.9) and (2.10); they have dropped out. We proceed in higher approximations in such a manner that this occurs in each order. Solving the above equations for S', we obtain

$$f_0 g_0 (S')^2 = fg \quad (2.11')$$

or

$$\int_{s_t}^{s} (-f_0 g_0)^{\frac{1}{2}} d\sigma = \int_{x_t}^{x} (-fg)^{\frac{1}{2}} d\rho. \quad (2.11)$$

We define S such that it is everywhere real and that S' is finite at the classical turning point x_t—defined to be that point where $f(x)$ vanishes. We therefore take the lower limit of the left-hand integral to be

$$S_t = l + 1.$$

Although we are primarily interested in the region $x > x_t$, Eq. (2.11) is of course valid for all x and S, including $x < x_t$ and $S < S_t$. An equation for $S(r)$ similar to Eq. (2.11) was also obtained by Good. He also chose the lower limits of his integrals to be x_t and $S_t = l + 1$, respectively, although for a different reason. The integrands in his equation had branch points at $\rho = x_t$ and $\sigma = l + 1$, respectively, which led him to associate them with each other. He obtained an $S(r)$ however that was, in general, not everywhere real.

We also find from Eqs. (2.9) and (2.10), that

$$(a_0/b_0)^2 = g_0 f/f_0 g. \quad (2.12)$$

To completely determine a_0 and b_0, it is necessary to take a look at the next higher approximation.

In first order, we cannot neglect the derivatives of a_0 and b_0; thus when we substitute Eq. (2.8) into Eq. (2.4), these introduce terms containing M_0 and N_0, respectively. However, if we write

$$M(s) \cong a_0 M_0(S) + a_1 N_0(S), \quad (2.13)$$
$$N(x) \cong b_0 N_0(S) + b_1 M_0(S), \quad (2.14)$$

and neglect the derivatives of a_1 and b_1, the functions M_0 and N_0 again drop out, resulting in the following set of equations:

$$\begin{bmatrix} -g_0 S' & f \\ g & -f_0 S' \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} a_0' \\ b_0' \end{bmatrix}. \quad (2.15)$$

The matrix on the left, however, is a singular one, its determinant vanishing according to Eq. (2.11'). Therefore, a_0' and b_0' must satisfy a solvability condition of the form

$$\alpha a_0' + \beta b_0' = 0,$$

and indeed we see that

$$f_0 S' a_0' + f b_0' = 0. \quad (2.16)$$

This, together with the fact that

$$a_0/b_0 = f/f_0 S' = g_0 S'/g$$

yields

$$a_0 b_0 = \text{const} = 1. \qquad (2.17)$$

We may set the constant equal to unity, as this affects only the normalization of the wavefunction. It follows that

$$a_0 = b_0^{-1} = (g_0 f/f_0 g)^{\frac{1}{2}} = (f/f_0 S')^{\frac{1}{2}} = (g_0 S'/g)^{\frac{1}{2}}. \qquad (2.18)$$

The amplitudes a_0 and b_0 are finite everywhere and in particular at the turning point, thus obviating the necessity of connection formulas. Furthermore, in most cases of interest, they deviate only slightly from unity.

The lowest-order wavefunction is now fully determined and we are in a position to calculate the lowest-order phase shift. The asymptotic solutions to Eq. (2.8), valid for large x, are

$$M \propto \cos [S - (l + \tfrac{3}{2})\tfrac{1}{2}\pi]$$

$$N \propto \cos [S - (l + \tfrac{5}{2})\tfrac{1}{2}\pi];$$

therefore, denoting the lowest-order phase shift by $\eta^{(0)}$, we have

$$\eta^{(0)} = \lim_{x \to \infty} (S - x - \gamma \ln 2x),$$

where

$$\gamma = Z\alpha,$$

Z is the atomic number of the target nucleus and α is the fine-structure constant. The logarithmic term, of course, takes into account the phase distortion at infinity due to the long tail of the Coulomb potential. From Eq. (2.11), we find

$$\lim_{x \to \infty} S = \lim_{x \to \infty} \int_{x_t}^{x} \left[(1 - v)^2 - \frac{(l + 1)^2}{\rho^2} \right]^{\frac{1}{2}} d\rho + (l + 1)\tfrac{1}{2}\pi, \qquad (2.19)$$

and hence

$$\eta^{(0)} = \lim_{x \to \infty} \left\{ \int_{x_t}^{x} \left[(1 - v)^2 - \frac{(l + 1)^2}{\rho^2} \right]^{\frac{1}{2}} d\rho - x - \gamma \ln 2x + (l + 1)\tfrac{1}{2}\pi \right\}. \qquad (2.20)$$

This is just the expression given by Baranger[3] for the WKB approximation to the phase shifts for the elastic scattering of high-energy electrons from a central potential. Our lowest-order phase shifts, then, are just those given by the usual WKB approximation.

A perhaps useful piece of information comes out of the above analysis. Calling the quantum number l which characterizes the solutions to Eq. (2.4) the

---

[3] E. Baranger, Phys. Rev. 93, 1127 (1954).

index of these functions, then the lowest-order phase shift is independent of the index we choose for the comparison functions M_0 and N_0. For example, if we denote the index of M_0 by k, then according to Eqs. (2.8) and (2.19)

$$M_0 \sim \cos \left\{ \left[\lim_{x \to \infty} \int_{x_t}^{x} \left((1 - v)^2 - \frac{(l + 1)^2}{\rho^2} \right)^{\frac{1}{2}} d\rho \right. \right.$$
$$\left. + (k + \tfrac{1}{2})\tfrac{1}{2}\pi \right] - (k + \tfrac{3}{2})\tfrac{1}{2}\pi \bigg\}$$
$$\sim \cos \left\{ \lim_{x \to \infty} \int_{x_t}^{x} \left[(1 - v)^2 - \left(\frac{l + 1}{\rho} \right)^2 \right]^{\frac{1}{2}} d\rho - \tfrac{1}{4}\pi \right\}.$$

This leads us once more to Eq. (2.20). There is therefore an infinite class of modified WKB approximations all giving the same phase shifts in lowest order, but which will not, in general, agree in higher orders. It seems reasonable to take k equal to l, but it is conceivable that with some other approach, it might be advantageous to choose k differently.

It is also of interest to see how condition (2.17) affects the function that plays the role of the Wronskian in Dirac scattering theory. Eq. (2.4) has two linearly independent sets of solutions which we denote by $(M^{(1)}, N^{(1)})$ and $(M^{(2)}, N^{(2)})$, respectively. The Wronskian is then given by

$$W(x) = (M^{(1)}N^{(2)} - M^{(2)}N^{(1)});$$

it follows directly from Eq. (2.4) that

$$W'(x) = 0. \qquad (2.21)$$

If we substitute Eq. (2.8) into Eq. (2.21), we obtain

$$d[a_0 b_0 (M_0^{(1)}N_0^{(2)} - M_0^{(2)}N_0^{(1)})]/dx$$
$$= d(M_0^{(1)}N_0^{(2)} - M_0^{(2)}N_0^{(1)})/dx = 0.$$

Thus, the Wronskian is independent of x both exactly and also in lowest order.

B. First-Order Phase Shift

We now go on to further consideration of Eq. (2.15). It is convenient to write (a_1, b_1) as the sum of two terms—one a solution of the homogeneous part of Eq. (2.15) and the other a particular solution:

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} s \\ t \end{bmatrix} + \begin{bmatrix} m \\ n \end{bmatrix},$$

where

$$\begin{bmatrix} -g_0 S' & f \\ g & -f_0 S' \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = 0. \qquad (2.22)$$

We can make the separation explicit by writing (a_1, b_1) in the following way:

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \frac{1}{(f_0 + g_0)S'} \left\{ \begin{bmatrix} g_0 S' & -f \\ -g & -f_0 S' \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \right. $$
$$\left. + \begin{bmatrix} f_0 S' & f \\ g & g_0 S' \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \right\}, \qquad (2.23)$$

from which we obtain

$$m = -\frac{a_0'}{(f_0 + g_0)S'}, \quad n = -\frac{b_0'}{(f_0 + g_0)S'}. \qquad (2.24)$$

From Eq. (2.22), we find

$$s = \rho(x)f, \qquad t = \rho(x)g_0 S', \qquad (2.25)$$

where the function $\rho(x)$ is determined only in the next higher order.

It is not difficult to see that as $x \to \infty$

$$a_0, \quad b_0 \to 1,$$
$$m, \quad n \to 0,$$

and

$$s \to -\rho(\infty), \quad t \to \rho(\infty).$$

Therefore

$$M \sim M_0 - \rho(\infty)N_0, \qquad (2.26)$$
$$N \sim N_0 + \rho(\infty)M_0,$$

from which, assuming $\rho(\infty)$ to be a small correction, we find

$$M(x) \sim \cos\left[S - (l + \tfrac{3}{2})\tfrac{1}{2}\pi + \rho(\infty)\right], \qquad (2.27)$$
$$N(x) \sim \cos\left[S - (l + \tfrac{5}{2})\tfrac{1}{2}\pi + \rho(\infty)\right].$$

We see that $\rho(\infty)$ is a correction to the lowest-order phase shift.

Going on to the next order, we write

$$M(x) = a_0 M_0 + a_1 N_0 + a_2 M_0,$$
$$N(x) = b_0 N_0 + b_1 M_0 + b_2 N_0,$$

and, proceeding as before, find

$$\begin{bmatrix} -f_0 S' & f \\ g & -g_0 S' \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1' \\ b_1' \end{bmatrix}.$$

The coefficients a_1 and b_1 must satisfy a solvability condition

$$g_0 S' a_1' + f b_1' = 0,$$

or

$$2fg_0 S' \rho'$$
$$+ \frac{d}{dx}(fg_0 S')\rho + g_0 S' \frac{dm}{dx} + f \frac{dn}{dx} = 0. \qquad (2.28)$$

Therefore

$$\rho(x) = \frac{-1}{2(fg_0 S')^{\frac{1}{2}}}$$
$$\times \int_{x_t}^{x} \frac{g_0 S'(dm/dx) + f(dn/dx)}{(fg_0 S')^{\frac{1}{2}}} dx + \frac{\text{const}}{(fg_0 S')^{\frac{1}{2}}}.$$

Now we want ρ to be real since our equations and all the quantities we have dealt with are real; we therefore require that the constant be zero. This also means that ρ is everywhere finite, which is necessary if our whole approach is to be meaningful. Thus

$$\rho(x) = \frac{-1}{2(fg_0 S')^{\frac{1}{2}}} \int_{x_t}^{x} \frac{g_0 S' \dfrac{dm}{dx} + f \dfrac{dn}{dx}}{(fg_0 S')^{\frac{1}{2}}} dx, \qquad (2.29)$$

and the first-order correction to the phase shift is

$$\eta^{(1)} = \lim_{x \to \infty} \rho(x).$$

C. Point Coulomb Phase Shifts

For the case of scattering from a point Coulomb field, the lowest-order phase shift is easily calculated. We have

$$\eta^{(0)} = \lim_{x \to \infty} \left[\int_{x_t}^{x} \left(1 + \frac{2\gamma}{\rho} - \frac{\Lambda^2}{\rho^2}\right)^{\frac{1}{2}} d\rho \right.$$
$$\left. - x - \gamma \ln 2x + (l + 1)\tfrac{1}{2}\pi \right] \qquad (2.30)$$
$$= \gamma - \Lambda \tan^{-1}\frac{\gamma}{\Lambda} - \gamma \ln(l + 1)$$
$$+ [(l + 1) - \Lambda]\tfrac{1}{2}\pi,$$

where

$$\Lambda^2 = (l + 1)^2 - \gamma^2$$

and we have used the fact that

$$x_t = (l + 1) - \gamma.$$

Now, the exact-point Coulomb phase shift η_c is given by[4]

$$\exp 2i\eta_C = \frac{(\Lambda - i\gamma)}{(l + 1)} \frac{\Gamma(\Lambda - i\gamma)}{\Gamma(\Lambda + i\gamma)} \exp[\pi i(l + 1 - \Lambda)].$$

If we use the asymptotic form of the gamma function

$$\ln \Gamma(x) = \tfrac{1}{2} \ln 2\pi$$
$$- x + (x - \tfrac{1}{2}) \ln x + O(1/x), \qquad (2.31)$$

we find

$$\eta_c \sim \gamma - \Lambda \tan^{-1}(\gamma/\Lambda) - \gamma \ln(l + 1)$$
$$+ [l + 1 - \Lambda]\tfrac{1}{2}\pi. \qquad (2.32)$$

[4] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London, 1949), p. 79.

The WKB phase shifts therefore are just those obtained by replacing the gamma functions in the exact expression by Stirling's approximation. We shall see that the contribution from $\eta^{(1)}$ is just that correction to Eq. (2.32) obtained by neglecting only terms $O(1/x^3)$ in Eq. (2.31).

To calculate the first-order correction to the phase shift, one has to evaluate the integral in Eq. (2.29); since ρ is a small correction, we do this only approximately. It is convenient to expand the quantities occurring in Eq. (2.29) to first order in the potential. We therefore let

$$S = (l + 1)[y + \beta(1 - y)U(y)], \qquad (2.33)$$

where

$$y = x/x_t, \qquad \beta = \gamma/(l + 1),$$

and, taking β to be a parameter of smallness, perform all further calculations only to first order in β. The equation for $U(y)$ is obtained by substituting Eq. (2.33) into Eq. (2.11') and neglecting terms of second and higher order in β. We find

$$y(y + 1)(y - 1)U'(y)$$
$$+ [y(y + 1) + 1]U(y) = y^2, \qquad (2.34)$$

from which we obtain

$$U(y) = \frac{y}{(y - 1)(y^2 - 1)^{\frac{1}{2}}}\left[\int_1^y dz\left(\frac{z - 1}{z + 1}\right)^{\frac{1}{2}} + \text{const}\right].$$

Since U must be real (because S is real) the constant is zero. Hence

$$U(y) = \frac{y}{(y - 1)(y^2 - 1)^{\frac{1}{2}}}\int_1^y dz\left(\frac{z - 1}{z + 1}\right)^{\frac{1}{2}}. \qquad (2.35)$$

The integral can now be done to first order in β and, after lengthy but straightforward manipulation, we obtain

$$\eta^{(1)} = \gamma/12(l + 1)^2. \qquad (2.36)$$

This is indeed the correction to Eq. (2.32) which is obtained if the $1/12x$ term in the asymptotic expansion of the gamma function is retained and higher terms are neglected.

D. Scattering from a Distributed Charge

In the case of scattering from an extended nucleus, the phase shift is given in terms of the Coulomb potential of the extended charge distribution. The potential may be written

$$v(x) = -(\gamma/x_t)\omega(y),$$

where

$$\omega(y) = \frac{4\pi}{y}\int_0^y t^2 q(t)\, dt$$
$$+ 4\pi\int_y^\infty tq(t)\, dt; \qquad y = \frac{x}{x_t} \qquad (2.37)$$

and the dimensionless reduced charge density q is normalized such that

$$4\pi\int_0^\infty q(t)t^2\, dt = 1. \qquad (2.38)$$

We can estimate the integral in Eq. (2.29) in a manner similar to that in the point Coulomb case; the function $U(y)$ is now given by

$$U(y)$$
$$= -\frac{y}{(y - 1)(y^2 - 1)^{\frac{1}{2}}}\int_1^y tW(t)\left(\frac{t - 1}{t + 1}\right)^{\frac{1}{2}} dt, \qquad (2.39)$$

where $W(y)$ is defined by

$$\omega(1) - \omega(y) = (1 - y)W(y).$$

Expressing $\eta^{(1)}$ in terms of the charge distribution, we find in a straightforward manner

$$\eta^{(1)} = \frac{\gamma}{12(l + 1)^2}$$
$$\times \left[1 + 2\pi\int_1^\infty \frac{2y^4 - y^2 + 2}{3(y^2 - 1)^{\frac{1}{2}}} q'(y)\, dy\right]. \qquad (2.40)$$

Now it is clear from Eq. (2.20) that $\eta^{(0)}$ is not strongly affected by the details of the charge distribution; the second term in $\eta^{(1)}$ is a correction which takes into account a more detailed feature of the distribution—its radial derivative—while the first term is seen to be a correction to the point Coulomb part of $\eta^{(0)}$.

To make clearer the dependence of $\eta^{(1)}$ on l we evaluate it explicitly for a simple case—that of a uniform charge distribution with radius R. We find

$$\eta^{(1)} \text{ (uniform)}$$

$$= \begin{cases} \dfrac{\gamma}{12(l + 1)^2}, & R < x_t \\[2ex] \dfrac{\gamma}{12(l + 1)^2}\left[1 - \dfrac{1}{2}\dfrac{2x_0^2 - x_0 + 2}{x_0[x_0(x_0 - 1)]^{\frac{1}{2}}}\right], & R > x_t, \end{cases}$$
$$(2.41)$$

where

$$x_0 = (R/x_t)^2,$$

and x_t is to a good approximation, proportional to $(l + 1)$. We see that as the turning point approaches

the surface of the charge distribution from the left, the correction becomes singular as $[-(R - x_t)^{-\frac{1}{2}}]$, but is finite and positive as the surface is approached from the right. This discontinuity is a consequence of the sharp discontinuity in the charge distribution; in the case of a smooth distribution $\eta^{(1)}$ is finite for all values of l—although remaining peaked and of opposite sign on opposite sides of the surface.

Rather than proceeding as we have done above, we could have written from the start, following the ideas sketched in the introduction,

$$M(x) = a(x)M_0(S) + c(x)N_0(S),$$

$$N(x) = b(x)N_0(S) + d(x)M_0(S).$$

Substituting these into Eq. (2.4) and equating the coefficients of M_0 and N_0 to zero, we obtain

$$a' + cS'g_0 = df, \qquad aS'f_0 + c' = bf,$$

$$bS'g_0 + d' = ag, \qquad b' + dS'f_0 = cg. \qquad (2.42)$$

If we take $S(r)$ to be as defined above, these form a set of four first-order equations for the four coefficients. Since the coefficients are generally slowly varying, these equations would lend themselves to numerical methods of solution. What we have actually done however, is equivalent to writing the coefficients as follows,

$$a = a_0 + a_2 + a_4 + \cdots,$$

$$b = b_0 + b_2 + b_4 + \cdots,$$

$$c = c_1 + c_3 + c_5 + \cdots,$$

$$d = d_1 + d_3 + d_5 + \cdots,$$

where the indices denote order of smallness. The coefficients a_0 and b_0 then satisfy Eqs. (2.9) and (2.10); c_1 and d_1 are just the coefficients we have previously denoted by a_1 and b_1, respectively.

If one follows through the analysis in Secs. 1 and 2, it is seen that, in all orders, the modified WKB phase shift depends only on the potential outside the turning point. Indeed it is straightforward to show that the even- and odd-order corrections, respectively, are given by

$$\rho_k(x) = -\frac{1}{2(ff_0S')^{\frac{1}{2}}}$$

$$\times \int_{x_t}^{x} \frac{f_0S'\dfrac{d}{dx}m_k + f\dfrac{d}{dx}n_k}{(ff_0S')^{\frac{1}{2}}} dx, \quad k \text{ even},$$

$$\rho_k(x) = -\frac{1}{2(fg_0S')^{\frac{1}{2}}}$$

$$\times \int_{x_t}^{x} \frac{g_0S'\dfrac{d}{dx}m_k + f\dfrac{d}{dx}n_k}{(fg_0S')^{\frac{1}{2}}} dx, \quad k \text{ odd}, \qquad (2.43)$$

where

$$m_k = -\frac{a'_{k-1}}{(f_0 + g_0)S'}, \qquad n_k = -\frac{b'_{k-1}}{(f_0 + g_0)S'}.$$

Thus, with this method, we can only obtain information about that part of the charge distribution which lies outside the turning point; e.g., if the turning point lies outside the charge distribution we could at best obtain pure point Coulomb phase shifts. This follows from the fact that

$$M \sim (1 - \rho_2(\infty) - \rho_4(\infty) - \cdots)M_0$$

$$- (\rho(\infty) + \rho_3(\infty) + \cdots)N_0,$$

$$N \sim (1 - \rho_2(\infty) - \rho_4(\infty) - \cdots)N_0$$

$$+ (\rho(\infty) + \rho_3(\infty) + \cdots)M_0.$$

This expansion may be asymptotic if the potential is analytic, and, of course, must be so if it is not. Indeed it is clear that the expansion eventually diverges if there is a discontinuity in, say, the nth derivative of the potential. For if $v^{(n)}$ is discontinuous or singular then so also are $a_0^{(n)}$ and $b_0^{(n)}$ and hence $m_1^{(n-1)}$, $n_1^{(n-1)}$, and $\rho_1(x)^{(n-1)}$. Each successively higher order coefficient becomes discontinuous or singular at one lower order of the derivative, so that eventually all the coefficients beyond a certain order are discontinuous or singular. From Eqs. (2.1) and (2.2), however, it is seen that if the potential contains a discontinuity or singularity in its nth derivative, then the exact functions contain a discontinuity or singularity only in the $(n + 1)$st and higher derivatives.

As an example, let us examine a case we have already treated above—that of a uniform distribution. Here $v(x)$ and $v'(x)$ are continuous but $v^{(2)}$ is discontinuous at the radius R of the charge distribution. We have seen that a_0 and b_0 are indeed continuous, as are also $a_0^{(1)}$ and $b_0^{(1)}$ (and consequently m_1 and n_1); $a_0^{(2)}$, $b_0^{(2)}$, $m_1^{(1)}$ and $n_1^{(1)}$ are discontinuous at $x = R$. We see also from Eq. (2.28) that $\rho_1^{(1)}(x)$ is discontinuous there and, moreover, its discontinuity is considerably enhanced if the turning point lies near the discontinuity. This accounts for the singular behavior of $\eta^{(1)}$ as a function of l in Eq. (2.41).

It should be noted that if we use the exact equations for a, b, c, and d, this singular behavior does not occur, for although their derivatives may be discontinuous, the coefficients themselves are not.

TABLE I. Point Coulomb phase shifts for scattering from gold; $\gamma = 0.5765$.

$(l + 1)$	WKB	Corrected WKB	Numerical[a]
1	+0.36175	+0.40979	+0.40736
2	−0.24973	−0.23772	−0.23797
3	−0.53832	−0.53298	−0.53303
4	−0.72957	−0.72657	−0.72659
5	−0.87289	−0.87097	−0.87098
6	−0.98756	−0.98623	−0.98623
7	−1.08316	−1.08218	−1.08218
8	−1.16513	−1.16438	−1.16438
9	−1.23688	−1.23629	−1.23628
10	−1.30068	−1.30020	−1.30020
11	−1.35811	−1.35771	−1.35772
12	−1.41034	−1.41001	−1.41001

[a] See Ref. 5.

The modified WKB phase shifts, in lowest order and with first-order corrections, together with phase shifts obtained by the numerical integration of the radial Dirac equations[5,6] are given for several different charge distributions in Tables I–III. The difference between the exact and the corrected WKB phase shifts are indicated in Fig. 1. It is seen that the error is appreciably smaller for the smoother Fermi-shaped distribution. The test however is in the angular distribution and in Fig. 2 are compared the WKB and the exact angular distributions for the scattering from gold for $k = 10^{13}$ cm^{-1} ($\cong 197$ MeV), using a

TABLE II. Phase shifts for scattering from gold, using a uniform distribution with $kR = 8.0$, $\gamma = 0.5765$.

$(l + 1)$	WKB	Corrected WKB	Numerical[a]
1	−0.83551	−0.83552	−0.83553
2	−0.85283	−0.85287	−0.85289
3	−0.88140	−0.88149	−0.88136
4	−0.92079	−0.92093	−0.92116
5	−0.97028	−0.97056	−0.97034
6	−1.02887	−1.02935	−1.02902
7	−1.09498	−1.09582	−1.09683
8	−1.16607	−1.16797	−1.16829
9	−1.23688	−1.23629	−1.23710
10	−1.30068	−1.30020	−1.30033
11	−1.35811	−1.35771	−1.35773
12	−1.41034	−1.41001	−1.41001

[a] See Ref. 5.

[5] D. G. Ravenhall and D. R. Yennie, Proc. Phys. Soc. London, 70A, 857 (1957).
[6] B. C. Clark, R. Herman, and D. G. Ravenhall, (private communication). The small differences between the exact phase shifts for the Fermi-shaped distribution given above and the earlier published set of Ravenhall and Yennie[5] are, to the accuracy quoted, due entirely to small but significant differences in the dimensionless parameters on which the calculation depends. Specifically, the earlier values and the present ones are as follows: γ was 0.5765, is here 0.5764854; kc was 7.5761421, is here 7.58; kt was 2.7883561, is here 2.79.

TABLE III. Phase shifts for scattering from gold, using a Fermi shape distribution with $kc = 7.58$, $kt = 2.79$ (c is the distance to the half point and t is the 90% to 10% distance), $\gamma = 0.5764854$.

$(l + 1)$	WKB	Corrected WKB	Numerical[a]
1	−0.83606	−0.83606	−0.83607
2	−0.85403	−0.85407	−0.85406
3	−0.88365	−0.88375	−0.88375
4	−0.92437	−0.92457	−0.92457
5	−0.97528	−0.97565	−0.97557
6	−1.03500	−1.03567	−1.03576
7	−1.10129	−1.10242	−1.10267
8	−1.17055	−1.17183	−1.17179
9	−1.23818	−1.23867	−1.23864
10	−1.30094	−1.30080	−1.30086
11	−1.35814	−1.35783	−1.35787
12	−1.41032	−1.41001	−1.41002

[a] See Ref. 5.

Fermi-shaped distribution.[6] The result is somewhat disappointing. One expects, of course, the modified WKB approximation to improve with increasing energy (and with more diffuse surfaces of the charge distributions), but on the other hand, the exact angular distribution drops off more steeply and necessitates greater accuracy in the phase shifts. Clearly a comparison of the WKB and exact angular distributions at higher energy is needed to check the usefulness of the modified WKB approximation in analyzing the elastic scattering of high-energy electrons from heavy nuclei. Although these results would discourage us from using the WKB phase shifts for a complete practical calculation, it is possible that one could use them in combination with phase shifts obtained by numerical integration of the radial wavefunctions. Since the greatest errors occur for those partial waves whose turning point lies in the surface, one might for example use
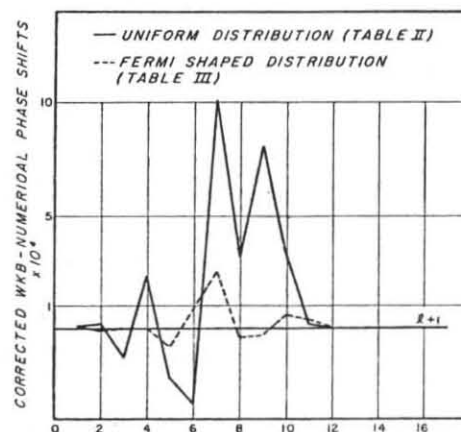

FIG. 1. Deviation of the corrected WKB phase shifts from the exact ones obtained by numerical integration of the radial equations.
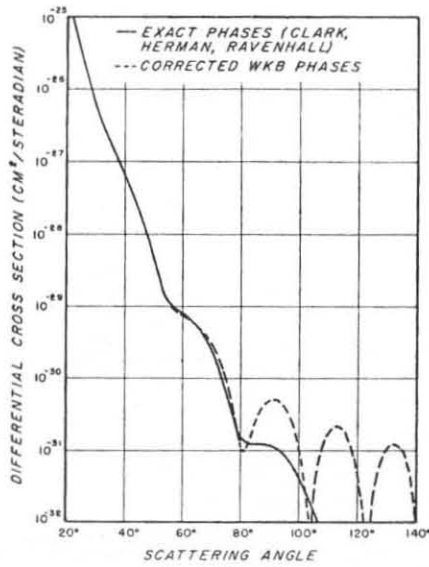
FIG. 2. Angular distributions at $k = 10^{13}$ cm^{-1} ($\cong$ 197 MeV) for scattering by the Fermi shape $kc = 7.58$, $kt = 2.79$, for gold with $\gamma = 0.5764854$.

the WKB phase shifts for small and large l:

$$(l + 1) \overset{\sim}{<} (kR - kt), \qquad (l + 1) \overset{\sim}{>} (kR + kt),$$

and the exact ones for the critical region. In this way, it might be possible to reduce considerably computing-time requirements while retaining reasonable accuracy in the angular distributions.

III. APPLICATION TO THE SCHRÖDINGER EQUATION

A. Modified WKB Wavefunction

In this section, we show that the essential ideas outlined in the previous sections can be used when dealing with the Schrödinger equation also.

We may write the radial wave equation

$$-\frac{\hbar^2}{2m} u''(r)$$

$$+ \left[V(r) + \frac{l(l + 1)\hbar^2}{2mr^2} \right] u(r) = E u(r), \qquad (3.1)$$

where

$$u(r) = r\psi(r),$$

l is the angular momentum quantum number, and E is the total energy (positive for positive energies and negative for bound states). Equation (3.1) can be rewritten in dimensionless form

$$u''(\rho) - q^2(\rho)u(\rho) = 0, \qquad (3.2)$$

where

$$q^2(\rho) = l(l + 1)/\rho^2 + v - 1 \qquad (E > 0),$$

$$\rho = kr, \qquad k^2 = \frac{2m |E|}{\hbar^2}, \qquad v(\rho) = \frac{V(r)}{|E|}.$$

If we could factor Eq. (3.2) into two linear equations having the same form as Eq. (2.4), we would be able to treat the radial Schrödinger equation in exactly the same way as we did the Dirac radial equations. Unfortunately, there is no natural factorization of Eq. (3.2) and we must proceed in a somewhat different manner.

As before, we want to approximate the solutions of the radial equations by functions of the same form as the unperturbed solutions to Eq. (3.2). We denote the unperturbed solutions by u_0; they satisfy the equation

$$u_0''(S) - q_0^2(S)u_0(S) = 0, \qquad (3.3)$$

where

$$q_0^2(S) = l(l + 1)/S^2 - 1.$$

We write therefore in lowest approximation,

$$u(\rho) = a_0(\rho)u_0(S). \qquad (3.4)$$

Substituting this into Eq. (3.2), we find

$$(a_0''u_0 + 2a_0'u_0'S' + a_0(S')^2u_0'' + a_0u_0'S'')$$

$$- a_0q^2u_0 = 0. \qquad (3.5)$$

By making use of Eq. (3.3), all derivatives of u_0 may be reduced to expressions containing no derivatives greater than the first. In picking out the dominant terms in Eq. (3.5), we take, u_0, u_0', and u_0'' to be of comparable order, but assume that derivatives of a_0 are relatively small. We also assume that second and higher derivatives of S are also small. We then obtain

$$q_0^2(S')^2 - q^2 = 0. \qquad (3.6)$$

This equation was also obtained by Good, but his treatment differs from ours in that he separates $l(l + 1)$ into two terms, $[(l + 1)^2 - (l + 1)]$ and assumes the second term to be small relative to the first.

Using Eq. (3.6) to define S, we proceed using the ideas described above, and write

$$u(\rho) = a(\rho)u_0(S) + b(\rho)u_0'(S).$$

Substituting this into the wave equation and equating the coefficients of u_0 and u_0', respectively, to zero, we find

$$a'' + 2(q_0^2S')^{\frac{1}{2}}(d/d\rho)[(q_0^2S')^{\frac{1}{2}}b] = 0, \qquad (3.7)$$

$$b'' + 2(S')^{\frac{1}{2}}(d/d\rho)[(S')^{\frac{1}{2}}a] = 0. \qquad (3.8)$$

As in the case of the Dirac equation, it might be interesting to solve these equations directly. Since the rapid oscillations of the wavefunction have been removed, they may lend themselves to an economical numerical procedure. However, we do not proceed along these lines, but instead write

$$a = a_0 + a_2 + a_4 + \cdots ,$$
$$b = b_1 + b_3 + b_5 + \cdots ,$$

where the indices indicate orders of smallness.

We are satisfied for our present purposes with calculating phase shifts only to the two lowest orders and, therefore, in determining only the coefficients a_0 and b_1. We find that they satisfy

$$2(S')^{\frac{1}{2}}(d/d\rho)[a_0(S')^{\frac{1}{2}}] = 0, \qquad (3.9)$$

$$a_0'' + 2(q_0^2 S')^{\frac{1}{2}}(d/d\rho)[b_1(q_0^2 S')^{\frac{1}{2}}] = 0. \qquad (3.10)$$

These equations, together with Eq. (3.6), are sufficient to determine the coefficients a_0 and b_1.

B. Calculation of Phase Shifts to First Order.

To find the lowest-order contribution to the phase shift, we need to know the behavior of S for large ρ. From Eq. (3.8)

$$\int_{s_t}^{s} (-q_0^2)^{\frac{1}{2}} ds = \int_{\rho_t}^{\rho} (-q^2)^{\frac{1}{2}} dr. \qquad (3.11)$$

As before, we choose the lower limit of the integral on the left in such a way as to keep S' everywhere finite and real. We therefore take

$$S_t = S(\rho_t) = [l(l + 1)]^{\frac{1}{2}} \equiv \Lambda,$$

where ρ_t, the classical turning point, is defined by

$$q^2(\rho_t) = 0.$$

From Eq. (3.11) we now find

$$S \sim \lim_{\rho \to \infty} \int_{\rho_t}^{\rho} (-q^2)^{\frac{1}{2}} dr + \Lambda \tfrac{1}{2}\pi. \qquad (3.12)$$

Since the field-free solution to Eq. (3.2) regular at the origin is the spherical Bessel function of order l, the lowest-order phase shift, $\eta^{(0)}$, is obtained from

$$\eta^{(0)} = \lim_{\rho \to \infty} (S - \rho - \alpha \ln 2\rho),$$

where the presence of the logarithmic term is due to the (assumed) presence of a Coulomb potential,

$$\alpha = ZZ'e^2/\hbar v,$$

Z and Z' are the atomic numbers of the nucleus and incident particles, respectively, and v is here the relative velocity of the incoming particle and the nucleus. Therefore,

$$\eta^{(0)} = \lim_{\rho \to \infty} \left[\int_{\rho_t}^{\rho} (-q^2)^{\frac{1}{2}} dr - \rho \right. $$
$$\left. - \alpha \ln 2\rho + \Lambda \tfrac{1}{2}\pi \right]. \qquad (3.13)$$

It might be noted that it is not true here, as it is with the Dirac equation, that the lowest-order phase shift is independent of the index of u_0.

The solution to Eq. (3.9) is

$$a_0 = \text{const } (S')^{-\frac{1}{2}}.$$

As the constant affects only the normalization of the wavefunction, we may set it equal to unity; thus

$$a_0 = (S')^{-\frac{1}{2}}.$$

This is finite everywhere, which means in particular that our wavefunction is finite at the turning point.

The solution to Eq. (3.10) that is everywhere real and finite is

$$b_1(\rho) = \frac{-1}{2(-q_0^2 S')^{\frac{1}{2}}} \int_{\rho_t}^{\rho} \frac{a_0''}{(-q_0^2 S')^{\frac{1}{2}}} dr. \qquad (3.14)$$

Now since in the asymptotic region u_0 and u_0' are $\tfrac{1}{2}\pi$ out of phase and a_0 approaches unity, we see that $b_1(\infty)$ is the first-order correction to the WKB phase shift.

We can write b_1 in a more convenient form. If we let

$$Q_0(S) = -q_0^2(S), \qquad Q(\rho) = -q^2(\rho),$$

we find

$$2b_1(\infty)$$
$$= \frac{1}{12} \int_{\Lambda}^{\infty} ds \frac{1}{Q_0^{\frac{1}{2}}} \frac{d}{ds} \left(\frac{Q_0''}{Q_0'} \right) - \frac{1}{12} \int_{\rho_t}^{\infty} d\rho \frac{1}{Q^{\frac{1}{2}}} \frac{d}{d\rho} \left(\frac{Q''}{Q'} \right)$$
$$+ \frac{5}{24} \left\{ \frac{Q_0'}{Q_0^{\frac{3}{2}}} \Big|_{\Lambda}^{\infty} - \frac{Q'}{Q^{\frac{3}{2}}} \Big|_{\rho_t}^{\infty} \right\} - \frac{1}{12} \left\{ \frac{Q_0''}{Q_0' Q_0^{\frac{1}{2}}} \Big|_{\Lambda}^{\infty} - \frac{Q''}{Q' Q^{\frac{1}{2}}} \Big|_{\rho_t}^{\infty} \right\}.$$

The integrals are convergent, but the integrated terms must be considered carefully. To investigate the behavior at the lower limit, we expand Eq. (3.8) about the turning point and find that in the neighborhood of ρ_t

$$S = \Lambda + C(\rho - \rho_t)$$
$$+ C \left[\frac{1}{10} \frac{Q_t''}{Q_t'} + \frac{3}{10} \frac{C}{\Lambda} \right] (\rho - \rho_t)^2 + \cdots , \qquad (3.16)$$

where

$$C^3 = \tfrac{1}{2}\Lambda Q_t'.$$

From this, it is easy to show that the integrated

terms cancel. We therefore have, after performing the integration over S,

$$b_1(\infty) = \frac{1}{8\Lambda}\frac{1}{2}\pi - \frac{1}{24}\int_{\rho_t}^{\infty}d\rho\,\frac{1}{Q^{\frac{3}{2}}}\frac{d}{d\rho}\left(\frac{Q''}{Q'}\right). \qquad (3.17)$$

It is remarkable that this reduces to an integral over ρ of only explicitly known functions of the potential. This was unfortunately not the case with the Dirac equation.

Note that Eq. (3.17) does not hold for S waves. Here the comparison function is purely oscillatory and of a very different character from u when a turning point exists. Indeed, we find that S' vanishes at the turning point and is no longer everywhere real; also a_0 and b_1 are singular at ρ_t.

As we have seen, the lowest order yields just the usual WKB approximation to the phase shift. Now an old rule of thumb, justified by Langer,[7] states that an improved result is obtained if one replaces $l(l+1)$ by $(l+\frac{1}{2})^2$. In order to introduce the $(l+\frac{1}{2})^2$ in a natural way, we make the following transformations

$$\rho = e^x \qquad -\infty \le x \le \infty$$

and then

$$u = e^{\frac{1}{2}x}w(x),$$

where $w(x)$ satisfies the equation

$$w'' + e^{2x}[1 - v(e^x) - (l+\tfrac{1}{2})^2 e^{-2x}]w = 0. \qquad (3.18)$$

Proceeding as with Eq. (3.2), we obtain

$$P_0(S)[S'(x)]^2 = P(x), \qquad (3.19)$$

$$\bar{a}_0(x) = (S')^{-\frac{1}{2}}, \qquad (3.20)$$

$$\bar{a}_0'' - 2(P_0 S')^{\frac{1}{2}}\frac{d}{dx}[\bar{b}_1(P_0 S')^{\frac{1}{2}}] = 0, \qquad (3.21)$$

where

$$P_0 = e^{2s}[1 - (l+\tfrac{1}{2})^2 e^{-2s}],$$
$$P = e^{2x}[1 - v(e^x) - (l+\tfrac{1}{2})^2 e^{-2x}],$$

and where we denote the $(l+\frac{1}{2})$ phase shifts and coefficients by a bar. Setting $\sigma = e^S$, we obtain from Eq. (3.19)

$$\int_{\bar{\Lambda}}^{\sigma}\left(1 - \frac{\bar{\Lambda}^2}{s^2}\right)^{\frac{1}{2}}ds = \int_{\rho_t}^{\rho}\left[1 - v(\rho) - \frac{\bar{\Lambda}^2}{\rho^2}\right]^{\frac{1}{2}}d\rho, \qquad (3.22)$$

where

$$\bar{\Lambda} = (l+\tfrac{1}{2});$$

and from Eq. (3.21)

[7] R. E. Langer, Phys. Rev. **51**, 669 (1937).

$$2\bar{b}_1(\infty) = \frac{1}{12}\left\{\int_{\ln\bar{\Lambda}}^{\infty}ds\,\frac{1}{P_0^{\frac{3}{2}}}\frac{d}{dS}\left(\frac{P_0''}{P_0'}\right)\right.$$
$$\left. - \int_{x_t}^{\infty}dx\,\frac{1}{P^{\frac{3}{2}}}\frac{d}{dx}\left(\frac{P''}{P'}\right)\right\}.$$

The first integral vanishes since $P_0'' = 2P_0'$; hence

$$\bar{b}_1(\infty) = -\frac{1}{24}\int_{x_t}^{\infty}dx\,\frac{1}{P^{\frac{3}{2}}}\frac{d}{dx}\left(\frac{P''}{P'}\right). \qquad (3.23)$$

The lowest-order phase shift is obtained from the asymptotic behavior of σ, [since $w_0(S) \sim J_{l+\frac{1}{2}}(\sigma)$], and is seen to be the same as the one previously obtained except that $l(l+1)$ has been replaced by $(l+\frac{1}{2})^2$. As a consequence, $\bar{\eta}^{(0)}$ is independent of the index of the comparison functions. Expressing $\bar{b}_1(\infty)$ explicitly as a function of the potential we find

$$\bar{b}_1(\infty) = -\frac{1}{24}\int_{\rho_t}^{\infty}d\rho\,\frac{1}{\rho Q^{\frac{3}{2}}}\frac{d}{d\rho}$$
$$\times\left(\frac{4(1-v) - 5\rho v' - \rho^2 v''}{2(1-v) - \rho v'}\right), \qquad (3.24)$$

where

$$Q(\rho) = 1 - v(\rho) - (l+\tfrac{1}{2})^2/\rho^2.$$

It is clear that the previous difficulty with S waves is not encountered here.

C. Point Coulomb Scattering

Consider Eq. (3.11) with

$$-q_0^2 = 1 - \frac{\Lambda^2}{S^2}, \qquad -q^2 = 1 - \frac{2\alpha}{\rho} - \frac{\Lambda^2}{\rho^2}, \qquad \alpha = \frac{ZZ'}{\hbar v}.$$

We find

$$S \sim \rho - \alpha\ln 2\rho - \alpha$$
$$+ \tfrac{1}{2}\alpha\ln(\alpha^2 + \Lambda^2) + \Lambda\tan^{-1}(\alpha/\Lambda),$$

and consequently,

$$\eta_C^{(0)} = -\alpha + \tfrac{1}{2}\alpha\ln(\alpha^2 + \Lambda^2)$$
$$+ \Lambda\tan^{-1}(\alpha/\Lambda); \qquad (3.25)$$

it is also easy to show that

$$\bar{\eta}_C^{(0)} = -\alpha + \tfrac{1}{2}\alpha\ln(\alpha^2 + \bar{\Lambda}^2)$$
$$+ \bar{\Lambda}\tan^{-1}(\alpha/\bar{\Lambda}). \qquad (3.26)$$

Now the exact Coulomb phase shift is given by

$$\eta_C = \arg\Gamma(l+1+i\alpha);$$

if we employ Stirling's approximation, we find

$$\eta_C\text{ (Stirling)} = \bar{\eta}^{(0)} + \frac{1}{8}\frac{\alpha}{\bar{\Lambda}^2 + \alpha^2} + o\left(\frac{\alpha}{\bar{\Lambda}^2 + \alpha^2}\right).$$

Including the next term in the asymptotic expansion of the gamma function, we obtain an additional contribution of $-\alpha/12(\bar{\Lambda}^2 + \alpha^2)$; thus

η_e ("extended" Stirling)

$$= \bar{\eta}^{(0)} + \frac{1}{24}\left(\frac{\alpha}{\bar{\Lambda}^2 + \alpha^2}\right) + o\left(\frac{\alpha}{\bar{\Lambda}^2 + \alpha^2}\right). \quad (3.27)$$

When $\bar{b}_1(\infty)$ is evaluated, we indeed find

$$\bar{b}_1(\infty) = \tfrac{1}{24}[\alpha/(\bar{\Lambda}^2 + \alpha^2)].$$

It is also easy to show that

$$b_1(\infty) = \frac{1}{24}\frac{\alpha}{\Lambda^2 + \alpha^2} + \frac{1}{8\Lambda}\tan^{-1}\frac{\alpha}{\Lambda}.$$

In Table IV, values of the exact and approximate point Coulomb phase shifts are given for special values of l. One may note that the $\eta^{(1)}$ and $\bar{\eta}^{(1)}$ terms bring substantial improvement in the WKB phase shifts and also that the Langer-type approximation is the better of the two. As one would expect, the approximation improves as l or α increases.

4. Further Applications

The modified WKB approximation outlined above can be applied, of course, to potentials other than the ones we have used. For instance, no essential changes need be made for a spin–orbit potential. The only difference would be that the potential would now be a function of l and j, the orbital and total angular momentum quantum numbers, respectively.

If however, $V(r)$ is no longer real, then the wave-function $u(\rho)$ is no longer required to be real, nor is the classical turning point on the real axis. Our discussion must therefore be suitably modified. Nodvik[2] has treated Eq. (3.6) (which he obtained in a somewhat different manner) in some detail for the case of a complex well, and so we restrict ourselves to a few descriptive remarks.

We allow ρ and S to assume complex values and understand the integrals in Eq. (3.11) to be contour integrals in the complex ρ and S planes, respectively; the square roots and integration paths can be defined in some convenient manner. In order that in the limit of real potentials the method reduce to that described above, the lower limits of the integrals are still to be defined by the zeros of q^2 and q_0^2, respectively—but here complications arise. $V(r)$, in general, is not analytic along the real r axis. Also,

TABLE IV. Nonrelativistic Point Coulomb phase shifts. The parameter of the Coulomb potential is $\alpha = ZZ'e^2/\hbar v$. The two different WKB approximations defined in Sec. 3 are shown.

l	α	WKB $\eta^{(0)}$	Corrected WKB $\eta^{(0)} + \eta^{(1)}$	Exact
1	$1/8$	0.04348	0.05386	0.05298
	$1/2$	0.18333	0.22263	0.21959
	2	1.14278	1.24111	1.23680
	8	10.73262	10.86105	10.85695
2	$1/8$	0.11204	0.11551	0.11540
	$1/2$	0.45137	0.46498	0.46457
	2	1.97980	2.02308	2.02219
	8	12.11383	12.18359	12.18277
l	α	$\bar{\eta}^{(0)}$	$\bar{\eta}^{(0)} + \bar{\eta}^{(1)}$	Exact
0	$1/8$	-0.08537	-0.06576	-0.07138
	$1/2$	-0.28059	-0.23892	-0.24406
	2	$+0.10983$	$+0.12944$	$+0.12965$
	8	9.40532	9.41050	9.41051
1	$1/8$	0.05083	0.05313	0.05298
	$1/2$	0.21170	0.22003	0.21959
	2	1.22352	1.23685	1.23680
	8	10.85191	10.85694	10.85695
2	$1/8$	0.11459	0.11542	0.11540
	$1/2$	0.46144	0.46465	0.46457
	2	2.01413	2.02226	2.02219
	8	12.17802	12.18276	12.18277

q^2 in general has more than one zero in the complex plane. Nodvik points out, however, that $V(r)$ usually is analytic on the real r axis except at a finite number of isolated points which thus divide the axis into a finite number of intervals; that in each interval there is a $V(r)$ defined which can be analytically continued into the complex r plane. He takes as r_t the zero of that particular q^2 which lies closest to the real axis. If the imaginary part of $V(r)$ is small, this will be near the "real" turning point. Further details may be found in Nodvik's report.

Particle Moderation: Random Functional Approach

Martin A. Leibowitz

Bellcomm Inc., Washington, D. C.

(Received 3 April 1964; final manuscript received 19 June 1964)

The slowing down of a particle by a homogeneous isotropic moderator is considered. It is shown that finding the collision density of the particle as a function of space and energy is equivalent to finding the probability distribution of a certain random functional. By means of this random functional, expressions for the spatial moments are obtained without imposing any restrictions on the variations of the scattering kernel or cross section with energy. These moments are then used to obtain the age equation, the derivation given here differing from others in that no *a priori* assumptions are made on the collision density itself. Finally, as a special case of the above, the time–energy moments are found.

1. INTRODUCTION

IN an earlier paper, random functionals were used to derive the moments of the neutron time–energy distribution.[1] The methods presented there can however be considerably simplified and extended. It is the purpose of this paper to make these extensions to the case where the spatial distribution of the particle (as a function of energy) is the quantity of interest. The class of scattering kernels and cross-sections covered is much larger than that allowed for in Ref. 1 (main restriction: infinite homogeneous medium at 0° temperature), and in fact is sufficiently broad that our more general title—"particle moderation" rather than just "neutron moderation"—appears justified.

Stochastic methods have been relatively ignored in recent years, moderation theory relying chiefly on the transport equation. Nevertheless, the classic paper of Fermi[2] used a stochastic approach to find the mean-square distance $\langle r^2(E)\rangle$ for a neutron to slow down to a given energy E with a hydrogen moderator. The determination of $\langle r^2(E)\rangle$ for a moderator with atomic mass greater than unity, however, has been done through a solution of the important "moment" equations of transport theory.[3]

While a general analytic solution of the moment equations has not yet been given, extensive computer programs for their numerical solution have been written.[4] For the one-dimensional case, $\frac{1}{2}[n(n+1)]$ coupled equations must be solved to obtain the first n moments $\langle x^i(E)\rangle (j = 1, \cdots, n)$ of the distance to slow down to energy E. Recently, group-theoretic techniques have been applied to neutron moderation notably by Guth and Inönü[5] and by Wigner.[6]

We proceed here, as in Ref. 1 by showing that finding the particle collision density as a function of space and energy (lethargy) is equivalent to finding the distribution of a certain random functional. This is based on the assumption of an infinite, homogeneous moderator at rest consisting of nuclei of only one kind (this last condition is removed in the Appendix), but no assumption is made on the cross section (scattering and absorbing) or on the scattering kernel (it need not be isotropic in the center of mass frame). In Sec. 6 we find the moments of this functional and hence of the collision density. The results are in the following form: each moment is expressed recursively in terms of the previous moments by means of an integral over lethargy. In the equations for the first n moments there appears n functions of lethargy which can be given as solutions of n integral equations. But in contrast, to the moment equations (which are $\frac{1}{2}[n(n+1)]$ in number), these equations are uncoupled and under special conditions (e.g., scattering-angle distribution independent of energy, scattering and absorption cross-sections proportional) permit an analytic solution. Moreover, since we do *not* require the particles' source direction to be isotropically distributed, these results should give information on the slowing down of a beam of particles.

In Sec. 7, the expressions for the moments are used to obtain the age equation. Though this result is well known, the proof given here differs from the usual one in that no prior assumptions are made concerning the collision density itself (e.g., it is not assumed that only the first two terms of the Legendre polynomial expansion of the collision density are of significance in passing to the age limit: this fact follows from the proof). Finally, in Sec. 8, we return to the time–energy moments which can now be readily found as a special case of the space moments.

It is of course true that any result obtained through the use of random functionals is, in princi-

[1] M. A. Leibowitz, J. Math. Phys. **4**, 446 (1963).
[2] E. Fermi, Ric. Sci., **7**, 13 (1936).
[3] R. E. Marshak, Rev. Mod. Phys. **19**, 185 (1947).
[4] H. Goldstein, *Fundamental Aspects of Reactor Shielding*, (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1959), Chap. 6.
[5] E. Guth and E. Inönü, J. Math. Phys. **2**, 451 (1961).
[6] E. Wigner, Phys. Rev. **94**, 17 (1954).

ple, obtainable by means of the transport equation.[7] The two methods are after all equivalent. Nevertheless, the use of random functionals can claim at least two advantages: (1) It is less formal in remaining closer to the physical situation, which is essentially a random walk; (2) It is keeping with a trend which has appeared in other parts of Mathematical Physics; namely, the formulation of problems in terms of random functionals.[8]

2. THE SCATTERING KERNEL

We consider an infinite, homogeneous isotropic moderator consisting of free atoms at rest, and suppose that at time $t = 0$, a particle with velocity v_0 enters the medium, at a point which will be taken to be at the origin of a rectangular coordinate system; through collisions with the atoms the particle will slow down. Let v be the particle velocity at any time and $u = \ln(v_0/v)$ its lethargy, so that u is a *nondecreasing* function of the time.

At scattering collisions, particles suffer an instantaneous change in direction and lethargy, the random nature of the change being fully described by the scattering kernel $K(\Omega \cdot \Omega', u, u')$ giving the probability density for a particle to have a direction of motion along the unit vector Ω and a lethargy u after collision assuming that it had direction Ω' and lethargy u' before collision. The scattering kernel may be factored as

$$K(\Omega \cdot \Omega', u, u') = K(\Omega \cdot \Omega' \mid u, u')f(u \mid u'), \quad (u' \leq u). \tag{1}$$

In (1), $f(u \mid u')$ is the probability density of a particle having lethargy u after collision given that it had lethargy u' before, and $K(\Omega \cdot \Omega \mid u, u')$ is then the probability density of Ω given u, u', and Ω'. Note that the scattering kernel, because of the isotropy of the medium depends on Ω and Ω' only through the cosine of the scattering angle θ.

For neutron moderation, which is the case uppermost in mind in this paper, the most important scattering kernel is the one resulting when the additional assumptions of elastic scattering and isotropic scattering in the center-of-mass frame are imposed. Then, if M is the atomic mass of the scattering nucleus, we have

$$f(u - u')$$
$$= \begin{cases} \dfrac{(M + 1)^2}{4M} e^{-(u-u')} & 0 \leq \Delta u \leq \ln\left(\dfrac{M + 1}{M - 1}\right)^2, \\ 0 & \text{otherwise,} \end{cases}$$

[7] E. Guth and E. Inönü, Phys. Rev. **118**, 899 (1960).
[8] I. M. Gelfand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).

$$K(\Omega \cdot \Omega' \mid u, u')$$
$$= \delta[\Omega \cdot \Omega' - M \sinh \tfrac{1}{2}(u - u') - \cosh \tfrac{1}{2}(u - u')]. \tag{2}$$

Note that in (2) the quantities u and u' enter only through their difference $\Delta u = u - u'$; this is the main advantage in introducing the lethargy variable. Implicit in the above is that there is only one kernel corresponding to the assumption of only one type of scatterer; this will be taken to be true for the present though it will be removed later.

The quantity of chief interest in transport theory is $F(\mathbf{r}, u, \Omega)$ the average number of particle collisions per unit lethargy and per unit solid angle and volume. For our purposes, $F(\mathbf{r}, u, \Omega)$ must be reinterpreted since its definition involves the phrase "number of particles" while here the functions of interest are probability densities over the ensemble of all possible ways a given particle may slow down after leaving a source.

Let $W(u \mid u')du$ be the probability that a particle with lethargy u' will at some later moment have a lethargy in the small interval $(u, u + du)$. For a particle leaving the origin with lethargy zero (i.e., $u' = 0$) $W(u \mid 0)$ is just the normalization factor for $F(\mathbf{r}, u, \Omega)$:

$$W(u \mid 0) = \int F(\mathbf{r}, u, \Omega) \, d\mathbf{r} \, d\Omega. \tag{3}$$

Hence

$$F(\mathbf{r}, \Omega \mid u) \equiv F(\mathbf{r}, u, \Omega)/W(u \mid 0) \tag{4}$$

is a probability density and has the following meaning: $F(\mathbf{r}, \Omega \mid u)d\mathbf{r}d\Omega$ is the probability that a particle with lethargy u while moving with a direction in the solid angle $[\Omega, \Omega + d\Omega]$ will make a collision in the volume $[\mathbf{r}, \mathbf{r} + d\mathbf{r}]$, *under the condition* that the particle does in fact have the lethargy u at some moment in its life. The simpler probability densities

$$F(\mathbf{r} \mid u) = \int F(\mathbf{r}, \Omega \mid u) \, d\Omega, \tag{5}$$

$$F(\Omega \mid u) = \int F(\mathbf{r}, \Omega \mid u) \, d\mathbf{r} \tag{6}$$

are also of importance and easier to obtain.

For the special case of the kernel (2), $W(u \mid u')$ was first introduced by Placzek.[9] In general, $W(u \mid u')$ satisfies the integral equation (assuming no absorption)

$$W(u \mid u') = \delta(u - u')$$
$$+ \int_{u'}^{u} f(u \mid u'')W(u'' \mid u') \, du'', \tag{7}$$

[9] G. Placzek, Phys. Rev. **69**, 423 (1946).

which simplifies to

$$W(u) = \delta(u) + \int_0^u f(u - u')W(u')\,du' \qquad (8)$$

when $f(u \mid u')$, and hence $W(u \mid u')$ depend only on the difference $u - u'$, i.e., $W(u \mid u') = W(u - u')$.

Equation (7) follows on noting that

$$W(u \mid u') = \sum_{j=0}^{\infty} W_j(u \mid u'),$$

where $W_j(u \mid u')$ is the probability density of the lethargy j collisions after the lethargy u' is attained. But

$$W_j(u \mid u') = \int_{u'}^{u} f(u \mid u'')W_{j-1}(u'' \mid u')\,du''.$$

3. RANDOM FUNCTIONAL FORMULATION[10]

Let $\mathbf{r}_j$ denote the position of the particle when it makes its jth collision so that

$$\mathbf{r}_j = \Delta\mathbf{r}_1 + \Delta\mathbf{r}_2 + \cdots + \Delta\mathbf{r}_j, \qquad (9)$$

where $\Delta\mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$. To make the dependence of $\mathbf{r}_j$ on the scattering cross section $\Sigma(u)$ explicit, we introduce random variables $\Delta\varphi_i$ related to $\Delta\mathbf{r}_i$ by

$$\mathbf{r}_i = \Delta\varrho_i / \Sigma(u_{i-1}), \qquad (10)$$

where u_i is the lethargy just after the ith collision. Then

$$\mathbf{r}_j = \frac{\Delta\varrho_1}{\Sigma(u_0)} + \frac{\Delta\varrho_2}{\Sigma(u_1)} + \cdots + \frac{\Delta\varrho_j}{\Sigma(u_{j-1})}. \qquad (11)$$

The variable $\Delta\varrho_i$ has a simple physical meaning: It is the change in position between the $(i - 1)$st and ith collision in a fictitious medium such that for all u $\Sigma(u) = 1$. $\Delta\varrho_i$, in turn, may be factored into

$$\Delta\varrho_i = \Omega_{i-1}\Delta l_i, \qquad (12)$$

with Ω_i being a unit vector in the direction of the particles' motion just after the ith collision, and Δl_i the distance (path length) traveled between the $(i - 1)$st and ith collision [if $\Sigma(u) = 1$]. The Δl_i are independent random variables with density $e^{-\Delta l}$. Then (11) becomes

$$\mathbf{r}_j = \frac{\Omega_0\Delta l_1}{\Sigma(u_0)} + \frac{\Omega_1\Delta l_2}{\Sigma(u_1)} + \cdots + \frac{\Omega_{j-1}\Delta l_j}{\Sigma(u_j)}. \qquad (13)$$

To proceed, it will be necessary to write (13) as an integral over lethargy. Define the random function $l(u')$ by

$$\begin{aligned} l(u') &= 0, \qquad u' = 0, \\ l(u') &= l_i, \qquad u_{i-1} < u' \le u_i, \end{aligned} \qquad (14)$$

[10] We restrict ourselves here to a moderator containing only one type of scatterer with no absorption.

where $l_i = \Delta l_1 + \Delta l_2 + \cdots + \Delta l_i$. $l(u')$ would be, if $\Sigma(u) \equiv 1$, just the distance traveled by the particle until it makes the collision which slows it down past the lethargy u'.

In addition, define the random function $\Omega(u')$ by

$$\begin{aligned} \Omega(u') &= \Omega_i, \qquad (u_i - \epsilon) \le u' \le (u_i + \epsilon), \\ \Omega(u') &= 0 \qquad \text{otherwise.} \end{aligned} \qquad (15)$$

Here ϵ may be any small positive number, greater than zero, chosen such that the intervals $u_i - \epsilon \le u \le u_i + \epsilon$ do not overlap. As will be seen, $\Omega(u')$ may be entirely arbitrary outside these intervals, but for definiteness we will suppose that it vanishes.

Now, assume that $u_{j-1} = u$, and write $\mathbf{r}_j$ as $\mathbf{r}_j(u)$; then

$$\mathbf{r}_j(u) = \int_0^u \frac{1}{\Sigma(u')}\Omega(u')\,dl(u'), \qquad (16)$$

where the integral *includes* the contributions due to the jumps at $u' = 0$ and $u' = u$.

To see the equivalence of (13) and (16) one need only observe that the function $l(u')$ has jumps of magnitude $\Delta l_{i+1} = l_{i+1} - l_i$ at the points $u' = u_i$, and at these points $\Omega(u_i) = \Omega_i$.

Note that in Eq. (16), the number of collisions, j, made by the particle no longer appears on the right side. Let us then drop the requirement that the lethargy u be attained at the jth collision and merely demand that there be some collision just after which the particle has the lethargy u. Then if $\mathbf{r}(u)$ is the position of such a particle when it makes its next collision, we have [dropping the subscript j in (16)]

$$\mathbf{r}(u) = \int_0^u \frac{1}{\Sigma(u')}\Omega(u')\,dl(u'), \qquad (17)$$

and $\mathbf{r}(u)$ has by definition the probability density $F(\mathbf{r} \mid u)$. The moments of the random variable $\mathbf{r}(u)$ may be obtained from (17):

$$\langle \mathbf{r}(u)\mathbf{r}(u) \cdots \mathbf{r}(u) \rangle = \int_0^u \cdots \int_0^u \frac{1}{\Sigma(u')} \cdots \frac{1}{\Sigma(u^{(n)})}$$
$$\times \langle \Omega(u') \cdots \Omega(u^{(n)}) \rangle \langle dl(u') \cdots dl(u^{(n)}) \rangle, \qquad (18)$$

where $\langle \mathbf{r}(u)\mathbf{r}(u) \cdots \mathbf{r}(u) \rangle$ represents any n-fold product of the three components of $\mathbf{r}(u)$, and similarly for $\langle \Omega(u') \cdots \Omega(u^{(n)}) \rangle$. For simplicity we restrict ourselves to the one-dimensional case. Then if μ denotes the x component of the vector Ω, (17) and (18) become

$$x(u) = \int_0^u \frac{1}{\Sigma(u')}\mu(u')\,dl(u') \qquad (19)$$

and

$$\langle x^n(u) \rangle = \int_0^u \cdots^n \int_0^u \frac{1}{\Sigma(u')} \cdots \frac{1}{\Sigma(u^{(n)})}$$

$$\times \langle \mu(u') \cdots \mu(u^{(n)}) \rangle \langle dl(u') \cdots dl(u^{(n)}) \rangle, \qquad (20)$$

with the probability density of $x(u)$ being denoted by $F(x \mid u)$.

Before any use can be made of (20) however, the quantities $\langle dl(u') \cdots dl(u^{(n)}) \rangle$ and $\langle \mu(u') \cdots \mu(u^{(n)}) \rangle$, must be obtained in an explicit form (which we proceed to do).

4. THE MOMENTS $\langle dl(u') \cdots dl(u^{(n)}) \rangle$

Consider first the quantity $\langle dl(u') \rangle$. We have

$$\langle dl(u') \rangle = \lim_{du' \to 0} \frac{\langle l(u' + du') - l(u') \rangle}{du'} du'. \qquad (21)$$

$l(u' + du') - l(u')$ vanishes unless the particle slows down past u' and $u' + du'$ at different collisions; or equivalently, unless the particle has, after some collision, a lethargy in the interval $(u', u' + du')$. Let us denote the probability of this event by $P(u \mid u' \mid 0)du'$ indicating by the notation that this probability is taken only over particles which at some moment have had the lethargy zero and would have the lethargy u at a later time. Because of this last condition, $P(u \mid u' \mid 0)$ is not given by $W(u' \mid 0)$ but rather by

$$P(u \mid u' \mid 0) = \frac{W(u \mid u')W(u' \mid 0)}{W(u \mid 0)}. \qquad (22)$$

Hence, since the mean distance between a collision if $\Sigma(u) \equiv 1$ is simply unity, one has

$$\langle dl(u') \rangle = \frac{W(u \mid u')W(u' \mid 0)}{W(u \mid 0)} du'. \qquad (23)$$

Now, let us consider $\langle dl(u')dl(u'') \rangle$ for $u'' \leq u'$. One has

$$\langle dl(u') \, dl(u'') \rangle$$

$$= k \frac{W(u \mid u')W(u' \mid u'')W(u'' \mid 0) \, du' \, du''}{W(u \mid 0)}, \quad (u' \geq u'') \qquad (24)$$

where

$$k = \begin{cases} 2 & u' = u'' \\ 1 & u' \neq u''. \end{cases}$$

In fact, for $u' \neq u''$ Eq. (24) is simply the probability of the particle having lethargies in the intervals $(u' + du')$, $(u'' + du'')$ [otherwise $dl(u')dl(u'')$ vanishes] given that it initially had lethargy zero and would later have lethargy u; multiplied by the

product of the mean distances traveled to the next collision when the particle has lethargy u'' and when it has lethargy u'. But this factor k is just unity. When $u'' = u'$, however, k is not the product of the mean of two different "intercollision" distances, but rather the mean square of a single intercollision distance:

$$k = \int e^{-\Delta l}(\Delta l)^2 \, d\Delta l = 2.$$

In general,

$$\langle dl(u') \, dl(u'') \cdots dl(u^{(n)}) \rangle$$

$$= k_1 ! \, k_2 ! \cdots k_j ! \, W(u \mid u')W(u' \mid u'') \cdots$$

$$\times W(u^{(n-1)} \mid u^{(n)})W(u^{(n)} \mid 0) \, du' \, du'' \cdots du^{(n)},$$

$$(k_1 + k_2 + \cdots + k_j = n; u' \geq u'' \geq \cdots \geq u^{(n)}), \qquad (25)$$

where j is the number of distinct values occurring among the numbers $u', u'', \cdots, u^{(n)}$ and k_i is the number of times the ith value appears. In obtaining (25) one uses the fact that

$$\int (\Delta l)^k e^{-(\Delta l)} \, d\Delta l = k!. \qquad (26)$$

5. THE MOMENTS $\langle \mu(u') \cdots \mu(u^{(n)}) \rangle$

Let $D(\mu(u) \mid \mu(u'))$ be the probability density of the direction $\mu(u)$ of a particle just after it obtains the lethargy u given that it had direction $\mu(u')$ at some previous lethargy $u'(u' \leq u)$. Implicit here is the condition that the particle does obtain the lethargy u at some collision.

Suppose that the probability density of $\mu(u')$ is $A(\mu)$[11] [if $\mu(u')$ is specified in advance then $A(\mu)$ reduces to a delta function]. After a collision at which the lethargy u_1 is attained, the particle's direction will have a different probability density $A_1(\mu \mid u_1; u')$ where

$$A_1(\mu \mid u_1; u') = \int \frac{1}{2\pi} K(\mathbf{\Omega} \cdot \mathbf{\Omega}' \mid u_1, u')A(\mu) \, d\Omega'. \qquad (27)$$

We introduce the Legendre polynomial expansions

$$K(\mathbf{\Omega} \cdot \mathbf{\Omega}' \mid u, u') = \sum_{n=0}^{\infty} \frac{2n + 1}{2} K_n(u, u')P_n(\mathbf{\Omega} \cdot \mathbf{\Omega}'), \qquad (28)$$

$$A(\mu) = \sum_{n=0}^{\infty} \frac{2n + 1}{2} a_n P_n(\mu). \qquad (29)$$

Substituting from (28) and (29) into (27), using the addition theorem for Legendre polynomials and

---

[11] We suppose the particle's initial direction has a symmetric distribution around the x axis: $A(\mu) = 1/2\pi \, A(\mathbf{\Omega})$.

integrating we find

$$A_1(\mu \mid u_1; u') = \sum_{n=0}^{\infty} \frac{2n+1}{2} K_n(u_1, u')a_n P_n(\mu). \quad (30)$$

Similarly, if $A_j(\mu \mid u_1, u_2, \cdots u_j; u')$ is the probability density of μ after j collisions giving the particle lethargies $u_1, u_2, \cdots, u_j$ (starting with lethargy u'),

then repeated use of (27) and (30) gives

$$A_j(\mu \mid u_1, u_2, \cdots, u_j; u') = \sum_{n=0}^{\infty} \frac{2n+1}{2}$$

$$\times [K_n(u_1, u')K_n(u_2, u_1) \cdots K_n(u_j, u_{j-1})]a_n P_n(\mu). \quad (31)$$

Hence, letting $u_i = u$ and defining $A_0(\mu) = A(\mu) = \delta(\mu - \mu(u'))$,

$$D(\mu(u) \mid \mu(u')) = \sum_{i=0}^{\infty} \frac{W_i(u \mid u')}{W(u \mid u')} \left[\frac{\int A_i(\mu \mid u_1, u_2, \cdots u_{i-1}, u; u')f(u_1, u') \cdots f(u, u_{i-1}) \, du_1 \cdots du_{i-1}}{W_i(u \mid u')} \right]. \quad (32)$$

The expression within the bracket is just the probability density of $\mu(u)$ conditional on the lethargy u being obtained j collisions after lethargy u, and the probability of this event is $W_j(u \mid u')/W(u \mid u')$. Substituting from (31) to (32) one finds that

$$D(\mu(u) \mid \mu(u'))$$

$$= \sum_{n=0}^{\infty} \frac{2n+1}{2} B_n(u \mid u')P_n(\mu) \Big/ W(u \mid u'), \quad (33)$$

where

$$B_n(u \mid u') = \delta(u - u') + f(u, u')K_n(u, u')$$

$$+ \sum_{i=1}^{\infty} \int f(u_1, u') \cdots f(u, u_i)K_n(u_1, u') \cdots$$

$$\times K_n(u, u_i) \, du_1 \cdots du_i.$$

Thus $B_n(u \mid u')$ is the solution of the integral equation,

$$B_n(u \mid u') = \delta(u - u')$$

$$+ \int_{u'}^{u} f(u, u'')K_n(u, u'')B_n(u'' \mid u') \, du'', \quad (34)$$

which simplifies to

$$B_n(u) = \delta(u - u')$$

$$+ \int_0^u f(u - u')K_n(u - u')B_n(u') \, du' \quad (35)$$

in the special case of a kernel like (2) which is a function only of $\Delta u = u - u'$, since then $B_n(u \mid u') = B_n(u - u')$.

The calculation of $\langle \mu(u') \cdots \mu(u^{(n)}) \rangle$ is most elegantly performed by the use of matrices. Let $\mathbf{P}(\mu)$ be the infinite-dimensional vector whose components are just the Legendre polynomials $P_n(\mu)$. Then, rather than just $\langle \mu(u')\mu(u'') \cdots \mu(u^{(n)}) \rangle$, we consider the more-general vector quantity

$$\langle \mathbf{P}[\mu(u)]\mu(u') \cdots \mu(u^{(n)}) \rangle,$$

where $P[\mu(u)]$ denotes the vector $P(\mu)$ evaluated at $\mu = \mu(u)$. The introduction of this vector moment is not only a matter of convenience, but permits us to obtain information on $F(x, \mu \mid u)$ instead of just $F(x \mid u)$.

In what follows, we use the fact that the random variables $\mu(u)$, $\mu(u')$, $\cdots$, $\mu(u^{(n)})$ are part of a Markoff process: this means that the probability distribution of $\mu(u^{(i-1)})$ given $\mu(u^{(i)})$ is independent of $\mu^{(j)}$, $\cdots$, $\mu^{(n)}$. Thus, let us first find

$$\langle \mathbf{P}[\mu(u)] \mid \mu(u') \rangle,$$

the average value of $\mathbf{P}[\mu(u)]$ given $\mu(u')$. Multiplying $D(\mu(u) \mid \mu(u'))$ by $P_n(\mu)$ integrating and using the orthogonality conditions, we find that

$$\langle P_n[\mu(u)] \mid \mu(u') \rangle = B_n(u \mid u')P_n[\mu(u')]/W(u \mid u'). \quad (36)$$

In matrix form,

$$\langle \mathbf{P}[\mu(u)] \mid \mu(u') \rangle = \mathbf{B}(u \mid u')\mathbf{P}[\mu(u')]/W(u \mid u'),$$

where $\mathbf{B}(u \mid u')$ is a diagonal matrix with elements $B_n(u \mid u')$. To obtain $\langle \mathbf{P}[\mu(u)]\mu(u') \mid \mu(u'') \rangle$ we multiply the right-hand side of Eq. (36) by $\mu(u')$. From the recursion formula for Legendre polynomials,

$$(i+1)P_{i+1}(\mu) - (2i+1)\mu P_i(\mu) + iP_{i-1}(\mu) = 0, \quad (37)$$

it follows that the vector $\mu \mathbf{P}(\mu)$ may be written

$$\mu \mathbf{P}(\mu) = \mathbf{A}\mathbf{P}(\mu), \quad (38)$$

where $\mathbf{A}$ is a matrix with elements a_{ij} given by

$$a_{ij} = \delta_{i,j-1}\left(\frac{i+1}{2i+1}\right) + \delta_{i,j+1}\left(\frac{i}{2i+1}\right)$$

$$(i, j = 0, 1, 2, \cdots). \quad (39)$$

Thus,

$$\langle \mathbf{P}[\mu(u)]\mu(u') \mid \mu(u'') \rangle$$

$$= \mathbf{B}(u \mid u')\mathbf{A}\langle \mathbf{P}[\mu(u')] \mid \mu(u'') \rangle/W(u \mid u')$$

$$= \mathbf{C}(u \mid u')\langle \mathbf{P}[\mu(u')] \mid \mu(u'') \rangle/W(u \mid u'), \quad (40)$$

where

$$\mathbf{C}(u \mid u') = \mathbf{B}(u \mid u')\mathbf{A} \quad (41)$$

is given by

$$
\mathbf{C}(u \mid u') = \begin{bmatrix}
0 & B_0 & 0 & 0 & \cdots & & \cdots & & \cdots \\
\frac{1}{3}B_1 & 0 & \frac{2}{3}B_1 & 0 & \cdots & & \cdots & & \cdots \\
& \frac{2}{5}B_2 & 0 & \frac{3}{5}B_2 & \cdots & & \cdots & & \cdots \\
& & \cdots & & & \cdots & & & \cdots \\
& & & \cdots & & & \cdots & & \cdots \\
& & & & [i/(2i+1)]B_i & 0 & [(i+1)/(2i+1)]B_i & 0 & \\
& & & & \cdots & & \cdots &
\end{bmatrix}, \quad (42)
$$

with $B_i = B_i(u \mid u')$. The elements c_{ij} of $\mathbf{C}$ are simply

$$c_{ij}(u \mid u') = B_i(u \mid u')a_{ij}. \quad (43)$$

Evaluating $\langle \mathbf{P}[\mu(u')] \mid \mu(u'')\rangle$ in analogy with (36) we have

$$\langle P[\mu(u)]\mu(u') \mid \mu(u'')\rangle$$

$$= \mathbf{C}(u \mid u')B(u' \mid u'')\mathbf{P}[\mu(u'')]/W(u \mid u')W(u' \mid u''). \quad (44)$$

Continuing in this way, we obtain

$$\langle P[\mu(u)]\mu(u') \cdots \mu(u^{(n-1)}) \mid \mu(u^{(n)})\rangle$$

$$\times W(u \mid u')W(u' \mid u'') \cdots W(u^{(n-1)} \mid u^{(n)})$$

$$= \mathbf{C}(u \mid u')\mathbf{C}(u' \mid u'^{l}) \cdots \mathbf{C}(u^{(n-1)} \mid u^{(n)})\mathbf{P}[\mu(u^{(n)})]. \quad (45)$$

$\langle \mathbf{P}[\mu(u^{(n)})]\rangle$, however, depends on the initial distribution of the particles' direction. Let us denote this initial direction by $\mathbf{\Omega}_0$ with μ_0 being the projection along the x axis and assume that μ_0 has the probability density $W(\mu_0)$ such that

$$W(\mu_0) = \sum_{i=0}^{\infty} \frac{2i+1}{2} w_i P_i(\mu_0). \quad (46)$$

Then, noting the initial lethargy is zero,

$$\langle \mathbf{P}[\mu(u^{(n)})]\rangle = \int \langle \mathbf{P}[\mu(u^{(n)})] \mid \mu_0\rangle W(\mu_0) \, d\mu_0$$

$$= \left[\mathbf{B}(u^{(n)} \mid 0) \int \mathbf{P}(\mu_0)W(\mu_0) \, d\mu_0 \right] \Big/ W(u^{(n)} \mid 0)$$

$$= \mathbf{B}(u^{(n)} \mid 0)\mathbf{W}_0/W(u^{(n)} \mid 0), \quad (47)$$

where $\mathbf{W}_0$ is a vector with components w_i. Thus, we have shown using (45) and (47) that

$$\langle \mathbf{P}[\mu(u)]\mu(u') \cdots \mu(u^{(n)})\rangle$$

$$\times W(u \mid u')W(u' \mid u'') \cdots W(u^{(n)} \mid 0)$$

$$= \mathbf{C}(u \mid u')\mathbf{C}(u' \mid u'') \cdots \mathbf{C}(u^{(n-1)} \mid u^{(n)})\mathbf{B}(u^{(n)} \mid 0)\mathbf{W}_0. \quad (48)$$

6. CALCULATION OF THE MOMENTS OF $x(u)$

We return to the expression (20) for $\langle x^n(u)\,\mathbf{P}[\mu(u)]\rangle$. Because of the symmetry of the integral with respect to $u', u'', \cdots, u^{(n)}$, the region of integration may be divided into all possible subregions of the form $u \geq u' \geq u'' \geq \cdots \geq u^{(n)}$; the total number of all such regions is $n!/k_1!k_2! \cdots k_j!$, where $k_1 + k_2 + \cdots + k_j = n$. Here j is the number of different values which appear among $u', u'', \cdots, u^{(n)}$, and $k_i(i = 1, \cdots, j)$ is the number of times which the ith value occurs. Note that, because of delta-function singularities which appear in them, the integrations over a region in which some of the $u', \cdots, u^{(n)}$ are equal cannot be ignored even though it is of less than n dimensions. Fortunately, the factor $[k_1! \cdots k_j!]^{-1}$ cancels the factor appearing in

$$\langle dl(u') \, dl(u'') \cdots dl(u^{(n)})\rangle$$

(cf. Eq. 25) so that we may write

$$\langle x^n(u)\mathbf{P}[\mu(u)]\rangle$$

$$= n! \int_0^u \int_0^{u'} \cdots \int_0^{u^{(n-1)}} \langle \mathbf{P}[\mu(u)]\mu(u') \cdots \mu(u^{(n)})\rangle$$

$$\times [\sum (u') \cdots \sum (u^{(n)})]^{-1}\langle dl(u') \, dl(u'') \cdots dl(u^{(n)})\rangle^*, \quad (49)$$

where the asterisk on $\langle dl(u')dl(u'') \cdots dl(u^{(n)})\rangle^*$ means that in evaluating it the factorial factors in (25) are to be ignored. In (49) the integration includes all singularities appearing at the lower limit and at the upper limit, but otherwise no special provision is necessary in the case of any equality among $u', u'', \cdots, u^{(n)}$.

Substituting from (25) and (48), we obtain

$$\langle x^n(u)\mathbf{P}[\mu(u)]\rangle W(u \mid 0)$$

$$= n! \int_0^u du' \int_0^{u'} du'' \cdots \int_0^{u^{(n-1)}} du^{(n)}$$

$$\times [\sum (u') \cdots \sum (u^{(n)})]^{-1}\mathbf{C}(u \mid u')\mathbf{C}(u' \mid u'') \cdots$$

$$\times \mathbf{C}(u^{(n-1)} \mid u^{(n)})\mathbf{B}(u^{(n)} \mid 0)\mathbf{W}_0. \quad (50)$$

Equation (50) is the main result of the moment calculation. The expression for $\langle x^n(u)P_j[\mu(u)]\rangle$ gives the nth moment of the jth Legendre coefficient of $F(x, \mu \mid u)$. It might appear that the use of Eq. (50) requires the multiplication of infinite matrices. Actually, this is not the case if one is content with only a finite number of moments of the initial terms of the Legendre expansion of $F(x, \mu \mid u)$. This is made clear below.

It might be noted that simplifications occur for the kernel (2). If the cross-section has special forms (e.g., a sum of exponentials) then (50) can be written as the inverse of a readily obtainable Laplace transform. This is more obvious from Eq. (51) below.

The special structure of the $\mathbf{C}$ matrices allow a number of general conclusions to be drawn concerning the dependence of the moments on the B functions and the initial vector $\mathbf{W}_0$. Consider any element $c_{ii}(u)$ of the matrix resulting from multiplying any of the $\mathbf{C}$ matrices together. Apart from a factor independent of the lethargy variables, $c_{ii}(u)$ will be a sum of terms of the form $B_iB_{j_1}B_{j_2}\cdots B_{j_{n-1}}$ (lethargy arguments have been omitted), where $j_1 = i \pm 1$, $j_k = j_{k-1} \pm 1$, $j = j_{n-1} \pm 1$. This follows from Eq. (42). Suppose now we consider a specific component of the moment vector $\langle x^n(u)\mathbf{P}[\mu(u)]\rangle$, say the ith. Then, referring to Eq. (50), $\langle x^n(u)P_i[\mu(u)]\rangle$ will be an integral over the lethargy variables of sums of products of the form $B_iB_{j_1}B_{j_2}\cdots B_{j_n}w_{i_n}$ with $j_k = j_{k+1} \pm 1$. Therefore,

(a) $\langle x^{(n)}(u)P_i[\mu(u)]\rangle$ can depend only on the functions B_{i-n}, B_{i-n+1}, $\cdots$, B_i, B_{i+1}, $\cdots$, B_{i+n}, and only on the $(i - n)$th to $(i + n)$th components of the initial vector $\mathbf{W}_0$. In particular, if the initial distribution of μ is isotropic, ($w_i = 0$, $i > 0$), then $\langle x^n(u)P_i[\mu(u)]\rangle$ vanishes for $i > n$.

(b) Conversely, any function B_i or any initial component w_i can influence only those moments $\langle x^n(u)P_k[\mu(u)]\rangle$ satisfying the inequalities $i - n \leq k \leq i + n$.

In particular, for $k = 0$, $\langle x^n(u)\rangle$ can depend only on the functions B_1, $\cdots$, B_n, and if all but w_0 vanishes, only on B_1, $\cdots$, B_n. As examples, we give (omitting arguments) the products which enter into the calculation of $\langle x^n(u)\rangle$ for $n = 1, 2, 3, 4$.

$n = 1 : B_0B_1w_1;$

$n = 2 : B_0B_1B_0w_0, \qquad B_0B_1B_2w_2;$

$n = 3 : B_0B_1B_2B_1w_1, \qquad B_0B_1B_0B_1w_1, \qquad B_0B_1B_2B_3w_3;$

$n = 4 : B_0B_1B_2B_1B_0w_0, \qquad B_0B_1B_0B_1B_0w_0,$

$\qquad B_0B_1B_2B_1B_2w_2, \qquad B_0B_1B_0B_1B_2w_2,$

$\qquad B_0B_1B_2B_3B_2w_2, \qquad B_0B_1B_2B_3B_4w_4.$

To find $\langle x^n(u)\rangle$, these products must be multiplied by suitable constant factors due to the $\mathbf{A}$ matrices, summed and integrated. For the special case of $n = 2$ and isotropic source (initial direction), we have the result found in Ref. (3) through the use of Laplace transforms.

Equation (50) may be written in a recursive form. We have

$$\langle x^n(u)\mathbf{P}[\mu(u)]\rangle W(u \mid 0)$$

$$= n \int_0^u \frac{1}{\Sigma(u')}\mathbf{C}(u \mid u')W(u' \mid 0)\langle x^{n-1}(u')\mathbf{P}[\mu(u')]\rangle \, du'$$

$$(n = 1, 2, \cdots), \qquad (51)$$

with

$$\langle x^0(u)\mathbf{P}[\mu(u)]\rangle = \mathbf{B}(u \mid 0)\mathbf{W}_0.$$

Though (51) is useful for theoretical purposes, the presence of delta-function singularities in $\mathbf{C}(u \mid u')$ makes it unsuitable for computation. To obtain a recursive expression free of such singularities we introduce a random variable $x^*(u)$ which is the x coordinate of a particle's position when it obtains the lethargy u. Hence,

$$x(u) = x^*(u) + [\Delta l/\Sigma(u)]\mu(u),$$

and

$$\langle x^n(u)\mathbf{P}[\mu(u)]\rangle$$

$$= \sum_{i=0}^n \frac{n!}{j!(n-j)!} \langle x^{*i}(u)\mu^{n-i}(u)\mathbf{P}[\mu(u)]\rangle \left\langle \left(\frac{\Delta l}{\Sigma(u)}\right)^{n-i}\right\rangle$$

$$= \sum_{i=0}^n \frac{n!}{j!} \frac{\mathbf{A}^{n-i}}{[\Sigma(u)]^{n-i}} \langle x^{*i}(u)\mathbf{P}[\mu(u)]\rangle. \qquad (52)$$

Here, in expressing $\mu^{n-i}\mathbf{P}(\mu)$, the relation $\mu\mathbf{P}(\mu) = \mathbf{A}\mathbf{P}(\mu)$ has been successively applied. In a more condensed notation, (52) reads

$$\langle x^n(u)\mathbf{P}[\mu(u)]\rangle W(u \mid 0) = \sum_{i=0}^n \frac{n!}{j!} \frac{\mathbf{A}^{n-i}}{[\Sigma(u)]^{n-i}} \mathbf{E}_i(u), \quad (53)$$

where

$$\mathbf{E}_i(u) = \langle x^{*i}(u)\mathbf{P}[\mu(u)]\rangle/W(u \mid 0). \qquad (54)$$

The expression for $\mathbf{E}_n(u)$ in terms of the $\mathbf{C}$ matrices is identical to the right-hand side of Eq. (50) except that the integration is only over $u' < u$, the contribution of the delta function in $\mathbf{C}(u \mid u')$ when $u' = u$ not being included. This fact leads to Eq. (53), with the term

$$\frac{n!}{j!} \frac{\mathbf{A}^{n-i}}{[\Sigma(u)]^{n-i}} \mathbf{E}_i$$

resulting from taking $u = u' = u'' = \cdots = u^{(n-i)}$ and $u^{(n-i)} > u^{(n-i-1)}$ in the integration in Eq. (50), and noting that the singular part of $\mathbf{C}(u \mid u')$ is just $\delta(u - u')\mathbf{A}$.

To obtain a recursion for $\mathbf{E}_n(u)$, consider the expression on the right of Eq. (50), with $u' < u$, so that this expression gives $\mathbf{E}_n(u)$. Now in performing the integrations, singularities can arise when $u' = u''$ with $u'' > u^{(3)}$, or $u' = u'' = u^{(3)}$ and $u^{(3)} > u^{(4)}$, etc. The singularity at $u = u''$ with $u'' > u^{(3)}$ gives the term

$$\frac{n!}{(n-1)!} \int_0^u \left[\frac{1}{\Sigma(u')}\right]^2 \mathbf{C}(u \mid u')\mathbf{A}\mathbf{E}_{n-1}(u') \, du'';$$

the singularity at $u' = u'' = u^{(3)}$ with $u^{(3)} > u^{(4)}$, the term

$$\frac{n!}{(n-2)!} \int_0^u \left[\frac{1}{\Sigma(u')}\right]^3 \mathbf{C}(u \mid u')\mathbf{A}^2\mathbf{E}_{n-2}(u') \, du'', \text{ etc.}$$

Considering all possible cases, we obtain

$$\mathbf{E}_n(u) = \frac{n! \, \mathbf{A}^n}{[\Sigma(u = 0)]^n}$$
$$+ \sum_{j=0}^{n-1} \frac{n!}{j!} \int_0^u \frac{\mathbf{C}(u \mid u')\mathbf{A}^{n-i}}{[\Sigma(u)]^{n-i}} \, \mathbf{E}_j(u') \, du', \quad (55)$$

$$\mathbf{E}_0(u) = \mathbf{B}(u \mid 0)\mathbf{W}_0. \quad (56)$$

In (55), the integration is only over $u' < u$, the singularity of $\mathbf{C}(u \mid u')$ at $u = u'$ not being included; and similarly in (56) the singularity of $B(u \mid 0)$ at $u = 0$ is to be ignored. The formulas (55) and (56) permit then a numerical evaluation to be made since they involve only functions which are always finite.

7. ASYMPTOTIC BEHAVIOR

Because of the difficulty in obtaining exact results and their complexity in the few cases in which they can be obtained, approximate theories play a key role in moderation theory. These are generally of the asymptotic type, becoming increasingly accurate as certain parameters or variables take on extreme values. Perhaps the one best known is age theory.

Under general conditions (see below) on the cross section and the scattering kernel, age theory provides a good description of the distribution of particles which have made a large number of collisions since leaving their source. It is closely related to ordinary diffusion and the central limit theorem and can be derived on the basis of the latter.[12] Our aim, here, is

---

[12] M. A. Leibowitz, *Rigorous Derivation of Fermi Age Theory*, Ph.D. thesis, Harvard University (1961).

to obtain the age result for an infinite homogeneous medium using the expression for the moments (50).

The condition of "many collisions" can be incorporated into our treatment (in which the collision number does not appear) by the following device. Let ξ be the mean lethargy gain at a collision (assume that it is lethargy independent) and n, the number of collisions suffered by a particle. Then by the law of large numbers, if n is large, $n \sim u/\xi$. For the important case of elastic scattering, ξ is asymptotically proportional to $1/M$ where M is the ratio of particle to scatterer masses. Thus we may write $\xi = \zeta/M$ where $\zeta = \zeta(M) = O(1)$ as $M \to \infty$, and $n \sim Mu/\zeta$.

The condition of large n is essentially equivalent to a condition of large M with u fixed, and in fact, age theory gives the first term of an asymptotic expansion of $F(x, \mu \mid u)$ in powers of $1/M^{\frac{1}{2}}$. On the other hand, if M is kept fixed, the requirement of large n can be achieved by letting the lethargy u become large. Actually this latter case can be transformed into the former by introducing a new lethargy variable u^* defined by $u = u^*U$ where u^* remains *fixed* as u and U approach infinity. Then $n \sim Mu^*U/\zeta$, so that u^* and U correspond here to u and M in the limiting case of u fixed, M tending to infinity. This explains the applicability of age theory to moderation in very light elements such as deuterium at sufficiently high lethargies. For definiteness, we shall suppose that it is the case of u fixed, M tending to infinity that is being considered.

We begin by obtaining the asymptotic behavior of the $\mathbf{B}$ matrix for large M. First, consider $B_0(u \mid u') = W(u \mid u')$. As above, it will be assumed that $f(u \mid u')$ actually depends only on $\Delta u = u - u'$; i.e., $f(u \mid u') = f(u - u')$. This is true then of $W(u \mid u') = W(u - u')$ so that

$$W(u) = \delta(0) + \int_0^u f(u - u')W(u') \, du'. \quad (57)$$

The dependence of $f(\Delta u)$ and $W(u)$ on M will be made explicit by writing $f(\Delta u, M)$ and $W(u, M)$. It will now be assumed that

Assumption (1). *The function* $(1/M)f(\Delta u/M, M)$ *approaches a probability density* $g(\Delta u)$ *which is not purely singular; i.e.,* $g(\Delta u)$ *has a continuous component as well as perhaps a delta-function part.*

(E.g., if $f(\Delta u, M) = Me^{-M\Delta u}$,

$$\xi = 1/M \quad \text{and} \quad g(\Delta u) = e^{-\Delta u}.)$$

The validity of our assumptions for the important kernel (2) is readily seen. It follows immediately

that $\xi \equiv \langle \Delta u \rangle = \xi_0/M + o(M)$ as $M \to \infty$, where ξ_0 is independent of M. Let

$$\varphi(p, M) = \int e^{-p\Delta u} f(\Delta u, M)\, d\Delta u$$

and

$$\psi(p) = \int e^{-\Delta u} g(\Delta u)\, d\Delta u.$$

Since (57) is an integral equation of the convolution type, it may be solved by Laplace transforms. We have

$$W(u, M) = \frac{1}{2\pi i} \int \frac{e^{pu}}{1 - \varphi(p, M)}\, dp, \qquad (58)$$

the integral being taken over a line to the left of all poles and parallel to the imaginary axis. Evaluating by residues,

$$W(u, M) = \frac{1}{\xi} + \sum_i R_i e^{p_i u} + \delta(u = 0). \qquad (59)$$

The term $1/\xi$ comes from the poles at $p = 0$ in (59), the terms $R_i e^{p_i u}$ from the poles at $p = p_i$, with $R_i = -1/\varphi'(p, M)$. From (58) it is clear that the real part of p_i is negative: the p_i and R_i are actually functions of M, and to determine their asymptotic behavior, Assumption (1) above is used. Now

$$\psi(p) = \lim_{M \to 0} \frac{1}{M} \int e^{-p\Delta u} f\left(\frac{\Delta u}{M}, M\right) d\Delta u$$

$$= \lim_{M \to 0} \int e^{-pM\Delta u} f(\Delta u, M)\, d\Delta u$$

$$= \lim_{M \to 0} \varphi(Mp, M).$$

It follows that if r_i are the nonzero roots of $\psi(p) = 1$ then there is a correspondence between the p_i and r_i such that

$$p_i = Mr_i + o(M) \qquad (60)$$

and

$$R_i = -M/\psi'(r_i) + o(M). \qquad (61)$$

Thus both R_i and p_i are of the order of M, as M tends to infinity.

Turning to the higher-order functions $B_i(u \mid u')$ $(i = 1, 2, \cdots)$, let us assume that $K_i(u \mid u')$ also depends only on the difference $\Delta u = u - u'$. Hence we may write $K_i(u \mid u') = K_i(u - u')$, and as well $B_i(u \mid u') = B_i(u - u')$ where

$$B_i(u) = \delta(u) + \int K_i(u - u') f(u - u') B_i(u')\, du'.$$
$$(62)$$

Again, Eq. (62) is of the convolution type and its solution is given by

$$B_i(u, M) = \int \frac{1}{1 - \varphi_i(p, M)} e^{-pu}\, dp, \qquad (63)$$

the integrand being taken over the same contour as in (58), and $\varphi_i(p, M)$ being defined by

$$\varphi_i(p, M) = \int_0^\infty e^{-pu} K_i(\Delta u, M) f(\Delta u, M)\, du. \qquad (64)$$

In (63) and (64) the dependence of the different functions on M is explicitly indicated. The integrand in (63) has poles at the points where

$$\varphi_i(p, M) = 1.$$

Let us denote these points by $p_{ij}(M)(j = 1, \cdots)$ with residues $R_{ij}(M)$. We will assume that the $p_{ij}(M)$ are bounded away from zero for all i, j, M:

$$|p_{ij}(M)| > \eta, \qquad (65)$$

where $\eta > 0$ is *not* dependent on M. This assumption is actually a consequence of a simpler one on the scattering. Note that for $\varphi_i(0, M)$ to equal unity it is necessary that $K_i(\Delta u, M) = 1$ for all Δu. This follows from (64) using the normalization condition for $f(\Delta u)$. But $K_i(\Delta u, M)$ by definition is the mean value of $P_i(\gamma)(\gamma = \mathbf{\Omega} \cdot \mathbf{\Omega'})$ for a change Δu in lethargy at a collision. Hence, since $P_i(\gamma) = 1$ only for $\gamma = 1$ or -1 (if i is even), inequality (65) is equivalent to

Assumption (2). There exist positive numbers $0 < \eta'$, $\eta'' < 1$ independent of M, such that the probability of the inequality $|\gamma| < \eta'$ is greater than η''.

In other words, scattering does not tend to become directly forward or directly backward as M tends to infinity. The effect of Assumption (2) is that after a large number of collisions all trace of the particle's initial direction is lost.

In addition, we make

Assumption (3). The limit of $K_i(\Delta u/M, M)$ as M goes to infinity exists.

Then by the same argument leading to Eq. (60) one can show that there are numbers r_{ij} with negative real parts, and numbers R_{ij}^* such that

$$p_{ij} = Mr_{ij} + o(M), \qquad R_{ij} = MR_{ij}^* + o(M). \qquad (61)$$

Again the validity of the above assumption can be readily verified for the elastic scattering kernel.

Finally we require

Assumption (4). For some function $L(M)$, $\lim_{M \to \infty} L(M) \Sigma (u, M)$ exists for all u and is positive.

The point of Assumption 4, is to exclude cross sections of the form $\Sigma(u, M) = e^{-\lambda u M}$. This case results if proton moderation at very large lethargies is considered (M then is the lethargy scale) because of the $1/v$ dependence of the cross section. For convenience, we take $L(M) = M^{\frac{1}{2}}$ which can always be achieved by a suitable transformation of the length scale.

Consider the moments $\langle x^n(u) \rangle$ for large M. According to Eq. (50), $\langle x^n(u) \rangle$ will be composed of integrals over different products of $n + 1$ of the B_i functions. Since B_0 is of the order of M (Eq. 59) while $B_i (i > 0)$ is of the order of unity,[13] the dominant contribution will be from the term containing the maximum possible number of the B_0. If we take n to be even, this is just the term consisting of the product $B_0 B_1 B_0 B_1 \cdots B_1 B_0$, where the B_0 alternate with the B_1. Let us consider $\langle x^2(u) \rangle$. Keeping only the dominant term,

$$W(u \mid 0)\langle x^2(u) \rangle = \frac{2}{3} \int_0^u \frac{G_0(u - u') \, du'}{\Sigma(u')}$$

$$\times \int_0^u \frac{B_1(u' - u'')B_0(u'')}{\Sigma(u'')} \, du''. \quad (66)$$

We have for the inner integral, with $B_0(u)$ replaced by $1/\xi$ and $B_1(u)$ by $\sum_i R_{i1} e^{p_i u} + \delta(u)$,

$$\int_0^u \frac{B_1(u - u')B_0(u')}{\Sigma(u')}$$

$$= \sum_i \frac{R_{i1}}{\xi} \int_0^u \frac{e^{p_i(u-u')} \, du'}{\Sigma(u')} + \frac{1}{\xi} \frac{1}{\Sigma(u)}$$

$$= \frac{1}{\xi} \frac{1}{\Sigma(u)} - \sum_i \frac{1}{\xi} \frac{R_{i1}}{p_{i1}} \frac{1}{\Sigma(u)} + \text{lower order terms.}$$

The last equation comes from integrating by parts and noting Eq. (61).[14] But

$$1 - \sum_i \frac{R_{i1}}{p_{i1}} = \frac{1}{1 - \varphi_1(p)}\bigg|_{p=0} = \frac{1}{1 - \langle \gamma \rangle}$$

$$(\gamma = \mathbf{\Omega} \cdot \mathbf{\Omega}').$$

Hence we finally have for $\langle x^2(u) \rangle$, with $B_0(u)$ taken to be $1/\xi$,

$$\langle x^2(u) \rangle = \frac{2}{3} \int_0^u \frac{1}{\xi(1 - \langle \gamma \rangle)} \frac{du'}{\Sigma^2(u)}. \quad (67)$$

Note that $B_1(u' - u'')$ in the integral in (66) acts just like the delta function $\delta(u' - u'')/(1 - \langle \gamma \rangle)$.

---

[13] Though the term $R_{ij} e^{p_{ij} u}$ tends to zero exponentially fast with increasing M for a fixed u, on integration one obtains R_{ij}/p_{ij} which is of order unity in M.

[14] The assumption on the cross section was used here. If, e. g., $\Sigma(u) \sim e^{-\lambda u M}$, then $\Sigma'(u) \sim M e^{-\lambda u M}$ and this would lead to additional terms of the same order as those retained.

Using this fact, one can establish (with $B_0(u) = 1/\xi$) the recursion for the largest terms,

$$\langle x^n(u) \rangle = \frac{n(n-1)}{3} \int_0^u \frac{1}{\Sigma^2(u')} \frac{1}{\xi(1 - \langle \gamma \rangle)} \langle x^{(n-2)}(u') \rangle \, du'.$$

Using (67), solving for $\langle x^n(u) \rangle$ explicitly, and taking into account the symmetry of the integral,

$$\langle x^n(u) \rangle = \frac{n!}{2^{n/2}(n/2)!} \left(\frac{2}{3}\right)^{\frac{1}{2}n} \left[\int_0^u \frac{du'}{\Sigma^2(u')\xi(1 - \langle \gamma \rangle)}\right]^{\frac{1}{2}n}$$

$$(n \text{ even}). \quad (68)$$

Note that since $\Sigma(u) \sim M^{-\frac{1}{2}}$, $\xi \sim 1/M$, the even moments are of the order of unity. On the other hand, the odd moments will be of the order of $M^{-\frac{1}{2}}$ since they will contain one less cross-section term and one less B_0 term then the next highest even moment. Thus $\langle x^n(u) \rangle \to 0$ (n odd) as $M \to \infty$. Hence to the highest order the moments of $F(x \mid u)$ are given by Eq. (68). But these are just the moments of the Gaussian distribution

$$F(x \mid u) = [2\pi\langle x^2(u)\rangle]^{-\frac{1}{2}} \exp\left[-x^2/2\langle x^2(u)\rangle\right]. \quad (69)$$

This is the age result for an infinite homogeneous moderator.

8. THE TIME MOMENTS

Let $F(u, t)$ be the average number of neutron collisions per unit time and lethargy interval. Corresponding to $F(u, t)$, we introduce the probability density $F(t \mid u)$ defined as the probability density of the time that a particle with lethargy u makes a collision, given that the particle does attain the lethargy u after some collision. We assume, as before, that the particle leaves the origin at time zero with lethargy 0. Then

$$F(t \mid u) = F(u, t)/W(u \mid 0). \quad (70)$$

The moments $\langle t^n(u) \rangle$ of $F(t \mid u)$ can be found immediately from the analysis of $\langle x^n(u) \rangle$.

Using the random functional

$$t(u) = \int_0^u \frac{1}{\Sigma(u')v'} \, dl(u') \quad (71)$$

(where v' is the velocity of a particle with lethargy u'), one obtains from Eq. (25)

$$\langle t^n(u) \rangle W(u \mid 0)$$

$$= n! \int_0^u du' \int_0^{u'} du'' \cdots \int_0^{u^{(n-1)}} du^{(n)} [\Sigma(u')v' \cdots$$

$$\times \Sigma(u^{(n)})v^{(n)}]^{-1} W(u \mid u')W(u' \mid u'') \cdots$$

$$\times W(u^{(n-1)} \mid u^{(n)})W(u^{(n)} \mid 0). \quad (72)$$

1526   MARTIN A. LEIBOWITZ

Just as before, we can introduce a singularity-free recursion formula for the moments. Following the notation of Ref. 1, we let $W^{(n)}(u)/W(u \mid 0)$ be the nth moment of $t^*(u)$, the position of the particle when it obtains the lethargy u. Then, by the identical argument that led to (55) and (56),

$$W^{(n)}(u) = \frac{n! \, W(u \mid 0)}{\Sigma(u = 0)v_0}$$
$$+ \sum_{i=0}^{n-1} \frac{n!}{i!} \int_0^u \frac{W(u \mid u')W^{(i)}(u')}{[\Sigma(u')v']^{n-i}} \, du' \quad (73)$$

and

$$\langle t^n(u) \rangle W(u \mid 0) = \sum_{i=0}^n \frac{n!}{i!} \frac{W^{(i)}(u)}{[\Sigma(u)v]^{n-i}}, \quad (74)$$

where in the integral in (73) the term $\delta(u - u')$ in $W(u \mid u')$ is to be *ignored*.

We can obtain the asymptotic expression for $F(t \mid u)$ in the age limit by substituting from Eq. (59) into (72), and in so doing retaining Assumption 1 above on $f(\Delta u, M)$. However, Assumption 4 on the cross section must be altered. In fact, it is required here that not $\Sigma(u, M)$ but rather $v\Sigma(u, M) = v_0 e^{-u/2} \Sigma(u, M)$ have an asymptotic behavior that is the same for all u. Thus the age approximation cannot be valid for both the space-lethargy and time-lethargy distribution of a slowing-down process.

In the age limit, $\langle t^n \rangle$ becomes then

$$\langle t^n \rangle = \left[\int_0^u \frac{du'}{\xi\Sigma(u')v'} \right]^n, \quad (75)$$

so that

$$F(t \mid u) = \delta(t - \langle t(u) \rangle), \quad (76)$$

i.e., there is no dispersion in the time to slow down to a given lethargy.

APPENDIX: EXTENSION TO THE CASE OF MANY SCATTERERS

Let us consider the case where more than one type of scattering occurs. Corresponding to the ith mode of scattering ($i = 1, 2, \cdots, n$) there will be a scattering cross section $\Sigma_i(u)$ and a scattering kernel $K_i(\Omega \cdot \Omega', u, u')$. The key result is that such a moderator is equivalent to one having only one type of scatterer with cross section $\Sigma(u)$ given by

$$\Sigma(u) = \Sigma_1(u) + \Sigma_2(u) + \cdots + \Sigma_n(u), \quad (77)$$

and kernel given by

$$K(\Omega \cdot \Omega', u, u') = \sum_{i=1}^n \frac{\Sigma_i(u)}{\Sigma(u)} K_i(\Omega \cdot \Omega', u, u'). \quad (78)$$

This follows immediately by comparison with the corresponding transport equation and can be seen readily using our stochastic point of view. In fact, the probability density of the distance s between any two collisions is $\exp[-\Sigma(u)s]$, while the factor $\Sigma_i(u)/\Sigma(u)$ in (78) gives the probability that if a collision occurs it will be with the ith type of scatterer.

An absorbing medium is just a special case of the above with the total cross section $\Sigma(u)$ being the sum of the absorbing and scattering cross sections, and the "scattering kernel" at a collision with an absorber taken to be identically zero.

ACKNOWLEDGMENT

Part of this work was done while the author was at Harvard University. Thanks are due to Professor G. Birkhoff for his advice, and encouragement and to the Office of Naval Research for their support.

Solution of the Initial-Value Transport Problem for Monoenergetic Neutrons in Slab Geometry*

ROBERT L. BOWDEN AND CLAYTON D. WILLIAMS

Physics Department, Virginia Polytechnic Institute, Blacksburg, Virginia
(Received 28 April 1964)

The initial-value transport problem of monoenergetic neutrons migrating in a thin slab is solved by applying the normal-mode expansion method of Case to the results of Lehner and Wing. Fredholm integral equations are derived for the expansion coefficients. In addition, exact expressions for the eigenvalues of the problem are derived and the results of calculations are presented. The solution is shown to have properties expected from elementary diffusion theory.

I. THE PROBLEM

SEVERAL solutions to neutron transport problems have been found with the use of the normal-mode expansion method of Case.[1] This paper presents an application of this method to an initial-value problem involving the monoenergetic neutron transport equation in slab geometry with isotropic scattering. We consider a slab of material extending from $x = -a$ to $x = a$ and surrounded by a vacuum. The neutron distribution $N(x, \mu, t)$ at the time t depends on the single position variable x (measured in units of mean-free paths) and corresponding direction cosine μ and satisfies the equation

$$\frac{\partial N}{\partial t} + \Sigma v \mu \frac{\partial N}{\partial x} + \Sigma v N$$

$$= \frac{c \Sigma v}{2} \int_{-1}^{1} N(x, \mu', t) \, d\mu', \qquad (1.1)$$

subject to the boundary conditions

$$N(\pm a, \mu, t) = 0, \quad \mu \lessgtr 0, \quad t > 0, \qquad (1.2)$$

where Σ is the total cross section of the medium, v the speed of the neutrons, and c the net number of neutrons produced per collision. The boundary conditions (1.2) simply express the fact that no neutrons enter the slab from outside. To complete the problem we specify that

$$N(x, \mu, 0) = f(x, \mu) \qquad (1.3)$$

and we wish to find the subsequent neutron distribution.

We may simplify the problem somewhat if we introduce

$$u(x, \mu, t) = N(x, \mu, t)e^{\Sigma v t} \qquad (1.4)$$

to obtain

$$\frac{\partial u(x, \mu, t)}{\partial t} = Au(x, \mu, t), \qquad (1.5)$$

$$u(\pm a, \mu, t) = 0, \quad \mu \lessgtr 0, \quad t > 0, \qquad (1.6)$$

$$u(x, \mu, 0) = f(x, \mu), \qquad (1.7)$$

where the operator A is defined as

$$A = -\mu \frac{\partial}{\partial x} \cdot + \frac{c}{2} \int_{-1}^{1} \cdot d\mu'. \qquad (1.8)$$

The formal application to Eq. (1.5) of the Laplace transformation with respect to t yields

$$[s - A]\bar{u}(x, \mu, s) = f(x, \mu), \qquad (1.9)$$

where

$$\bar{u}(x, \mu, s) = \int_{0}^{\infty} u(x, \mu, t)e^{-st} \, dt. \qquad (1.10)$$

Proceeding formally, we write the solution of Eq. (1.9) as

$$\bar{u}(x, \mu, s) = [s - A]^{-1}f(x, \mu), \qquad (1.11)$$

where $[s - A]^{-1}$ is the operator inverse to $[s - A]$. Finally, we apply the inverse Laplace transformation operator to Eq. (1.11) to obtain

$$u(x, \mu, t) = \lim_{\omega \to \infty} \frac{1}{2\pi i} \int_{b-i\omega}^{b+i\omega} [s - A]^{-1}f(x, \mu)e^{st} \, ds, \qquad (1.12)$$

where b is to the right of all the singularities of $[s - A]^{-1}f(x, \mu)$. These singularities occur when the operator $[s - A]^{-1}$ fails to exist, that is, when s is an eigenvalue α_i of the operator A. Hopefully these eigenvalues are poles of the integrand of Eq. (1.12). Assuming that $[s - A]^{-1}f(x, \mu)$ is otherwise sufficiently well behaved, we may move the line of integration in Eq. (1.12) to the left and pick up the residue contributions. For many years, it was assumed on the basis of physical argument and analogy with other mathematical

[1] K. M. Case, Ann. Phys. (N. Y.) 9, 1 (1960).

physics problems that the eigenvalues $\{\alpha_i\}$ formed a denumerably infinite set and by deforming the contour of Eq. (1.12) arbitrarily far to the left a formal expansion of the form

$$u(x, \mu, t) = \sum_{i=0}^{\infty} g_i(x, \mu)e^{\alpha_i t} \qquad (1.13)$$

would be obtained.

Several years ago, Lehner and Wing[2,3] showed that the form of Eq. (1.13) is not correct. They found that the set $\{\alpha_i\}$ is, in fact, finite in number, and that the integrand of Eq. (1.12) is not sufficiently well behaved to move the line of integration arbitrarily far to the left. They proceeded to show that the expansion analogous to Eq. (1.13) should consist of a finite series plus a residual or continuum term.

To be more precise let us introduce the following notation. Let H be the Hilbert space of complex functions $g(x, \mu)$ square integrable on $|x| \leq a$, $|\mu| \leq 1$:

$$\int_{-a}^{a} \int_{-1}^{1} |g(x, \mu)|^2 \, dx \, d\mu < \infty, \qquad (1.14)$$

and let the domain d_A of the operator A be the linear manifold of those functions $g \in H$ such that

$$Ag \in H, \qquad g(\pm a, \mu) = 0, \qquad \mu \lessgtr 0. \qquad (1.15)$$

The inner product of functions g and h in H is defined by

$$(g, h) = \int_{-a}^{a} \int_{-1}^{1} g(x, \mu)h^*(x, \mu) \, dx \, d\mu, \qquad (1.16)$$

where the $*$ indicates complex conjugation. The adjoint operator $A^\dagger$ is given by

$$A^\dagger = \mu \frac{\partial}{\partial x} \cdot + \frac{c}{2} \int_{-1}^{1} \cdot d\mu' \qquad (1.17)$$

with domain $d_A^\dagger$ consisting of those functions $g^\dagger \in H$ such that

$$A^\dagger g^\dagger \in H, \qquad g^\dagger(\pm a, \mu) = 0, \qquad \mu \gtrless 0. \qquad (1.18)$$

The principal result of Lehner and Wing is the following:

$$u(x, \mu, t) = \sum_{j=0}^{M} (f, \Psi_j^\dagger)\Psi_j(x, \mu)e^{\alpha_j t} + \zeta(x, \mu, t), \qquad (1.19)$$

$$\zeta(x, \mu, t) = \lim_{\omega \to \infty} \frac{1}{2\pi i} \int_{\gamma - i\omega}^{\gamma + i\omega} R(x, \mu, s)e^{st} \, ds,$$

$$0 < \gamma < \alpha_M, \qquad (1.20)$$

where

$$[s - A]R(x, \mu, s) = f(x, \mu), \qquad R(x, \mu, s) \in d_A. \quad (1.21)$$

The function $R(x, \mu, s)$ is analytic (for fixed x and μ) in the right half complex plane of s (Re $s > 0$) except for simple poles at the eigenvalues $\{\alpha_i\}$ of

$$A\Psi_i(x, \mu) = \alpha_i \Psi_i(x, \mu), \qquad \Psi_i(x, \mu) \in d_A. \quad (1.22)$$

These eigenvalues are real, positive, distinct, and finite in number, and we order them as $0 < \alpha_M < \cdots < \alpha_0$. The eigenfunctions $\Psi_i^\dagger(x, \mu)$ are the solutions to the adjoint eigenvalue problem

$$A^\dagger \Psi_i^\dagger(x, \mu) = \alpha_i \Psi_i^\dagger(x, \mu), \qquad \Psi_i^\dagger(x, \mu) \in d_A \quad (1.23)$$

and have the same eigenvalues as the eigenvalue problem (1.22). We have assumed that the eigenfunctions are not degenerate and that

$$(\Psi_i, \Psi_i^\dagger) = 1. \qquad (1.24)$$

Although Lehner and Wing obtain an explicit expression for $R(x, \mu, s)$ [Ref. 3, Eq. (1.17)], their analysis of the slab problem does not suggest either the shape of the eigenfunctions $\Psi_i(x, \mu)$ or the character of the distribution of the eigenvalues $\{\alpha_i\}$ as functions of Σ, c, and a. Schlesinger[4] has devised a numerical scheme based on the above results to approximate some of these quantities.

Below we apply Case's method directly to the results of Lehner and Wing to obtain a more complete solution to the initial-value problem. That is, we find "elementary" solutions of the equations

$$[s - A]\psi(x, \mu, s) = 0 \qquad (1.25)$$

and

$$[s - A^\dagger]\psi^\dagger(x, \mu, s) = 0, \qquad (1.26)$$

then construct solutions to the problems (1.21), (1.22), and (1.23) as superpositions of these "elementary" solutions.

II. ELEMENTARY SOLUTIONS

The form of Eq. (1.25) suggests the solution

$$\psi(x, \mu, s) = \phi(\mu, \nu, s)e^{-sx/\nu}, \qquad (2.1)$$

where the spectrum of ν and the shape of $\phi(\mu, \nu, s)$ for each value of s are to be determined. With this assumption Eq. (1.25) reduces to

$$-\frac{\mu}{\nu} s\phi(\mu, \nu, s) + s\phi(\mu, \nu, s)$$

$$= \frac{c}{2} \int_{-1}^{1} \phi(\mu', \nu, s) \, d\mu'. \qquad (2.2)$$

[2] J. Lehner and G. M. Wing, Commun. Pure Appl. Math. 8, 217 (1955).

[3] J. Lehner and G. M. Wing, Duke Math. J. 23, 125 (1956).

[4] S. I. Schlesinger, LA-1908 (1955) (unpublished).

The normalization of $\phi(\mu, \nu, s)$ is at our disposal, and it is convenient to normalize so that

$$\int_{-1}^{1} \phi(\mu', \nu, s) \, d\mu' = s. \qquad (2.3)$$

We have then

$$[\nu - \mu]\phi(\mu, \nu, s) = \tfrac{1}{2}c\nu. \qquad (2.4)$$

The general solution of Eq. (2.4) is

$$\phi(\mu, \nu, s) = \frac{c}{2} P \frac{\nu}{\nu - \mu} + \lambda(\nu, s)\delta(\nu - \mu), \qquad (2.5)$$

where the P indicates that the principal value is to be taken in any integral involving this term and $\delta(\nu - \mu)$ is the usual Dirac delta function. Two cases must be distinguished: (a) $\nu \notin (-1, 1)$ and (b) $\nu \in (-1, 1)$.

(a) Discrete solutions: For $\nu \notin (-1, 1)$ the solution of Eq. (2.4) is

$$\phi(\mu, \nu, s) = \frac{c}{2} \frac{\nu}{\nu - \mu}. \qquad (2.6)$$

The normalization condition (2.3) yields

$$s = c\nu \tanh^{-1}(1/\nu), \qquad (2.7)$$

and $\phi(\mu, \nu, s)$ of Eq. (2.6) will have the proper shape only if Eq. (2.7) is satisfied. This situation can be expressed as follows. If we define $\Omega(\nu, s)$ such that

$$\Omega(\nu, s) = s - c\nu \tanh^{-1}(1/\nu), \qquad (2.8)$$

the ν of Eq. (2.6) will be correct only if it is a zero of $\Omega(\nu, s)$ which lies in the complex plane of ν, cut along the real interval $(-1, 1)$.

We now list some properties of the zeros of $\Omega(\nu, s)$. Since $\Omega(\nu, s)$ is even with respect to ν, the zeros occur in pairs as $\pm\nu_0$. Consider the counterclockwise contour C_s (Fig. 1) in the complex plane s defined by

$$C_s = \left\{ s = \alpha + i\beta \,\middle|\, \alpha = \frac{2}{\pi} \beta \tanh^{-1} \frac{2\beta}{\pi c} \right\}. \qquad (2.9)$$

We label by S_i the set of s values lying to the left of, but not on the contour C_s. Similarly, the set of all s such that $s \notin S_i \cup C_s$ is denoted by S_e. The zeros of $\Omega(\nu, s)$ may then be enumerated as follows. If $s \in S_i \cup C_s$, there are two zeros of $\Omega(\nu, s)$. If $s \in S_i$, the two zeros $\pm\nu_0$ do not lie in the real interval $(-1, 1)$. However, if $s \in C_s$, then the two zeros are $\pm\nu_0 = 2\beta/\pi c$ and lie in the real interval $(-1, 1)$. Finally, if $s \in S_e$, the function $\Omega(\nu, s)$ has no zeros. A proof of these facts is sketched in Appendix A.

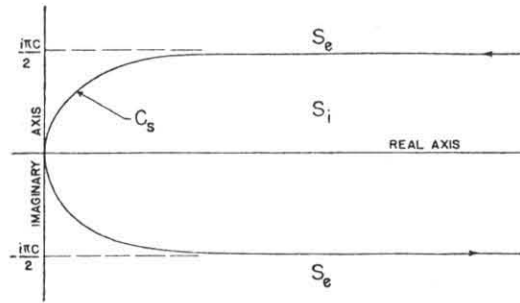Thus it is found that Eq. (2.7) represents the

FIG. 1. The contour C_s.

condition for discrete solutions. In fact, from the above results, we have discrete solutions of the form

$$\psi_\pm(x, \mu, s) = \phi_\pm(\mu, s)e^{\mp sx/\nu_0}, \qquad s \in S_i, \qquad (2.10)$$

where

$$\phi_\pm(\mu, s) \equiv \phi(\mu, \pm\nu_0, s) = \frac{c}{2} \frac{\nu_0}{\nu_0 \mp \mu} \qquad (2.11)$$

and we have no discrete solutions if $s \in S_e$.

(b) Continuum solution: For $\nu \in (-1, 1)$, Eq. (2.5) represents the solution of Eq. (2.4) at a point of the continuous spectrum of ν. We can determine $\lambda(\nu, s)$ from the normalization condition (2.3), which yields

$$\lambda(\nu, s) = s - c\nu \tanh^{-1} \nu. \qquad (2.12)$$

Therefore, the continuum solution is

$$\psi(x, \mu, \nu, s) = \phi(\mu, \nu, s)e^{-sx/\nu}, \qquad (2.13)$$

with $\phi(\mu, \nu, s)$ given by Eq. (2.5) and $\lambda(\nu, s)$ by Eq. (2.12).

The functions (distributions) $\phi_\pm(\mu, s)$ and $\phi(\mu, \nu, s)$ $(-1 \le \nu \le 1)$ are orthogonal on the interval $-1 \le \mu \le 1$ with respect to the weight function μ. That is,

$$\int_{-1}^{1} \mu\phi(\mu, \nu, s)\phi(\mu, \nu', s) \, d\mu = 0, \qquad \nu \ne \nu'. \qquad (2.14)$$

The proof of Eq. (2.14) follows closely that of Case (Ref. 1, Theorem I) and is omitted here.

We can evaluate the last integral when $\nu = \nu'$. In the case of the discrete solutions we obtain

$$N_\pm(s) \equiv \int_{-1}^{1} \mu\phi_\pm^2(\mu, s) \, d\mu$$

$$= \pm\frac{c\nu_0}{2}\left(\frac{c\nu_0}{\nu_0^2 - 1} - s\right), \qquad s \in S_i, \qquad (2.15)$$

and for the continuum case we find

$$\int_{-1}^{1} \mu\phi(\mu, \nu, s)\phi(\mu, \nu', s) \, d\mu = \nu\lambda(\nu, s)\delta(\nu - \nu'). \qquad (2.16)$$

Another very useful formula is

$$\int_{-1}^{1} \mu \phi(\mu, \nu, s) \int_{-1}^{1} Q(\nu') \phi(\mu, \nu', s) \, d\mu \, d\nu'$$

$$= Q(\nu) N(\nu, s), \quad (2.17)$$

where

$$N(\nu, s) = \nu[c^2 \pi^2 \nu^2 / 4 + \lambda^2(\nu, s)] \quad (2.18)$$

and $Q(\nu)$ is some sufficiently well behaved but otherwise arbitrary function defined on $-1 \le \nu \le 1$.

The functions $\phi_{\pm}(\mu, s)$ and $\phi(\mu, \nu, s)$ are complete for any sufficiently well-behaved function defined on $-1 \le \mu \le 1$. The restrictions on this function seem to be very weak. However, for our purposes, a sufficient condition is that its product with μ obey an H^* condition[5] on $(-1, 1)$. A function $Q(\mu)$ is said to obey an H^* condition on the interval (a, b) if (i) with μ_1 and μ_2 belonging to any closed interval contained in the open interval (a, b), there exists a constant C and a positive number ξ such that

$$|Q(\mu_1) - Q(\mu_2)| \le C \, |\mu_1 - \mu_2|^{\xi} \quad (2.19)$$

and (ii) near the endpoints a and b, $Q(\mu)$ behaves as

$$Q(\mu) = \frac{\tilde{Q}(\mu)}{(\mu - d)^{\delta}}, \quad 0 \le \delta < 1, \quad (2.20)$$

where d stands for a or b and $\tilde{Q}(\mu)$ satisfies the inequality (2.19) on the closed interval (a, b).

By completeness, then, we mean that if $\mu F(\mu)$ $(-1 \le \mu \le 1)$ obeys an H^* condition on $(-1, 1)$, the following expansion is possible

$$F(\mu) = a_+(s)\phi_+(\mu, s) + a_-(s)\phi_-(\mu, s)$$

$$+ \int_{-1}^{1} A(\nu, s)\phi(\mu, \nu, s) \, d\nu, \quad s \in S_i,$$

$$= \int_{-1}^{1} A(\nu, s)\phi(\mu, \nu, s) \, d\nu, \quad s \in S_e. \quad (2.21)$$

With such an expansion we can evaluate the coefficients from the orthogonality relations. For example, multiplying Eq. (2.21) by $\mu \phi_{\pm}(\mu, s)$ and integrating on μ over $(-1, 1)$ we get

$$a_{\pm}(s) = \frac{1}{N_{\pm}(s)} \int_{-1}^{1} \mu F(\mu)\phi_{\pm}(\mu, s) \, d\mu, \quad s \in S_i, \quad (2.22)$$

where $N_{\pm}(s)$ are defined by Eqs. (2.15).

Let us define $F'(\mu)$ by

$$F'(\mu) = F(\mu) - a_+(s)\phi_+(\mu, s)$$

$$- a_-(s)\phi_-(\mu, s), \quad s \in S_i,$$

$$= F(\mu), \quad s \in S_e, \quad (2.23)$$

where $a_{\pm}(s)$ are given by Eqs. (2.22). The expansion (2.21) is then rewritten as

$$F'(\mu) = \lambda(\mu, s)A(\mu, s)$$

$$+ P \int_{-1}^{1} \frac{c\nu}{2} \frac{A(\nu, s)}{\nu - \mu} \, d\nu, \quad s \in S_i \cup S_e, \quad (2.24)$$

where we have used the explicit expression for $\phi(\mu, \nu, s)$. The last equation has the form of a standard singular integral equation. Thus the expansion (2.21) is valid if we can demonstrate the existence of the solution of this singular integral equation for arbitrary $F'(\mu)$ (subject to the H^* condition). The proof[6] of the existence of this solution essentially parallels that given by Case (Ref. 1, Theorem II) and will not be presented here.

The validity of the expansion (2.21) allows us to expand any solution of Eq. (1.25) in terms of $\phi_{\pm}(\mu, s)$ and $\phi(\mu, \nu, s)$ with coefficients depending on x. These coefficients must be proportional to $e^{\pm sx/\nu_0}$ and $e^{-sx/\nu}$, as can be seen by substituting directly into Eq. (1.25). Therefore, the general solution of Eq. (1.25) is of the form

$$\Psi(x, \mu, s) = a_+(s)\psi_+(x, \mu, s) + a_-(s)\psi_-(x, \mu, s)$$

$$+ \int_{-1}^{1} A(\nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad s \in S_i,$$

$$= \int_{-1}^{1} A(\nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad s \in S_e. \quad (2.25)$$

III. THE EIGENFUNCTIONS $\Psi_j(x, \mu)$

Since the eigenvalue problem (1.22) has solutions only for positive real values of s, we can write its general solution as

$$\Psi(x, \mu, s) = a_1(s)\psi_+(x, \mu, s) + a_2(s)\psi_-(x, \mu, s)$$

$$+ \int_{-1}^{1} A(\nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad (3.1)$$

where the expansion coefficients are obtained from the boundary conditions

$$\Psi(\pm a, \mu, s) = 0, \quad \mu \lessgtr 0. \quad (3.2)$$

These boundary conditions imply that

$$0 = a_1(s)\phi_+(\mu, s)e^{sa/\nu_0} + a_2(s)\phi_-(\mu, s)e^{-sa/\nu_0}$$

$$+ \int_{-1}^{1} A(\nu, s)\phi(\mu, \nu, s)e^{sa/\nu} \, d\nu, \quad \mu > 0, \quad (3.3)$$

$$0 = a_1(s)\phi_+(-\mu, s)e^{-sa/\nu_0} + a_2(s)\phi_-(-\mu, s)e^{sa/\nu_0}$$

$$+ \int_{-1}^{1} A(\nu, s)\phi(-\mu, \nu, s)e^{-sa/\nu} \, d\nu, \quad \mu > 0. \quad (3.4)$$

[5] N. I. Muskhelishvili, *Singular Integral Equations* (C. P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

[6] R. L. Bowden, Ph.D. thesis, Virginia Polytechnic Institute (unpublished).

It is convenient to let

$$A(\nu, s) = A_1(\nu, s)H(\nu) + A_2(-\nu, s)H(-\nu), \quad (3.5)$$

where

$$H(\nu) = 1, \quad \nu > 0$$
$$= 0, \quad \nu < 0. \quad (3.6)$$

Then by noting that

$$\phi_+(\mu, s) = \phi_-(-\mu, s), \quad \phi(\mu, \nu, s) = \phi(-\mu, -\nu, s) \quad (3.7)$$

and defining

$$a_1(s) \pm a_2(s) = a_\pm(s),$$
$$[A_1(\nu, s) \pm A_2(\nu, s)]e^{sa/\nu} = A_\pm(\nu, s), \quad (3.8)$$

we add and subtract Eqs. (3.3) and (3.4) to obtain

$$0 = a_\pm(s)[\phi_+(\mu, s)e^{sa/\nu_0} \pm \phi_-(\mu, s)e^{-sa/\nu_0}]$$
$$+ \int_0^1 A_\pm(\nu, s)[\phi(\mu, \nu, s)$$
$$\pm \phi(\mu, -\nu, s)e^{-2sa/\nu}] \, d\nu, \quad \mu > 0. \quad (3.9)$$

We note that for $\mu > 0$, $\phi(\mu, -\nu, s)$ is nonsingular, so that Eqs. (3.9) are singular integral equations whose kernels have singular and nonsingular parts. These equations can be reduced to equivalent Fredholm integral equations. This reduction is indicated in Appendix B. The results are

$$A_+'(\mu, s) = -h_+(\mu, s)$$
$$- \int_0^1 K(\mu, \nu, s)A_+'(\nu, s) \, d\nu, \quad s \in S_i, \quad (3.10a)$$

$$A_-'(\mu, s) = -h_-(\mu, s)$$
$$+ \int_0^1 K(\mu, \nu, s)A_-'(\nu, s) \, d\nu, \quad s \in S_i, \quad (3.10b)$$

provided the following conditions are fulfilled:

$$n_+(s) = -\frac{c}{2} \int_0^1 \nu X(-\nu, s)A_+'(\nu, s)e^{-2sa/\nu} \, d\nu, \quad s \in S_i, \quad (3.11a)$$

$$n_-(s) = \frac{c}{2} \int_0^1 \nu X(-\nu, s)A_-'(\nu, s)e^{-2sa/\nu} \, d\nu, \quad s \in S_i, \quad (3.11b)$$

where

$$a_\pm(s)A_\pm'(\nu, s) = A_\pm(\nu, s). \quad (3.12)$$

The kernel $K(\mu, \nu, s)$ is given by

$$K(\mu, \nu, s)$$
$$= \frac{c\nu(\nu_0^2 - \mu^2)(s - c)X(-\mu, s)X(-\nu, s)p(\mu, s)e^{-2sa/\nu}}{2(\mu + \nu)},$$
$$(3.13)$$

and

$$h_\pm(\mu, s) = (\nu_0^2 - \mu^2)(s - c)X(-\mu, s)p(\mu, s)$$
$$\times [X(\nu_0, s)\psi_+(-a, \mu, s)$$
$$\pm X(-\nu_0, s)\psi_-(-a, \mu, s)], \quad (3.14)$$

$$n_\pm(s) = -\tfrac{1}{2}(c\nu_0)X(\nu_0, s)e^{sa/\nu_0}$$
$$\pm \tfrac{1}{2}(c\nu_0)X(-\nu_0, s)e^{-sa/\nu_0}, \quad (3.15)$$

where

$$p(\mu, s) = 1/[\lambda^2(\mu, s) + \tfrac{1}{4}\pi^2c^2\mu^2]. \quad (3.16)$$

The function $X(z, s)$ is analytic in z on the complex plane cut along the real interval $(0, 1)$ and is defined by

$$X(z, s) = \frac{\exp \Gamma(z, s)}{1 - z}, \quad s \in S_i, \quad (3.17)$$

where

$$\Gamma(z, s) = \int_0^1 \frac{\ln T(\mu', s)}{\mu' - z} \, d\mu', \quad s \in S_i \cup S_e, \quad (3.18)$$

$$T(\mu, s) = \frac{\lambda(\mu, s) + \tfrac{1}{2}i\pi c\mu}{\lambda(\mu, s) - \tfrac{1}{2}i\pi c\mu}. \quad (3.19)$$

Equations (3.10) are Fredholm integral equations of the second kind which have bounded kernels (Re $s > 0$), and they determine $A_\pm'(\nu, s)$. It is apparent from the above that $a_\pm(s)$ are arbitrary at this point. This is entirely proper because $a_\pm(s)$ represent the amplitudes of the eigenfunctions and should depend on how these eigenfunctions are "excited" by the initial distribution $f(x, \mu)$. We shall assume that $a_\pm(s)$ are determined from Eq. (1.24). Since they depend only on s, c, and a, Eqs. (3.11) are determining expressions for the eigenvalues $\{\alpha_i\}$. In fact, we know from the results of Lehner and Wing that these equations have solutions at only a finite number of values of s and these values are real and positive. However, suppose Eqs. (3.10a) and (3.11a) have a nontrivial solution at $s = \alpha_i$; then it is not apparent whether Eqs. (3.10b) and (3.11b) also have a nontrivial solution at the same value of s.

We proceed by setting

$$a_-(\alpha_i) = A_-'(\nu, \alpha_i) = 0 \quad (3.20a)$$

in Eqs. (3.8) to obtain

$$a_1(\alpha_i) = a_2(\alpha_i) = a_+(\alpha_i)/2 \equiv a_{i+}, \quad (3.20b)$$

$$A_1(\nu, \alpha_i) = A_2(\nu, \alpha_i)$$
$$= A_+(\nu, \alpha_i)e^{-\alpha_i a/\nu}/2 \equiv a_{i+}A_{i+}(\nu). \quad (3.20c)$$

Substituting these values into Eq. (3.1) and denoting

$\Psi_i(x, \mu) \equiv \Psi(x, \mu, \alpha_i)$ yields

$$\Psi_i(x, \mu) = a_{i+}\Big\{\psi_+(x, \mu, \alpha_i) + \psi_-(x, \mu, \alpha_i)$$

$$+ \int_0^1 A_{i+}(\nu)[\psi(x, \mu, \nu, \alpha_i)$$

$$+ \psi(x, \mu, -\nu, \alpha_i)] \, d\nu\Big\}. \quad (3.21)$$

Similarly, if Eqs. (3.10b) and (3.11b) have a non-trivial solution at $s = \alpha_k$, then we set

$$a_+(\alpha_k) = A_+'(\nu, \alpha_k) = 0 \quad (3.22a)$$

to obtain

$$a_1(\alpha_k) = -a_2(\alpha_k) = a_-(\alpha_k)/2 \equiv a_{k-}, \quad (3.22b)$$

$$A_1(\nu, \alpha_k) = -A_2(\nu, \alpha_k)$$

$$= A_-(\nu, \alpha_k)e^{-\alpha_k a/\nu}/2 \equiv a_{k-}A_{k-}(\nu). \quad (3.22c)$$

Substituting these values into Eq. (3.1), we get

$$\Psi_k(x, \mu) = a_{k-}\Big\{\psi_+(x, \mu, \alpha_k) - \psi_-(x, \mu, \alpha_k)$$

$$+ \int_0^1 A_{k-}(\nu)[\psi(x, \mu, \nu, \alpha_k) - \psi(x, \mu, -\nu, \alpha_k)] \, d\nu\Big\}.$$

$$(3.23)$$

These eigenfunctions have exactly the required symmetry properties noted by Lehner and Wing, that is,

$$\Psi_i(x, \mu) = \pm\Psi_i(-x, -\mu), \quad (3.24)$$

where the plus sign goes with Eq. (3.21) and the negative sign with (3.23).

IV. THE ADJOINT EIGENFUNCTIONS $\Psi_i{}^\dagger(x, \mu)$

We look for solutions of Eq. (1.26) by separation of variables in the form

$$\psi^\dagger(x, \mu, s) = \phi^\dagger(\mu, \nu, s)e^{sx/\nu}, \quad (4.1)$$

where again the spectrum of ν and the shape of $\phi^\dagger(\mu, \nu, s)$ are to be determined for every value of s. We find that Eq. (1.26) becomes

$$-\frac{\mu}{\nu}s\phi^\dagger(\mu, \nu, s) + s\phi^\dagger(\mu, \nu, s)$$

$$= \frac{c}{2}\int_{-1}^1 \phi^\dagger(\mu', \nu, s) \, d\mu', \quad (4.2)$$

and normalizing so that

$$\int_{-1}^1 \phi^\dagger(\mu, \nu, s) \, d\mu = s, \quad (4.3)$$

we get

$$[\nu - \mu]\phi^\dagger(\mu, \nu, s) = \tfrac{1}{2}c\nu, \quad (4.4)$$

which is the same as Eq. (2.4). Therefore, the spectrum of ν is the same as that obtained in Sec. II and

$$\phi^\dagger(\mu, \nu, s) = \phi(\mu, \nu, s). \quad (4.5)$$

Thus, Eq. (1.26) has the discrete solutions

$$\psi_\pm^\dagger(x, \mu, s) = \phi_\pm(\mu, s)e^{\pm sx/\nu_0} \quad (4.6)$$

and the continuum solutions

$$\psi^\dagger(x, \mu, \nu, s) = \phi(\mu, \nu, s)e^{sx/\nu}. \quad (4.7)$$

We use the completeness property of the ϕ's to express the general solution of the eigenvalue problem (1.23) as

$$\Psi^\dagger(x, \mu, s) = a_1^\dagger(s)\psi_+^\dagger(x, \mu, s) + a_2^\dagger(s)\psi_-^\dagger(x, \mu, s)$$

$$+ \int_{-1}^1 A^\dagger(\nu, s)\psi^\dagger(x, \mu, \nu, s) \, d\nu, \quad (4.8)$$

where the expansion coefficients are obtained from the boundary conditions

$$\Psi^\dagger(\pm a, \mu, s) = 0, \qquad \mu \gtrless 0. \quad (4.9)$$

These boundary conditions imply that

$$0 = a_1^\dagger(s)\phi_+(\mu, s)e^{sa/\nu_0} + a_2^\dagger(s)\phi_-(\mu, s)e^{-sa/\nu_0}$$

$$+ \int_{-1}^1 A^\dagger(\nu, s)\phi(\mu, \nu, s)e^{sa/\nu} \, d\nu, \quad \mu > 0, \quad (4.10)$$

and

$$0 = a_1^\dagger(s)\phi_+(-\mu, s)e^{-sa/\nu_0} + a_2^\dagger(s)\phi_-(-\mu, s)e^{sa/\nu_0}$$

$$+ \int_{-1}^1 A^\dagger(\nu, s)\phi(-\mu, \nu, s)e^{-sa/\nu} \, d\nu, \quad \mu > 0. \quad (4.11)$$

But these equations have exactly the same form as Eqs. (3.3) and (3.4). Therefore, the expansion coefficients of the adjoint eigenvalue problem (1.23) are the same as those of the eigenvalue problem (1.22). Likewise, the eigenvalues of these problems coincide. This agrees, of course, with the results of Lehner and Wing. From the general form of the solutions we readily find that

$$\Psi_i^\dagger(x, \mu) = \Psi_i^*(-x, \mu); \quad (4.12)$$

this result was also noted by Lehner and Wing, and, in fact, could have been used to find $\Psi_i^\dagger(x, \mu)$.

V. THE RESOLVENT $R(x, \mu, s)$

We can obtain an expression for the resolvent function $R(x, \mu, s)$ of Eq. (1.21) from the Green's

function $G(x, \mu, s; x_0)$ which satisfies the equation

$$[s - A]G(x, \mu, s; x_0) = \delta(x - x_0)f(x_0, \mu) \qquad (5.1)$$

and the boundary conditions

$$G(\pm a, \mu, s; x_0) = 0, \quad \mu \lessgtr 0, \qquad (5.2)$$

viz.,

$$R(x, \mu, s) = \int_{-a}^{a} G(x, \mu, s; x_0) \, dx_0. \qquad (5.3)$$

For $x \neq x_0$, $G(x, \mu, s; x_0)$ obviously satisfies Eq.

(1.25). We can determine the behavior of G near $x = x_0$ by integrating Eq. (5.1) on x over $(x_0 - \epsilon, x_0 + \epsilon)$ and passing to the limit as $\epsilon \to 0$ to obtain the jump condition of the Green's function:

$$G(x_0+, \mu, s; x_0) - G(x_0-, \mu, s; x_0)$$
$$= f(x_0, \mu)/\mu. \qquad (5.4)$$

We look for a linear combination of the elementary solutions of Eq. (1.25) which will satisfy the jump condition (5.4) and the boundary conditions (5.2). First we introduce

$$G_0(x - x_0, \mu, s) = c_+(x_0, s)\psi_+(x - x_0, \mu, s) + \int_0^1 C(x_0, \nu, s)\psi(x - x_0, \mu, \nu, s) \, d\nu, \qquad x > x_0, \qquad s \in S_i,$$

$$= -c_-(x_0, s)\psi_-(x - x_0, \mu, s) - \int_{-1}^0 C(x_0, \nu, s)\psi(x - x_0, \mu, \nu, s) \, d\nu, \qquad x < x_0, \qquad s \in S_i,$$
$$\hspace{12cm} (5.5)$$

$$= \int_0^1 C(x_0, \nu, s)\psi(x - x_0, \mu, \nu, s) \, d\nu, \qquad x > x_0, \qquad s \in S_e,$$

$$= -\int_{-1}^0 C(x_0, \nu, s)\psi(x - x_0, \mu, \nu, s) \, d\nu, \qquad x < x_0, \qquad s \in S_e,$$

and insert it into the jump condition to obtain

$$f(x_0, \mu)/\mu = c_+(x_0, s)\phi_+(\mu, s) + c_-(x_0, s)\phi_-(\mu, s)$$
$$+ \int_{-1}^1 C(x_0, \nu, s)\phi(\mu, \nu, s) \, d\nu, \quad s \in S_i,$$

$$= \int_{-1}^1 C(x_0, \nu, s)\phi(\mu, \nu, s) \, d\nu, \quad s \in S_e. \qquad (5.6)$$

For arbitrarily fixed x_0, we know from the completeness property that the expansion (5.6) is possible. In fact, from Eq. (2.22) we have

$$c_\pm(x_0, s) = \frac{1}{N_\pm(s)} \int_{-1}^1 f(x_0, \mu)\phi_\pm(\mu, s) \, d\mu, \quad s \in S_i. \quad (5.7)$$

Multiplying Eq. (5.6) by $\mu\phi(\mu, \nu', s)$ and integrating on μ over $(-1, 1)$, we obtain, with the use of Eqs. (2.17) and (2.18),

$$C(x_0, \nu, s) = \frac{1}{N(\nu, s)} \int_{-1}^1 f(x_0, \mu)\phi(\mu, \nu, s) \, d\mu,$$
$$s \in S_i \cup S_e. \qquad (5.8)$$

In order to satisfy the boundary conditions (5.2), we add to $G_0(x - x_0, \mu, s)$ an appropriately adjusted solution of Eq. (1.25) to obtain

$$G(x, \mu, s; x_0) = G_0(x - x_0, \mu, s)$$
$$+ d_1(x_0, s)\psi_+(x, \mu, s) + d_2(x_0, s)\psi_-(x, \mu, s)$$
$$+ \int_{-1}^1 D(x_0, \nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad s \in S_i,$$

$$= G_0(x - x_0, \mu, s)$$
$$+ \int_{-1}^1 D(x_0, \nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad s \in S_e, \quad (5.9)$$

where the expansion coefficients are to be determined from these boundary conditions. However, it is more convenient to work with $R(x, \mu, s)$ directly, which according to the prescription (5.3) is then of the form

$$R(x, \mu, s) = \eta(x, \mu, s)$$
$$+ b_1(s)\psi_+(x, \mu, s) + b_2(s)\psi_-(x, \mu, s)$$
$$+ \int_{-1}^1 B(\nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad s \in S_i,$$

$$= \eta(x, \mu, s) + \int_{-1}^1 B(\nu, s)\psi(x, \mu, \nu, s) \, d\nu, \quad s \in S_e, \quad (5.10)$$

where in this case the expansion coefficients $b_1(s)$, $b_2(s)$, and $B(\nu, s)$ are determined from the boundary conditions

$$R(\pm a, \mu, s) = 0, \quad \mu \lessgtr 0 \qquad (5.11)$$

and $\eta(x, \mu, s)$ is given by

$$\eta(x, \mu, s) = \int_{-a}^{a} G_0(x - x_0, \mu, s) \, dx_0. \qquad (5.12)$$

It can be readily shown that $R(x, \mu, s)$ given by Eq. (5.10) is the general solution to Eq. (1.21). Next we define

$$\eta_\pm(\mu, s) \equiv -\eta(\pm a, \mp\mu, s) = \tau_\pm(s)\psi_-(-a, \mu, s)$$

$$+ \int_0^1 \tau(\pm\nu, s)\psi(-a, \mu, -\nu, s)\, d\nu, \quad s \in S_i$$

$$= \int_0^1 \tau(\pm\nu, s)\psi(-a, \mu, -\nu, s)\, d\nu, \quad s \in S_e, \quad (5.13)$$

where

$$\tau_\pm(s) = \frac{1}{N_-(s)} \int_{-a}^a \int_{-1}^1 \psi_\pm^\dagger(x_0, \mu_0, s)f(x_0, \mu_0)\, dx_0\, d\mu_0$$

$$(5.14)$$

and

$$\tau(\nu, s) = \frac{1}{N(-\nu, s)}$$

$$\times \int_{-a}^a \int_{-1}^1 \psi^\dagger(x_0, \mu_0, \nu, s)f(x_0, \mu_0)\, dx_0\, d\mu_0. \quad (5.15)$$

If we now let

$$B(\nu, s) = B_1(\nu, s)H(\nu) + B_2(-\nu, s)H(-\nu), \quad (5.16)$$

where $H(\nu)$ is defined by Eq. (3.6), the boundary conditions (5.11) require that

$$\eta_-(\mu, s) = b_1(s)\psi_+(-a, \mu, s) + b_2(s)\psi_-(-a, \mu, s)$$

$$+ \int_0^1 \{B_1(\nu, s)\psi(-a, \mu, \nu, s)$$

$$+ B_2(\nu, s)\psi(-a, \mu, -\nu, s)\}\, d\nu, \quad \mu > 0, \quad s \in S_i,$$

$$= \int_0^1 \{B_1(\nu, s)\psi(-a, \mu, \nu, s)$$

$$+ B_2(\nu, s)\psi(-a, \mu, -\nu, s)\}\, d\nu, \quad \mu > 0, \quad s \in S_e, \quad (5.17)$$

$$\eta_+(\mu, s) = b_1(s)\psi_-(-a, \mu, s) + b_2(s)\psi_+(-a, \mu, s)$$

$$+ \int_0^1 \{B_1(\nu, s)\psi(-a, \mu, -\nu, s)$$

$$+ B_2(\nu, s)\psi(-a, \mu, \nu, s)\}\, d\nu, \quad \mu > 0, \quad s \in S_i$$

$$= \int_0^1 \{B_1(\nu, s)\psi(-a, \mu, -\nu, s)$$

$$+ B_2(\nu, s)\psi(-a, \mu, \nu, s)\}\, d\nu, \quad \mu > 0, \quad s \in S_e. \quad (5.18)$$

Equations (5.17) and (5.18) are coupled singular integral equations involving the expansion parameters $b_1(s)$, $b_2(s)$, $B_1(\nu, s)$, and $B_2(\nu, s)$ for the resolvent $R(x, \mu, s)$. We find that these equations are sufficient to determine these parameters. However, the problem can be uncoupled if we do not solve for the parameters directly, but for a combination of them by a procedure similar to that used in Sec. III.

By adding and subtracting Eqs. (5.17) and (5.18) and letting

$$b_1(s) \pm b_2(s) = b_\pm(s),$$

$$[B_1(\nu, s) \pm B_2(\nu, s)]e^{sa/\nu} = B_\pm(\nu, s), \quad (5.19)$$

we obtain

$$\eta_-(\mu, s) \pm \eta_+(\mu, s)$$

$$= b_\pm(s)[\phi_+(\mu, s)e^{sa/\nu_0} \pm \phi_-(\mu, s)e^{-sa/\nu_0}]$$

$$+ \int_0^1 B_\pm(\nu, s)[\phi(\mu, \nu, s)$$

$$\pm \phi(\mu, -\nu, s)e^{-2sa/\nu}]\, d\nu, \quad \mu > 0, \quad s \in S_i,$$

$$= \int_0^1 B_\pm(\nu, s)[\phi(\mu, \nu, s)$$

$$\pm \phi(\mu, -\nu, s)e^{-2sa/\nu}]\, d\nu, \quad \mu > 0, \quad s \in S_e. \quad (5.20)$$

Again, for $\mu > 0$ Eqs. (5.20) are singular integral equations whose kernels have singular and nonsingular parts. These equations can also be reduced to equivalent Fredholm integral equations. As these reductions are quite lengthy, we state the results here and outline the reductions in Appendix B. We find that Eqs. (5.20) reduce to the relations

$$B_\pm(\mu, s) = g_\pm(\mu, s) - b_\pm(s)h_\pm(\mu, s)$$

$$\mp \int_0^1 K(\mu, \nu, s)B_\pm(\nu, s)\, d\nu, \quad s \in S_i,$$

$$= g_\pm(\mu, s) \mp \int_0^1 K(\mu, \nu, s)B_\pm(\nu, s)\, d\nu, \quad s \in S_e, \quad (5.21)$$

plus the conditions

$$b_\pm(s)n_\pm(s) = m_\pm(s)$$

$$\mp \frac{c}{2}\int_0^1 \nu X(-\nu, s)B_\pm(\nu, s)e^{-2sa/\nu}\, d\nu, \quad s \in S_i, \quad (5.22)$$

where for $s \in S_i$ the kernel $K(\mu, \nu, s)$ and the functions $h_\pm(\mu, s)$, $n_\pm(s)$, and $X(z, s)$ are defined by Eqs. (3.13)–(3.16) and, for $s \in S_e$,

$$K(\mu, \nu, s)$$

$$= \frac{c\nu X(-\mu, s)[X(-\mu, s) + (s - c)p(\mu, s)X(-\nu, s)]e^{-2sa/\nu}}{2(\mu + \nu)}$$

$$(5.23)$$

and

$$X(z, s) = \exp \Gamma(z, s) \quad (5.24)$$

with $\Gamma(z, s)$ defined by Eq. (3.18). Finally,

$$g_\pm(\mu, s) = (\nu_0^2 - \mu^2)(s - c)p(\mu, s)X(-\mu, s)\Big\{X(-\nu_0, s)[\tau_-(s) \pm \tau_+(s)]\psi_-(-a, \mu, s)$$

$$+ \int_0^1 X(-\nu, s)[\tau(-\nu, s) \pm \tau(\nu, s)]\psi(-a, \mu, -\nu, s)\, d\nu\Big\}, \qquad s \in S_i,$$

$$= X^2(-\mu, s)[\eta_-(\mu, s) \pm \eta_+(\mu, s)] - (s - c)p(\mu, s)X(-\mu, s)$$

$$\times \int_0^1 X(-\nu, s)[\tau(-\nu, s) \pm \tau(\nu, s)]\psi(-a, \mu, -\nu, s)\, d\nu, \qquad s \in S_e.$$

$$(5.25)$$

and

$$m_\pm(s) = \frac{c\nu_0}{2} [\tau_-(s) \pm \tau_+(s)]X(-\nu_0, s)e^{-sa/\nu_0}$$

$$- \frac{c}{2}\int_0^1 \nu[\tau(-\nu, s) \pm \tau(\nu, s)]X(-\nu, s)e^{-sa/\nu}\, d\nu, \qquad s \in S_i.$$

$$(5.26)$$

Equations (5.21) are Fredholm integral equation of the second kind with bounded kernels (Re $s > 0$); they determine $B_\pm(\mu, s)$ apart from the parameters $b_\pm(s)$. Equations (5.22) then determine $b_\pm(s)$.

If we next define $B'_\pm(\mu, s)$ so that

$$B'_\pm(\mu, s) = g_\pm(\mu, s)$$

$$\mp \int_0^1 K(\mu, \nu, s)B'_\pm(\nu, s)\, d\nu, \qquad s \in S_i \cup S_e, \quad (5.27)$$

then

$$B_\pm(\mu, s) = B'_\pm(\mu, s) + b_\pm(s)A'_\pm(\mu, s), \qquad s \in S_i, \quad (5.28)$$

where $A'_\pm(\mu, s)$ satisfy Eqs. (3.10). We then solve for $b_\pm(s)$ to obtain

$$b_\pm(s) = \frac{m_\pm(s) \mp \dfrac{c}{2}\displaystyle\int_0^1 \nu X(-\nu, s)B'_\pm(\nu, s)e^{-2sa/\nu}\, d\nu}{n_\pm(s) \pm \dfrac{c}{2}\displaystyle\int_0^1 \nu X(-\nu, s)A'_\pm(\nu, s)^{-2sa/\nu}\, d\nu}. \quad (5.29)$$

Our original expansion coefficients in Eq. (5.10) may now be expressed in terms of $b_\pm(s)$ and $B_\pm(\mu, s)$ with the use of Eqs. (5.16) and (5.19). Note that the singularities of $R(x, \mu, s)$ occur when the denominators of $b_\pm(s)$ vanish; comparing Eqs. (5.29) with Eqs. (3.11), we see that these singularities occur at the eigenvalues $\{\alpha_i\}$, just as expected.

The appearance of terms containing $p(\mu, s)$ in $R(x, \mu, s)$ may seem to require special attention in the integral of Eq. (1.20), since $1/p(\mu, s)$, $s \in C_e$, vanishes for some value of μ (depending on the value of s). However, this difficulty must be illusional since $R(x, \mu, s)$ is defined by Eq. (5.28) for $s \in S_i \cup S_e$ and the requirement for the existence of the integral in Eq. (1.20) is that $R(x, \mu, s)$ tend to a definite limit as s approaches a point on C_e. This convergence

is guaranteed by the analytic property of $R(x, \mu, s)$ in the right half complex plane of s with the set of eigenvalues $\{\alpha_i\}$ deleted from it.

VI. CALCULATION OF THE EIGENVALUES

It is evident that the above analysis does not give an explicit expression for the distribution of eigenvalues as functions of c and a. However, the expressions (3.11) determining the eigenvalues are quite amenable to numerical calculation. These eigenvalue conditions can be made to depend only on the ratio s/c and the product ca, and it is therefore convenient to represent the distribution of eigenvalues as a sequence of curves in α_i/c versus ca. These curves have the following properties:

(1) $\alpha_0/c \to 0$ as $ca \to 0$.
(2) $\alpha_i/c \to 0$ as $ca \to k_j$, $j = 1, 2, \cdots$, where $k_{i+i} > k_i > 0$ and the sequence $\{k_i\}$ has no finite accumulation point.
(3) the α_i/c are continuous and increasing functions of ca.
(4) $\alpha_i/c \to 1$ as $ca \to \infty$.

Properties (1)–(3) follow from the results of Lehner and Wing. Property 4 follows from the fact that for s/c real, $A'_\pm(\mu, s)$ of Eqs. (3.10) vanish in the limits as $ca \to \infty$ and $s/c \to 1$. This is true since for large values of ca, the Neumann series solutions of Eqs. (3.12) converge uniformly, and in passing to the limits above, the series converges to zero since each term of the series vanishes. The eigenvalue conditions (3.13) are then identically satisfied in these limits since each side of the equations vanishes.

For s real and $s/c < 1$, we write

$$X(z, s) = \frac{1}{1 - z} \exp \int_0^1 \frac{\theta(\mu')}{\mu' - z}\, d\mu', \quad (6.1)$$

where

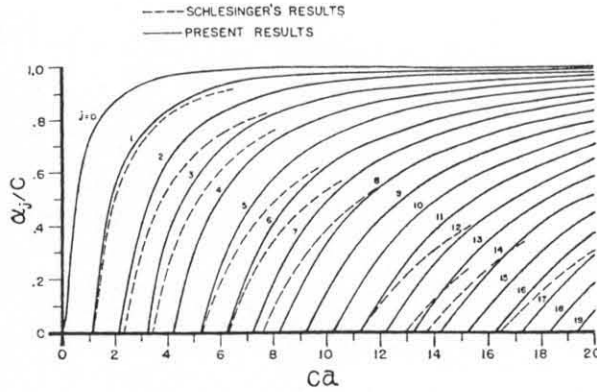$$\theta(\mu) = \frac{1}{\pi} \tan^{-1}\left[\frac{\pi\mu/2}{s/c - \mu \tanh^{-1}\mu}\right]. \quad (6.2)$$

FIG. 2. The eigenvalues α_j.

Now if we let

$$I_1 = \int_0^1 \frac{\mu' \theta(\mu')}{\mu'^2 - \nu_0^2} \, d\mu', \qquad (6.3)$$

$$I_2 = \int_0^1 \frac{\theta(\mu')}{\mu'^2 - \nu_0^2} \, d\mu', \qquad (6.4)$$

and

$$Y(\mu, s) = \left[\frac{\mu(\mu^2 - \nu_0^2)(1 - s/c)}{(s/c - \mu \tanh^{-1} \mu)^2 + \pi^2 \mu^2/4} \right]^{\frac{1}{2}} \qquad (6.5)$$

and define $D_\pm(\mu, s)$ by

$$\frac{i Y(\mu, s)}{\mu(1 - \nu_0^2)} D_+(\mu, s) e^{I_1 + sa/\mu} = A'_+(\mu, s), \qquad (6.6)$$

$$\frac{Y(\mu, s)}{\mu(1 - \nu_0^2)} D_-(\mu, s) e^{I_1 + sa/\mu} = A'_-(\mu, s), \qquad (6.7)$$

Eqs. (3.10) may be rewritten for the above range of s as

$$D_\pm(\mu, s) = d_\pm(\mu, s) \mp \int_0^1 K'(\mu, \nu, s) D_\pm(\nu, s) \, d\nu, \quad (6.8)$$

where the symmetric kernel $K'(\mu, \nu, s)$ is given by

$$K'(\mu, \nu, s)$$
$$= \frac{Y(\mu, s) Y(\nu, s) X(-\mu, s) X(-\nu, s) e^{-sa/\mu} e^{-sa/\nu}}{2(\mu + \nu)} \quad (6.9)$$

and

$$d_+(\mu, s) = \frac{Y(\mu, s) X(-\mu, s)}{\mu^2 + |\nu_0|^2} [(\mu - |\nu_0|^2) \sin \sigma$$
$$+ |\nu_0| (1 + \mu) \cos \sigma] e^{-sa/\mu}, \quad (6.10)$$

$$d_-(\mu, s) = \frac{Y(\mu, s) X(-\mu, s)}{\mu^2 + |\nu_0|^2} [(\mu - |\nu_0|^2) \cos \sigma$$
$$- |\nu_0| (1 + \mu) \sin \sigma] e^{-sa/\mu}, \quad (6.11)$$

with

$$\sigma = |\nu_0| I_2 - sa/|\nu_0|. \qquad (6.12)$$

We now rewrite the eigenvalue conditions (3.11) as

$$\Delta_\pm(s) = 0, \qquad (6.13)$$

where

$$\Delta_+(s) = |\nu_0| \cos \sigma + \sin \sigma$$
$$- \frac{1}{2} \int_0^1 Y(\mu, s) X(-\mu, s) D_+(\mu, s) \, d\mu, \quad (6.14a)$$

$$\Delta_-(s) = -\cos \sigma + |\nu_0| \sin \sigma$$
$$- \frac{1}{2} \int_0^1 Y(\mu, s) X(-\mu, s) D_-(\mu, s) \, d\mu. \quad (6.14b)$$

Equations (6.8) and (6.14) were used to calculate the sequence $\{\alpha_i/c\}$ as a function of ca by numerically solving for the zeros of $\Delta_\pm(s)$.[6] The results for $ca \leq 20$ are displayed in Fig. 2. The curves correspond alternately to eigenfunctions given by Eq. (3.21) and (3.23), the first corresponding to Eq. (3.23). For comparison purposes, the results of Schlesinger (Ref. 4, calculated from Table III) are also shown in Fig. 2.

VII. ASYMPTOTIC SOLUTION

With the transformation (1.4) we have for the actual neutron distribution at time t the expansion

$$N(x, \mu, t) = \sum_{j=0}^M (f, \Psi_j^\dagger) \Psi_j(x, \mu) e^{-(1 - \alpha_j) \Sigma_v t}$$
$$+ \zeta(x, \mu, t) e^{-\Sigma_v t}. \quad (7.1)$$

For large values of t we expect the $j = 0$ term to dominate. Therefore, in the limit of large times, the asymptotic solution is

$$N_0(x, \mu, t) = f_0 \Psi_0(x, \mu) e^{-(1 - \alpha_0) \Sigma_v t}, \qquad (7.2)$$

where

$$f_0 = (f, \Psi_0^\dagger), \qquad (7.3)$$

and $\Psi_0(x, \mu)$ is given by Eq. (3.21) with $j = 0$. Let us assume that $A'_{0+}(\nu, s) = 0$. Since this is strictly true for $ca \to \infty$, we expect this assumption to be valid sufficiently far from the slab edges in large systems. Then

$$N_0(x, \mu, t) = f_0 a_{0+} [\phi_+(\mu, \alpha_0) e^{-\alpha_0 x/\nu_0}$$
$$+ \phi_-(\mu, \alpha_0) e^{\alpha_0 x/\nu_0}] e^{-(1 - \alpha_0) \Sigma_v t} \quad (7.4)$$

and the neutron density

$$\rho(x, t) = \int_{-1}^1 N_0(x, \mu, t) \, d\mu \qquad (7.5)$$

is given in this approximation by

$$\rho(x, t) = f_0 a_{0+} \alpha_0 \cos(\alpha_0 x/|\nu_0|) e^{-(1 - \alpha_0) \Sigma_v t} \quad (7.6)$$

Letting

$$\rho_0(x) = f_0 a_{0+\alpha_0} \cos(\alpha_0 x/|\nu_0|), \qquad (7.7)$$

we note that $\rho_0(x)$ obeys the standard diffusion equation

$$d^2\rho_0(x)/dx^2 + (\alpha_0/|\nu_0|)^2\rho_0(x) = 0. \qquad (7.8)$$

where $(\alpha_0/|\nu_0|)^2$ plays the role of the "buckling" parameter. Also, in this approximation the eigenvalue condition (3.11) becomes

$$X(\nu_0, \alpha_0)/X(-\nu_0, \alpha_0) = \exp(-2a\alpha_0/\nu_0). \qquad (7.9)$$

Now from Eq. (6.1) we find

$$\frac{X(-\nu_0, \alpha_0)}{X(\nu_0, \alpha_0)} = \frac{1-\nu_0}{1+\nu_0} \exp\left[-2\nu_0 \int_0^1 \frac{\theta(\mu')}{\mu'^2 - \nu_0^2} d\mu'\right],$$

$$= \frac{1-\nu_0}{1+\nu_0} \exp\left[2 \int_0^1 \theta(\mu')\left\{\frac{d}{d\mu'}\tanh^{-1}\frac{\mu'}{\nu_0}\right\} d\mu'\right]. \qquad (7.10)$$

Integrating by parts we find

$$X(-\nu_0, \alpha_0)/X(\nu_0, \alpha_0) = \exp(-2\alpha_0 z_0/\nu_0), \qquad (7.11)$$

where

$$z_0 = \int_0^1 \frac{c}{2}\frac{\nu_0}{\alpha_0}\left[\frac{\alpha_0 + c\mu'/(1-\mu'^2)}{\lambda^2(\mu', \alpha_0) + \pi^2 c^2 \mu'^2/4}\right]\left\{\tanh^{-1}\frac{\mu'}{\nu_0}\right\} d\mu', \qquad (7.12)$$

so that Eqs. (7.9) and (7.11) yield the condition

$$\cos\{[\alpha_0/|\nu_0|][\pm(a+z_0)]\} = 0. \qquad (7.13)$$

Therefore, z_0 is the so-called extrapolation distance for the time-dependent slab problem in this approximation.

Let us now assume that the medium is free from fissionable material so that

$$c = \Sigma_s/\Sigma, \qquad (7.14)$$

where

$$\Sigma = \Sigma_a + \Sigma_s, \qquad (7.15)$$

with Σ_s and Σ_a the total scattering and absorption cross sections respectively. From Eq. (7.2) we note that the time behavior of the system after a long time can be characterized by the decay constant

$$\lambda_0 = (1 - \alpha_0)\Sigma v. \qquad (7.16)$$

Let us expand α_0 from Eq. (2.7) in a power series in the buckling $(\alpha_0/|\nu_0|)^2$:

$$\alpha_0\Sigma = \Sigma_s - \frac{\Sigma^2}{3\Sigma_s}\left(\frac{\alpha_0}{|\nu_0|}\right)^2 + \frac{\Sigma^4}{45\Sigma_s^2}\left(\frac{\alpha_0}{|\nu_0|}\right)^4 + \cdots. \qquad (7.17)$$

Then

$$\lambda_0 = \Sigma_a v + Dv\Sigma^2\left(\frac{\alpha_0}{|\nu_0|}\right)^2 - \frac{Dv\Sigma^4}{15\Sigma_s^3}\left(\frac{\alpha_0}{|\nu_0|}\right)^4 + \cdots, \qquad (7.18)$$

where

$$D = \tfrac{1}{3}\Sigma_s \qquad (7.19)$$

is the diffusion coefficient from elementary diffusion theory. Now since

$$\lim_{a\to\infty} \alpha_0/|\nu_0| = 0. \qquad (7.20)$$

we find in this limit the well-known result

$$\lambda_0 \sim \Sigma_a v. \qquad (7.21)$$

The second term in the expansion (7.18) is due to leakage of neutrons from the slab and, in fact, is just the result that one would expect from elementary diffusion theory. Finally, the third term is the first transport correction term to diffusion theory (cf. Nelkin[7]).

VIII. CONCLUSION

In the solution of the initial-value problem of monoenergetic neutrons migrating in slab geometry, as displayed in Eqs. (1.19) and (1.20), it has been shown that the eigenfunctions $\Psi_j(x, \mu)$ and $\Psi_j^\dagger(x, \mu)$ can be exactly represented by an expansion of elementary solutions of Eqs. (1.25) and (1.26) in the forms of Eqs. (3.1) and (4.8), respectively, and $R(x, \mu, s)$ can be represented by an expansion of the form of Eq. (5.10). Furthermore, integral equations have been derived which determine the expansion parameters. In addition, exact expressions for the distribution of eigenvalues $\{\alpha_i\}$ as functions and c and a have been derived and calculations presented for $ca \leq 20$. Finally, in the limit of long time and large slab widths the results have been shown to have properties expected from elementary diffusion theory.

APPENDIX A

The number of zeros m of $\Omega(\nu, s)$ in the complex plane of ν cut along $(-1, 1)$ is given by the principle of the argument in the form

$$m = \frac{1}{2\pi}\Delta_{C_1}\arg\Omega(\nu, s) = \frac{1}{2\pi i}\int_{C_1}\frac{\Omega'(\nu, s)}{\Omega(\nu, s)}d\nu, \qquad (A1)$$

where the prime indicates differentation with respect to ν and $\Delta_{C_1}\arg\Omega(\nu, s)$ represents the change of the argument of $\Omega(\nu, s)$ around the contour C_1 (Fig. 3) generated by letting $\rho \to 0$. This is true unless $s = c$, in which case $|\nu_0| = \infty$. We, therefore,
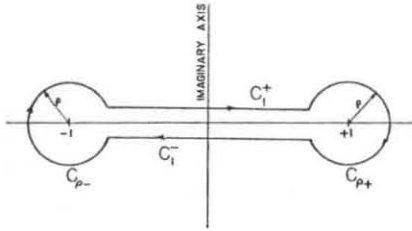
[7] M. Nelkin, Nucl. Sci. Engr. 7, 210 (1960).

FIG. 3. The integration contour C^1.

assume for the rest of this appendix that $s \neq c$. Since $\Omega(\nu, s)$ is sectionally analytic.

$$\lim_{\rho \to 0} \int_{C_{\rho \pm}} \frac{\Omega'(\nu, s)}{\Omega(\nu, s)} \, d\nu = 0, \quad (A2)$$

so that

$$m = \frac{1}{2\pi} \{\Delta_{C_1^+} \arg \Omega^+(\mu, s) + \Delta_{C_1^-} \arg \Omega^-(\mu, s)\}, \quad (A3)$$

where $\Omega^\pm(\mu, s)$ are the boundary values of $\Omega(\nu, s)$ as ν approaches the cut $(-1, 1)$ from the left $(+)$ and from the right $(-)$. From Plemelj's formulas,[5] we find

$$\Omega^\pm(\mu, s) = s - c\mu \tanh^{-1} \mu \pm \tfrac{1}{2} i\pi c\mu. \quad (A4)$$

Let $s = \alpha + i\beta$, so that

$$\arg \Omega^\pm(\mu, s) = \arg \{\alpha - c\mu \tanh^{-1} \mu + i(\beta \pm \tfrac{1}{2}\pi c\mu)\} \quad (A5)$$

and, in particular,

$$\arg \Omega^\pm(\pm 1, s) = \arctan (0). \quad (A6)$$

We first note that if $\alpha < 0$ or $|\beta| > \tfrac{1}{2}\pi c$, then $\Omega(\nu, s)$ has no zeros. Thus to complete the proof of the behavior of the zeros of $\Omega(\nu, s)$ stated in Sec. II, we need be concerned only with $\alpha \geq 0$ and $|\beta| \leq \tfrac{1}{2}\pi c$. Let us consider $s \notin C_s$ and denote by μ_i and μ_r a zero of the imaginary and real part of the $\arg \Omega^\pm(\mu, s)$, respectively. Common zeros occur only for $s \in C_s$, which we have excluded for the moment. We now note that if $|\mu_i| > |\mu_r|$ then $\Omega(\nu, s)$ has no zeros. But it is simple to show that $|\mu_i| > |\mu_r|$ if and only if $s \in S_s$. For $s \in S_i$, we have

$$\arg \Omega^+(\mp 1, s) = \arg \Omega^-(\pm 1, s) = \mp \pi, \quad (A7)$$

Hence from Eq. (A3), we find that

$$m = \frac{1}{2\pi} \{\pi - (-\pi) + \pi - (-\pi)\} = 2, \quad s \in S_i. \quad (A8)$$

Finally, we consider $s \in C_s$, in which case

$$\Omega(\nu, s) = \frac{2\beta}{\pi} \tanh^{-1} \frac{2\beta}{\pi c} + i\beta$$

$$- c\nu \tanh^{-1} (1/\nu), \quad s \in C_s. \quad (A9)$$

Setting $\Omega(\nu, s)$ equal to zero and solving for $1/\nu_0$, we obtain

$$\frac{1}{\nu_0} = \tanh \left\{ \frac{2}{\pi c \nu_0} \tanh^{-1} \frac{2\beta}{\pi c} + \frac{i\beta}{c\nu_0} \right\}, \quad s \in C_s. \quad (A10)$$

This equation has the solutions $\pm \nu_0 = 2\beta/\pi c$, as may be easily seen by direct substitution, and these two solutions, which are zeros of $\Omega(\nu, s)$ for $s \in C_s$, lie in the interval $(-1, 1)$. We now show that these are the only zeros of $\Omega(\nu, s)$ for $s \in C_s$. Consider the contour C_2 (Fig. 4) as $\rho \to 0$ and $\tau \to 0$. As above, the number of zeros to the left of C_2 is

$$m = \Delta_{C_2} \arg \Omega(\nu, s) = \int_{C_2} \frac{\Omega'(\nu, s)}{\Omega(\nu, s)} \, d\nu, \quad s \in C_s. \quad (A11)$$

The contribution from the segments $C_{\rho \pm}$ will be zero as before. Since $\Omega'(\nu, s)/\Omega(\nu, s)$ has simple poles at $\pm 2\beta/\pi c$,

$$\lim_{\tau \to 0} \int_{C_{\tau \pm}} \frac{\Omega'(\nu, s)}{\Omega(\nu, s)} \, d\nu$$

$$= -i\pi \operatorname{Re} s (\pm 2\beta/\pi c) = -i\pi. \quad (A12)$$

The contribution to the change in $\arg \Omega(\nu, s)$ from the contour $C_{\tau \pm}$ is then

$$(-i\pi - i\pi)/2\pi i = -1. \quad (A13)$$

It remains only to evaluate

$$\Delta_{C_2 \pm} \arg \Omega^\pm(\nu, s) = \Delta_{C_2 \pm} \arg \{\alpha - c\mu \tanh^{-1} \mu$$

$$+ i(\beta \pm \pi c\mu/2)\}, \quad s \in C_s. \quad (A14)$$

We find that

$$\Delta_{C_2 \pm} \arg \Omega^\pm(\nu, s) = \pi, \quad s \in C_s, \quad (A15)$$

so that

$$m = -1 + (\pi + \pi)/2\pi = 0 \quad (A16)$$

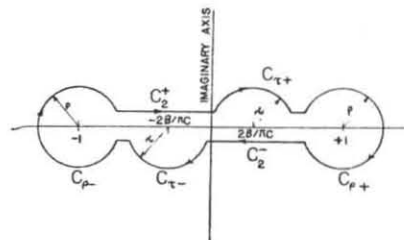as was to be shown. We note here the well-known result that if s is real and $0 < s/c < 1$, then ν_0 is pure imaginary.

FIG. 4. The integration contour C_2.

APPENDIX B

In order to reduce Eqs. (5.20), let us define

$$\zeta_{\pm}(\mu, s) = -b_{\pm}(s)\xi_{\pm}(\mu, s) + \eta_{-}(\mu, s) \pm \eta_{+}(\mu, s)$$

$$\mp \int_0^1 B_{\pm}(\nu, s)\phi(\mu, -\nu, s)e^{-2sa/\nu}\, d\nu, \quad s \in S_i,$$

$$= \eta_{-}(\mu, s) \pm \eta_{+}(\mu, s)$$

$$\mp \int_0^1 B_{\pm}(\nu, s)\phi(\mu, -\nu, s)e^{-2sa/\nu}\, d\nu, \quad s \in S_e, \quad (B1)$$

where

$$\xi_{\pm}(\mu, s) = \phi_{+}(\mu, s)e^{sa/\nu_0} \pm \phi_{-}(\mu, s)e^{-sa/\nu_0}, \quad s \in S_i. \quad (B2)$$

We can then write Eqs. (5.10) as

$$\zeta_{\pm}(\mu, s) = \lambda(\mu, s)B_{\pm}(\mu, s)$$

$$+ P \int_0^1 \frac{c}{2} \frac{\nu B_{\pm}(\nu, s)}{\nu - \mu}\, d\nu, \quad s \in S_i \cup S_e. \quad (B3)$$

These types of singular integral equations may be solved with a method treated extensively by Muskhelishvili.[5] We shall assume that $s \in S_i \cup S_e$ only. Let us also assume that $B_{\pm}(\mu, s)$ exist and obey an H^* condition on $(0, 1)$ and introduce the sectionally analytic functions

$$H_{\pm}(z, s) = \frac{1}{2\pi i} \int_0^1 \frac{c}{2} \frac{\nu B_{\pm}(\nu, s)}{\nu - z}\, d\nu, \quad (B4)$$

which vanish as $1/z$ as $|z| \to \infty$. Applying Plemelj's formulas to $H_{\pm}(z, s)$ we get

$$H_{\pm}^{(+)}(\mu, s) + H_{\pm}^{(-)}(\mu, s) = \frac{1}{\pi i} P \int_0^1 \frac{c}{2} \frac{\nu B_{\pm}(\nu, s)}{\nu - z}\, d\nu, \quad (B5)$$

$$H_{\pm}^{(+)}(\mu, s) - H_{\pm}^{(-)}(\mu, s) = \tfrac{1}{2}c\mu B_{\pm}(\mu, s). \quad (B6)$$

We can write the singular integral equations (B3) in terms of $H_{\pm}^{(\pm)}(\mu, s)$ as

$$T(\mu, s)H_{\pm}^{(+)}(\mu, s) - H_{\pm}^{(-)}(\mu, s)$$

$$= \frac{\tfrac{1}{2}c\mu\zeta_{\pm}(\mu, s)}{\lambda(\mu, s) - \tfrac{1}{2}i\pi c\mu}, \quad (B7)$$

where $T(\mu, s)$ is defined by Eq. (3.19). Thus the problems are reduced to the following nonhomogeneous Hilbert problems in the case of an arc[5]: to find the sectionally analytic functions $H_{\pm}(z, s)$ vanishing as $1/z$ at infinity which satisfy the complex boundary value problems (B7).

Let $[X(z, s)]^{-1}$ be the sectionally analytic function which is the solution of the homogeneous part of Eqs. (B7):

$$T(\mu, s) = X^{+}(\mu, s)/X^{-}(\mu, s). \quad (B8)$$

The solution of this problem is readily found to be that given by Eqs. (3.17) and (5.24). We can now write Eqs. (B7) as

$$X^{+}(\mu, s)H_{\pm}^{(+)}(\mu, s) - X^{-}(\mu, s)H_{\pm}^{(-)}(\mu, s)$$

$$= \gamma(\mu, s)\zeta_{\pm}(\mu, s), \quad (B9)$$

where

$$\gamma(\mu, s) = \frac{c}{2} \frac{\mu X^{-}(\mu, s)}{\lambda(\mu, s) - \tfrac{1}{2}i\pi c\mu}, \quad (B10)$$

and look for functions $X(z, s)H_{\pm}(z, s)$ which are solutions of Eqs. (B9). From Plemelj's formulas, we see that the solutions are

$$X(z, s)H_{\pm}(z, s)$$

$$= \frac{1}{2\pi i} \int_0^1 \frac{\gamma(\mu', s)\zeta_{\pm}(\mu', s)}{\mu' - z}\, d\mu' + \frac{P_{k\pm}(z)}{X(z, s)}, \quad (B11)$$

where $P_{k\pm}(z)$ are arbitrary polymonials. Therefore,

$$H_{\pm}(z, s) = \frac{1}{2\pi i} \frac{1}{X(z, s)}$$

$$\times \int_0^1 \frac{\gamma(\mu', s)\zeta_{\pm}(\mu', s)}{\mu' - z}\, d\mu' + \frac{P_{k\pm}(z)}{X(z, s)}. \quad (B12)$$

We have arrived at Eqs. (B12) by assuming that $B_{\pm}(\mu, s)$ existed. On the other hand, given our $X(z, s)$, if $H_{\pm}(z, s)$ defined by Eqs. (B12) are sectionally analytic functions which vanish as $1/z$ at infinity, then $B_{\pm}(\mu, s)$ defined by Eqs. (B6) will be the solutions of the singular integral equations (B3). The only property which gives us any difficulty is the behavior of $H_{\pm}(z, s)$ at infinity. In order that $H_{\pm}(z, s)$ vanish as $|z| \to \infty$, we must set $P_{k\pm}(z)$ equal to zero. We also find that for $s \in S_i$ the following conditions must hold:

$$\int_0^1 \gamma(\mu', s)\zeta_{\pm}(\mu', s)\, d\mu' = 0. \quad (B13)$$

Assuming this is true, we use Eqs. (B6) to write the solutions of Eqs. (B3) as

$$B_{\pm}(\mu, s) = \lambda(\mu, s)p(\mu, s)\zeta_{\pm}(\mu, s)$$

$$- q(\mu, s)P \int_0^1 \frac{\gamma(\mu', s)\zeta_{\pm}(\mu', s)}{\mu' - \mu}\, d\mu', \quad (B14)$$

where $p(\mu, s)$ is given by Eq. (3.16),

$$q(\mu, s) = \{X^{-}(\mu, s)[\lambda(\mu, s) + i\pi c\mu/2]\}^{-1} \quad (B15)$$

and Eqs. (B13) must hold if $s \in S_i$.

The functions $\zeta_{\pm}(\mu, s)$ contain the unknowns $B_{\pm}(\mu, s)$ and $b_{\pm}(s)$ and we can write the proposed integral equations for the expansion coefficient by

substituting Eqs. (B1) into Eqs. (B14) and (B13). However, the final form of the equations is greatly simplified by the following identities. Similar results were used by Mitsis[8] in the solution of the one-velocity critical problem.

The first identity is

$$X(z, s) = \int_0^1 \frac{\gamma(\mu', s)}{\mu' - z} d\mu', \quad z \notin (0, 1), \quad s \in S_i,$$

$$= \int_0^1 \frac{\gamma(\mu', s)}{\mu' - z} d\mu' + 1, \quad z \notin (0, 1), \quad s \in S_e. \quad \text{(B16)}$$

To show this, we note that the function

$$R(z, s) = X(z, s) - \int_0^1 \frac{\gamma(\mu', s)}{\mu' - z} d\mu', \quad s \in S_i,$$

$$= X(z, s) - \int_0^1 \frac{\gamma(\mu', s)}{\mu' - z} d\mu' - 1, \quad s \in S_e, \quad \text{(B17)}$$

is analytic in the complex plane of z except perhaps for a cut along $(0, 1)$. We then find that

$$R^+(\mu, s) - R^-(\mu, s) = X^-(\mu, s)$$
$$\times \left[\frac{X^+(\mu, s)}{X^-(\mu, s)} - 1 - \frac{i\pi c\mu}{\lambda(\mu, s) - i\pi c\mu/2} \right] = 0, \quad \text{(B18)}$$

and

$$\lim_{z \to \infty} R(z, s) = 0, \quad \text{(B19)}$$

so that $R(z) \equiv 0$ and Eq. (B16) is proved.

The second identity is

$$X(z, s)X(-z, s) = \frac{\Omega(z, s)}{(\nu_0^2 - z^2)(s - c)},$$
$$z \notin (0, 1), \quad s \in S_i,$$

$$= \frac{\Omega(z, s)}{(s - c)}, \quad z \notin (0, 1), \quad s \in S_e, \quad \text{(B20)}$$

where $\Omega(z, s)$ is defined by Eq. (2.8). To show this, consider the function

$$J(z, s) = \frac{\Omega(z, s)}{X(z, s)X(-z, s)(\nu_0^2 - z^2)(s - c)}, \quad s \in S_i,$$

$$= \frac{\Omega(z, s)}{X(z, s)X(-z, s)(s - c)}, \quad s \in S_e, \quad \text{(B21)}$$

[8] G. J. Mitsis, ANL-6459 (1961) (unpublished).

which is analytic in the complex plane of z except perhaps for a cut along $(-1, 1)$. We find then that

$$J^+(\mu, s)/J^-(\mu, s) = 1 \quad \text{(B22)}$$

and

$$\lim_{z \to \infty} J(z, s) = 1, \quad \text{(B23)}$$

so that $J(z) \equiv 1$, and the identity is proved.

Let us define

$$PX(\mu, s) = P \int_0^1 \frac{\gamma(\mu', s)}{\mu' - \mu} d\mu', \quad \mu \in (0, 1). \quad \text{(B24)}$$

We then find from the second identity that

$$X(-\mu, s)PX(\mu, s) = \frac{\lambda(\mu, s)}{(\nu_0^2 - \mu^2)(s - c)}, \quad s \in S_i,$$

$$= \frac{\lambda(\mu, s)}{(s - c)} - X(-\mu, s), \quad s \in S_e, \quad \text{(B25)}$$

and

$$q(\mu, s) = (\nu_0^2 - \mu^2)(s - c)X(-\mu, s)p(\mu, s), \quad s \in S_i,$$

$$= (s - c)X(-\mu, s)p(\mu, s), \quad s \in S_e. \quad \text{(B26)}$$

When we substitute Eqs. (B1) into Eqs. (B13) and (B14) we find integrals of the following forms which are evaluated by decomposing by partial fractions and applying Eqs. (B16) and (B24):

$$P \int \frac{\phi_\pm(\mu, s)\gamma(\mu', s)}{\mu' - \mu} d\mu' = \phi_\pm(\mu, s)[PX(\mu, s)$$
$$- X(\pm\nu_0, s)], \quad s \in S_i, \quad \text{(B27)}$$

$$P \int_0^1 \frac{\gamma(\mu', s)}{\mu' - \mu} \int_0^1 Q(\nu)\phi(\mu, -\nu, s) d\nu d\mu'$$

$$= \int_0^1 Q(\nu)\phi(\mu, -\nu, s)[PX(\mu, s) - X(-\nu, s)] d\nu. \quad \text{(B28)}$$

Equations (5.21) and (5.22) now follow by using the corollaries (B25) and (B26) and canceling common terms.

Finally, we note that Eqs. (3.19) are just Eqs. (5.18) with $\eta_\pm(\mu, s) = 0$ and $s \in S_i$, and their reductions to Eqs. (3.10) and (3.11) follow directly from the above results.

On Scattering of Waves by Objects Imbedded in Random Media: Stochastic Linear Partial Differential Equations and Scattering of Waves by Conducting Sphere Imbedded in Random Media

YUNG MING CHEN

Division of Mathematical Sciences Purdue University Lafayette, Indiana
(Received 30 June, 1964)

A new inhomogeneous linear partial differential equation satisfied by the mean value of the solution of the corresponding inhomogeneous stochastic linear partial differential equation is derived. This new equation has the interesting phenomenon that the differential operators couple with the inhomogeneous terms to form new inhomogeneous terms of the equation. Physically, this means that the randomness of medium and source are coupled together to form new sources. The above approach is then used to derive the equation characterizing wave motions in random media due to random sources. Finally, the problem of scattering of a plane wave by a perfectly conducting sphere of radius a imbedded in a random medium is considered. By utilizing a "pseudopotential" to incorporate the effect of the boundary condition into the reduced wave equation and by the above result, for both ka large and small it is found that up to and including terms of order ϵ^2 (ϵ = perturbation parameter) the mean value of the scattered field can be calculated from the same deterministic scattering problem with k replaced by an effective propagation constant $k\bar{n}$. The specialization of the new formulation to the problem of scattering of a plane wave by a perfectly conducting semi-infinite space checks with a previous result of Chen.

1. INTRODUCTION

THE subject of wave propagation in random media plays an important role in many branches of modern science and engineering concerning wave motions. It may be used to study the following problems in wave propagation. First, if one may wish to consider the case in which the known medium is very complex and the determination of the associated wave motion is impractical, one can choose a random medium in which certain statistical properties of wave motion may be closely related to the actual properties of the wave motion. Secondly, if one may wish to consider the case in which the medium is not known precisely, but in which the probability that the medium is any one of the family of media is known, then one can determine the probability that the wave motion is any one of the associated family of wave motions. One can also determine the mean wave motion and its other statistics. These statistical informations can be then utilized for estimating what is likely to be the wave motion.

Up to now, wave propagation in radom media has been studied quite extensively by many mathematicians and physicists. Some of the interesting results on continuous approach have been obtained by Keller,[1,2] Chen,[3,4] Karal and Keller,[5] etc. Only the most recent works on continuous approach are referred here and the complete bibliographies can be found in the above references. Very little, however, has been done on the subject of scattering phenomenon in random media by the "honest" methods.[1]

In this paper, scattering of waves by an object imbedded in a random medium is investigated by the "honest" method. A result on the mean value of the solution of stochastic partial differential equations more general than that of Ref. 2, is obtained. Then by utilizing the method of "pseudopotential" which incorporates the boundary condition into the stochastic partial differential equation, the above result is applied to the case of scattering of plane waves by a perfectly conducting sphere (the total field vanishes on the surface of the sphere) imbedded in a random medium.

2. STOCHASTIC LINEAR PARTIAL DIFFERENTIAL EQUATION

The stochastic linear partial differential equation $L(q) u = g(q)$ is a family of linear partial differential equations depending upon a parameter q which ranges over a space Ω in which a probability density $P(q)$ is defined. The probability density $P(q)$ determines the probability of a given value of q and, therefore, that of the corresponding linear partial

[1] J. B. Keller, in *Proceedings of the Thirteenth Symposium on Applied Mathematics*, (American Mathematical Society, New York, 1960).

[2] J. B. Keller, *Proceedings of the Seventeenth Symposium on Applied Mathematics* (American Mathematical Society, New York, 1964).

[3] Y. M. Chen, J. Math. & Phys. (to be published).

[4] Y. M. Chen, J. Res. Natl. Bur. Std. (U. S.) "D-Radio Propagation" (to be published).

[5] F. C. Karal and J. B. Keller, J. Math. Phys. (to be published).

differential equation of the family. If the unique solution $u(q)$ of the linear partial differential equation is a single-valued function of q, then $u(q)$ is a random variable and its probability density is $P'(q)$. The mean value or the expectation value of $u(q)$ is then given as $\langle u \rangle = \int_\Omega u(q) P'(q)\, dq$.

Now, if $L(q)$ and $g(q)$ can be expanded into power series in ϵ, where ϵ is a small parameter measuring the randomness in $L(q)$ and $g(q)$, then we have

$$[L_0 + \epsilon L_1(q) + \epsilon^2 L_2(q)]u$$
$$= g_0 + \epsilon g_1(q) + \epsilon^2 g_2(q) + O(\epsilon^3). \qquad (1)$$

Our main purpose of this section is to derive the linear partial differential equation satisfied by $\langle u \rangle$ up to and including terms of order ϵ^2. In the following, we will achieve this by using Keller's approach.[2] Upon letting

$$L_0 u_0 = g_0 \qquad (2)$$

and assuming that L_0^{-1} is defined, we solve formally for u from (1) and (2) as

$$u = u_0 + \epsilon L_0^{-1}(g_1 - L_1 u)$$
$$+ \epsilon^2 L_0^{-1}(g_2 - L_2 u) + O(\epsilon^3). \qquad (3)$$

It is also reasonable to suppose that u can be represented as a power series in ϵ,

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + O(\epsilon^3). \qquad (4)$$

Upon inserting (4) into the right-hand side of (3), we obtain

$$u = u_0 + \epsilon L_0^{-1}(g_1 - L_1 u_0)$$
$$+ \epsilon^2 L_0^{-1}(g_2 - L_2 u_0 - L_1 u_1) + O(\epsilon^3). \qquad (5)$$

In order to find the expression of u_1 in terms of u_0, we substitute (4) into both sides of (3) and equate to zero the coefficient of each power of ϵ, then

$$u_1 = L_0^{-1} g_1 - L_0^{-1} L_1 u_0. \qquad (6)$$

By combining (5) and (6) we arrive at the following result,

$$u = u_0 + \epsilon L_0^{-1}(g_1 - L_1 u_0) + \epsilon^2 L_0^{-1}(g_2 - L_2 u_0$$
$$- L_1 L_0^{-1} g_1 + L_1 L_0^{-1} L_1 u_0) + O(\epsilon^3). \qquad (7)$$

Now the expection value of u is

$$\langle u \rangle = u_0 + \epsilon L_0^{-1}(\langle g_1 \rangle - \langle L_1 \rangle u_0) + \epsilon^2 L_0^{-1}(\langle g_2 \rangle - \langle L_2 \rangle u_0$$
$$- \langle L_1 L_0^{-1} g_1 \rangle + \langle L_1 L_0^{-1} L_1 \rangle u_0) + O(\epsilon^3). \qquad (8)$$

From (8), it is found that

$$u_0 = \langle u \rangle + \epsilon(L_0^{-1} \langle L_1 \rangle \langle u \rangle - L_0^{-1} \langle g_1 \rangle) + O(\epsilon^2). \qquad (9)$$

Upon inserting (9) into the terms of $O(\epsilon)$ and $O(\epsilon^2)$ of (8) only, we obtain

$$\langle u \rangle = u_0 + \epsilon L_0^{-1}(\langle g_1 \rangle - \langle L_1 \rangle \langle u \rangle)$$
$$+ \epsilon^2 L_0^{-1}(\langle L_1 \rangle L_0^{-1} \langle g_1 \rangle + \langle g_2 \rangle$$
$$- \langle L_1 L_0^{-1} g_1 \rangle - \langle L_1 \rangle L_0^{-1} \langle L_1 \rangle \langle u \rangle$$
$$- \langle L_2 \rangle \langle u \rangle + \langle L_1 L_0^{-1} L_1 \rangle \langle u \rangle) + O(\epsilon^3). \qquad (10)$$

Finally, by applying L_0 to both sides of (10) and collecting terms, we obtain the "key" equation of this section as

$$[L_0 + \epsilon \langle L_1 \rangle + \epsilon^2(\langle L_2 \rangle + \langle L_1 \rangle L_0^{-1} \langle L_1 \rangle$$
$$- \langle L_1 L_0^{-1} L_1 \rangle)] \langle u \rangle$$
$$= g_0 + \epsilon \langle g_1 \rangle + \epsilon^2(\langle g_2 \rangle + \langle L_1 \rangle L_0^{-1} \langle g_1 \rangle$$
$$- \langle L_1 L_0^{-1} g_1 \rangle) + O(\epsilon^3). \qquad (11)$$

The above result contains that of Keller's,[2] which can be obtained by simply setting g_1 and g_2 to zero. However, an interesting phenomenon arises here because of the presence of g_1 and g_2. It is well known that if the solution of a partial differential equation describes a certain physical phenomenon, the inhomogeneous term in general corresponds to the source of the particular physical problem under consideration. By examining Eq. (11) one finds that the source term is not just $g_0 + \epsilon \langle g_1 \rangle + \epsilon^2 \langle g_2 \rangle$ as one might have guessed but there are additional terms of $O(\epsilon^2)$, $\langle L_1 \rangle L_0^{-1} \langle g_1 \rangle$, and $\langle L_1 L_0^{-1} g_1 \rangle$. This shows the coupling between the differential operators and the source terms. If L_1 and g_1 are statistically dependent, it can be decoupled if and only if one or both of L_1 and g_1 equal to zero. If L_1 and g_1 are statistically independent, it can be decoupled if and only if one or both of $\langle L_1 \rangle$ and $\langle g_1 \rangle$ equal to zero.

3. EQUATIONS CHARACTERIZING WAVE MOTION IN RANDOM MEDIA DUE TO RANDOM SOURCES

Let us now consider a simple application of (11) which describes the wave motion due to a random source in a random medium. Consider the following inhomogeneous wave equation,

$$\nabla^2 u - c^{-2}[1 + \epsilon f(\mathbf{r}, t)]^2 [1 + \epsilon w(\mathbf{r}, t, q)]^2 u_{tt}$$
$$= g_0(\mathbf{r}, t) + \epsilon g_1(\mathbf{r}, t, q) + \epsilon^2 g_2(\mathbf{r}, t, q) + O(\epsilon^3). \qquad (12)$$

Then upon comparing (12) with (11), L_0, L_1, and L_2 are given by

$$L_0 = \nabla^2 - c^{-2} \partial_t^2, \qquad (13)$$

$$L_1 = -2c^{-2}(f + w) \partial_t^2, \qquad (14)$$

and

$$L_2 = -c^{-2}(f^2 + 4fw + w^2)\partial_t^2. \tag{15}$$

The inverse operator of L_0 is given as

$$L_0^{-1}F(\mathbf{r}, t) = -(4\pi)^{-1} \int r^{-1}F(\mathbf{r}, t') \, d\mathbf{r}', \tag{16}$$

where $|\mathbf{r} - \mathbf{r}'| = r$ and $t - r(c)^{-1} = t - s = t'$. With some manipulation we obtain

$$\langle L_1 \rangle \langle u \rangle = -2c^{-2}f\langle u \rangle_{tt}, \tag{17}$$

$$\langle L_2 \rangle \langle u \rangle = -c^{-2}(f^2 + \langle w^2 \rangle)\langle u \rangle_{tt}, \tag{18}$$

$$\langle L_1 \rangle L_0^{-1} \langle L_1 \rangle \langle u \rangle = -(\pi c^4)^{-1} f(\mathbf{r}, t)$$
$$\times \int [f_{tt}(\mathbf{r}, t')\langle u(\mathbf{r}', t')\rangle_{tt} + 2f_t(\mathbf{r}, t')\langle u(\mathbf{r}', t')\rangle_{ttt}$$
$$+ f(\mathbf{r}', t')\langle u(\mathbf{r}', t')\rangle_{tttt}]r^{-1} \, d\mathbf{r}', \tag{19}$$

and

$$\langle L_1 L_0^{-1} L_1 \rangle \langle u \rangle = -(\pi c^4)^{-1}\{\langle f(\mathbf{r}, t) \int [(f(\mathbf{r}', t')$$
$$+ w(\mathbf{r}', t'))\langle u(\mathbf{r}', t')\rangle_{tttt} + 2(f_t(\mathbf{r}', t')$$
$$+ w_t(\mathbf{r}', t'))\langle u(\mathbf{r}', t')\rangle_{ttt} + (f_{tt}(\mathbf{r}', t')$$
$$+ w_{tt}(\mathbf{r}', t'))\langle u(\mathbf{r}', t')\rangle_{tt}]r^{-1} \, d\mathbf{r}'\rangle$$
$$+ \int [C(r, s)\langle u(\mathbf{r}', t')\rangle_{tttt}$$
$$+ 2C_s(r, s)\langle u(\mathbf{r}', t')\rangle_{ttt}$$
$$+ C_{ss}(r, s)\langle u(\mathbf{r}', t')\rangle_{tt}]r^{-1} \, d\mathbf{r}'$$
$$+ \langle w(\mathbf{r}, t) \int [f(\mathbf{r}', t')\langle u(\mathbf{r}', t')\rangle_{tttt}$$
$$+ 2f_t(\mathbf{r}', t')\langle u(\mathbf{r}', t')\rangle_{ttt}$$
$$+ f_{tt}(\mathbf{r}', t')\langle u(\mathbf{r}', t')\rangle_{tt}]r^{-1} \, d\mathbf{r}'\rangle\}, \tag{20}$$

where

$$C(r, |t - \bar{t}'|) = \langle w(\mathbf{r}, t)w(\mathbf{r}', \bar{t}')\rangle \tag{21}$$

is the correlation function for the medium being statistically homogeneous and isotropic in space and stationary in time, and also

$$\langle L_1 \rangle L_0^{-1} \langle g_1 \rangle = (2\pi c^2)^{-1} f(\mathbf{r}, t) \int \langle g_1(\mathbf{r}', t')\rangle_{tt} r^{-1} \, d\mathbf{r}', \tag{22}$$

$$\langle L_1 L_0^{-1} g_1 \rangle = (2\pi c^2)^{-1} [f(\mathbf{r}, t) \int \langle g_1(\mathbf{r}', t')\rangle_{tt} r^{-1} \, d\mathbf{r}'$$
$$+ \langle w(\mathbf{r}, t) \int g_{1tt}(\mathbf{r}', t')r^{-1} \, d\mathbf{r}'\rangle]. \tag{23}$$

Upon collecting above results and substituting them in (11) we finally obtain

$$\{\nabla^2 - (c^2)^{-1}[1 + \epsilon 2f + \epsilon^2(f^2 + \langle w^2 \rangle)]\partial_t^2\}\langle u \rangle$$
$$+ \epsilon^2(\pi c^4)^{-1}\{\int [C(r, s)\langle u(\mathbf{r}', t')\rangle_{tttt}$$
$$+ 2C_s(r, s)\langle u(\mathbf{r}', t')\rangle_{ttt}$$
$$+ C_{ss}(r, s)\langle u(\mathbf{r}', t')\rangle_{tt}]\cdot r^{-1} \, d\mathbf{r}'$$
$$+ \int [(f(\mathbf{r}, t)w(\mathbf{r}', t') + f(\mathbf{r}', t')w(\mathbf{r}, t))$$

$$\times \langle u(\mathbf{r}', t')\rangle_{tttt} + 2(f(\mathbf{r}, t)w_t(\mathbf{r}', t')$$
$$+ f_t(\mathbf{r}', t')w(\mathbf{r}, t))\langle u(\mathbf{r}', t')\rangle_{ttt}$$
$$+ (f(\mathbf{r}, t)w_{tt}(\mathbf{r}', t') + f_{tt}(\mathbf{r}', t')w(\mathbf{r}, t))$$
$$\times \langle u(\mathbf{r}', t')\rangle_{tt}r^{-1} \, d\mathbf{r}' = g_0 + \epsilon\langle g_1 \rangle$$
$$+ \epsilon^2[\langle g_2 \rangle - (2\pi c^2)^{-1}\langle w(\mathbf{r}, t)$$
$$\times \int g_{1tt}(\mathbf{r}', t')r^{-1} \, d\mathbf{r}'\rangle] + O(\epsilon^3). \tag{24}$$

The above result gives Eq. (8) of Keller[2] if f, g_0, g_1, and g_2 are set to zero. Also, if one lets $f(\mathbf{r}, t, q) = f(\mathbf{r})$, $w(\mathbf{r}, t, q) = w(\mathbf{r}, q)$, $\langle u(\mathbf{r}, t, q) \rangle = \langle v(\mathbf{r}, q)\rangle e^{-i\omega t}$, $g_0 = -\delta(\mathbf{r})$, and $g_1 = g_2 = 0$, then (24) gives the same result as Eq. (24) of Chen.[3] It is also interesting to notice that the coupling between the randomness of medium and the randomness of source constitutes a new kind of source.

4. SCATTERING OF A PLANE WAVE BY A PERFECTLY CONDUCTING SPHERE IMBEDDED IN A RANDOM MEDIUM

The problem of scattering of a plane wave by a perfectly conducting sphere of radius a imbedded in a random medium is considered here. Let the center of the sphere be the origin of a spherical coordinate system (r, θ, ϕ) and the incident field be independent of ϕ. This physical problem can be formulated mathematically as the following boundary value problem.

The total field satisfies the reduced wave equation

$$\nabla^2 u + k^2[1 + \epsilon w(\mathbf{r}, q)]^2 u = 0, \quad \text{for } r < a, \tag{25}$$

$$u(\mathbf{r}) = 0, \quad \text{for } r \le a, \tag{26}$$

and

$$\lim_{r \to \infty} r\left[\frac{\partial u}{\partial r} - ik(1 + \epsilon w)u\right] = 0, \tag{27}$$

where

$$\mathbf{r} = (r, \theta, \phi) \quad \text{and} \quad |\mathbf{r}| = r.$$

For simplicity $\langle w \rangle = 0$ is assumed. In order to use the result of (11) of Sec. 2 we must incorporate the effect of boundary condition (26) into the partial differential equation (25) such that the new partial differential equation with no boundary condition gives the same solution as that of (25) and (26) in the exterior region of the sphere. For this purpose we shall employ the "pseudopotential" as an equivalent for the boundary condition (26).

Pseudopotential was first introduced by Fermi[6] in scattering problem, but he limited its use to the

---

[6] E. Fermi, Ric. Sci. 7, 13 (1936).

Born approximation. Huang and Yang[7] generalized it to include all the partial waves, but the form of their pseudopotential is rather complicated. Later, Liu and Wong[8] have derived a simpler form of the generalized pseudopotential. From above results it is obvious that the representation of the pseudopotential is not unique in general and one can always make the choice which suits him best.

Now, let L_{0B} be the pseudopotential which incorporates the effect of the boundary condition (26) into (25), and then we obtain

$$\{[\nabla^2 + L_{0B} + k^2] + \epsilon 2k^2 w(\mathbf{r}, q)$$
$$+ \epsilon^2 k^2 w^2(\mathbf{r}, q)\} u(\mathbf{r}, q) = 0, \quad \text{for} \quad r \geq a \quad (28)$$

and

$$\lim_{r \to \infty} r\left[\frac{\partial u}{\partial r} - ik(1 + \epsilon w)u\right] = 0. \quad (29)$$

{If $w(\mathbf{r}, q) = w(r, q)$,

then

$$L_{0B} = -a^{-1}\delta(r - a)[(\partial/\partial r)r]_{r=a_+}$$

where

$$a_+ = \lim_{\delta \to 0}(a + \delta) \quad \text{and} \quad \delta > 0.[8]\}$$

By comparing Eqs. (1) and (28) we obtain

$$L_0 = \nabla^2 + k^2 + L_{0B} \quad (30)$$

$$L_1 = 2k^2 w(\mathbf{r}, q), \quad (31)$$

and

$$L_2 = k^2 w^2(\mathbf{r}, q), \quad (32)$$

therefore,

$$\langle L_1 \rangle = 0 \quad (33)$$

and

$$\langle L_2 \rangle = k^2 \langle w^2 \rangle. \quad (34)$$

Then $\langle u \rangle$ satisfies

$$[L_0 + \epsilon^2(\langle L_2 \rangle - \langle L_1 L_0^{-1} L_1 \rangle)]\langle u \rangle = O(\epsilon^3) \quad (35)$$

and its appropriate radiation condition. The Green's function of differential oerator L_0 is well known as

$$G(\mathbf{r}, \mathbf{r}') = G_{in}(\mathbf{r}, \mathbf{r}') + G_s(\mathbf{r}, \mathbf{r}'), \quad (36)$$

where

$$G_{in}(\mathbf{r}, \mathbf{r}') = e^{ik|\mathbf{r}-\mathbf{r}'|}/4\pi |\mathbf{r} - \mathbf{r}'| \quad (37)$$

is the part of the Green's function due to incident field and

$$G_s(\mathbf{r}, \mathbf{r}') = -\frac{ik}{4\pi}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\epsilon_m(2n+1)\frac{(n-m)!}{(n+m)!}$$
$$\times \cos[m(\phi - \phi')]P_n^m(\cos\theta)P_n^m(\cos\theta')$$

[7] K. Huang and C. N. Yang, Phys. Rev. **105**, 767 (1957).
[8] L. Liu and K. W. Wong, Phys. Rev. **132**, 3, 1349 (1963).

$$\times [j_n(ka)/h_n^{(1)}(ka)]h_n^{(1)}(kr_>)h_n^{(1)}(kr_<), \quad (38)$$

$$r_> = \max(r, r'), \quad r_< = \min(r, r'),$$

$$\epsilon_0 = 1, \quad \epsilon_1 = \epsilon_2 = \epsilon_3 = \cdots = 2,$$

is the part of Green's function due to the scattered field. Then

$$L_0^{-1}F(\mathbf{r}) = (L_{0in}^{-1} + L_{0s}^{-1})F(\mathbf{r}), \quad (39)$$

where

$$L_{0in}^{-1}F(\mathbf{r}) = -\int G_{in}(\mathbf{r}, \mathbf{r}')F(\mathbf{r}')\,d\mathbf{r}' \quad (40)$$

and

$$L_{0s}^{-1}F(\mathbf{r}) = -\int G_s(\mathbf{r}, \mathbf{r}')F(\mathbf{r}')\,d\mathbf{r}'. \quad (41)$$

Upon using above results, we have

$$\langle L_1 L_0^{-1} L_1 \rangle\langle u \rangle$$

$$= -4k^4 \int G_{in}(\mathbf{r}, \mathbf{r}')C(|\mathbf{r} - \mathbf{r}'|)\langle u(\mathbf{r}', q)\rangle\,d\mathbf{r}'$$

$$- 4k^4 \int G_s(\mathbf{r}, \mathbf{r}')C(|\mathbf{r} - \mathbf{r}'|)\langle u(\mathbf{r}', q)\rangle\,d\mathbf{r}', \quad (42)$$

where

$$C(|\mathbf{r} - \mathbf{r}'|) = \langle w(\mathbf{r})w(\mathbf{r}')\rangle \quad (43)$$

is the correlation function for the medium being statistically homogeneous and isotropic in space. By employing the mean-value theorem for any solution of the reduced wave equation,[1-3] we find

$$4k^4 \int G_{in}(\mathbf{r}, \mathbf{r}')C(|\mathbf{r} - \mathbf{r}'|)\langle u(\mathbf{r}')\rangle\,d\mathbf{r}'$$

$$= -i2k^3\left[\int_0^\infty (e^{i2k\alpha} - 1)C(\alpha)\,d\alpha\right]\langle u \rangle \quad (44)$$

with

$$\alpha = |\mathbf{r} - \mathbf{r}'|.$$

Then (35) becomes

$$[\nabla^2 + k^2\tilde{n}^2 + L_{0B}]\langle u \rangle$$

$$+ \epsilon^2 4k^4 \int G_s(\mathbf{r}, \mathbf{r}')C(|\mathbf{r} - \mathbf{r}'|)\langle u(\mathbf{r}', q)\rangle\,d\mathbf{r}' = O(\epsilon^3), \quad (45)$$

where

$$\tilde{n}^2 = 1 + \epsilon^2\left(\langle w^2 \rangle - i2k \int_0^\infty (e^{i2k\alpha} - 1)C(\alpha)\,d\alpha\right). \quad (46)$$

The integral operator of (45) represents the effect of interaction between the boundary of the scatterer and the random medium. Obviously, it is a function of (r, θ, ϕ), and because of $h_n^{(1)}(kr)$ it behaves like $1/r$ for $kr \gg 1$.

For $kr > 1$ and $ka \ll 1$, we can asymptotically expand $j_n(ka)/h_n^{(1)}(ka)$ into power series of ka, and then (38) yields

$$G_s(\mathbf{r}, \mathbf{r}') = (4\pi)^{-1}k^2 ah_0^{(1)}(kr_>)h_0^{(1)}(kr_<) + O(k^2a^2). \quad (47)$$

Upon inserting (47) into (45) we obtain

$$[\nabla^2 + k^2\tilde{n}^2 + L_{0B}]\langle u\rangle$$

$$+ \epsilon^2\pi^{-1}k^6a \int h_0^{(1)}(kr_>)h_0^{(1)}(kr_<)$$

$$\times C(|\mathbf{r} - \mathbf{r}'|)\langle u(\mathbf{r}')\rangle \, d\mathbf{r}' = O(\epsilon^3). \quad (48)$$

From Eq. (48) it is obvious that if $ka \to 0$, the integral operator can be neglected. Or better, if the correlation distance is short in comparison with the wavelength and $ka = O(\epsilon)$, it can be again neglected.

For $r > a$ and $ka \gg 1$, we can utilize the well-known Watson transformation or Poisson summation formula to change $G_s(\mathbf{r}, \mathbf{r}')$ into a new series representation such that the leading term of the asymptotic expansion of the new series is enough to give an excellent asymptotic representation of (38). Without loss of generality we shall assume that $G_s(\mathbf{r}, \mathbf{r}')$ is independent of ϕ and $\mathbf{r}' = (r', 0)$, because we are only interested in the functional behavior of $G_s(\mathbf{r}, \mathbf{r}')$ with respect to ka. Now, upon applying Watson transformation to (38) and evaluating it asymptotically[9-11] we obtain

$$G_s(\mathbf{r}, \mathbf{r}')$$

$$\cong \frac{1}{ka}\left\{\frac{a^2\sin 2\psi_3}{\left[\left(1+\frac{r'^2}{a^2}-2\frac{r'}{a}\cos\psi_1\right)^{\frac{1}{2}}\left(\frac{r^2}{a^2}-\sin^2\psi_3\right)^{\frac{1}{2}}-\left(1+\frac{r^2}{a^2}-2\frac{r}{a}\cos\psi_2\right)^{\frac{1}{2}}\left(\frac{r'^2}{a^2}-\sin^2\psi_3\right)^{\frac{1}{2}}\right]2rr'\sin(\psi_1+\psi_2)}\right\}^{\frac{1}{2}}$$

$$\times \exp\{ik[(r^2 + a^2 - 2ar\cos\psi_2)^{\frac{1}{2}} + (r'^2 + a^2 - 2ar'\cos\psi_1)^{\frac{1}{2}}]\}$$

$$+ \frac{1}{12[A'(P_m)]^2}\left(\frac{36}{ka}\right)^{\frac{1}{3}}\left[\frac{\pi}{2rr'\sin\theta(r^2/a^2 - 1)^{\frac{1}{2}}(r'^2/a^2 - 1)^{\frac{1}{2}}}\right]^{\frac{1}{2}}$$

$$\times \frac{\exp[iv_m(\theta + 2\pi) - \frac{1}{4}i\pi] + \exp[iv_m(2\pi - \theta) + \frac{1}{4}i\pi]}{1 + e^{i2\pi v_m}}$$

$$\times \exp\{ik[(r^2 - a^2)^{\frac{1}{2}} + (r'^2 - a^2)^{\frac{1}{2}}]\} - iv_m[\cos^{-1}(a/r) + \cos^{-1}(a/r')] \quad (49)$$

for (r, θ) in the lit region, and

$$G_s(\mathbf{r}, \mathbf{r}') \cong \frac{-1}{12[A'(P_m)]^2}\left(\frac{-36}{ka}\right)^{\frac{1}{3}}\left[\frac{\pi}{2rr'\sin\theta(r^2/a^2 - 1)^{\frac{1}{2}}}\right]^{\frac{1}{2}}\left(\frac{r'^2}{a^2} - 1\right)^{-\frac{1}{4}}$$

$$\times \frac{\exp[+iv_m(-\theta + 2\pi) - \frac{1}{4}i\pi] + \exp[iv_m\theta + \frac{1}{4}i\pi]}{1 + e^{i2\pi v_m}}$$

$$\times \exp\{ik[(r^2 - a^2)^{\frac{1}{2}} + (r'^2 - a^2)^{\frac{1}{2}}]\} - iv_m\left[\cos^{-1}\left(\frac{a}{r}\right) + \cos^{-1}\left(\frac{a}{r'}\right)\right] \quad (50)$$

for (r, θ) in the shadow region, where $v_m = ka + e^{i\pi/3}p_m(ka/6)^{\frac{1}{3}}$, p_m is the smallest root of the Airy function, $A(p_m)$. The physical meaning of ψ_1, ψ_2, and ψ_3 are given in Fig. 1. By carefully examining Eqs. (49) and (50), we find

$$G_s(\mathbf{r}, \mathbf{r}') = O(ka)^{-1}. \quad (51)$$

Then upon inserting (51) into (45) we obtain

$$[\nabla^2 + k^2\tilde{n}^2 + L_{0B}]\langle u\rangle + \epsilon^2 O(ka)^{-1} = O(\epsilon^3). \quad (52)$$

It is obvious from (52) that if $ka \to \infty$, the term involving integral operator can be neglected or even better, if the correlation distance is short in comparison with the wave length and $ka = O(\epsilon^{-1})$, it can be also neglected. The case of scattering of a plane wave by a perfectly conducting semi-infinite space is equivalent to (52) with $ka \to \infty$. Then the solution of (52) is

$$\langle u\rangle = \exp[ik\tilde{n}x\sin\bar{\theta} - ik\tilde{n}y\cos\bar{\theta}]$$

$$+ R\exp[ik\tilde{n}x\sin\bar{\theta} + ik\tilde{n}y\cos\bar{\theta}] + O(\epsilon^3) \quad (53)$$

with

$$R = -1. \quad (54)$$

This is exactly the same result given in the last section of Chen[3] (Fig. 2).

[9] B. R. Levy and J. B. Keller, Commun. Pure Appl. Math. **12**, 159 (1959).
[10] W. Franz, Z. Naturforsch. **9a**, 705 (1954).
[11] H. Bremmer, *Terrestrial Radio Waves* (Elsevier Publishing Company, Inc., New York, 1949).
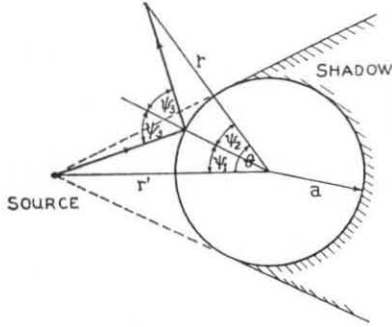
Fig. 1. The geometry of the scattering problem is shown. In addition, the physical meaning of r, r', a, ψ_1, ψ_2, and ψ_3 are given.

In summing up our results above, we have attained for both small ka and large ka that up to and including terms of order ϵ^2, the expectation value of u can be calculated from an equivalent deterministic scattering problem with the propagation constant k replaced by the effective propagation constant $k\tilde{n}$.

For the case that the term in (45) involving integral operator is small but not negligible, we can
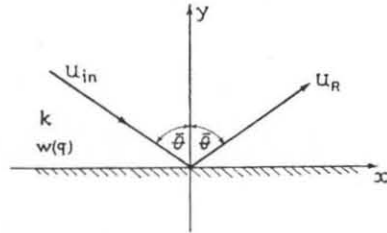

Fig. 2. The directions of propagation of incident and reflected waves are shown.

again employ the method of iterations to obtain the solution. We shall assume that

$$4k^4 \int G_a(\mathbf{r}, \mathbf{r}') C(|\mathbf{r} - \mathbf{r}'|) \langle u(\mathbf{r}', q) \rangle \, d\mathbf{r}' = M(\beta)\langle u \rangle \quad (55)$$

depends upon a small nondimensional parameter β. Upon expanding $M(\beta)$ in powers of β and omitting the $O(\epsilon^3)$ term we may write (45) as

$$[\nabla^2 + k^2 \tilde{n}^2 + L_{0B}]\langle u \rangle$$
$$+ \epsilon^2 (M_0 + \beta M_1 + \beta^2 M_2 + O(\beta^3))\langle u \rangle = 0. \quad (56)$$

It seems also reasonable to suppose that $\langle u \rangle$ can be represented as a power series in β,

$$\langle u \rangle = \langle u \rangle_0 + \beta \langle u \rangle_1 + \beta^2 \langle u \rangle_2 + O(\beta^3). \quad (57)$$

Upon inserting (57) into (56) and equating to zero the coefficient of each power of β, we obtain

$$[\nabla^2 + k^2 \tilde{n}^2 + \epsilon^2 M_0 + L_{0B}]\langle u \rangle_0 = 0, \quad (58)$$

$$[\nabla^2 + k^2 \tilde{n}^2 + \epsilon^2 M_0 + L_{0B}]\langle u \rangle_1 = -\epsilon^2 M_1 \langle u \rangle_0, \quad (59)$$

and

$$[\nabla^2 + k^2 \tilde{n}^2 + \epsilon^2 M_0 + L_{0B}]\langle u \rangle_2$$
$$= -\epsilon^2 M_2 \langle u \rangle_0 - \epsilon^2 M_1 \langle u \rangle_1. \quad (60)$$

Now, the differential–integral equation (45) is transformed into a system of many differential equations, therefore in principle we know how to solve (45) systematically. In our case the parameter β can be ka if $ka < 1$ or can be $(ka)^{-1}$ if $ka > 1$.

Some Remarks Concerning a Pathological Matrix of Interest in the Inverse-Scattering Problem*

Peter J. Redmond

Defense Research Corporation, Santa Barbara, California
(Received 4 February 1964; final manuscript received 30 June 1964)

A Hermitian matrix which occurs in the theory of the quantum-mechanical inverse-scattering problem has apparently contradictory properties. It has a well-behaved inverse in spite of having zero as one of its eigenvalues. The properties of the matrix are investigated and the relevance of the results to the theory are discussed.

INTRODUCTION

IN an elegant paper, Newton[1] has developed a theory for determining a potential from a knowledge of all phase shifts at a given energy. Newton showed that the solution of this problem is not unique and that there is at least one nontrivial central potential which leads to zero phase shifts at all energies. The lack of uniqueness follows from the properties of the matrix

$$N_{l,l'} = (l'^2 - l^2)^{-1} \quad l' - l \quad \text{odd} \tag{1}$$
$$= 0 \quad l' - l \quad \text{even} \quad \text{(N38)}$$

introduced by Newton.[2] Newton exhibited a matrix N^{-1} which was both a right and left inverse of N. The components of N^{-1} are[3]

$$(N^{-1})_{l,l'} = \begin{cases} \dfrac{16}{\pi^2} \dfrac{l'^2}{l^2 - l'^2} & l \text{ even}, \ l' \text{ odd}, \ l \neq 0 \\[2ex] \dfrac{16}{\pi^2} \dfrac{l^2}{l^2 - l'^2} & l \text{ odd}, \ l' \text{ even}, l' \neq 0 \\[2ex] -\dfrac{8}{\pi^2} & l = 0, \ l' \text{ odd} \\[2ex] +\dfrac{8}{\pi^2} & l' = 0, \ l \text{ odd} \\[2ex] 0 & \text{otherwise.} \end{cases} \tag{2}$$
$$\text{(N39)}$$

Newton also showed that a column vector

$$\hat{o}_l = \begin{cases} \tfrac{1}{2} & l = 0 \\ 1 & l = \text{even} \\ 0 & l = \text{odd} \end{cases} \tag{3}$$
$$\text{(N51)}$$

exists which is annihilated by N. That is,

$$\sum_{l'=0}^{\infty} N_{l,l'} \hat{o}_{l'} = 0. \tag{4}$$
$$\text{(N50)}$$

These properties impressed the author as being rather extraordinary. The matrix $-iN$ is Hermitian. Normally one expects an Hermitian matrix to have a set of eigenvectors associated with the matrix which are orthogonal and which can be used to resolve the unit matrix and the original matrix. If none of the eigenvalues of the matrix are zero then this complete set of eigenvectors can be used to resolve the inverse of the matrix whereas, if zero is one of the eigenvalues the inverse of the matrix will not exist. The Hermitian matrix $-iN$, however, has a well-behaved inverse *and* a state which it annihilates.

It was felt that these paradoxical properties were worthy of further investigation. The results of such an investigation are the subject of this paper and can be summarized as follows:

(1) The secular equation, $f(\lambda^2) = 0$, for the eigenvalues $i\lambda$ of the matrix N is found. The function $f(\lambda^2)$ has an infinite set of simple zeros and associated with each such zero there are two normalizable column vectors $A_l(\lambda)$ and $A_l(-\lambda)$. The point $\lambda^2 = 0$ is not in this set. The function $f(\lambda^2)$ has an essential singularity at $\lambda^2 = 0$. Although $f(0)$ is not defined, $f(\lambda^2) \to 0$ as $\lambda^2 \to 0$ along any straight line *except* the positive real axis. As $\lambda^2 \to 0$ the ratios of the components $A_l(\lambda)$ approach the vector o_l and it can be verified directly[1] that $N\hat{o} = 0$.

(2) The set of vectors $A(\pm\lambda)$ is complete and any normalizable vector can be expanded in this set. The unit matrix, the matrix N, and the matrix N^{-1} are resolvable in terms of this set and we can calculate the elements of N^{-1} in terms of such a resolution.

(3) The vector $\hat{o}$ is orthogonal to every member

of the complete set of vectors A. It therefore cannot be expanded in terms of this set. This does not contradict the remarks under 2 since $\hat{o}$ is not normalizable.[4]

(4) The vector $\hat{o}$ satisfying $\mathbf{N}\hat{o} = 0$ is unique.

(5) It is shown that the existence of the unique vector $\hat{o}$ gives rise to a one dimensional infinity of potentials corresponding to any set of phase shifts specified at one energy.

THE SECULAR EQUATION

The equation $NA = +i\lambda A$ when written in detail becomes

$$i\lambda A_{2n}(\lambda) = \sum_{k=0}^{\infty} [(2k+1)^2 - (2n)^2]^{-1} A_{2k+1}(\lambda), \quad (5)$$

and

$$i\lambda A_{2m+1}(\lambda) = \sum_{j=0}^{\infty} [(2j)^2 - (2m+1)^2]^{-1} A_{2j}(\lambda). \quad (6)$$

The A_l with odd l can be eliminated and we obtain[5]

$$\lambda^2 A_{2n}(\lambda) = \sum_{m=0}^{\infty} F_{n,m} A_{2m}(\lambda). \quad (7)$$

The symmetric, real matrix F has components given by

$$F_{n,m} = \sum_{k=0}^{\infty} [(2k+1)^2 - (2n)^2]^{-1}$$
$$\times [(2k+1)^2 - (2m)^2]^{-1}. \quad (8)$$

The summations over k can be evaluated using the formulas in the Appendix.[6] In particular, when

$$n \neq m, \quad F_{n,m} = \tfrac{1}{4}(n^2 - m^2)^{-1}(U_n - U_m),$$

where

$$U_n = \sum_{k=0}^{\infty} [(2k+1)^2 - (2n)^2]^{-1}.$$

and this can be evaluated by substituting $z = n\pi$

[4] As we see, the source of the paradox is the existence of an eigenfunction $\hat{o}$ which has many of the expected properties of an eigenfunction (e.g., it is orthogonal to all other eigenfunctions) but which is not normalizable and which lies outside of a complete set of eigenfunctions. From this point of view the fact that the eigenvalue is zero is not essential, (consider the matrix $N + 1$ for example). However, since the eigenfunctions of a matrix are unchanged by adding a multiple of the unit matrix, the paradox can always be expressed as we have in the text.

[5] In order that the eigenvalue equations (5) and (6) be meaningful, it is necessary to define the infinite sums in a way that forbids unrestricted rearrangements of the terms. When this is done it is possible to prove that the change in the order of summation involved in going from 5 and 6 to 7 is always allowed. A proof of this is contained in an Appendix.

[6] We evaluate such sums by relating them to trigonometric functions. Newton had to evaluate similar sums in his investigation which he summed by purely algebraic methods.

into Eq. (A1) of the Appendix with the result that $u_n = 0$ if $n \neq 0$ and $u_0 = (\pi^2/8)$. When $n = m$ the expressions in the Appendix can be used directly with $z = n\pi$ and the result of the calculation is

$$F_{n,m} = \begin{cases} -(\pi^2/32m^2) & n = 0, \quad m \neq 0 \\ -(\pi^2/32n^2) & n \neq 0, \quad m = 0 \\ (\pi^2/64n^2) & n = m \neq 0 \\ (\pi^4/96) & n = m = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

The matrix F has a very simple form and the associated eigenvalue problem is easily solved. Explicitly we have

$$n \neq 0 : \lambda^2 A_{2n}(\lambda) = (\pi^2/64n^2) A_{2n}(\lambda) - (\pi_s^2/32n^2) A_0(\lambda) \quad (10)$$

and

$$\lambda^2 A_0(\lambda) = (\pi^4/96) A_0(\lambda) - \sum_{n=1}^{\infty} (\pi^2/32n^2) A_{2n}(\lambda). \quad (11)$$

By eliminating A_{2n} $(n \neq 0)$ from these equations we find a necessary condition for the existence of nontrivial solutions is that the secular equation

$$f(\lambda^2) = 0 \quad (12)$$

is satisfied with

$$f(z) = z - (\pi^4/96)$$
$$- (\pi^2/32)^2 \sum_{n=1}^{\infty} n^{-4}[z - (\pi^2/64n^2)]^{-1}, \quad (13a)$$

$$= z - (\pi^2/4) \sum_{n=1}^{\infty} [(2n)^2 - (\pi^2/16z)]^{-1}, \quad (13b)$$

$$= (\pi^2 \sqrt{z}/4) \cot(\pi^2/8\sqrt{z}) - z, \quad (13c)$$

where we have used $\sum_1^{\infty} n^{-2} = (\pi^2/6)$ and Eq. (A3) from the Appendix.

From Eq. (13a) it is readily seen that if z is complex that Im $f(z)$ and Im z have the same sign and that Im $f(z) \neq 0$ if Im $z \neq 0$. The roots of $f(z) = 0$ are therefore all real. From Eq. (13b) it is obvious that if z is negative $f(z)$ is negative so that all the roots λ^2 are positive. Since there are an infinite number of poles of $f(z)$ in the vicinity of $z = 0$, $f(0)$ is not defined. However, direct substitution of $z = 0$ into either Eq. (13a) or Eq. (13b) would give the result $f(0) = 0$. Substituting $\lambda = 0$ into Eq. (10) yields the vector $\hat{o}$ and the equation $N\hat{o} = 0$ is readily verified with the help of (A3).

Our derivation of Eq. (7) is not valid when $\lambda = 0$. The result is still true, however, as can

be seen by multiplying Eq. (6) (with $\lambda = 0$) by $[(2n)^2 - (2m + 1)^2]^{-1}$ and summing over m. When $\lambda = 0$ the A_l with l odd and l even are decoupled so it might be possible that there is another vector annihilated by N containing only components A_l with l odd. We now prove that no such vector exists and that the vector $\hat{o}$ is therefore the unique solution of $N\hat{o} = 0$. Since little additional labor is involved by doing so, we derive an equation satisfied by $A_{2k+1}(\lambda)$ for any eigenvalue λ.

By eliminating the A_{2n} we get an equation of the form

$$\lambda^2 A_{2k+1}(\lambda) = \sum_{l=0}^{\infty} G_{k,l} A_{2l+1}(\lambda), \tag{14}$$

where the symmetric matrix G has components given by

$$G_{k,l} = \sum_{j=0}^{\infty} [(2j)^2 - (2k + 1)^2]^{-1} [(2j)^2 - (2l + 1)^2]^{-1}.$$

As in the even case, the summations can be expressed in terms of trigonometric functions with simple arguments and

$$G_{k,l} = \tfrac{1}{2}(2k + 1)^{-2}(2l + 1)^{-2}, \qquad k \neq l$$
$$= \tfrac{1}{2}(2k + 1)^{-4} + (\pi^4/16)(2k + 1)^{-2},$$
$$k = l. \tag{15}$$

The matrix G is the direct sum of a diagonal matrix and a separable matrix, so that the eigenfunctions are readily found. Explicitly,

$$\lambda^2 A_{2k+1}(\lambda) = (\pi^2/16)(2k + 1)^{-2} A_{2k+1}(\lambda)$$
$$+ (2k + 1)^{-2} K(\lambda), \tag{16}$$

with

$$K(\lambda) = \tfrac{1}{2} \sum_{l=0}^{\infty} (2l + 1)^{-2} A_{2l+1}(\lambda). \tag{17}$$

Comparing these equations we find that the secular equation

$$g(\lambda^2) = 0 \tag{18}$$

must be satisfied with

$$g(z) = 1 - \tfrac{1}{2} \sum_{l=0}^{\infty} (2l + 1)^{-4}$$
$$\times \{z - [(\pi^2/16)(2l + 1)^{-2}]\}^{-1} \tag{19a}$$
$$= 2 - [8(z)^{\frac{1}{2}}/\pi^2] \tan [\pi^2/8(z)^{\frac{1}{2}}]. \tag{19b}$$

Since

$$g(z) = [8/\pi^2(z)^{\frac{1}{2}}] \tan [\pi^2/8(z)^{\frac{1}{2}}] f(z),$$

the roots of $g(\lambda^2) = 0$ are the same as those of $f(\lambda^2) = 0$ except that $g(z) \to 2$ as z approaches zero along a path which avoids the real axis. If

we substitute $\lambda = 0$ into Eq. (16), this implies that all the $A_{2l+1}(0)$ are equal. Direct substitution [and Eq. (A1)] shows that this is *not* annihilated by N. Although this is more than enough to prove the uniqueness of the vector $\hat{o}$ previously found, we also point out that the form G is positive definite with

$$\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} B_k^* G_{k,l} B_l$$
$$= \frac{1}{2} \left| \sum_{k=0}^{\infty} \frac{B_k}{2k + 1} \right|^2 + \frac{\pi^2}{16} \sum_{k=0}^{\infty} \left| \frac{B_k}{2k + 1} \right|^2. \tag{20}$$

ORTHONORMALITY OF THE A (λ)

It is a straightforward matter to verify that the A's form an orthogonal set. It is convenient to consider the components A_l with l even and with l odd separately.

Thus,

$$\sum_{n=0}^{\infty} A_{2n}(\lambda) A_{2n}(\lambda') = A_0(\lambda) A_0(\lambda') \Big\{ 1 + (\pi^2/32)^2$$
$$\times \sum_{k=1}^{\infty} k^{-4} [\lambda^2 - (\pi^2/64k^2)]^{-1} [\lambda'^2 - (\pi^2/64k^2)]^{-1} \Big\}$$

and if $\lambda^2 \neq \lambda'^2$,

$$= A_0(\lambda) A_0(\lambda') [f(\lambda^2) - f(\lambda'^2)]/(\lambda^2 - \lambda'^2),$$
$$= 0.$$

When $\lambda^2 = \lambda'^2$ we get

$$\sum_{n=0}^{\infty} A_{2n}^2(\lambda) = A_0^2(\lambda) f'(\lambda^2),$$

where $f'(\lambda^2) = [(d/dz)f(z)]$, $(z = \lambda^2)$. We choose

$$A_0(\lambda) = [2f'(\lambda^2)]^{-\frac{1}{2}}. \tag{21}$$

For the $A_l(\lambda)$ with odd l, one finds that Eqs. (6) and (8) imply that

$$\sum_{k=0}^{\infty} A_{2k+1}(\lambda) A_{2k+1}^*(\lambda')$$
$$= (\lambda\lambda')^{-1} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} A_{2n}(\lambda) F_{n,m} A_{2m}(\lambda).$$

We may now summarize the orthogonality properties:

$$\sum_{n=0}^{\infty} A_{2n}(\lambda) A_{2n}(\lambda') = \tfrac{1}{2}, \qquad \lambda^2 = \lambda'^2,$$
$$= 0, \qquad \lambda^2 \neq \lambda'^2,$$
$$\sum_{=0}^{\infty} A_{2n+1}(\lambda) A_{2n+1}^*(\lambda) = \tfrac{1}{2}, \qquad \lambda = \lambda', \tag{22}$$
$$= -\tfrac{1}{2}, \qquad \lambda = -\lambda',$$
$$= 0, \qquad \lambda^2 \neq \lambda'^2.$$

A very simple expression for $A_{2n}(\lambda)$ is obtained by using Eq. (10). It is then possible to use Eq. (6) to obtain an expression for $A_{2n+1}(\lambda)$. This will involve a fairly complicated sum. On the other hand, Eq. (16) suggests that $A_{2k+1}(\lambda)$ depends on k in a very simple way. We now demonstrate that these results are consistent.

From Eqs. (6), (10), and (21), we obtain

$$[2f'(\lambda^2)]^{\frac{1}{2}}A_{2n+1}(\lambda) = (i/\lambda)\left\{(2n+1)^{-2} + \sum_{k=1}^{\infty} (\pi^2/32k^2)\right.$$

$$\left. \times [(2k)^2 - (2n+1)^2]^{-1}[\lambda^2 - (\pi^2/64k^2)]^{-1}\right\},$$

$$= 2i\lambda \sum_{k=1}^{\infty} [(2k)^2 - (2n+1)^2]^{-1}[\lambda^2 - (\pi^2/64k^2)]^{-1},$$

$$= (i/2\lambda)[(\pi^2/4\lambda)\cot(\pi^2/8\lambda)][(2k+1)^2 - (\pi^2/16\lambda)]^{-1}. \quad (23)$$

In going from the first line to the second, we have used

$$\sum_{j=1}^{\infty} [(2j)^2 - (2k+1)^2]^{-1} = \tfrac{1}{2}(2k+1)^{-2},$$

which can be obtained from Eq. (A3). The last line is a consequence of the identity given by Eq. (A5). Finally, the first quantity in square brackets is unity since λ must be a root of the secular equation. The result is seen to be consistent with Eq. (16).

In summary we now have the simple equations for the A_l

$$[2f'(\lambda^2)]^{\frac{1}{2}}A_0(\lambda) = 1,$$

$$[2f'(\lambda^2)]^{\frac{1}{2}}A_{2n}(\lambda) = -(\pi^2/32n^2)[\lambda^2 - (\pi^2/64n^2)]^{-1},$$
$$n \neq 0 \quad (24)$$

and

$$[2f'(\lambda^2)]^{\frac{1}{2}}A_{2n+1}(\lambda) = (i\lambda/2)(2n+1)^{-2}$$

$$\times [\lambda^2 - (\pi^2/16)(2n+1)^{-2}]^{-1}. \quad (25)$$

COMPLETENESS

In order to verify completeness it is necessary to consider sums over the eigenvalues λ. These sums may be evaluated by the method of contour integration. The necessary summation formula are listed in the Appendix. Because $A_{2n}(\lambda) = A_{2n}(-\lambda)$ and $A_{2n+1}(\lambda) = -A_{2n+1}(-\lambda)$ the sums can be changed into sums over λ^2 and several terms are zero by parity arguments.

The completeness relationship we wish to prove is

$$\delta_{l,l'} = \sum_{\lambda} A_l(\lambda)A_{l'}^*(\lambda). \quad (26)$$

Now

$$\sum_{\lambda} A_l(\lambda)A_{l'}^*(\lambda)$$

$$= \begin{cases} 0, & l'-l \quad \text{odd} \\ C_0, & l = l' = 0 \\ -(\pi^2/32n^2)C_1(2n), & \begin{cases} l = 0 & l' = 2n \\ l = 2n & l' = 0 \end{cases} \\ \tfrac{1}{4}C_6(2j+1, 2k+1), & l = 2j+1, \quad l' = 2k+1 \\ (\pi^2/32n^2)(\pi^2/32m^2)C_2(2n, 2m), & \begin{array}{l} l = 2n \neq 0 \\ l' = 2m \neq 0, \end{array} \end{cases} \quad (27)$$

and by referring to the results in the Appendix it can be verified that these are equivalent to Eq. (26). Hence, the set is complete.

The set of vectors resolve N if

$$N_{l,l'} = i \sum_{\lambda} \lambda A_l(\lambda)A_{l'}^*(\lambda). \quad (28)$$

We find that

$$N_{l,l'} = i \begin{cases} 0, & l'-l \text{ even} \\ -(\tfrac{1}{2}i)C_5(2k+1), & l = 0, \quad l' = 2k+1 \\ +(\tfrac{1}{2}i)C_5(2k+1), & l = 2k+1, \quad l' = 0 \\ (\tfrac{1}{2}i)(\pi^2/32n^2)C_8(2k+1, 2n), \\ \qquad l = 2n, \quad l' = 2k+1 \\ -(\tfrac{1}{2}i)(\pi^2/32n^2)C_8(2k+1, 2n), \\ \qquad l = 2k+1, \quad l' = 2n, \end{cases} \quad (29)$$

and this is equivalent to the definition of $N_{ll'}$ given by Eq. (1).

Finally, it is expected that

$$N_{l,l'}^{-1} = -i \sum_{\lambda} \lambda^{-1} A_l(\lambda)A_{l'}^*(\lambda), \quad (30)$$

and by inserting the expressions for A_l in the right-hand side, we obtain

$$N_{l,l'}^{-1} = \begin{cases} 0, & l'-l \text{ even} \\ -(\tfrac{1}{2}i)C_4(2k+1), & l = 0, \quad l' = 2k+1 \\ +(\tfrac{1}{2}i)C_4(2k+1), & l = 2k+1, \quad l' = 0 \\ +(\tfrac{1}{2}i)(\pi^2/32n^2)C_7(2j+1, 2n), \\ \qquad l = 2n \neq 0, \quad l' = 2j+1 \\ -(\tfrac{1}{2}i)(\pi^2/32n^2)C_7(2j+1, 2n), \\ \qquad l = 2j+1, \quad l' = 2n \neq 0, \end{cases} \quad (31)$$

and this agrees with the known result.

The completeness of the set $A_i(\lambda)$ implies that a vector b_l can be expanded in terms of the $A_i(\lambda)$ it it is possible to change the order of summation in

$$b_l = \sum_\lambda A_i(\lambda) \sum_{l'} A_{l'}^*(\lambda) b_{l'}. \qquad (32)$$

It is easily seen that $\hat{o}$ does not belong to this class. In fact, $\hat{o}$ is orthogonal to the A's since

$$\sum_l \hat{o}_l A_i(\lambda) = \tfrac{1}{2} \lambda^{-2} f(\lambda^2) = 0. \qquad (33)$$

It is interesting to note that by adding $\hat{o}$ to the basis, the domain of vectors which can be expanded is enlarged.

DISCUSSION

The uniqueness of the solution of $N\hat{o} = 0$ permits a determination of the dimensionality of the family of potentials consistent with a given set of phase shifts at one energy. The general solution of the matrix equation $Nx = y$ is

$$x = N^{-1}y + \alpha\hat{o}. \qquad (34)$$

One of the steps in Newton's argument involved the inversion of such a matrix equation. The solution is then used to determine a set of members C_l which define

$$f(r, r') = \sum_{l=0}^{\infty} C_l r^2 j_l(r) j_l(r'). \qquad \begin{array}{c}(35)\\(N1)\end{array}$$

The potential $V(r)$ is then obtained by finding the function $K(r, r')$ which is uniquely determined once $f(r, r')$ is known by solving the integral equation

$$K(r,r') = f(r,r') - \int_0^r dr'' r''^{-2} K(r, r'') f(r'', r'), \qquad \begin{array}{c}(36)\\(N5)\end{array}$$

and then

$$V(r) = -2r^{-1}(d/dr)[r^{-1}K(r,r)]. \qquad \begin{array}{c}(37)\\(N7)\end{array}$$

If one takes note of the form we have obtained for the solution x of $Nx = y$ and traces Newton's argument back, it is found that C_l has the form

$$C_l = (X_l + \alpha Y_l)/(Z_l + \alpha W_l). \qquad (38)$$

There is therefore a one-dimensional infinity of potentials $V(\alpha, r)$ and it is evident that the potentials depend on the parameter α in a rather complicated way.

ACKNOWLEDGMENTS

APPENDIX A

In the course of the analysis it was necessary to evaluate several infinite sums. We present below a list of formulas sufficient to perform all the sums encountered.

$$\sum_{j=0}^{\infty} [(2j+1)^2 - (4z^2/\pi^2)]^{-1} = (\pi^2/8z)\tan z, \qquad (A1)$$

$$\sum_{j=0}^{\infty} [(2j+1)^2 - (4z^2/\pi^2)]^{-2}$$
$$= (\pi^2/8z^2)^2(z^2 \sec^2 z - z \tan z), \qquad (A2)$$

$$\sum_{j=1}^{\infty} [(2j)^2 - (4z^2/\pi^2)]^{-1} = (\pi^2/8z^2)(1 - z \cot z), \qquad (A3)$$

$$\sum_{j=1}^{\infty} [(2j)^2 - (4z^2/\pi^2)]^{-2}$$
$$= (\pi^2/8z^2)^2(z \cot z + z^2 \csc^2 z - 2), \qquad (A5)$$

$$\sum_{j=1}^{\infty} (2j)^2[(2j)^2 - (2k+1)^2]^{-1}[(2j)^2 - (4z^2/\pi^2)]^{-1}$$
$$= \tfrac{1}{2}z \cot z[(2k+1)^2 - (4z^2/\pi^2)]^{-1}. \qquad (A5)$$

The proof of the above results is straightforward. For example, the last result is derived by considering the contour integral

$$\oint dw\{w^2 \cot w[4w^2 - (2k+1)^2\pi^2]^{-1}[w^2 - z^2]^{-1}\}.$$

It is easy to show that $\cot w$ is bounded on a rectangular contour with vertices at the points $\pm(2N+1)(\pi/2) \pm iV$. By letting $N \to \infty$ through integer values and $V \to \infty$ it is easy to see that the integral vanishes. An evaluation of the integral by the method of residues immediately gives Eq. (A5).

It is also necessary to perform sums over the eigenvalues of N in order to demonstrate completeness. In the following, the summations range over all values of λ^2.

$$C_0 = \sum [1/f'(\lambda^2)] = 1, \qquad (A6)$$

$$C_1(2n) = \sum [f'(\lambda^2)]^{-1}[\lambda^2 - (\pi^2/64n^2)]^{-1} = 0, \qquad (A7)$$

$$C_2(2n, 2m)$$
$$= \sum [f'(\lambda^2)]^{-1}[\lambda^2 - (\pi^2/64n^2)]^{-1}[\lambda^2 - (\pi^2/64m^2)]^{-1}$$
$$= 0 \qquad n \neq m$$
$$= (32n^2/\pi^2)^2, \qquad n = m; \qquad (A8)$$

$$C_3(2n, 2m) = \sum \lambda^2[f'(\lambda^2)]^{-1}[\lambda^2 - (\pi^2/64n^2)]^{-1}$$
$$\times [\lambda^2 - (\pi^2/64m^2)]^{-1},$$
$$= 0, \qquad n \neq m$$
$$= 16n^2/\pi^2, \qquad n = m \qquad (A9)$$

$$C_4(2k + 1) = \sum [f'(\lambda^2)]^{-1}[(2k + 1)^2\lambda^2 - (\pi^2/16)]^{-1}$$
$$= (16/\pi^2), \qquad (A10)$$

$$C_5(2k + 1) = \sum \lambda^2 [f'(\lambda^2)]^{-1}[(2k + 1)^2\lambda^2$$
$$- (\pi^2/16)]^{-1} = 2(2k + 1)^{-2}, \qquad (A11)$$

$$C_6(2j + 1, 2k + 1) = \sum \lambda^2[f'(\lambda^2)]^{-1}[(2j + 1)^2\lambda^2$$
$$- (\pi^2/16)]^{-1}[(2k + 1)^2\lambda^2 - (\pi^2/16)]^{-1}$$
$$= 0, \qquad j \neq k$$
$$= 4, \qquad j = k, \qquad (A12)$$

$$C_7(2j + 1, 2n) = \sum [f'(\lambda^2)]^{-1}[(2j + 1)^2\lambda^2$$
$$- (\pi^2/16)]^{-1}[\lambda^2 - (\pi^2/64n^2)]^{-1}$$
$$= (2j + 1)^2(2n)^2[(2n)^2 - (2j + 1)^2]^{-1}(16/\pi^2)^2, \qquad (A13)$$

$$C_8(2j + 1, 2n) = \sum \lambda^2[f'(\lambda^2)]^{-1}[(2j + 1)^2\lambda^2$$
$$- (\pi^2/16)]^{-1}[\lambda^2 - (\pi^2/64n^2)]^{-1}$$
$$= (2n)^2(16/\pi^2)[(2n)^2 - (2j + 1)^2]^{-1}. \qquad (A14)$$

The above results are easily proven. The properties of $f(z)$ needed are the following:

(1) As $z \to \infty$ $f(z)/z \to 1$.

(2) It is possible to find a sequence of contours C_N, such that $|z| \to 0$ on the contours as N goes to infinity and $\cot [\pi^2/8(z)^{\frac{1}{2}}]$ remains bounded on each of the contours. An example of such a contour C_N is one made up of two segments of cardoids defined as follows. On the first segment

$$\mathrm{Re} \, [\pi^2/8(z)^{\frac{1}{2}}] = (N + \tfrac{1}{2})\pi$$
$$|\mathrm{Im} \, [\pi^2/8(z)^{\frac{1}{2}}]| \leq (N + \tfrac{1}{2})\pi,$$

and on the other segment

$$|\mathrm{Im} \, [\pi^2/8(z)^{\frac{1}{2}}]| = (N + \tfrac{1}{2})\pi$$

and

$$\mathrm{Re} \, [\pi^2/8(z)^{\frac{1}{2}}] \leq (N + \tfrac{1}{2})\pi.$$

On such a contour $f(z)/(z)^{\frac{1}{2}}$ remains bounded.

We can now consider an integral over a contour consisting of a large circle of radius R_N such that $\lim_{N\to\infty} R_N = \infty$ which is traversed counterclockwise and the contour C_N traversed in a clockwise sense. By integrating appropriately chosen functions over this contour and then evaluating the integrals by the method of residues, the above results are obtained.

For example: consideration of

$$\int_C dz[f(z)]^{-1}[z - (\pi^2/16)(2j + 1)^{-2}]^{-1}$$
$$\times [z - (\pi^2/64n^2)]^{-1} = 0$$

gives the expression for $C_7(2j + 1, 2n)$. The integrand has poles at the roots of $f(z) = 0$ and at $z = (\pi^2/16)(2j + 1)^{-2}$ but not at $z = \pi^2/64n^2$.

APPENDIX B

The eigenvalue equation

$$\sum_{l'=0}^{\infty} N_{l,l'} A_{l'}(\lambda) = i\lambda A_l(\lambda), \qquad (B1)$$

is not meaningful if an unrestricted rearrangement of the terms in the infinite sum is allowed. For one could choose *any* vector $A_l(\lambda)$ such that the sum converged but did not converge absolutely. Since such a series can be assigned any value by re-ordering the terms[7] one could choose *any* λ and make Eq. (B1) valid by taking a suitable rearrangement of the terms for each value of l. In order to make the eigenvalue meaningful we adopt the following definition of the infinite sum

$$\sum_{l=0}^{\infty} a_l \equiv \lim_{L\to\infty} \sum_{l=0}^{L} a_l. \qquad (B2)$$

When such a definition is adopted all the manipulations in the paper can be rigorously justified. In particular, with this definition of the infinite sum the solution of the equation

$$\sum_{l'=0}^{\infty} N_{l,l'} b_{l'} = 0 \qquad (B3)$$

is unique and $b_l = \delta_l$.

In order to justify the formal manipulations performed in the paper it is necessary to show that the change in the order of summation in going from Eqs. (5) and (6) to Eqs. (7) and (14) is allowed. We show how to rigorously obtain Eq. (7), the proof for Eq. (14) is similar. In order to do this it suffices to prove the following theorem:

Theorem.

Hypothesis: The sum

$$s_k = \sum_{j=0}^{\infty} [(2k + 1)^2 - (2j)^2]^{-1} a_j \qquad (B4)$$

converges.

Conclusions:

---

[7] T. J. I'a Bromwich, *Theory of Infinite Series* (The MacMillan Company, New York, 1949), 2nd ed., p. 74.

Let us define $T_{\mu\nu}(n)$ by

$$T_{\mu,\nu}(n) = \sum_{j=0}^{\mu} \sum_{k=0}^{\nu} [(2k+1)^2 - (2n)^2]^{-1}$$
$$\times [(2k+1)^2 - (2j)^2]^{-1} a_i. \quad (B5)$$

Then

(i) $\lim_{\mu\to\infty} \lim_{\nu\to\infty} T_{\mu\nu}(n)$ exists; call it $T(n)$;

(ii) $\lim_{(\mu,\nu\to\infty)} T_{\mu\nu}(n)$ exists and is equal to $T(n)$;

(iii) $\lim_{\nu\to\infty} \lim_{\mu\to\infty} T_{\mu\nu}(n)$ exists and is also equal to $T(n)$.

We prove parts (i) and (ii) of the theorem. Part (iii) of the theorem follows immediately from part (ii) according to Pringsheim's theorem.[8]

Notice that no restriction is placed on the a_i[9] other than the convergence of the sum (B4). We need Abel's test which provides a powerful tool for discussing the convergence of series.

Abel's Test. A convergent sequence $\sum a_n$ (which need not converge absolutely) remains convergent if its terms are multiplied by a factor u_n, provided that the sequence u_n is monotonic and that $|u_n|$ is less than a constant k.

This result is proven in all standard texts on infinite series[10] so we omit the proof.

We also need a lemma which is a slight generalization of Abel's Lemma.[11]

Lemma. Given a sequence of positive numbers v_i which is initially monotonically increasing and finally monotonically decreasing so that it reaches a maximum value when $j = J$, and given a sequence a_i such that

$$H \geq \sum_{i=1}^{n} a_i \geq h \quad \text{for} \quad n \leq p,$$

then

$$Hv_J - h(v_J - v_1) \geq \sum_{i=1}^{p} v_i a_i \geq -H(v_J - v_1) + hv_J.$$

Proof. The proof is straightforward. We consider $p > J$. The modifications necessary when $p \leq J$ are trivial.

---

[8] Pringsheims theorem states that if the double limit $\lim (\mu, \nu \to \infty) S_{\mu\nu}$ exists, and if the single limit $\lim (\nu \to \infty) S_{\mu\nu}$ exists, then the iterated limit $\lim (\mu \to \infty) \lim (\nu \to \infty) S_{\mu\nu}$ also exists and it has the same value as the double limit. Reference 7, p. 58.

[9] In the following, we take all quantities to be real. The following theorem can be proven with a_i complex simply by considering real and imaginary parts.

[10] Ref. 7, p. 58.

[11] Ref. 7, p. 57.

Let

$$s_1 = a_1, \quad s_2 = a_1 + a_2, \text{ etc.}$$

Then

$$\sum a_i v_i = s_1 v_1 + (s_2 - s_1)v_2 + \cdots + (s_p - s_{p-1})v_p,$$
$$= [s_1(v_1 - v_2) + \cdots + s_{J-1}(v_{J-1} - v_J)]$$
$$+ [s_J(v_J - v_{J+1}) + \cdots + s_p v_p],$$
$$= A + B.$$

In the first square bracket, the coefficients of s_k are all nonpositive so that a lower (upper) limit on A is obtained by replacing the s_k by $H(h)$. Thus

$$-h(v_J - v_1) \geq A \geq -H(v_J - v_1).$$

Similarly,

$$Hv_J \geq B \geq hv_J$$

and the Lemma follows.

We now return to the proof of the theorem. We proceed as follows. The sum over k in Eq. (B5) can be done explicitly and defines a function which has the properties of the function v_j of the lemma. Abel's test can be used to prove the convergence of the remaining sum over j and the lemma is used to place limits on the difference between the two methods of summation to show that the difference vanishes.

We now proceed to prove the theorem. Now

$$\lim_{\mu\to\infty} \lim_{\nu\to\infty} T_{\mu,\nu}(n) = \sum_{m=0}^{\infty} F_{n,m} a_m, \quad (B6)$$

where $F_{n,m}$ is given by Eq. (8) of the main text. The sum over m contains only two terms if $n \neq 0$ and therefore converges. When $n = 0$

$$\sum_{j=0}^{\infty} F_{0,j} a_i = (\pi^4/96)a_0 - \sum_{j=1}^{\infty} (\pi^2/32j^2)a_i. \quad (B7)$$

Comparing this with Eq. (B4) we see that the terms in the sum for (B7) can be obtained by multiplying the term of (B4) by the factor

$$(\pi^2/32j^2)[(2k+1)^2 - (2j)^2].$$

Since this factor defines a monotonic and bounded sequence, Abel's test implies that the series (B7) converges. The first part of the theorem is proven.

To prove the second part of the theorem we first note that

$$T_{\mu,\nu}(n) - \sum_{j=0}^{\mu} F_{n,j} a_i = a_n W_n(\nu)$$
$$+ \sum_{j=0}^{\mu}{}' a_i [(2j)^2 - (2n)^2]^{-1} [u_n(\nu) - u_i(\nu)], \quad (B8)$$

where

$$u_i(\nu) = \sum_{k=\nu+1}^{\infty} [(2k+1)^2 - (2j)^2]^{-1}, \qquad (B9)$$

$$= \sum_{k=0}^{\nu} [(2j)^2 - (2k+1)^2]^{-1} \text{ if } j \neq 0, \quad (B10)$$

$$= (4j)^{-1}[(2j - 2\nu - 1)^{-1} + (2j - 2\nu + 1)^{-1}$$
$$+ \cdots + (2j + 2\nu + 1)^{-1}] \qquad (B11)$$

if $j \neq 0$, and

$$W_n(\nu) = \sum_{k=\nu+1}^{\infty} (2k+1)^{-2} \sim (4\nu)^{-1}. \qquad (B12)$$

We obtain (B10) by noting that the right-hand side of (B9) is equal to u_i of the main text when the lower limit of the sum is zero, and that $u_i = 0$ if $j \neq 0$.

The prime on the summation sign in (B8) indicates that the term with $n = j$ is to be omitted. It is now necessary to show that the right-hand side of (B8) goes to zero as μ and ν go to infinity independently. We have indicated the asymptotic behavior of $W_n(\nu)$.[12] The contribution $a_n W_n(\nu)$ is independent of μ and goes to zero as $\nu \to \infty$. We next show that the contribution

$$u_n(\nu) \sum_{j=0}^{\mu}{}' a_j[(2j)^2 - (2n)^2]^{-1} \to 0. \qquad (B13)$$

For large ν, and $n \neq 0$, $u_n(\nu) \sim (\tfrac{1}{4}\nu)$ since all but $2n$ of the terms in (B11) cancel and each of these terms behaves like $(4n)^{-1}(2\nu)^{-1}$. It is easily seen from (B9) that $u_0(\nu) \sim \tfrac{1}{2}(2\nu + 3)^{-1}$. Also the sum over j in (B13) converges since the general term is related to the comparison series (B4) by a factor $[(2k+1)^2 - (2j)^2]/[(2j)^2 - (2n)^2]$ which is bounded ($j \neq n$) and is ultimately monotonic so that Abel's test applies. Since the infinite sum converges the

---

[12] We obtain estimates of the asymptotic behavior of the sums here, and in the following, by replacing the sum by an integral. Such a procedure is easily justified in the cases considered.

partial sums have finite upper and lower bounds independent of μ and the product in (B13) goes to zero.

To complete the proof it is necessary to show that

$$\sum_{j=0}^{\mu}{}' a_j[(2j)^2 - (2n)^2]^{-1} u_i(\nu) \to 0. \qquad (B14)$$

To do this we show that $u_i(\nu)$ is a function of the type considered in the lemma. The lemma provides upper and lower bounds on the sum and it is found that these go to zero in the limit.

For $j \leq \nu + 1$, Eq. (B9) represents $u_i(\nu)$ as a sum of positive terms. When $j \leq \nu$, it is easily seen that increasing j increases each of these terms so that $u_i(\nu)$ is initially an increasing function of j.

For $j \geq \nu + 1$, Eq. (B10) represents $u_i(\nu)$ as a sum of positive terms. It is easily seen that increasing j decreases each of these terms so that $u_i(\nu)$ is ultimately decreasing. The maximum value of $u_i(\nu)$ occurs when $j = \nu + 1$ and

$$u_{\nu+1}(\nu) = \frac{1}{4(\nu+1)} [1^{-1} + 3^{-1} + \cdots (4\nu + 3)^{-1}]$$

$$\sim \frac{1}{8\nu} \ln \nu, \qquad (B15)$$

where we have used (B11).

We have already noted that

$$H \geq \sum_{j=0}^{\mu}{}' a_j[(2j)^2 - (2n)^2]^{-1} \geq h,$$

with H and h independent of μ. The lemma then gives

$$H u_{\nu+1}(\nu) - h(u_{\nu+1}(\nu) - u_0(\nu))$$

$$\geq \sum_{j=0}^{\mu}{}' a_j[(2j)^2 - (2n)^2]^{-1} u_i(\nu)$$

$$\geq h u_{\nu+1}(\nu) - H(u_{\nu+1}(\nu) - u_0(\nu))$$

and the upper and lower bounds both go to zero as $\mu, \nu \to \infty$.

The Effective Resistance of Passive Networks*

GEORGE L. MONTET

Argonne National Laboratory, Argonne, Illinois

(Received 12 June 1964)

A recently found solution to the problem of a random walk over a lattice with reflecting boundaries is used to evaluate the resistances of several passive electrical networks. The solutions are particularly simple for rectangular planar networks and various types are considered. The extension of the solution to the corresponding three-dimensional array, the simple cubic lattice, is given.

INTRODUCTION

A RECENTLY found solution[1] to the problem of a random walk executed on a lattice with reflecting boundaries makes possible the calculation of the effective resistances of several interesting passive networks. It is the purpose of this article to exhibit the method of solution.

The method is applicable to a large class of networks, but here attention will be concentrated on the physically meaningful example of a finite planar network made up of squares bounded by resistors; the two cases where the horizontal and vertical resistors have equal and unequal values will be considered.

A brief discussion of the extension of the method to three dimensional networks is presented. In particular, the three-dimensional analogy to the square planar network, that is, the simple cubic lattice, is discussed in some detail.

1. RANDOM WALK WITH INTERNAL TRAP

Consider the movement of a walker over an array of n rows of m lattice points surrounded by a border of points from which there is no return; let the border consist of the points $(p, 0)$, $(p, n + 1)$, $(0, q)$, and $(m + 1, q)$. If $F(p, q)$ is the expectation that the walker will visit the point (p, q) after release from the source point (a, b) before capture at a boundary point, it is clear that the partial difference equation

$$F(p, q) = \delta_{pa}\delta_{qb} + \tfrac{1}{4}[F(p - 1, q) + F(p + 1, q)$$
$$+ F(p, q - 1) + F(p, q + 1)] \quad (1.1)$$

is satisfied. Since the expectation that the walker will leave the boundary is, by hypothesis, zero, the accessory conditions

$$F(0, q) = F(m + 1, q) = 0 \ (0 \leq q \leq n + 1)$$
$$F(p, 0) = F(p, n + 1) = 0 \ (0 \leq p \leq m + 1), \quad (1.2)$$

must be satisfied. Solutions to Eqs. (1.1) and (1.2) have been given by McCrea and Whipple.[2]

Now let it be supposed that there is an internal trap at the point (k, l) in addition to the traps at the boundaries; Eqs. (1.1) and (1.2) apply with the additional condition

$$F(k, l) = \delta_{ka}\delta_{lb}. \quad (1.3)$$

It may be shown[3] that the probability of capture of the walker at (k, l) is given by

$$P(k, l; a, b) = \frac{F(k, l; a, b) - \delta_{ka}\delta_{lb}}{F(k, l; k, l)}, \quad (1.4)$$

where it is now necessary to indicate explicitly the position of the source. The expectation of visit to (p, q) in the presence of a trap at (k, l) is given by

$$F(p, q; k, l; a, b) = F(p, q; a, b)$$
$$- P(k, l; a, b)F(p, q; k, l). \quad (1.5)$$

As relations (1.4) and (1.5) show, these solutions are expressible in terms of previously known solutions.[2,4] In situations where the solutions $F(p, q; a, b)$ tend to infinity, as when the boundaries are reflecting or the lattice is of infinite extent, it is necessary to analyze the manner in which $P(k, l; a, b)$ approaches unity and Eq. (1.5) must be modified to read

$$F(p, q; k, l; a, b) = F(p, q; a, b) - F(k, l; a, b)$$
$$+ F(k, l; k, l) - F(p, q; k, l). \quad (1.6)$$

2. AN ELECTRICAL NETWORK ANALOG

The solutions discussed in Sec. 1 are closely related to the effective resistance of a passive network, as will be demonstrated forthwith. Let an electrical network be constructed by connecting

[1] The pertinent solutions are given in the Appendix.

[2] W. H. McCrea and F. J. W. Whipple, Proc. Roy. Soc. (Edinburgh) 60, 281 (1940).
[3] E. W. Montroll, in *Proceedings of the Sixteenth Symposia in Applied Mathematics of the American Mathematical Society* (to be published).
[4] E. M. Keberle and G. L. Montet, J. Math. Anal. and Appl. 6, 1 (1963).

all the lattice points through resistances R and let current I be introduced at the source (a, b) and removed at the sink (k, l). [Note that this requirement limits the discussion to cases of an infinite network or a finite network with all boundaries reflecting; otherwise, not all the current is removed at the sink. It is evident, therefore, that (1.6) is the pertinent relation.] The passage of the current results in a potential distribution in the network. If $V(p, q)$ denotes the voltage at point (p, q), the current flowing from (p, q) to $(p + 1, q)$ is given by $R^{-1}[V(p, q) - V(p + 1, q)]$; since the current flowing from any point, other than (a, b), must be zero, the difference equation

$$R^{-1}[V(p, q) - V(p + 1, q) + V(P, q)$$
$$- V(p - 1, q) + V(p, q) - V(p, q + 1)$$
$$+ V(p, q) - V(p, q - 1)] \cdots = I\delta_{pa}\delta_{qb} \quad (2.1)$$

results. Comparison of (2.1) with (1.1) shows that

$$F(p, q; k, l; a, b) = 4(RI)^{-1}V(p, q). \quad (2.2)$$

Now $V(a, b) - V(k, l)$ is the potential drop from (a, b) to (k, l) and, since the current I flows through an effective resistance R_{eff} in going from (a, b) to (k, l), it is evident that

$$V(a, b) - V(k, l) = IR_{eff} = \tfrac{1}{4}RI[F(a, b; k, l; a, b)$$
$$- F(k, l; k, l; ab)];$$

or

$$R_{eff}/R = \tfrac{1}{4}F(a, b; k, l; a, b). \quad (2.3)$$

In obtaining (2.3) use has been made of the fact that, according to (1.6), $F(k, l; k, l; a, b) = 0$.

3. THE INFINITE SQUARE NETWORK

The simpler case of (2.3) occurs in considering the passage of current from one point to another in an infinite network; Eqs. (2.3) and (1.6) then yield

$$R_{eff}/R = \tfrac{1}{2}[F(a, b; a, b) - F(k, l; a, b)]$$
$$\equiv \tfrac{1}{2}G(|k - a|, |l - b|), \quad (3.1)$$

a result first found by Van der Pol.[5] Similar results have been given by Davies,[6] who also discusses the corresponding formula for a simple cubic grid. Values of the difference function, G, are known for the square net[2] and for the triequiangular net.[4] For both these arrays the resistance diverges logarithmically as the distance between the source and the sink tends to infinity.

[5] B. Van Der Pol in *Probability and Related Topics in Physical Sciences*, edited by M. Kac (Interscience Publishers, Inc., New York, 1959).
[6] H. Davies, Quart. J. Appl. Math. **6**, 232 (1955).

4. THE FINITE SQUARE NETWORK

The more interesting case, and the one which corresponds more closely to reality, is that of a finite electrical network; this is analogous to an array with four rectilinear reflecting boundaries. In applying (1.6) to this situation, care must be exercised in regards to the relative magnitudes of l and b; choosing $b \leq l$, it is found that

$$R_{eff}/R = \tfrac{1}{4}[F(a, b; a, b) - F_2(k, l; a, b)$$
$$+ F(k, l; k, l) - F_1(a, b; k, l)]. \quad (4.1)$$

The functions $F(k, l; a, b)$ are given in the Appendix; $F_1(k, l; a, b)$ is the solution when $l \leq b$, $F_2(k, l; a, b)$ is the solution when $l \geq b$. Eq. (4.1) holds for all values of the variables: $1 \leq a \leq m$, $1 \leq k \leq m$, $1 \leq b \leq l$, $1 \leq l \leq n$; however, the discussion will be limited to the important case where the current enters and leaves at diagonally opposite corners of the rectangle and the values $a = b = 1$, $k = m$, $l = n$ will be chosen. For this configuration it is evident that the resistance may be written as

$$\frac{R_{eff}}{R} = \frac{n - 1}{m} + \frac{1}{m} \sum_{\nu=1}^{m-1} \frac{1 + \cos(\nu\pi/m)}{1 - \cos(\nu\pi/m)}$$
$$\times \left[1 - \frac{\sinh(n - 1)\beta_\nu + (-1)^\nu \sinh\beta_\nu}{\sinh n\beta_\nu}\right], \quad (4.2)$$

where

$$\cosh\beta_\nu = 2 - \cos(\nu\pi/m), \quad \nu = 1, 2, \cdots, m - 1. \quad (4.3)$$

For small values of m and n the resistances may be evaluated explicitly; values so found are given as fractions in Table I. For somewhat larger values tables of trignometric functions may be used; values found in this manner are listed in decimal form in Table I. It is obvious, however, that an asymptotic formula is desirable. For large n the quantity in brackets in (4.2) becomes $1 - \cosh\beta_\nu + \sinh\beta_\nu$, so that

$$\frac{R_{eff}}{R} \xrightarrow[n\,\text{large}]{} \frac{n - 1}{m} - \frac{1}{m} \sum_{\nu=1}^{m-1}\left(1 + \cos\frac{\nu\pi}{m}\right)$$
$$+ \frac{1}{m} \sum_{\nu=1}^{m-1} \frac{1 + \cos(\nu\pi/m)}{1 - \cos(\nu\pi/m)}\left[\left(1 - \cos\frac{\nu\pi}{m}\right)\left(3 - \cos\frac{\nu\pi}{m}\right)\right]^{\frac{1}{2}}$$
$$= \frac{n}{m} - 1 + \frac{1}{m} \sum_{\nu=1}^{m-1}\left(1 + \cos\frac{\nu\pi}{m}\right)\left[\frac{3 - \cos(\nu\pi/m)}{1 - \cos(\nu\pi/m)}\right]^{\frac{1}{2}}. \quad (4.4)$$

Values found using (4.4) are given in Table II.

TABLE I. Values of R_{eff}/R [from Eqs. (4.2) and (4.3)].

	1	2	3	4	5	6	7	8	9	10	11	12
6						$\frac{1171}{495}$	2.488	2.628	2.778	2.935	3.096	3.259
5					$\frac{47}{22}$	$\frac{2749}{1205}$	2.457	2.642	2.833	3.028	3.226	3.425
4				$\frac{13}{7}$	$\frac{2089}{1023}$	$\frac{37873}{16744}$	2.504	2.747	2.994	3.242	3.492	3.741
3			$\frac{3}{2}$	$\frac{121}{69}$	$\frac{430}{209}$	$\frac{1047}{440}$	$\frac{7148}{2639}$	$\frac{33647}{11067}$	3.733	3.707	4.040	4.3737
2		1	$\frac{7}{5}$	$\frac{15}{8}$	$\frac{45}{19}$	$\frac{43}{15}$	$\frac{239}{71}$	$\frac{433}{112}$	$\frac{1157}{265}$	$\frac{1017}{209}$	5.366	5.866
1	0	1	2	3	4	5	6	7	8	9	10	11
↑ $m/n \to$	1	2	3	4	5	6	7	8	9	10	11	12

TABLE II. Asymptotic values of R_{eff}/R [from Eq. (4.4)].

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
6						2.248	2.415	2.582	2.748	2.915	3.182	3.248	3.415	3.582
5					2.017	2.217	2.417	2.617	2.817	3.017	3.217	3.417	3.617	3.817
4				1.735	1.985	2.235	2.485	2.735	2.985	3.235	3.485	3.735	3.985	4.235
3			1.373	1.706	2.039	2.373	2.706	3.039	3.372	3.706	4.039	4.372	4.706	5.039
2		0.866	1.366	1.866	2.366	2.866	3.366	3.866	4.366	4.866	5.366	5.866	6.366	6.866
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
↑ $m/n \to$	1	2	3	4	5	6	7	8	9	10	11	12	13	14

5. A GENERALIZATION

A slight generalization of this result may be obtained by considering a square mesh made up of two different resistances, R_h for the resistances in the horizontal rows and R_v for those in the vertical columns. Reasoning similar to that in Sec. 2 shows that the voltage distribution is governed by the equation

$$R_h^{-1}[V(p, q) - V(p - 1, q) + V(p, q)$$
$$- V(p + 1, q)] + R_v^{-1}[V(p, q) - V(p, q - 1)$$
$$+ V(p, q) - V(p, q + 1)] = I\delta_{pa}\delta_{qb}. \quad (5.1)$$

The random walk analog to this equation is found by considering a random walk over a rectangular lattice in which the walker has a probability γ_p of moving in the p direction and a probability $\gamma_q = \frac{1}{2} - \gamma_p$ of moving in the q direction. The expectation of visit $F(p, q)$ is then given by the difference equation

$$F(p, q) = \delta_{pa}\delta_{qb} + \gamma_p[F(p - 1, q) + F(p + 1, q)]$$

$$+ \gamma_q[F(p, q - 1) + F(p, q + 1)]. \quad (5.2)$$

Equations (5.1) and (5.2) are equivalent if

$$V(p, q) = \frac{\rho RI}{2(1 + \rho)} F(p, q; k, l; a, b),$$
$$\gamma_p = \frac{\rho}{2(1 + \rho)}, \qquad \gamma_q = \frac{1}{2(1 + \rho)}, \quad (5.3)$$

where the notation has been simplified by introducing the ratio $\rho \equiv R_v/R_h$ and setting $R_h \equiv R$. The effective resistance between the point (a, b) where the current is introduced and the point (k, l), where it is removed is given by

$$\frac{R_{eff}}{R} = \frac{\rho}{2(1 + \rho)} F(a, b; k, l; a, b). \quad (5.4)$$

The function $F(a, b; k, l; a, b)$ is obtained by carrying out calculations similar to those given in the Appendix for the case $\gamma_p = \gamma_q = \frac{1}{4}$. After some not quite trivial computations [see also Ref. 3] it is found that

$$F(a, b; k, l; a, b) = \frac{1}{4\gamma_q m} \sum_{\lambda=0}^{m-1} \left\{ \frac{2 - \delta_{\lambda 0}}{\sinh \beta_\lambda (\cosh \beta_\lambda - 1) \sinh n\beta_\lambda} \right.$$

$$\times \left[\left(1 + \cos (2a - 1)\frac{\lambda\pi}{m}\right)(\sinh b\beta_\lambda - \sinh (b - 1)\beta_\lambda)(\sinh (n + 1 - b)\beta_\lambda - \sinh (n - b)\beta_\lambda) \right.$$

$$- 2\left(\cos (k - a)\frac{\lambda\pi}{m} + \cos (k + a - 1)\frac{\lambda\pi}{m}\right)(\sinh b\beta_\lambda - \sinh (b - 1)\beta_\lambda)(\sinh (n + 1 - l)\beta_\lambda - \sinh (n - l)\beta_\lambda$$

$$\left. \left. + \left(1 + \cos (2k - 1)\frac{\lambda\pi}{m}\right)(\sinh l\beta_\lambda - \sinh (l - 1)\beta_\lambda)(\sinh (n + 1 - l)\beta_\lambda - \sinh (n - l)\beta_\lambda) \right] \right\}, \quad (5.5)$$

TABLE III. Values of R_{eff}/R [from Eqs. (5.8) and (5.9)].

4	3	$\dfrac{6+8\rho+\rho^2}{4(1+\rho)}$	$\dfrac{9+40\rho+50\rho^2+20\rho^3+2\rho^4}{9+30\rho+26\rho^2+4\rho^3}$	$\dfrac{3+25\rho+63\rho^2+63\rho^3+25\rho^4+3\rho^5}{4+24\rho+42\rho^2+24\rho^3+4\rho^4}$	
3	2	$\dfrac{2+4\rho+\rho^2}{2+3\rho}$	$\dfrac{2+10\rho+10\rho^2+2\rho^3}{3+10\rho+3\rho^2}$	$\dfrac{2+20\rho+50\rho^2+40\rho^3+9\rho^4}{4+26\rho+30\rho^2+9\rho^3}$	$\dfrac{2+32\rho+130\rho^2+172\rho^3+82\rho^4+12\rho^5}{5+50\rho+91\rho^2+54\rho^3+9\rho^4}$
2	1	$\tfrac{1}{2}(1+\rho)$	$\dfrac{1+4\rho+2\rho}{3+2\rho}$	$\dfrac{1+8\rho+6\rho^2}{4(1+\rho)}$	$\dfrac{1+14\rho+22\rho^2+8\rho^3}{5+10\rho+4\rho^2}$
1	0	ρ	2ρ	3ρ	4ρ
$\uparrow$					
$m/n \to 1$	2	3		4	5

where

$$\cosh \beta_\lambda = \frac{1}{2\gamma_q} - \frac{\gamma_p}{\gamma_q} \cos \frac{\lambda \pi}{m}, \quad \lambda = 0, 1, \cdots, m-1,$$

$$\gamma_p + \gamma_q = \tfrac{1}{2}. \quad (5.6)$$

In using Eqs. (5.4) and (5.5) only the specialized case, $a = b = 1$, $k = m$, $l = n$, will be considered. It is readily shown that

$$\frac{R_{eff}}{R} = \frac{\rho}{2m} \sum_{\lambda=0}^{m-1} \left\{ \frac{2-\delta_{\lambda 0}}{(\cosh \beta_\lambda - 1)\sinh n\beta_\lambda} \cdots \right.$$

$$\times [(1+\cos\lambda\pi/m)(\sinh n\beta_\lambda - \sinh(n-1)\beta_\lambda$$

$$\left. - (-1)^\lambda \sinh \beta_\lambda)] \right\}, \quad (5.7)$$

where

$$\cosh \beta_\lambda = 1 + \rho - \rho\cos\lambda\pi/m, \quad \rho = \gamma_p/\gamma_q. \quad (5.8)$$

It is instructive to consider the two extreme cases of (5.7); that is, $\rho = 0$ and $\rho = \infty$. In the former case, there are n resistors, each of resistance $(m-1)R$ in parallel so that $R_{eff} = [(m-1)/n]R$; in the latter case, there are m resistors each of resistance $(n-1)\rho R$, in parallel so that $R_{eff} = [(n-1)/m]\rho R$. It is easily seen that (5.7) reduces to the proper results in these two limiting cases.

For the purpose of computation it is convenient to evaluate the term $\lambda = 0$; the result $2(n-1)\rho$ is obtained so that

$$\frac{R_{eff}}{R} = \left(\frac{n-1}{m}\right)\rho + \frac{1}{m}\sum_{\lambda=1}^{m-1}\left[\frac{1+\cos\lambda\pi/m}{1-\cos\lambda\pi/m}\right.$$

$$\left. \times \left(1 - \frac{\sinh(n-1)\beta_\lambda + (-)^\lambda \sinh\beta_\lambda}{\sinh n\beta_\lambda}\right)\right], \quad (5.9)$$

TABLE IV. Values of R_{eff}/R [from Eqs. (6.1) and (6.2) with $l = m = 2$] [asymptotic values from Eq. (6.3)].

n	1	2	3	4	5	6
R_{eff}/R	1	5/6	1	167/134	811/551	3407/1980
R_{eff}/R	0.470	0.720	0.970	1.220	1.470	1.720

where (5.8) continues to apply. Some values of the resistance obtained where m and n are small are given in Table III. An asymptotic formula analogous to (4.4) may be derived; it is

$$\frac{R_{eff}}{R} \xrightarrow[n\text{ large}]{} \left(\frac{n-m}{m}\right)\rho + \frac{(\rho^{\frac{1}{2}})}{m}\sum_{\lambda=1}^{m-1}\left(1 + \cos\frac{\lambda\pi}{m}\right)$$

$$\times \left[\frac{2+\rho-\rho\cos\lambda\pi/m}{1-\cos\lambda\pi/m}\right]^{\frac{1}{2}}. \quad (5.10)$$

6. THREE-DIMENSIONAL LATTICES

The method is readily extended to the three-dimensional counterpart of the plane square net; that is, the simple cubic lattice of points (p, q, r) with $1 \le p \le l$, $1 \le q \le m$, $1 \le r \le n$. The needed functions $F(p, q, r)$ are given in Ref. 3. Work along the lines used in Secs. 2 and 4 leads to

$$\frac{R_{eff}}{R} = \frac{1}{4lm}\sum_{\lambda=0}^{l-1}\sum_{\mu=0}^{m-1}\left\{(2-\delta_{\lambda 0})(2-\delta_{\mu 0}) \cdots\right.$$

$$\times \left(1 + \cos\frac{\lambda\pi}{l}\right)\left(1 + \cos\frac{\mu\pi}{m}\right)$$

$$\left. \times \frac{\sinh n\gamma_{\lambda\mu} - \sinh(n-1)\gamma_{\lambda\mu} - (-1)^{\lambda+\mu}\sinh\gamma_{\lambda\mu}}{\sinh n\gamma_{\lambda\mu}(\cosh\gamma_{\lambda\mu}-1)}\right\}, \quad (6.1)$$

where

$$\cosh \gamma_{\lambda\mu} = 3 - \cos\lambda\pi/l - \cos\mu\pi/m. \quad (6.2)$$

This expression is easily evaluated for small values of the parameters l, m, and n; results for $l = m = 2$, $n = 2$ to 6 are given in Table IV. An asymptotic formula is easily derived for this simple case; the result is

$$R_{eff}/R \xrightarrow[n\text{ large}]{} \tfrac{1}{4}(n-1) + \tfrac{1}{4}[2(\sqrt{3}-1)+\sqrt{2}-1]$$

$$= \tfrac{1}{4}(n+0.87824), \quad l = m = 2. \quad (6.3)$$

Resistances found by use of (6.3) are also listed in Table IV.

It is a simple matter to extend these considerations to determine the effective resistance of a

simple cubic lattice of infinite extent. When this is done the value

$$R_{eff}/R \text{ (simple cubic)} = 0.252731$$

is found in agreement with the results of Vineyard[7] who has also evaluated the resistances of the body-centered-cubic and face-centered-cubic lattices of infinite extent.

APPENDIX

To evaluate the effective resistance of a finite electrical network, the solutions to the problem of a random walk on a rectangular array with four rectilinear reflecting boundaries are needed. Using the expectation of visit as defined in Sect. 1 and the lattice described therein, an extension of the methods developed by Kerberle and Montet (Ref. 4) leads to

$$F_1(p, q) = \frac{1}{m} \sum_{k=0}^{m-1} \left\{ [2 - \delta_{k0}] \left[\cos (p - a) \frac{k\pi}{m} + \cos (p + a - 1) \frac{k\pi}{m} \right] \right.$$
$$\left. \times \cdots \frac{[\sinh q\beta_k - \sinh (q - 1)\beta_k][\sinh (n + 1 - b)\beta_k - \sinh (n - b)\beta_k]}{\sinh \beta_k(\cosh \beta_k - 1) \sinh n\beta_k} \right\}, \qquad (A1)$$

with

$$\cosh \beta_k = 2 - \cos k\pi/m, \quad k = 0, 1, 2, \cdots, m - 1. \qquad (A2)$$

This solution is equivalent to the solution[3] found for a random walk on a rectangular $(m \times n)$ net with periodic boundary conditions

$$F(p, q) = F(p + m, q) = F(p, q + n),$$

if m and n are odd integers. If m or n is even, the solutions have somewhat different forms; however, these differences become negligible for reasonably large values of m or n.[8] Now, when $k = 0$, Eq. (A2) shows that $\beta = 0$, and it is evident that the first term in Eq. (A1) varies as β^{-2} for β small; hence, the expectation [Eq. (A1)] diverges when all four boundaries are reflecting. This divergence is similar to that found when absorbing boundaries are allowed to recede to infinity.

Using (A1) in (1.6) shows that, for the case $q \leq l \leq b$,

$$F_1(p, q; k, l; a, b) = \frac{1}{m} \sum_{\nu=0}^{m-1} \frac{2 - \delta_{\nu 0}}{\sinh \beta_\nu(\cosh \beta_\nu - 1) \sinh n\beta_\nu} \cdots \left\{ [\sinh (n + 1 - b)\beta_\nu - \sinh (n - b)\beta_\nu] \right.$$

$$\times \left[\left(\cos (p - a) \frac{\nu\pi}{m} + \cos (p + a - 1) \frac{\nu\pi}{m} \right) \cdots (\sinh q\beta_\nu - \sinh (q - 1)\beta_\nu) \right.$$

$$\left. - \left(\cos (k - a) \frac{\nu\pi}{m} + \cos (k + a - 1) \frac{\nu\pi}{m} \right) (\sinh l\beta_\nu - \sinh (l - 1)\beta_\nu) \right]$$

$$+ [\sinh (n + 1 - l)\beta_\nu - \sinh (n - l)\beta_\nu] \left[\left(1 + \cos (2k - 1) \frac{\nu\pi}{m} \cdots (\sinh l\beta_\nu - \sinh (l - 1)\beta_\nu) \right） \right.$$

$$\left. \left. - \left(\cos (p - k) \frac{\nu\pi}{m} + \cos (p + k - 1) \frac{\nu\pi}{m} \right) (\sinh q\beta_\nu - \sinh (q - 1)\beta_\nu) \right] \right\}, \qquad (A3)$$

with (A2) applying.

Equation (A3) reduces to (4.2) of the text when the designated values of the variables are inserted and some simple trigonometric identities are applied.

[7] G. H. Vineyard, J. Math. Phys. 4, 1191 (1963).
[8] G. L. Montet (unpublished).

Characters of Irreducible Representations of the Simple Groups. II.
Application to Classical Groups

J.-P. Antoine and D. Speiser

Université de Louvain, Centre de Physique Nucléaire, Héverlé, Belgium
(Received 6 July 1964)

The general formulas found in a preceding paper for the characters of irreducible representations of simple Lie groups are developed in the case of the classical groups: the four series A_l, B_l, C_l, D_l, plus the exceptional case G_2. Groups of rank 1 and 2, plus $A_3 \simeq D_3$ are studied in detail.

INTRODUCTION

IN a previous paper,[1] hereafter referred to as I, we have given a geometrical construction of the characters of irreducible representations of all simple compact Lie groups. In the following paper, we apply this procedure to the classical groups; in particular, we specialize our general formulas for the characters [formulas (19), (20), and (21) of I], to the four series A_l, B_l, C_l, D_l, and to G_2. Furthermore, we examine in detail all groups of rank 1 and 2 and one group of rank $3 : A_3 \simeq D_3$. Indeed, these groups seem the most important ones for present day physics.

A. GROUPS OF TYPE A_l ($\sim SU_{l+1}$)

If referred to an orthonormal basis $\{e_i\}$ in the space E_{l+1}, the roots of A_l are the vectors[2–4]

$$e_{ik} = e_i - e_k, \quad \text{thus} \quad m = \tfrac{1}{2}l(l+1).$$

They all belong to the l-dimensional hyperplane:

$$\sum_{i=1}^{l+1} x_i = 0. \tag{1}$$

This space E_l contains the diagram Γ.

We now construct the *affine coordinate system* $p_1 \cdots p_l$ adapted to g^c by the method indicated in I (Sec. 3A),

$$R_0 = \tfrac{1}{2}\sum \alpha_i$$
$$= \tfrac{1}{2}(l, l-2, l-4, \cdots, -l+2, -l); \tag{2}$$

thus,

$$(R_0)_j = \tfrac{1}{2}l - j + 1, \quad j = 1, 2, \cdots, l+1.$$

Singular hyperplanes

$$\vartheta_{ik} = x_i - x_k = 0, \quad i \neq k.$$

Distance from the extremity of R_0 to ϑ_{ik}

$$\delta_{ik}(R_0, \vartheta_{ik}) = 1/\sqrt{2}\,|k - i|,$$

and thus

$$\min \delta_{ik} = \delta_{i,i+1}, \quad i = 1, 2, \cdots, l.$$

Hence, the surfaces of D_0 are the hyperplanes $x_i - x_{i+1} = 0$ and one finds

$$p_i = \lambda_i(x_i - x_{i+1}),$$

where $\lambda_i = 1$ because of (2). Thus, finally,

$$p_i = x_i - x_{i+1}. \tag{3a}$$

This means that the l roots $e_{i,i+1}$ are the outermost or *elementary* roots.

Inverting the system (3a) one obtains

$$x_1 = \frac{l}{l+1}p_1 + \frac{l-1}{l+1}p_2 + \cdots + \frac{1}{l+1}p_l,$$

$$x_2 = -\frac{1}{l+1}p_1 + \frac{l-1}{l+1}p_2 + \cdots + \frac{1}{l+1}p_l,$$

$$\vdots$$

$$x_{l+1} = -\frac{1}{l+1}p_1 - \frac{2}{l+1}p_2 - \cdots - \frac{l}{l+1}p_l. \tag{3b}$$

The *group* S is generated by reflections in the hyperplanes ϑ_{ik}, i.e., by permutations $x_i \leftrightarrow x_k$. Therefore, it is the symmetrical group S_{l+1} (group of all permutations of $l+1$ elements) which acts here on the coordinates x_i. The order of S_{l+1} is $(l+1)!$

A vector $\mathbf{V} = \sum_i x_i e_i$ is called *dominant* if $\mathbf{V} \in D_0$, i.e., if $x_i \geq x_{i+1}$ (of course, only vectors which fulfil the supplementary condition $\sum_i x_i = 0$ are considered).

The dimension of the irreducible representation $D(K_0)$ is given by Weyl's formula[5]

$$\dim D(K_0) = \prod_{\alpha_i} \frac{(\alpha_i, K_0)}{(\alpha_i, R_0)}. \tag{4}$$

[1] J. P. Antoine and D. Speiser, J. Math. Phys., 5, 1226 (1964).

[2] E. Cartan, thesis, Paris (1834); Bull. Soc. Math. 41, 53 (1913); Ann. Math. 4, 209 (1929); [in *Oeuvres complètes* (I), Gaumthier-Villars, Paris, 1952.]

[3] L. S. Pontrjagin, *Topologische Gruppen* (Teubner, Leipzig, 1957), 2nd ed.

[4] G. Racah, "Group Theory and Spectroscopy," Princeton Lecture Notes, CERN (unpublished).

[5] H. Weyl, Math. Zs. 23, 271 (1925); 24, 328, 377, 789 (1926); See IV, Article 3 in *Selecta* (Birkhäuser, Basel, 1956).

For A_l, (4) yields

$$\dim D(K_0) = \frac{1}{1!\,2!\,\cdots\,l!} \prod_{i<k} (x_i - x_k),$$

or

$$\dim D(K_0) = \prod_{j=0}^{l-1} \frac{1}{(j+1)!} \prod_{k=1}^{l-j} (p_k + \cdots + p_{k+j}). \quad (5)$$

The irreducible representations of A_l fall into $l + 1$ classes, which form a group, isomorphic to the cyclic group Z_{l+1}. Indeed, consider the representation SU_{l+1} (set of all unitary unimodular matrices of dimension $l + 1$), by which A_l is usually defined. From Schur's lemma and the unimodularity property, it follows that the center of SU_{l+1} consists of all elements $a \cdot I$, $a^{l+1} = 1$. Thus it is isomorphic to Z_{l+1}. The lattice g^e, therefore, may be decomposed into $(l + 1)$ sublattices similar to g^e (see Fig. 3). That is, the crystal class of every sublattice is the group S (of course, a different origin has to be assigned to each of them). The points of the different sublattices represent, respectively, the elements 1, ϵ, $\epsilon^2, \cdots, \epsilon^l$ of the toroid (see I, Sec. 2). We call these sublattices the classes $0, 1, 2, 3, \cdots, l$. We now define the association

$$0 \to 1 : \text{class } 0 \quad (\text{sublattice } g^e),$$
$$p_i \to \epsilon^i : \text{class } j, \quad j = 1, \cdots, l.$$

(This association is not unique, we could have defined as well: $p_i \to \epsilon^{i-j}$: class j.)

Thereby, the class of every lattice point is determined. Indeed, to the sum $P_a + P_b$ corresponds in the toroid the product $\epsilon^a \cdot \epsilon^b = \epsilon^{a+b}$, thus, to the vector $P = p^i P_i$, the element

$$\epsilon^{p_1} \epsilon^{2p_2} \cdots \epsilon^{lp_l} = \epsilon^{\sum_{1}^{l} kp},$$

which belongs to class j if and only if $\sum_1^l kp_k \equiv j \bmod (l + 1)$. Thus clearly the $l + 1$ classes form a group, isomorphic to the cyclic group $Z_{l+1} = \{1, \epsilon, \epsilon^2, \cdots, \epsilon^l\}$. It follows that also the classes of irreducible representations form a group with the direct product as group operation, since to every lattice point inside D_0 one and only one representation is associated.

Let the representation D be characterized through the vector ($\equiv$ lattice point) $K_0 = (p_1 \cdots p_l)$. Its highest weight then is the vector $L_0 = (p_1 - 1 \cdots p_l - 1)$. We say: D belongs to class j if its highest weight does so, i.e., if

$$\sum_1^l k(p_k - 1) = \sum_1^l kp_k - \frac{l(l+1)}{2} \equiv j \bmod (l + 1).$$

Formula (I.6) then shows that all its weights belong to the same sublattice, because the same property
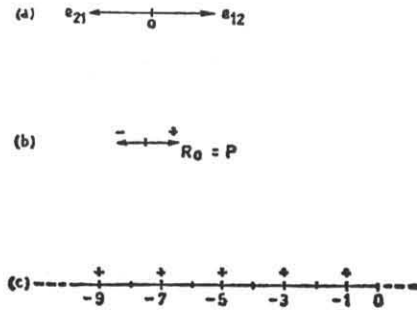

FIG. 1 The case $A_1 = B_1 = C_1$: (a) root diagram; (b) Δ; (c) $1/\Delta$.

holds for the points of Δ and of $X(K_0)$, by (I.7) and (I.9).

Thus, the unit representation belongs to class 0, the jth fundamental representation ($p_k = 1$, $k \neq j$, $p_j = 2$, i.e., $L_0 = P_j$) to class j; the adjoint representation $(1, 0, \cdots, 0, 1)$ again to class 0, and so on.

1. $A_1 = B_1 = C_1$

Positive root : e_{12} [see Fig. 1(a)],

$$R_0 = \tfrac{1}{2} e_{12} = P.$$

Coordinates : $x_1 = \tfrac{1}{2} p,$

$$x_2 = -\tfrac{1}{2} p,$$
$$\Delta = [1] - [-1] \quad \text{[see Fig. 1(b)]}.$$

$$\frac{1}{\Delta} : \sum_{e_{21}} [-1] = \sum_{k=0}^{\infty} [-2k - 1], \quad \text{[see Fig. 1(c)]}.$$

Construction of a character:

$$X(p) = [p] - [-p],$$
$$= \sum_{e_{21}} (X(p)[-1])$$
$$= \sum_{e_{21}} ([p - 1] - [-p - 1]),$$
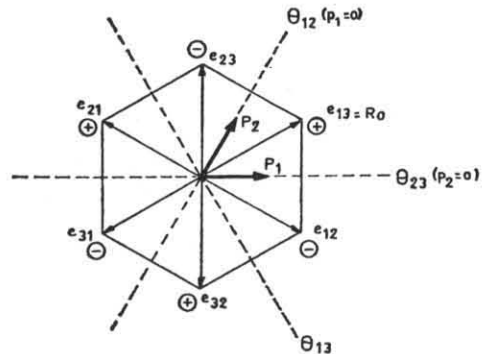$$= [p - 1] + [p - 3] + \cdots + [-p + 1].$$
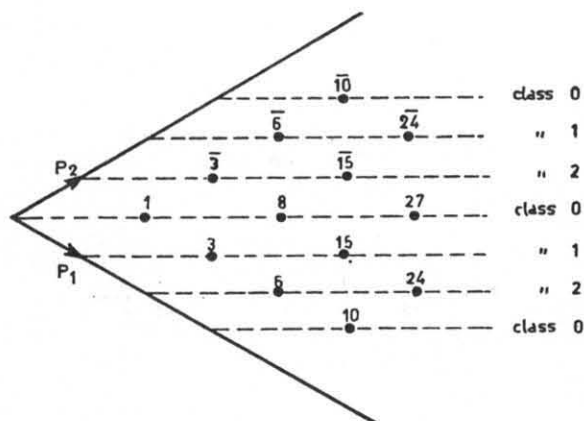

FIG. 2. Root diagram and Δ of A_2.

FIG. 3. A_2, the repartition of the three classes of representations into D_0; next to every lattice point K_0 of g^c stands the dimension of the representation associated with it ($\bar{d}$ denotes the complex conjugate representation of d).

In more traditional notation, putting $p = 2m + 1$ and $\varphi = \frac{1}{2}\psi$, this becomes

$$\chi = [m] + \cdots + [-m],$$

$$\chi = \frac{e^{i(m+\frac{1}{2})\psi} - e^{-i(m+\frac{1}{2})\psi}}{e^{i(\psi/2)} - e^{-i(\psi/2)}} = \sum_{k=-m}^{+m} e^{ik\psi}.$$

Dimension of D : dim $D(p) = p = 2m + 1$

As is well known, there are two classes of representations : $p - 1 \equiv 0$ or $1 \mod 2$:

class 0 : p odd, m integer,

class 1 : p even, m half integer.

2. A_2

Positive roots : e_{12}, e_{13}, e_{23}, where $e_{13} = e_{12} + e_{23}$, thus $R_0 = e_{13}$.

Coordinates:

$$x_1 = \tfrac{1}{3}(2p_1 + p_2),$$
$$x_2 = \tfrac{1}{3}(-p_1 + p_2),$$
$$x_3 = \tfrac{1}{3}(-p_1 - 2p_2).$$

Δ is a regular hexagon (Fig. 2).

Dimension[6]: dim $D(p_1 p_2) = \frac{1}{2} p_1 p_2 (p_1 + p_2)$.

There are three classes of representations. [This classification is basic for the Sakata model[7] : class 1 describes baryons; class 2, antibaryons; and class 0, the mesons in accordance with $B \otimes B \simeq M$, etc. The eightfold way[8] uses only representations of

---

[6] One often writes: $p_i = 1 + \lambda_i$ in the dimension formula (e. g., when the representations are presented as tensorial representations); λ_i are the components of the highest weight L_0.

[7] S. Sakata, Progr. Theoret. Phys. (Kyoto) **16**, 686 (1956). M. Ikeda, S. Ogawa and Y. Ohnuki, Progr. Theoret. Phys. (Kyoto) **22**, 715 (1959); **23**, 1073 (1960).

[8] M. Gell-Mann, (unpublished); Phys. Rev. **125**, 1067 (1962). Y. Ne'eman, Nucl. Phys. **26**, 222 (1961). D. Speiser and J. Tarski, Possible global symmetries (unpublished); J. Math. Phys. **4**, 588 (1963).

---

class 0.]:

$$p_1 + 2p_2 \equiv j \mod 3,$$

or

$$p_1 - p_2 \equiv j \mod 3, \qquad j = 0, 1, 2.$$

Figure 3 shows the repartition of the 3 classes into D_0.

Note: Δ and $1/\Delta$ belong to class 0.

The diagram $1/\Delta$ is constructed in the standard way in Fig. 4. However, it may also be considered as the "sum" (superposition) of an infinite number of similar diagrams ω_k (Fig. 5). All weights on the ω_k are simple. ω_1 has the same position as $1/\Delta$, ω_2 is shifted by the length $-R_0$ in the direction e_{31}, ω_3 by $-2R_0$, and so on. This interpretation permits to write down at once every character explicitly.[9]

Indeed, let $X(p_1, p_2)$ be a characteristic. Computing then the product $X(p_1, p_2) \cdot 1/\Delta$ one sees that the character χ may be considered as the superposition of a finite series of hexagonal figures (precisely: figures invariant under S), all points of it having multiplicity 1. Symbolically:

$$\chi(p_1, p_2) = F(p_1 - 1, p_2 - 1) + F(p_1 - 2, p_2 - 2)$$
$$+ \cdots + F(p_1 - p_2, 0) \qquad (p_1 \geq p_2).$$

Here $F(i, k)$ represents a diagram which contains the point (i, k), its five equivalents by S and all points of the same sublattice located inside or on this hexagon, all of them with multiplicity 1. The sum on the rhs then means the superposition of the different diagrams $F(i, k)$. A better understanding of this procedure may be provided through an example; Fig. 6 shows the construction of $\chi(6, 3)$.
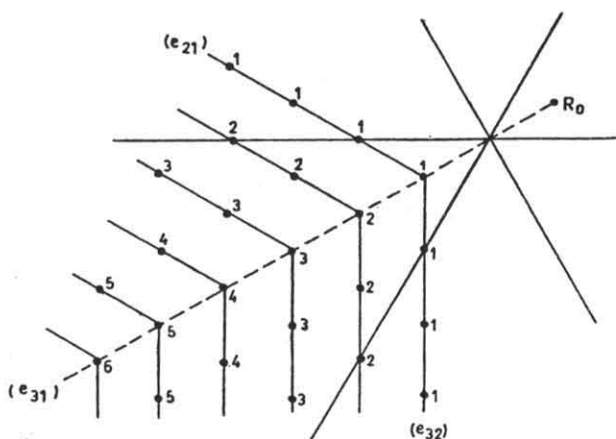

FIG. 4. The $1/\Delta$ of A_2.

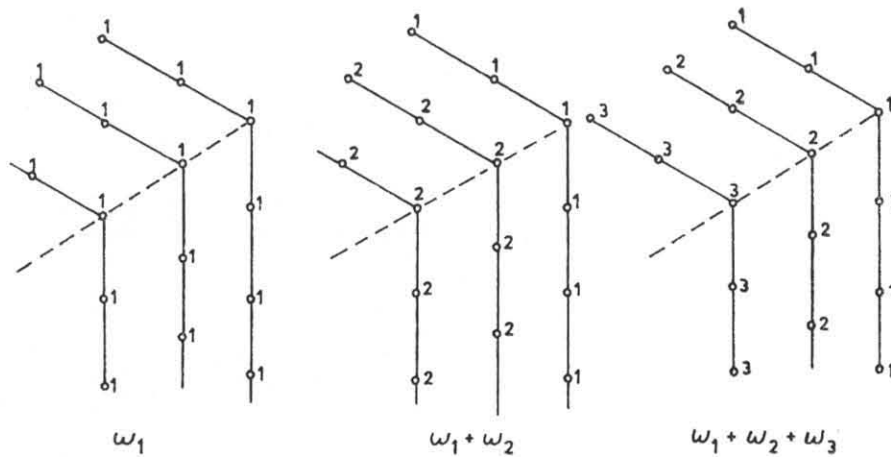[9] J. P. Antoine, Ann. Soc. Sci. Bruxelles **77** III, 150 (1963).

FIG. 5. Decomposition of $1/\Delta$ of A_2 into the diagrams ω_i.

On this example, one also sees the following property of the WD's: The boundary of a WD consists of 2 types of edges: one finds—starting from the highest weight—two edges, one of type p_1 (perpendicular to P_1) carrying p_2 weights and one of type p_2 (perpendicular to P_2) carrying p_1 weights. This property is in fact valid for all A_l (see below).

3. $A_3 = D_3$

The root diagram is the set of vectors pointing to the centers of the 12 edges of a cube (oriented this way, the diagram shows the characteristic properties of the group D_3 rather than of A_3) (Fig. 7). Positive roots:

$$e_{12}, e_{13}, e_{14}, e_{23}, e_{24}, e_{34},$$

where

$$e_{ij} + e_{jk} = e_{ik} \qquad i < j < k;$$

thus,

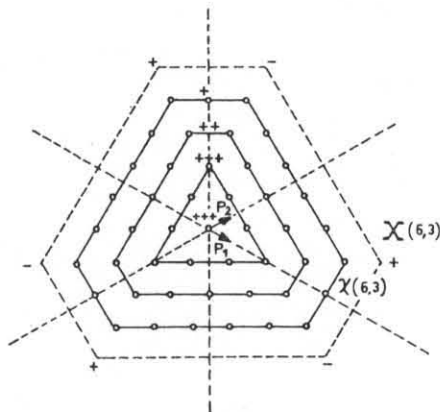$$R_0 = (-\tfrac{3}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}).$$

FIG. 6. Construction of the WD of $D(6, 3)$ of A_2 (dim = 81) $\chi(6,3) = F(5, 2) + F(4, 1) + F(3, 0)$ (see the text).

Coordinates:

$$x_1 = \tfrac{1}{4}(3p_1 + 2p_2 + p_3),$$
$$x_2 = \tfrac{1}{4}(-p_1 + 2p_2 + p_3),$$
$$x_3 = \tfrac{1}{4}(-p_1 - 2p_2 + p_3),$$
$$x_4 = \tfrac{1}{4}(-p_1 - 2p_2 - 3p_3).$$

Δ is a convex polyhedron with 24 corners, 36 edges
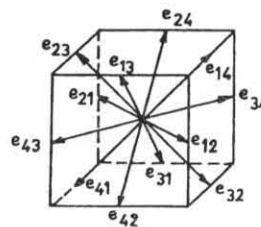

FIG. 7. Root diagram of A_3.

and 14 surfaces. 8 of these are regular hexagons and 6 are squares (Fig. 8).

Dimension:

$$\dim D(p_1 p_2 p_3)$$
$$= \tfrac{1}{12} p_1 p_2 p_3 (p_1 + p_2)(p_2 + p_3)(p_1 + p_2 + p_3).$$

There are 4 classes of representations according to:

$$p_1 + 2p_2 + 3p_3 - 6 \equiv j \bmod 4, \qquad j = 0, 1, 2, 3.$$

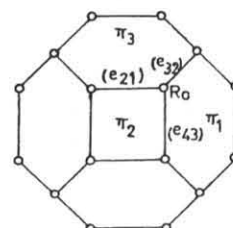Their distribution into D_0 is shown in Fig. 9.

FIG. 8. Δ of A_3 as viewed along P_2 (projection on a plane π_2).
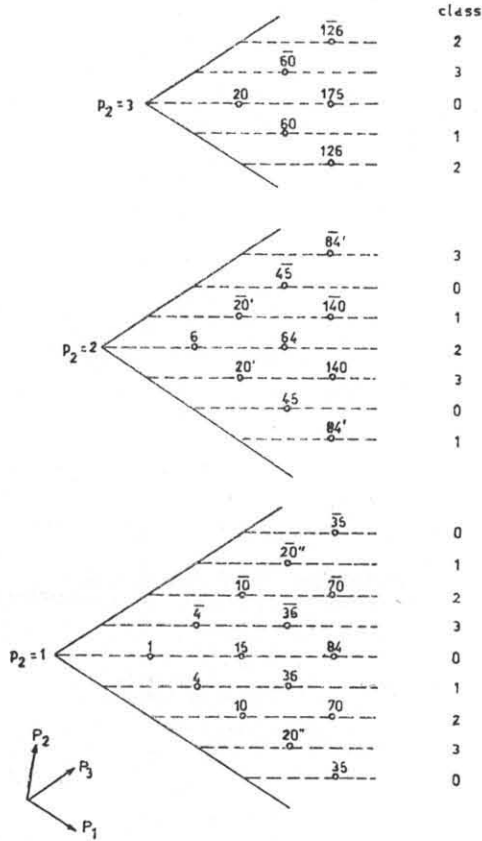
FIG. 9. A_3, the repartition of the 4 classes of representations into D_0; one proceeds as for A_2, taking one by one the intersections with the planes $p_2 =$ constant.

$1/\Delta$ is an infinite pyramid with 3 edges and 3 surfaces. The edges are parallel to the elementary roots e_{21}, e_{32}, e_{43}.

surface $\pi_1 = (e_{43}, e_{32}) \perp P_1$
    the angle between e_{43} and $e_{32} = 120°$.

surface $\pi_2 = (e_{21}, e_{43}) \perp P_2$
    the angle between e_{43} and $e_{21} = 90°$.

surface $\pi_3 = (e_{32}, e_{21}) \perp P_3$
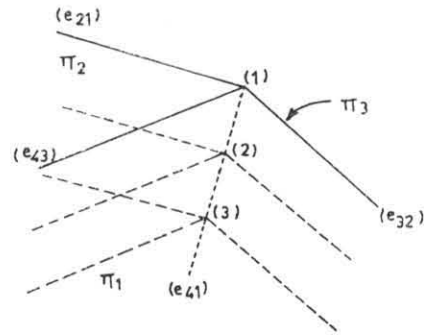   the angle between e_{32} and $e_{21} = 120°$.

FIG. 10. The $1/\Delta$ of A_3. General view.

All weights on the edges are simple. The multiplicities of the weights depend only upon the angles between the different roots, as was shown explicitly through the construction of $1/\Delta$ in I; whence follows that all weights on surface π_2 have multiplicity 1 whereas surfaces π_1 and π_3 are identical to $1/\Delta$ of A_2. (Fig. 10).

Here one may also decompose $1/\Delta$ in successive "shells" or trihedra, all similar but with increasing multiplicity. The summits of the trihedra lie on a ray pointing along the root e_{41}. On each "shell" the multiplicities are ordered similarly to the outermost "shell." Figure 11 shows the projections of the 3 first "shells" on planes parallel to π_2. When $1/\Delta$ is presented in this geometric way, the characters may be easily obtained also as geometric figures, e.g., by proceeding through successive layers parallel to π_2. One finds a result analogous to the one found above for A_2: $\chi(p_1 p_2 p_3)$ is a polyhedron invariant under S, with 3 types of surfaces ($\pi_i \perp P_i$):

surface $\pi_1 =$ figure $\chi(p_2 p_3)$ of A_2, its edges having respectively, p_2 and p_3 weights.

surface $\pi_2 =$ rectangle, its edges have respectively, p_1 and p_3 weights, and all weights on it are simple.

surface $\pi_3 =$ figure $\chi(p_1 p_2)$ of A_2, its edges having respectively, p_1 and p_2 weights.
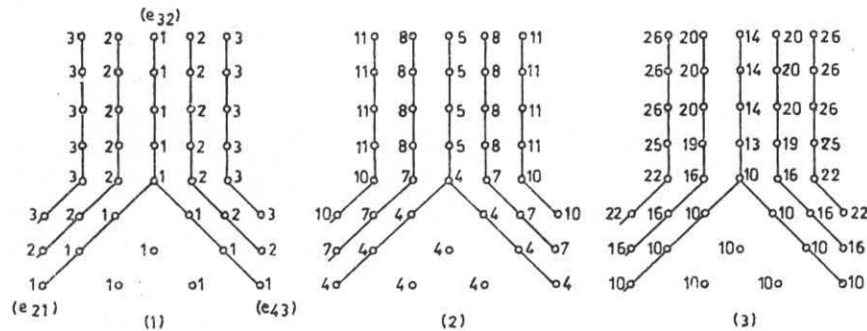

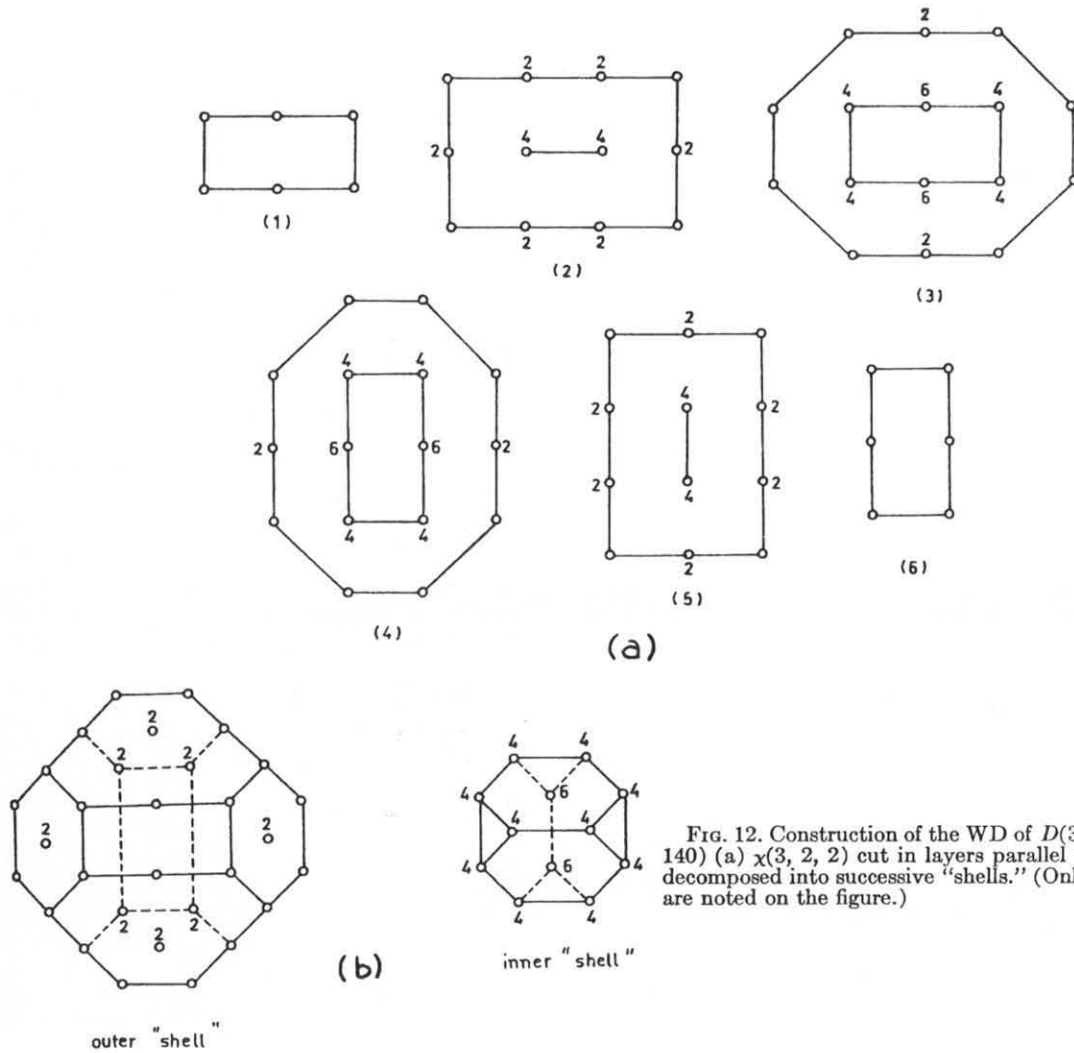FIG. 11. The $1/\Delta$ of A_3, projection of the first 3 "shells" on the plane π_2.

FIG. 12. Construction of the WD of $D(3, 2, 2)$ of A_3 (dim = 140) (a) $\chi(3, 2, 2)$ cut in layers parallel to π_2; (b) $\chi(3, 2, 2)$ decomposed into successive "shells." (Only multiplicities > 1 are noted on the figure.)

Thus from every corner start 3 edges, the edge (π_i, π_j) having p_k weights ($i, j, k = 1, 2, 3 + $ cycl.).

The weights inside this boundary are distributed on "shells" in much the same way, but the multiplicities are higher. All weights of course are on lattice points of the same class. As an example, Fig. 12 shows $\chi(3, 2, 2)$ cut in layers parallel to π_2 and then decomposed in successive "shells". For more examples of WD from A_3, see the work of Wigner.[10]

4. A_l, The General Case

It is clearly impossible to use a graphical method for treating groups of rank > 3. Therefore, the general formula must be expressed in the coordinate system $(x_1, x_2, \cdots, x_{l+1})$ best suited for groups A_l, but geometrical analogies will provide valuable guidance. In order to do so the m summations in

[10] E. P. Wigner, Phys. Rev. **51**, 106 (1937).

(I. 18) may conveniently be carried out in two stages of, respectively, l and $(m - l)$ summations.

In the first stage, we sum along l independent roots: this yields a pyramid ω, which has l edges, l surfaces [$(l - 1)$-dimensional hyperplanes] and multiplicity 1 at all its points. These roots will be selected in such a way that the l surfaces of ω obey equations as simple as possible.

We choose to sum first along $e_{l+1, j}, j = 1, 2, \cdots, l$. This yields for the surfaces of ω the l hyperplanes:

$$x_i = (-R_0)_i = -l/2 + j - 1, \quad j = 1, 2, \cdots, l,$$

which all pass through the point $- R_0$ (summit of ω).

Indeed, by definition: (m_{ik} is the parameter indicating the summation along e_{ik}):

$$\omega \equiv \omega(-R_0) = \sum_{e_{l+1, i}} \cdots \sum_{e_{l+1, l}} [-R_0],$$

$$= \sum_{m_{l+1, i}=0}^{\infty} \cdots \sum_{m_{l+1, l}=0}^{\infty} \left[-R_0 + \sum_{i=1}^{l} m_{l+1, i} e_{l+1, i} \right].$$

For every $j = 1, 2, \cdots, l$, one has

$$\left(-R_0 + \sum_{i=1}^{l} m_{i+1,i} e_{i+1,i}\right)_j$$

$$= (-R_0)_j - m_{i+1,j} \leq (-R_0)_j,$$

where, as in all formulas below, the inequality sign means that both sides of the inequality differ by an integer. Thus, $\omega(-R_0)$ is the set of all lattice points Q whose coordinates $x_1 \cdots x_l$ satisfy the l inequalities (with the convention on inequality signs)

$$x_j \leq (-R_0)_j, \qquad j = 1, 2, \cdots, l, \qquad (6)$$

i.e., the pyramid described above.

Note: For A_2, we did not follow this prescription in order to get ω, we summed along e_{21} and e_{32} rather than along e_{31} and e_{32}; in this particular case, the geometric construction was somewhat simplified in this way.

Performing the $m - l = \frac{1}{2}l(l + 1)$ remaining summations, one gets:

$$1/\Delta = \sum_{e_{21}} \cdots \sum_{e_{l,l-1}} \omega(-R_0),$$

$$= \sum_{m_{21}=0}^{\infty} \cdots \sum_{m_{l,l-1}=0}^{\infty} \omega(-R_0 + m_{21}e_{21} + \cdots$$

$$+ m_{l,l-1}e_{l,l-1}).$$

We write this relation in the following form:

$$1/\Delta = \sum_{m_{21}=0}^{\infty} \cdots \sum_{m_{l,l-1}=0}^{\infty} \omega(m_{21} \cdots m_{l,l-1}). \qquad (7)$$

In this summation ($\equiv$ superposition, as usual), the term $\omega(m_{21} \cdots m_{l,l-1})$ clearly represents a pyramid identical to ω, with its summit shifted up to the point:

$$-R_0 + m_{21}e_{21} + \cdots + m_{l,l-1}e_{l,l-1}.$$

For every $j = 1, 2, \cdots, l$, we have

$$(-R_0 + m_{21}e_{21} + \cdots + m_{l,l-1}e_{l,l-1})_j$$

$$= (-R_0)_j - \sum_{k=0}^{l-j-1} m_{l-k,j} + \sum_{k=1}^{j-1} m_{j,j-k}.$$

Thus, as above, we may write

$$\omega(m_{21} \cdots m_{l,l-1}) = \sum_Q [Q(m_{21} \cdots m_{l,l-1})], \qquad (8a)$$

where the right-hand side means the set of all lattice-points Q satisfying the l conditions ($j = 1, 2, \cdots, l$)

$$Q_j(m_{21} \cdots m_{l,l-1}) \equiv x_j \leq (-R_0)_j$$

$$- \sum_{k=0}^{l-j-1} m_{l-k,j} + \sum_{k=1}^{j-1} m_{j,j-k}. \qquad (8b)$$

Obviously $\omega(0, \cdots, 0) = \omega(-R_0)$. Summing up

conditions (8b) for $j = 1, 2, \cdots, l$, one gets

$$-x_{l+1} = \sum_{j=1}^{l} x_j \leq -\frac{l}{2}, \quad \text{i.e.,} \quad x_{l+1} \geq \frac{l}{2},$$

since every index $m_{\alpha\beta}$ appears in exactly two lines: in the line α with sign $+$ and in the line β with sign $-$ ($\alpha > \beta > 0$).

Remembering now that $S = S_{l+1}(x_1 \cdots x_{l+1})$ (permutation group), a characteristic $X(K_0)$ will be written as

$$X(K_0) = \sum_S \delta_*[sK_0] = \sum_{\text{perm}} \delta_P[PK_0],$$

where

$$(PK_0)_i = (K_0)_{i_i} \quad \text{with} \quad P \equiv \begin{vmatrix} 1 & 2 & \cdots & l+1 \\ i_1 & i_2 & \cdots & i_{l+1} \end{vmatrix}. \qquad (9)$$

This gives for the corresponding characters:

$$\chi = \sum_{m_{21}=0}^{\infty} \cdots \sum_{m_{l,l-1}=0}^{\infty} \left\{ \sum_{\text{perm}} \delta_P[PK_0] \cdot \right.$$

$$\left. \times \sum_{Q \in (8b)} [Q(m_{21} \cdots m_{l,l-1})] \right\},$$

$$= \sum_{m_{21}=0}^{\infty} \cdots \sum_{m_{l,l-1}=0}^{\infty} \left\{ \sum_{\text{perm}} \delta_P \right.$$

$$\left. \times \sum_{Q \in (8b)} [Q + PK_0] \right\}. \qquad (10)$$

χ_0 is of the same form, with the additional condition

$$Q + PK_0 \in D_0,$$

i.e.,

$$z_k \geq z_{k+1}, \quad \text{where} \quad z_k \equiv (Q + PK_0)_k$$

$$= x_k + (K_0)_{i_k}. \qquad (11)$$

Conditions (8b) and (11) restrict the range of the m_{jk}, such that—to obtain the character—it suffices to enumerate their possible values. To show how the method works, we shall construct the l fundamental representations of A_l:

$$K_0^{(\lambda)} = (1, 1, \cdots, 1, 2, 1, \cdots, 1)$$

(one 2 at the λth place, all others 1).

According to the place which 2 occupies, i.e., according to whether $\lambda = 1, 2, \cdots, l$, one obtains the λth fundamental representation.

The $(l + 1)$ Cartesian coordinates of $K_0^{(\lambda)}$ are by (3b)

$$(K_0^{(\lambda)})_i = \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - i + 1,$$

$$\text{if} \quad i = 1, \cdots, \lambda,$$

$$= \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - i,$$

$$\text{if} \quad i = \lambda + 1, \cdots, l + 1.$$

Condition (11) together with $\sum_1^{l+1} z_i = 0$ implies that $z_1 \geq 0$ and $z_{l+1} \leq 0$. Let in (9)

$$(K_0^{(\lambda)})_{i_1} = \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - q \qquad (q \geq 0),$$

whence

$$z_1 = x_1 + \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - q.$$

With the help of (8b) and $z_1 \geq 0$ one finds

$$-\frac{l}{2} - \frac{l - \lambda + 1}{l + 1} + q \leq x_1 \leq -\frac{l}{2} - \sum_{k=0}^{l-2} m_{l-k,1}.$$

This equation yields

$$q = 0,$$

$$m_{l-k,1} = 0, \quad \text{for} \quad k = 0, 1, \cdots, l - 2.$$

Thus, $x_1 = -\frac{1}{2}l$ and $z_1 = (l - \lambda + 1)/(l + 1)$. On the other hand, let

$$(K_0^{(\lambda)})_{i_{l+1}} = -\frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - 1 + p \qquad (p \geq 0)$$

and thus:

$$z_{l+1} = x_{l+1} - \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - 1 + p.$$

Therefore, again with the help of (8b) and $z_{l+1} \leq 0$ one obtains

$$\frac{l}{2} \leq x_{l+1} \leq \frac{l}{2} - \frac{l - \lambda + 1}{l + 1} + 1 - p$$

from which follows that $p = 0$, whence

$$x_{l+1} = \frac{l}{2}, \quad \text{and} \quad z_{l+1} = \frac{l - \lambda + 1}{l + 1} - 1.$$

Thus one has

$$z_1 - z_{l+1} = 1.$$

Since $z_k - z_{k+1}$ is a positive integer one finds

$$z_1 = z_2 = \cdots = z_\alpha = \frac{l - \lambda + 1}{l + 1},$$

$$z_{\alpha+1} = z_{\alpha+2} = \cdots = z_{l+1} = \frac{l - \lambda + 1}{l + 1} - 1.$$

The condition $\sum z_i = 0$ shows that $\alpha = \lambda$.

Then proceeding step by step one gets, first,

$$z_2 = \frac{l - \lambda + 1}{l + 1},$$

$$(K_0^{(\lambda)})_{i_2} = (K_0^{(\lambda)})_2, \quad \text{and} \quad x_2 = -\frac{1}{2}l + 1,$$

whence

$$m_{l-k,2} = 0, \quad k = 0, 1, \cdots, l - 3.$$

Indeed, let

$$(K_0^{(\lambda)})_{i_2} = \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - q \quad \text{with} \quad q \geq 1,$$

whence

$$z_2 = x_2 + \frac{l}{2} + \frac{l - \lambda + 1}{l + 1} - q = \frac{l - \lambda + 1}{l + 1}.$$

With the help of (8b) then follows

$$-\frac{l}{2} + q = x_2 \leq -\frac{l}{2} + 1 - \sum_{k=0}^{l-3} m_{l-k,2},$$

and, therefore, since $q \geq 1$,

$$q = 1, \quad \text{i.e.,} \quad (K_0^{(\lambda)})_{i_2} = (K_0^{(\lambda)})_2,$$

$$m_{l-k,2} = 0, \quad k = 0, 1, \cdots, l - 3.$$

Then in the same way one finds

$$z_3 = \frac{l - \lambda + 1}{l + 1}$$

$$(K_0^{(\lambda)})_{i3} = (K_0^{(\lambda)})_3 \quad \text{and} \quad x_3 = -\frac{1}{2}l + 2,$$

whence

$$m_{l-k,3} = 0 \quad k = 0, 1, \cdots, l - 4,$$

and so on for $z_4 \cdots z_{l+1}$.

Finally, one finds a unique choice: $m_{jk} = 0$ for all admissible j and k, and therefore exactly one weight vector in D_0:

$$z = \left(\frac{l - \lambda + 1}{l + 1}, \cdots, \frac{l - \lambda + 1}{l + 1}, \frac{l - \lambda + 1}{l + 1} - 1, \right.$$

$$\left. \cdots, \frac{l - \lambda + 1}{l + 1} - 1 \right),$$

$$= (0, 0, \cdots, 0, 1, 0, \cdots, 0) \quad \text{in coordinates } q_i.$$

The WD thus contains the vector z and all its equivalents by S, so that the dimension of the λth fundamental representation is $\binom{\lambda+1}{l}$.

This special example, of course, is somewhat trivial. For by comparing the dimension of the representation (which is known already before this computation) with the number of weights equivalent to the highest one, one sees that there are no other weights. In particular, there are no multiple weights. Nevertheless, the example illustrates the principle as well as the main details of this method.

B. GROUPS OF TYPE B_l ($\sim O_{2l+1}$)

Referred to an orthonormal basis $\{e_i\}$ in E_l the roots are the vectors
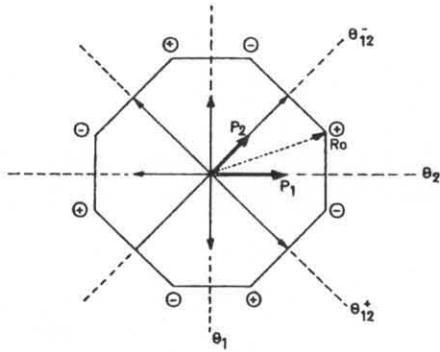
$$\pm e_i, \quad \pm e_i \pm e_k \qquad \text{(cf. Refs. 2–4).}$$

FIG. 13. Root diagram and Δ of B_2.

Thus,

$$R_0 = \tfrac{1}{2} \sum \alpha_i = \tfrac{1}{2}(2l - 1, 2l - 3, \cdots, 1),$$

i.e.,

$$(R_0)_i = (2l + 1 - 2i). \tag{12}$$

Affine coordinates. A computation similar to the one effected for case A_l yields

$$p_i = x_i - x_{i+1}, \qquad i = 1, 2, \cdots, l - 1,$$

$$p_l = 2x_l,$$

whence by inversion

$$
\begin{aligned}
x_1 &= p_1 + p_2 + p_3 + \cdots + p_{l-1} + \tfrac{1}{2}p_l, \\
x_2 &= \qquad\ \ p_2 + p_3 + \cdots + p_{l-1} + \tfrac{1}{2}p_l, \\
x_{l-1} &= \qquad\qquad\qquad\qquad\quad p_{l-1} + \tfrac{1}{2}p_l, \\
x_l &= \qquad\qquad\qquad\qquad\qquad\qquad \tfrac{1}{2}p_l.
\end{aligned}
\tag{13}
$$

The *group* S is generated by reflections in the singular hyperplanes $\vartheta_{ik}^{\pm}(x_i \pm x_k = 0)$ and $\vartheta_i(x_i = 0)$, i.e., by transformations

$$x_i \leftrightarrow \pm x_k \quad \text{and} \quad x_i \leftrightarrow -x_i.$$

Thus S is the group of all permutations of the coordinates x_i, plus permutations of the coordinates supplemented by an arbitrary change of their signs: its order is therefore $s = 2^l l!$

A vector $V = \sum_i x_i e_i$ is *dominant* if $x_1 \geq x_2 \geq \cdots \geq x_l \geq 0$.

The *dimension* of an irreducible representation is given by Weyl's formula (4), which here gives

$$\dim D(K_0)$$

$$= \frac{2^l}{(2l - 1)!!} \frac{\prod\limits_{i<k} (x_i - x_k)(x_i + x_k) \prod\limits_i x_i}{\prod\limits_{j=1}^{l-1} (2l - 2j)!}, \tag{14}$$

where $K_0 = (x_1, x_2 \cdots x_l)$.

Classes of representations. The representation O_{2l+1}

(set of all orthonormal matrices of dimension $2l + 1$; by this representation B_l is usually defined) has only the unit element in its center, but its underlying space is doubly connected. Whence the universal covering group has two elements in the center, and the vectors of g^c, therefore, belong to either of two classes. These classes form the group Z_2. The same holds for the representations which the lattice points determine.

O_{2l+1} and the adjoint representation belong to the same class which is spanned by the roots (class 0).

Thus, one sees immediately the distribution of the lattice points into the two classes (sublattices):

class 0: x_i integer, $q_l \equiv 0 \bmod 2$,
class 1: x_i half-integer, $q_l \equiv 1 \bmod 2$.

A representation $D(K_0(p_1 \cdots p_l))$ belongs to class 0 (1) if its highest weight $L_0(p_1 - 1, \cdots, p_l - 1)$ belongs to class 0(1). Thus one sees, that the $(l - 1)$ first fundamental representations belong to class 0, but the lth, the "spinor" representation, to class 1. Expressed in Cartesian and affine coordinates, its highest weight is the vector:

$$L_0^{(sp)} = (\tfrac{1}{2}, \tfrac{1}{2}, \cdots, \tfrac{1}{2}) = (0, 0, \cdots, 0, 1).$$

1. $B_2 = C_2$

Positive roots: $e_1 \pm e_2, e_1, e_2.$
Coordinates: $x_1 = p_1 + \tfrac{1}{2}p_2$
                              : tetragonal lattice.
                  $x_2 = \tfrac{1}{2}p_2$

Root diagram and Δ: See Fig. 13.
Dimension: $\dim D(p_1 p_2) = \tfrac{1}{6} p_1 p_2 (p_1 + p_2)(2p_1 + p_2).$
Classes: class 0: p_2 odd,
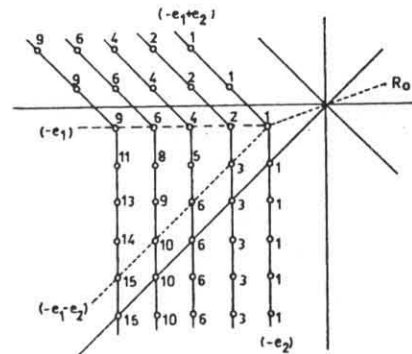   class 1: p_2 even.
The $1/\Delta$ is given in Fig. 14.

FIG. 14. The $1/\Delta$ of B_2. Note that except in the sector between the rays $(-e_1)$ and $(-e_1 - e_2)$ the multiplicities are constant on the lines parallel to the roots $(-e_2)$ and $(-e_1 + e_2)$, respectively.

In order to construct a character χ one may

(a) either construct χ_0 applying $1/\Delta$ on every point of the characteristic with the appropriate sign, and then superpose the different contributions to D_0,

(b) or even simpler, use formula (I.20): in this case only four summations occur,

(c) or use the following rule (found empirically but rigorously based on the structure of $1/\Delta$) which is analogous to the one given for A_2. Let $F(j, k)$ be defined as for A_2. $F(j, k)$ is here an octagonal figure. One finds for the character $\chi(p_1, p_2)$ [all $F(j, k)$ must be superposed]:

1. p_2 odd (class 0):

$$\chi(p_1, p_2) = F(p_1 - 1, p_2 - 1) + F(p_1 - 2, p_2 - 1)$$
$$+ \cdots + F(0, p_2 - 1)$$
$$+ F(p_1 - 1, p_2 - 3) + F(p_1 - 2, p_2 - 3)$$
$$+ \cdots + F(0, p_2 - 3)$$
$$\vdots$$
$$+ F(p_1 - 1, 0) + F(p_1 - 3, 0)$$
$$+ \cdots + F(\eta, 0).$$

$\eta = 0$ for odd p_1, $\eta = 1$ for even p_1.

2. p_2 even (class 1):

$$\chi(p_1, p_2) = F(p_1 - 1, p_2 - 1) + F(p_1 - 2, p_2 - 1)$$
$$+ \cdots + F(0, p_2 - 1)$$
$$+ F(p_1 - 1, p_2 - 3) + F(p_1 - 2, p_2 - 3)$$
$$+ \cdots + F(0, p_2 - 3)$$
$$\vdots$$
$$+ F(p_1 - 1, 1) + F(p_1 - 2, 1)$$
$$+ \cdots + F(0, 1).$$

This procedure may be better pursued on the lattice (p_1, p_2) [see Fig. 15(a, b) for two examples].

2. B_l: The General Case

General formulas must be used if $l > 3$, as was done for the groups A_l. First we sum over the roots $-e_1, -e_2, \cdots, -e_l$. This yields a domain ω bounded by the hyperplanes

$$x_i = (-R_0)_i = -\tfrac{1}{2}(2l + 1 - 2j), \quad j = 1, 2, \cdots, l.$$

If $m_{jk}^{\pm}$ now denotes a parameter corresponding to the summation along the root $-e_j \pm e_k$ ($1 \leq j <$
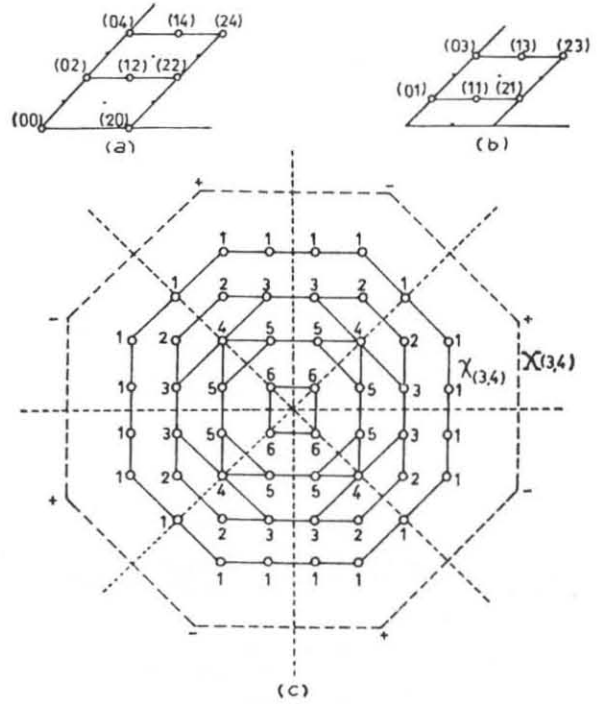

FIG. 15. Construction of WD of B_2: (a) The general rule in the case of $D(3, 5)$; (b) id. for $D(3, 4)$; (c) the WD of $D(3, 4)$; like the WD of A_2, $\chi(p_1, p_2)$ has two sorts of edges: an edge $p_1 \perp P_1$ carrying $p_2 = 4$ weights and an edge $p_2 \perp P_2$ carrying $p_1 = 3$ weights.

$k \leq l$), we get

$$1/\Delta = \sum_{m_{12}^+ = 0}^{\infty} \cdots \sum_{m_{l-1, l}^- = 0}^{\infty} \omega(m_{12}^+ \cdots m_{l-1, 1}^-) \quad (15)$$

with $\omega(0 \cdots 0) = \omega$,

$$\omega(m_{12}^+ \cdots m_{l-1, l}^-) = \sum_Q [Q(m_{12}^+ \cdots m_{l-1, l}^-)], \quad (16a)$$

where the right-hand side means the set of all lattice points Q whose coordinates satisfy the l conditions:

$$Q_i(m_{12}^+ \cdots m_{l-1, l}^-) \equiv x_i \leq (-R_0)_i$$
$$- \sum_{i < k \leq l} (m_{ik}^+ + m_{ik}^-) + \sum_{1 \leq k < i} (m_{ki}^+ - m_{ki}^-); \quad (16b)$$

e.g., for $l = 3$ these conditions read

$$x_1 \leq -\tfrac{5}{2} - m_{12}^+ - m_{13}^+ - m_{12}^- - m_{13}^-,$$
$$x_2 \leq -\tfrac{3}{2} - m_{23}^+ - m_{23}^- + m_{12}^+ - m_{12}^-,$$
$$x_3 \leq -\tfrac{1}{2} + m_{13}^+ + m_{23}^+ - m_{13}^- - m_{23}^-.$$

The condition that χ_0 belongs to D_0 yields, like in case A_l, lower bounds for the values of the x_i's.

C. GROUPS OF TYPE C_l ($\sim Sp(l)$ = SYMPLECTIC GROUPS).

Referred to an orthonormal basis in E_l the roots

are $\pm 2e_i$, $\pm e_i \pm e_k$ (Refs. 2–4), thus

$$R_0 = \tfrac{1}{2} \sum \alpha_i = (l, l-1, \cdots, 1),$$

$$\text{i.e.,} \quad (R_0)_i = l - i + 1. \quad (17)$$

Affine coordinates. The same method, used for the groups A_l and B_l, yields

$$p_i = x_i - x_{i+1}, \quad i = 1, 2, \cdots, l-1.$$

$$p_l = x_l.$$

Whence by inversion

$$x_1 = p_1 + p_2 + \cdots + p_l,$$
$$x_2 = \qquad p_2 + \cdots + p_l, \quad (18)$$
$$\vdots$$
$$x_l = \qquad\qquad\qquad p_l.$$

The *group S* of C_l is the same as the group S of B_l. The *dimension* of an irreducible representation is given by

$$\dim D(K_0) = \frac{\prod\limits_{i<k}(x_i - x_k)(x_i + x_k)\prod\limits_i x_i}{\prod\limits_{i=1}^{l}(2l - 2i + 1)!}, \quad (19)$$

$$K_0 = (x_1 x_2 \cdots x_l).$$

Classes of representations. The center of the representation $Sp(l)$ (whose underlying space is simply connected) is Z_2. The same reasoning made for case B_l shows also here that the points of g^e belong to either of two sublattices:

class 0:

$$\sum_1^l x_i \equiv 0 \bmod 2, \quad \text{i.e.,} \quad \sum_\nu q_{2\nu+1} \equiv 0 \bmod 2;$$

class 1:

$$\sum_1^l x_i \equiv 1 \bmod 2, \quad \text{i.e.,} \quad \sum_\nu q_{2\nu+1} \equiv 1 \bmod 2.$$

A representation $D(K_0(p_1 \cdots p_l))$ belongs to class 0 (1) if its highest weight $L_0(p_1 - 1, \cdots, p_l - 1)$ belongs to class 0(1). These two classes form a group ($\sim Z_2$), class 0 being its unit element. The l fundamental representations ($L_0 = (0 \cdots 0, 1, 0 \cdots 0)$) belong alternately to class 1 and 0.

1. $C_2 = B_2$

The diagrams are the same up to a rotation of $45°$.

2. C_l, The General Case

$1/\Delta$ is constructed in exactly the same way as for B_l [cf., (15), (16)]. The only difference is that

the equations of the surfaces of ω are now

$$x_i = (-R_0)_i = -(l - j + 1).$$

D. GROUPS OF TYPE D_l ($\sim O_{2l}$)

Referred to an orthonormal basis $\{e_i\}$ in E_l the roots are the vectors: $\pm e_i \pm e_k$ (cf., Refs. 2–4). Thus,

$$R_0 = \tfrac{1}{2} \sum \alpha_i = (l-1, l-2, \cdots, 0),$$

$$\text{i.e.,} \quad (R_0)_i = l - j. \quad (20)$$

Affine coordinates.

$$p_i = x_i - x_{i+1}, \quad i = 1, 2, \cdots, l-1,$$

$$p_l = x_{l-1} + x_l,$$

whence by inversion

$$x_1 = p_1 + p_2 + \cdots + p_{l-2} + \tfrac{1}{2}(p_{l-1} + p_l),$$
$$x_2 = \qquad p_2 + \cdots + p_{l-2} + \tfrac{1}{2}(p_{l-1} + p_l), \quad (21)$$
$$\vdots$$
$$x_{l-1} = \qquad\qquad\qquad\qquad \tfrac{1}{2}(p_{l-1} + p_l),$$
$$x_l = \qquad\qquad\qquad\qquad \tfrac{1}{2}(-p_{l-1} + p_l).$$

The *group S* is generated by reflections in the hyperplanes $\vartheta_{ik}^{\pm}(x_i \pm x_k = 0)$, i.e., by transpositions $x_i \leftrightarrow \pm x_k$: it is the group of all permutations of coordinates supplemented by an arbitrary but *even* number of changes of signs, whence its order is $2^{l-1} l!$.

The *dimension* of an irreducible representation is given by the formula

$$\dim D(K_0) = \frac{\prod\limits_{i<k}(x_i - x_k)(x_i + x_k)}{(l-1)! \prod\limits_{i=1}^{l-1}(2l - 2i - 1)!}, \quad (22)$$

where

$$K_0 = (x_1 x_2 \cdots x_l).$$

Classes of representations. Here a distinction must be made between odd and even l.

1. l odd: The center of the representation O_{2l}, whose underlying space is doubly connected, is Z_2. The center of the universal covering group therefore is of order 4. Since the direct product of a spin representation with itself does *not* contain the identity, the group of the classes contains an element of order 4. Thus this group is Z_4. Class 0 is spanned by the roots ($\sum x_i \equiv 0 \bmod 2$). This yields for the 4 classes:

$$\text{class } j: \quad 2\sum_1^l x_i \equiv j \bmod 4, \quad j = 0, 1, 2, 3.$$

Distribution of the fundamental representations:

$$\lambda: \quad 1 \; 2 \; 3 \cdots l - 2 \; l - 1 \; l,$$

$$l = 4k + 1 : \text{class:} \quad 2 \; 0 \; 2 \cdots \quad 2 \quad \quad 3 \quad 1,$$

$$l = 4k + 3 : \text{class:} \quad 2 \; 0 \; 2 \cdots \quad 2 \quad \quad 1 \quad 3.$$

The last ones are the two (complex conjugate) spinor representations (half spinors).

2. l even: Also in this case the center of the universal covering group is of order 4. But since the direct product of either spinor representation with itself contains the identity, the group has only elements of order 2 (besides the identity). Thus it is the Klein 4-group. This group admits no faithful irreducible representations. Therefore also the universal covering group of D_l (l even) whose center the 4-group is, does not have faithful irreducible representations.

The remaining 4 classes of irreducible representations are conveniently labelled e, a, b, c. Class e (0) is spanned by the roots.

The classes are:

$$\text{class } e: x_i \text{ integer,} \quad \sum x_i \equiv 0 \bmod 2,$$

$$\text{class } a: x_i \text{ integer,} \quad \sum x_i \equiv 1 \bmod 2,$$

$$\text{class } b: x_i \text{ half-integer,} \quad \sum x_i \equiv 0 \bmod 2,$$

$$\text{class } c: x_i \text{ half-integer,} \quad \sum x_i \equiv 1 \bmod 2.$$

One verifies the multiplication table of the 4-group:

$$a^2 = b^2 = c^2 = e \quad ab = ba = c + \text{cycl.}$$

Distribution of the fundamental representations into the classes:

$$\lambda: \quad 1 \; 2 \; 3 \cdots l - 2 \; l - 1 \; l;$$

$$l = 4k \quad \quad : \text{class:} \quad a \; e \; a \cdots \quad e \quad \quad c \quad \; b;$$

$$l = 4k + 2 : \text{class:} \quad a \; e \; a \cdots \quad e \quad \quad b \quad \; c.$$

The two last ones are the two spinorial representations (half-spinors).

1. $D_3 = A_3$

If $(q_1 q_2 q_3)$ and $(y_1 y_2 y_3 y_4)$ are affine and Cartesian coordinates, respectively, related by (3), one may pass from A_3 to D_3 with help of the transformation

$$p_1 = q_2, \quad y_1 = \tfrac{1}{2}(x_1 + x_2 + x_3),$$

$$p_2 = q_3, \quad y_2 = \tfrac{1}{2}(x_1 - x_2 - x_3),$$

$$p_3 = q_1, \quad y_3 = \tfrac{1}{2}(-x_1 + x_2 - x_3),$$

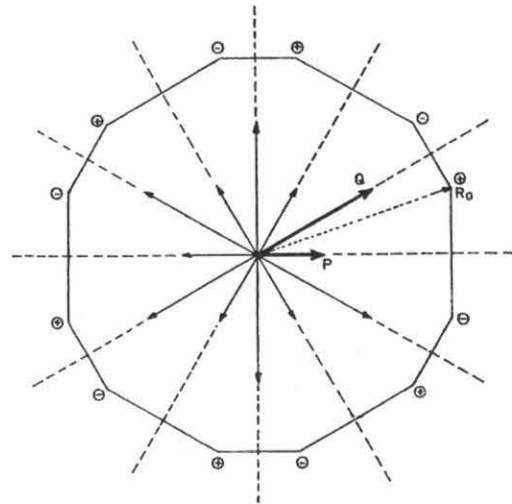$$y_4 = \tfrac{1}{2}(-x_1 - x_2 + x_3).$$

FIG. 16. Root diagram and Δ of G_2.

This shows that the group S is the set of permutations of the $y_i(A_3)$, as well as the permutations of the x_i together with change of an even number of signs (D_3).

2. D_l, The General Case

Here it is less straightforward to calculate $1/\Delta$ in the same simple way as was done for A_l, B_l, C_l; for one cannot construct in the system x_i a pyramid ω with faces parallel to the coordinate hyperplanes. The method can still work but calculations will be more intricate.

E. EXCEPTIONAL GROUPS OF TYPE G_2

Referred to the orthogonal basis $\{e_i\}$ in the space E_3 the roots are the vectors: $e_i - e_k$, $e_i - 2e_j + e_k$ which all belong to the hyperplane $\sum x_i = 0$ (cf., Ref. 2–4). Thus, $R_0 = \frac{1}{2} \sum \alpha_i = (3, -1, -2)$.

Affine coordinates.

$$p = x_2 - x_3, \quad x_1 = +p + 2q,$$

$$x_2 = -q, \quad \quad (23)$$

$$q = -x_2, \quad x_3 = -p - q.$$

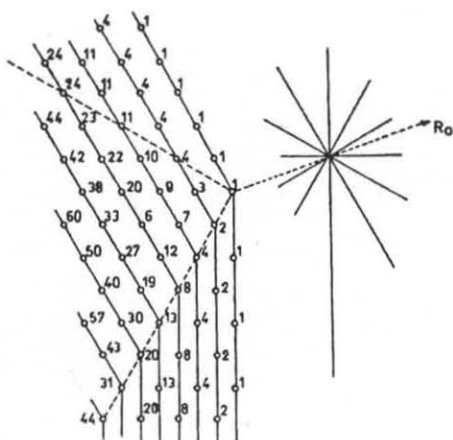The *group S* is d_6 (= dihedral group of order 12, symmetry group of a hexagon).

The root diagram and Δ are shown in Fig. 16.
Dimension:

$$\dim D(pq)$$

$$= \tfrac{1}{120} pq(p + q)(p + 2q)(p + 3q)(2p + 3q). \quad (24)$$

Classes of representations. The center of the universal covering group consists of the unit element alone. Therefore, all representations belong to the same class.

Fig. 17. The $1/\Delta$ of G_2.

The $1/\Delta$ is shown in Fig. 17. It has the same structure as the two diagrams obtained for A_2 and $B_2 = C_2$, since they are all constructed in the same way. Here also as for the other groups of rank 2, the simplest way to construct a character is to use formula (I. 20).

Note added in proof: After this work was completed, we discovered a paper by B. Kostant,[11] where this author solves the multiplicity problem stated in I, but by purely abstract algebraic methods and without consideration to concrete examples. For this reason his paper is complementary to the present one and does not interfere with it.

[11] B. Kostant, Trans. Amer. Math. Soc. 93, 53 (1959).

Kinetic Approach to Thermal Transport Phenomena. I. Classical Theory*

N. Hashitsume† and S. Fujita

Faculté des Sciences de l'Université Libre de Bruxelles, Brussels, Belgium

(Received 21 April 1964; final manuscript received 7 July 1964)

Assuming noninteracting particles, which are scattered by randomly arranged static potentials, such as free electrons in a low-temperature semiconductor containing impurity atoms, the expressions of thermal transport coefficients are derived by means of the kinetic approach established by Prigogine *et al.* In this part, classical law of mechanics is assumed, and the case of zero magnetic field is discussed. It is shown that the thermal transport coefficients thus obtained can be expressed in the terms of the correlation functions of suitable currents in the same way as the Kubo formula for the electric conductivity.

1. INTRODUCTION

FOR the last few decades, the expression for transport coefficients in terms of correlation functions has been proposed or suggested by many authors.[1] The present situation is different in the thermal transport coefficients from the mechanical transport coefficients. For the transport phenomena induced by mechanical disturbances, the formulas for the corresponding transport coefficients have been well established by Kubo[2] and others. However, in the derivation of expressions for transport coefficients in the case of thermal disturbances or internal disturbances, an additional assumption was necessary. For instance, Kubo *et al.*[3] made use of Onsager's assumption concerning the average regression of spontaneous fluctuations, and Nakajima[4]

[1] For example: J. G. Kirkwood, J. Chem. Phys. 14, 180 (1946); M. S. Green: J. Chem. Phys. 20, 1281 (1952).

[2] R. Kubo, J. Phys. Soc. Japan 12, 570 (1957).

[3] R. Kubo, M. Yokota, and S. Nakajima, J. Phys. Soc. Japan 12, 1203 (1957).

[4] S. Nakajima, Progr. Theoret. Phys. (Kyoto) 20, 948 (1958). See, also, R. Kubo, "Some Aspects of the Statistical-Mechanical Theory of Irreversible Processes," in *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York 1959), Vol.1.

FIG. 17. The $1/\Delta$ of G_2.

The $1/\Delta$ is shown in Fig. 17. It has the same structure as the two diagrams obtained for A_2 and $B_2 = C_2$, since they are all constructed in the same way. Here also as for the other groups of rank 2, the simplest way to construct a character is to use formula (I. 20).

Note added in proof: After this work was completed, we discovered a paper by B. Kostant,[11] where this author solves the multiplicity problem stated in I, but by purely abstract algebraic methods and without consideration to concrete examples. For this reason his paper is complementary to the present one and does not interfere with it.

[11] B. Kostant, Trans. Amer. Math. Soc. 93, 53 (1959).

Kinetic Approach to Thermal Transport Phenomena. I. Classical Theory*

N. Hashitsume† AND S. Fujita

Faculté des Sciences de l'Université Libre de Bruxelles, Brussels, Belgium

(Received 21 April 1964; final manuscript received 7 July 1964)

Assuming noninteracting particles, which are scattered by randomly arranged static potentials, such as free electrons in a low-temperature semiconductor containing impurity atoms, the expressions of thermal transport coefficients are derived by means of the kinetic approach established by Prigogine *et al.* In this part, classical law of mechanics is assumed, and the case of zero magnetic field is discussed. It is shown that the thermal transport coefficients thus obtained can be expressed in the terms of the correlation functions of suitable currents in the same way as the Kubo formula for the electric conductivity.

1. INTRODUCTION

FOR the last few decades, the expression for transport coefficients in terms of correlation functions has been proposed or suggested by many authors.[1] The present situation is different in the thermal transport coefficients from the mechanical transport coefficients. For the transport phenomena induced by mechanical disturbances, the formulas for the corresponding transport coefficients have been well established by Kubo[2] and others. However, in the derivation of expressions for transport coefficients in the case of thermal disturbances or internal disturbances, an additional assumption was necessary. For instance, Kubo *et al.*[3] made use of Onsager's assumption concerning the average regression of spontaneous fluctuations, and Nakajima[4]

[1] For example: J. G. Kirkwood, J. Chem. Phys. 14, 180 (1946); M. S. Green: J. Chem. Phys. 20, 1281 (1952).

[2] R. Kubo, J. Phys. Soc. Japan 12, 570 (1957).

[3] R. Kubo, M. Yokota, and S. Nakajima, J. Phys. Soc. Japan 12, 1203 (1957).

[4] S. Nakajima, Progr. Theoret. Phys. (Kyoto) 20, 948 (1958). See also, R. Kubo, "Some Aspects of the Statistical-Mechanical Theory of Irreversible Processes," in *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York 1959), Vol.1.

solved the stationary Liouville equation by making use of the local equilibrium distribution function determined by the entropy-maximum principle as the zeroth approximation. By this reason, the formulas for the thermal coefficients expressed in terms of correlation functions were regarded as questionable. Whether such formulas exist or not is one of the fundamental questions in statistical mechanics. Montroll[5] has made attempts to replace thermal or internal disturbances by some mechanical interference.

Recent development of the kinetic approach has led us to a point, from which we can shed some light on this question. Indeed the nonequilibrium statistical mechanics due to Prigogine and his co-workers[6] has been extended by Severne[7] to include the case of nonuniform systems. In this way we have now general kinetic equations for the one-body phase-space distribution function and we may discuss the transport processes by a method similar to the Enskog–Chapman method used in conjunction with the Boltzmann equation. Using these results Prigogine, Severne, and Résibois[8] have discussed the validity of the correlation function formalism in the case of classical mechanics. In particular, Résibois[9] has given a general proof for the equivalence between the formal expressions derived from the kinetic approach and those given by the correlation function method. According to the idea of the Enskog–Chapman method, the local equilibrium distribution function is determined by the kinetic equation itself. We need neither the introduction of generalized entropy, the functional form of which is not precisely known, nor the assumption of the entropy-maximum principle. We do not make use of the local distribution function as an initial form, starting from which we trace the temporal change of the distribution function for a short time interval as was done in the work by Mori.[10] The local equilibrium function is the zeroth

approximation for the solution of the kinetic equation in the hydrodynamical stage. We do not discuss the validity of the Enskog–Chapman method in this paper.

We assume for the sake of mathematical simplicity the following model: a system of noninteracting particles contained in a sufficiently large volume, which are scattered by randomly arranged obstacles represented by static potentials. The density of obstacles is assumed to be not very high, so that we may neglect the correlation of positions of any two obstacles. This type of model is often used in the simple theory of conductors: for example, free electron gas scattered by impurity atoms. In Part I (this paper), we assume that our system is governed by the classical mechanics, and that there is no magnetic field, for the sake of simplicity. Kasuya and Nakajima[4] already pointed out that the Einstein relation does not hold between the antisymmetric parts of electric conductivity and of diffusion coefficient for a quantum system in a magnetic field. It is important to generalize our theory to the case of a quantized system under a magnetic field. We shall do this in Part II (to be published).

In Sec. 2 we formulate our problem, and in Sec. 3 derive the kinetic equation for the one-particle distribution function averaged over all possible arrangements of scatterers. This averaging would be, in principle, not necessary, but we postpone a formulation without the averaging procedure to a future occasion, because the inclusion of averaging procedure is considered as to make our theory have a simpler appearance. In Sec. 4 we solve the kinetic equation by a method based on the idea of Enskog and Chapman. We limit ourselves to the linear-response theory thereafter. We find in Sec. 5 that the linear relations between currents and gradients hold, and that the transport coefficients appearing in those relations are expressed in terms of correlation functions of currents in the same way as in the Kubo formula for the electric conductivity.

[5] E. W. Montroll, Rend. Scuola Intern. Fis. "Enrico Fermi" 10, 217 (1959).

[6] I. Prigogine, *Non Equilibrium Statistical Mechanics*, (John Wiley & Sons, Inc.–Interscience Publishers, Inc., New York, 1962); R. Balescu, *Statistical Mechanics of Charged Particles*, (John Wiley & Sons, Inc.–Interscience Publishers, Inc., New York, 1964); P. Résibois, "Irreversible Processes in Classical Gases," in *Many-Particle Physics*, edited by E. Meeron (Gordon & Breash, New York, to appear in 1964).

[7] G. Severne, thesis, Brussels, 1963; G. Severne, Physica (to be published).

[8] I. Prigogine and G. Severne, Physics Letters 8, 173 (1963). I. Prigogine, P. Résibois, and G. Severne, Physics Letters, 9, (to be published in 1964); I. Prigogine, P. Résibois, and G. Severne, *Proceedings of the International Seminar on Transport Processes* (Brown University, Providence, Rhode Island, 1964).

[9] P. Résibois (unpublished).

[10] H. Mori, Phys. Rev. 112, 1829 (1958).

2. AVERAGED DISTRIBUTION FUNCTIONS

As was stated in the introduction, we consider a system of noninteracting particles with one-particle Hamiltonian of the form

$$\mathcal{H}(\mathbf{r}, \mathbf{p}) = \mathbf{p}^2/2m + \lambda U(\mathbf{r}). \tag{2.1}$$

m denotes the mass of a particle, $\mathbf{p}$ the momentum, and $\mathbf{r}$ the position. The scattering potential $\lambda U(\mathbf{r})$ is composed of those due to scattering centers located at $\mathbf{R}_\alpha$, $\alpha = 1, 2, \cdots, N$,:

$$U(\mathbf{r}) = \sum_{\alpha=1}^{N_s} u(\mathbf{r} - \mathbf{R}_\alpha). \qquad (2.2)$$

λ is a parameter, which measures the intensity of scattering potential, and is assumed to be sufficiently small to allow the use of perturbation theoretical expansion. Our system is contained in a volume Ω, and the positions of scattering centers are arranged in this volume Ω randomly and uniformly. We assume that the density of scatterers $n_s = N_s/\Omega$ is not very high, so that we may neglect correlations between any two $\mathbf{R}_\alpha$'s. The temporal behavior of our system will be nearly the same for almost all arrangements of scattering centers, if our volume contains a very large number of scatterers, i.e., in the limit

$$N_s \to \infty, \quad \Omega \to \infty \text{ with } n_s = N_s/\Omega \text{ fixed.} \quad (2.3)$$

We may trace in place of individual behavior for a specified arrangement of scatterers the averaged behavior, assuming that $\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_{N_s}$ are a set of independent stochastic variables with uniform probability distribution over the volume Ω. We calculate the average of an observable dynamical variable $g(\mathbf{r}, \mathbf{p}, t; \mathbf{R}_1, \cdots, \mathbf{R}_{N_s})$ by the formula

$$\langle g(\mathbf{r}, \mathbf{p}, t) \rangle = \int_\Omega \frac{d\mathbf{R}_1}{\Omega} \int_\Omega \frac{d\mathbf{R}_2}{\Omega} \cdots$$

$$\times \int_\Omega \frac{d\mathbf{R}_{N_s}}{\Omega} g(\mathbf{r}, \mathbf{p}, t; \mathbf{R}_1, \cdots, \mathbf{R}_{N_s}). \qquad (2.4)$$

Since we neglect the particle–particle interaction, a state of our system at a time t is described by the one-particle phase-space distribution function $f(\mathbf{r}, \mathbf{p}, t)$, which contains $\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_{N_s}$ as parameters. The temporal behavior is governed by the one-particle Liouville equation

$$i\partial_t f = \mathcal{L}f, \qquad (2.5)$$

where $\mathcal{L}$ denotes the Liouville operator defined by

$$\mathcal{L} = \mathcal{L}_0 + \lambda\delta\mathcal{L}, \qquad (2.6)$$

$$\mathcal{L}_0 = -i\mathbf{v}\cdot\nabla, \quad \delta\mathcal{L} = i\{\nabla U(\mathbf{r})\}\cdot\partial. \qquad (2.7)$$

$\mathbf{v} = \mathbf{p}/m$ is the velocity of a particle. We have introduced usual notations

$$\partial_t = \partial/\partial t, \quad \nabla = \partial/\partial\mathbf{r}, \quad \partial = \partial/\partial\mathbf{p}. \qquad (2.8)$$

For the use of perturbational expansion it is convenient to make the Fourier transformation

$$f(\mathbf{r}, \mathbf{p}, t) = \sum_k e^{i\mathbf{k}\cdot\mathbf{r}} f_k(\mathbf{p}, t), \qquad (2.9)$$

$$f_k(\mathbf{p}, t) = \frac{1}{\Omega} \int_\Omega d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r}, \mathbf{p}, t), \qquad (2.10)$$

assuming the periodic boundary condition as usual. Then the Liouville operator can be expressed by the matrix

$$[\mathcal{L}f]_k = \sum_{k'} (\mathbf{k}|\mathcal{L}|\mathbf{k}') f_{k'}, \qquad (2.11)$$

where $[\cdots]_k$ stands for the $\mathbf{k}$ component. In fact, corresponding to Eqs. (2.7), we obtain

$$(\mathbf{k}|\mathcal{L}_0|\mathbf{k}') = \mathbf{k}\cdot\mathbf{v}\,\delta_{k,k'}, \qquad (2.12)$$

$$(\mathbf{k}|\delta\mathcal{L}|\mathbf{k}')$$

$$= \Omega^{-1} \sum_{\alpha=1}^{N_s} e^{-i(k-k')\cdot\mathbf{R}_\alpha} \tilde{u}(\mathbf{k}-\mathbf{k}')(\mathbf{k}-\mathbf{k}')\cdot\partial, \quad (2.13)$$

where

$$\tilde{u}(\mathbf{k}) = \int_\Omega d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r}). \qquad (2.14)$$

We may assume without loss of generality that

$$\tilde{u}(0) = 0. \qquad (2.15)$$

The observable quantities which we want to discuss are the diffusion current and the energy flow. Both occur because of the migration of particles through the scattering centers. The corresponding microscopic variables are the particle velocity and the energy flow vector

$$\mathbf{w} = \mathbf{v}\{\mathbf{p}^2/2m + \lambda U(\mathbf{r})\}, \qquad (2.16)$$

and hence the diffusion current is given by

$$\mathbf{j}(\mathbf{r}, t) = \int d\mathbf{p}\,\mathbf{v}\langle f(\mathbf{r}, \mathbf{p}, t)\rangle \qquad (2.17)$$

and the energy flow by

$$\mathbf{q}(\mathbf{r}, t) = \int d\mathbf{p}\,\mathbf{w}\langle f(\mathbf{r}, \mathbf{p}, t)\rangle. \qquad (2.18)$$

We see that for the calculation of diffusion current we need only the averaged distribution function $\langle f(\mathbf{r}, \mathbf{p}, t)\rangle$. By making the Fourier transform

$$\mathbf{q}_k(t) = \frac{1}{\Omega} \int_\Omega d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{q}(\mathbf{r}, t)$$

$$= \int d\mathbf{p}\,\mathbf{v}\left\{\frac{\mathbf{p}^2}{2m}\langle f_k(\mathbf{p}, t)\rangle + \sum_l \frac{\lambda\tilde{u}(1)}{\Omega} g_{1,k}(\mathbf{p}, t)\right\}, \qquad (2.19)$$

we find that a kind of averaged 'two-body'' distribution function

$$g_{1,k}(\mathbf{p}, t) = \langle \sum_\alpha e^{i1\cdot\mathbf{R}_\alpha} f_{k+1}(\mathbf{p}, t)\rangle \qquad (2.20)$$

is sufficient for the calculation of energy flow.

3. KINETIC EQUATIONS

Let us first construct an equation satisfied by the averaged "one-body" distribution function

$\langle f(\mathbf{r}, \mathbf{p}, t) \rangle$. The formal solution of Eq. (2.5) is given by

$$f_{\mathbf{k}}(\mathbf{p}, t) = [e^{-i\mathcal{L}t} f(0)]_{\mathbf{k}}$$

$$= \frac{-1}{2\pi i} \oint_{\Gamma} dz\, e^{-izt} \sum_{\mathbf{k}'} \left(\mathbf{k} \left| \frac{1}{\mathcal{L} - z} \right| \mathbf{k}' \right) f_{\mathbf{k}'}(\mathbf{p}, 0), \quad (3.1)$$

where the contour Γ should encircle the whole real axis in the complex z plane counterclockwise. We have to allow in general an initial distribution function $f(\mathbf{r}, \mathbf{p}, 0)$ depending on the stochastic variables $\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_{N_s}$. From Eq. (3.1) we obtain the equation

$$(\partial_t + i\mathbf{k}\cdot\mathbf{v}) f_{\mathbf{k}}(\mathbf{p}, t) = \frac{-1}{2\pi i} \oint_{\Gamma} dz\, e^{-izt} i$$

$$\times \sum_{\mathbf{k}'} \left\langle \left(\mathbf{k} \left| \sum_{n=0}^{\infty} \left(-\lambda\delta\mathcal{L} \frac{1}{\mathcal{L}_0 - z} \right)^n \right| \mathbf{k}' \right) f_{\mathbf{k}'}(\mathbf{p}, 0) \right\rangle, \quad (3.2)$$

where we have made use of the well-known formula for the perturbation theoretical expansion of the resolvent operator $(\mathcal{L} - z)^{-1}$.

To see the structure of the sum on the right-hand side of Eq. (3.2), it is convenient to make use of the diagram technique. We write a solid-line segment horizontally corresponding to the unperturbed resolvent operator $(\mathcal{L}_0 - z)^{-1}$, and a vertex corresponding to $-\lambda\delta\mathcal{L}$, from which a dotted line starts and ends at a dot representing the location of a scattering center. As was illustrated by Edwards,[11] the average over our stochastic variables $\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_{N_s}$ means to connect at least two dotted lines at one dot; the contributions from dotted lines ending singly vanish because of our assumption (2.15). By this procedure we can write a diagram uniquely corresponding to a summand in the sum of Eq. (3.2). When the initial distribution function $f(\mathbf{r}, \mathbf{p}, 0)$ depends on $\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_{N_s}$ through the potentials, we write a vertical solid-line segment upward at the right end of the horizontal-line segment at the extreme right, and a set of dotted lines starting from some points on the vertical solid line. We have to connect the free ends of these dotted lines with each other or with those starting from horizontal solid line. A typical diagram is given in Fig. 1. Since the matrix of $\mathcal{L}_0$, Eq. (2.12), is diagonal in the $\mathbf{k}$ representation, the propagator $(\mathcal{L}_0 - z)^{-1}$ transfers the particle from a "state" $\mathbf{k}$ to the same "state" $\mathbf{k}$, and we can attach this $\mathbf{k}$ value to a horizontal solid-line segment in the diagram. The vertex $-\lambda\delta\mathcal{L}$ transfers the particle from a "state" $\mathbf{k}' = \mathbf{k} + \mathbf{l}$ to a different "state" $\mathbf{k}$, $\mathbf{l} \neq 0$. Since the

[11] S. F. Edwards, Phil. Mag. **3**, 1020 (1958).
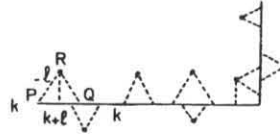
FIG. 1. A typical diagram.

matrix element of $-\lambda\delta\mathcal{L}$, Eq. (2.13), depends only on the difference $\mathbf{k} - \mathbf{k}' = -\mathbf{l}$, we can attach $-\mathbf{l}$ to the dotted line starting from that vertex. The sum of such wave vectors attached to the dotted lines ending at the same dot should vanish.[11]

We can distinguish two classes of diagrams according to whether we can separate a subdiagram by cutting a single horizontal solid-line segment or not. In Fig. 1, we can separate a subdiagram by cutting the segment on the right side of the vertex Q. This separated subdiagram is "connected", i.e., connected through the dotted lines; we cannot further separate the subdiagram by cutting horizontal solid-line segments. The matrix corresponding to such a connected subdiagram should be diagonal in the $\mathbf{k}$ representation. It is convenient to make the partial sum of such separated connected subdiagrams and to introduce the operator

$$i\Psi(\mathbf{k}, z)$$

$$= \sum_{n=1}^{\infty} \left\langle \left(\mathbf{k} \left| \left(-\lambda\delta\mathcal{L} \frac{1}{\mathcal{L}_0 - z} \right)^n (-\lambda\delta\mathcal{L}) \right| \mathbf{k} \right) \right\rangle^{\text{conn}}. \quad (3.3)$$

The sum of the remaining parts of the diagrams has the same structure appearing in the perturbational expansion of Eq. (3.1). The partial sum over diagrams of the second class contributes to the "destruction" part

$$\mathfrak{D}_{\mathbf{k}}[z, f(0)]$$

$$= \sum_{\mathbf{k}'} \left\langle \sum_{n=1}^{\infty} \left(\mathbf{k} \left| \left(-\lambda\delta\mathcal{L} \frac{1}{\mathcal{L}_0 - z} \right)^n \right| \mathbf{k}' \right) f_{\mathbf{k}'}(\mathbf{p}, 0) \right\rangle^{\text{conn}}. \quad (3.4)$$

Since all the diagrams corresponding to the summands of the sum in Eq. (3.2) belong either to the first class or to the second class, we obtain

$$(\partial_t + i\mathbf{k}\cdot\mathbf{v})\langle f_{\mathbf{k}}(\mathbf{p}, t) \rangle$$

$$= -\int_0^t dt'\, G(\mathbf{k}, t')\langle f_{\mathbf{k}}(\mathbf{p}, t - t')\rangle + D_{\mathbf{k}}[t, f(0)] \quad (3.5)$$

or

$$(\partial_t + \mathbf{v}\cdot\nabla)\langle f(\mathbf{r}, \mathbf{p}, t)\rangle = -\int_0^t dt' \int d\mathbf{r}'\, K(\mathbf{r}', t')$$

$$\times \langle f(\mathbf{r} - \mathbf{r}', t - t')\rangle + D[\mathbf{r}, t, f(0)], \quad (3.6)$$

where we have introduced

$$D[\mathbf{r}, t, f(0)] = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} D_{\mathbf{k}}[t, f(0)], \qquad (3.7)$$

$$D_{\mathbf{k}}[t, f(0)] = \frac{-1}{2\pi i} \oint_{\Gamma} dz e^{-izt} i \mathfrak{D}_{\mathbf{k}}[z, f(0)], \qquad (3.8)$$

and

$$K(\mathbf{r}, t) = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} G(\mathbf{k}, t), \qquad (3.9)$$

$$G(\mathbf{k}, t) = \frac{-1}{2\pi i} \oint_{\Gamma} dz e^{-izt} i \Psi(\mathbf{k}, z). \qquad (3.10)$$

The kinetic equation is the equation, which is valid in the kinetic stage, i.e., after the initial mixing is finished. If the scattering potential $u(\mathbf{r})$ is of short-ranged nature, the integral kernel $K(\mathbf{r}, t)$ vanishes for a distance of the order of force range a and for a time of the order of duration of a collision $t_c \sim a/v$. This can be inferred from the diagram in the $\mathbf{r}$ representation, in which the unperturbed propagator corresponds to the free path and the vertex to the position, where the particle feels the scattering force from a center expressed by a dot in the diagram. Then the length of a dotted line should be at most of the order of force range a, and a connected part of the diagram corresponds to a process, which has spatial extension of the order of a multiple of the force range a. If we neglect bound states of the particle around a scatterer, such a process terminates within a time of a multiple of the order of collision time t_c. Since we are interested in the distribution function, which does not differ much from the equilibrium distribution function, particles with small speeds v do not contribute appreciably to the diffusion current and to the energy flow. We may thus assume that t_c is finite. By a similar but a little more complicated argument we can see that the "destruction" term D in Eqs. (3.5) and (3.6) vanishes after a time of the order of t_c, provided that the initial distribution function $f(0)$ is suitably chosen. Thanks to these properties of K and D, we may drop D, and replace the upper limit of the time integral in Eqs. (3.5) and (3.6) by $+\infty$ for a time such that

$$t \gg t_c. \qquad (3.11)$$

We thus arrive at the equation

$$(\partial_t + i\mathbf{k}\cdot\mathbf{v})\langle f_{\mathbf{k}}(\mathbf{p}, t)\rangle$$
$$= -\lim_{\epsilon\to+0} \Psi(\mathbf{k}, i\partial_t + i\epsilon)\langle f_{\mathbf{k}}(\mathbf{p}, t)\rangle, \qquad (3.12)$$

or

$$(\partial_t + \mathbf{v}\cdot\nabla)\langle f(\mathbf{r}, \mathbf{p}, t)\rangle$$
$$= -\lim_{\epsilon\to+0} \Psi(-i\nabla, i\partial_t + i\epsilon)\langle f(\mathbf{r}, \mathbf{p}, t)\rangle. \qquad (3.13)$$

These are the required kinetic equations.

As was mentioned in the preceding section, we need the function $g_{1,\mathbf{k}}(\mathbf{p}, t)$ defined in Eq. (2.20) in the calculation of energy flow. Inserting the formal solution (3.1) into Eq. (2.20), we have the expression

$$g_{1,\mathbf{k}}(\mathbf{p}, t) = \frac{-1}{2\pi i} \oint_{\Gamma} dz e^{-izt}$$

$$\times \sum_{\mathbf{k}'} \left\langle \sum_{\alpha} e^{i\mathbf{l}\cdot\mathbf{R}_{\alpha}} \left(\mathbf{k} + \mathbf{1} \left| \frac{1}{\mathfrak{L}_0 - z} \right.\right.\right.$$

$$\times \left.\left.\sum_{n=0}^{\infty} \left(-\lambda\delta\mathfrak{L} \frac{1}{\mathfrak{L}_0 - z}\right)^n \right| \mathbf{k}'\right) f_{\mathbf{k}'}(\mathbf{p}, 0)\right\rangle. \qquad (3.14)$$

Comparing this with the right-hand side of Eq. (3.2), we see that only the first $-i\lambda\delta\mathfrak{L}$ in Eq. (3.2) is replaced by $\sum_{\alpha} e^{i\mathbf{l}\cdot\mathbf{R}_{\alpha}}$. The structure of diagrams corresponding to the summands in Eq. (3.14) is the same as that for Eq. (3.2), except the first vertex, i.e., the dotted line at extreme left (PR in Fig. 1). Thus we introduce in place of $\Psi(\mathbf{k}, z)$ defined by Eq. (3.3) another operator

$$\mathfrak{C}(\mathbf{1}, \mathbf{k}, z) = \sum_{n=1}^{\infty} \left\langle \sum_{\alpha} e^{i\mathbf{l}\cdot\mathbf{R}_{\alpha}} \right.$$

$$\times \left. \left(\mathbf{k} + \mathbf{1} \left| \left\{ \frac{1}{\mathfrak{L}_0 - z} (-\lambda\delta\mathfrak{L}) \right\}^n \right| \mathbf{k}\right)\right\rangle^{\text{conn}}. \qquad (3.15)$$

Then by the same argument used to derive Eq. (3.12) we arrive at the relation

$$g_{1,\mathbf{k}}(\mathbf{p}, t) = \lim_{\epsilon\to+0} \mathfrak{C}(\mathbf{1}, \mathbf{k}, i\partial_t + i\epsilon)\langle f_{\mathbf{k}}(\mathbf{p}, t)\rangle \qquad (3.16)$$

for $t \gg t_c$.

4. LINEAR-RESPONSE SOLUTION

In the hydrodynamic stage the phase-space distribution function $\langle f(\mathbf{r}, \mathbf{p}, t)\rangle$ becomes a functional of hydrodynamical variables, such as temperature and chemical potential or particle density, and its temporal change is governed by the hydrodynamical equations determined consistently with the kinetic equation. This is the basic idea of the method of Enskog and Chapman.[12] For the present model those equations are the laws of conservation of mass and of energy

$$\partial_t n(\mathbf{r}, t) = -\nabla\cdot\mathbf{j}(\mathbf{r}, t), \qquad (4.1)$$

$$\partial_t E(\mathbf{r}, t) = -\nabla\cdot\mathbf{q}(\mathbf{r}, t), \qquad (4.2)$$

[12] S. Chapman and T. Cowling: *The Mathematical Theory of Non-Uniform Cases*, (Cambridge University Press, Cambridge, 1939).
N. N. Bogolubov, "Problems of a Dynamical Theory in Statistical Physics," (1946). The English translation is given in *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962), Vol. 1.

where $n(\mathbf{r}, t)$ stands for the number density of particles

$$n(\mathbf{r}, t) = \int d\mathbf{p} \langle f(\mathbf{r}, \mathbf{p}, t) \rangle \qquad (4.3)$$

and $E(\mathbf{r}, t)$ for the energy density

$$E(\mathbf{r}, t) = \int d\mathbf{p} \left\langle \left\{ \frac{\mathbf{p}^2}{2m} + \lambda U(\mathbf{r}) \right\} f(\mathbf{r}, \mathbf{p}, t) \right\rangle . \qquad (4.4)$$

Equations (4.1) and (4.2) can be derived from the Liouville equation (2.5). Equation (4.1) can also be derived from the kinetic equation for $\langle f(\mathbf{r}, \mathbf{p}, t) \rangle$, Eq. (3.13). The derivation of Eq. (4.2) from Eqs. (3.13) and (3.16) is more complicated, and is not given here.

Hereafter, we limit ourselves to the linear response theory, and keep only such terms as are linear with respect to the gradients of hydrodynamical variables. When we solve the kinetic equation (3.13) or (3.12), we may put

$$\langle f \rangle = f^0 + f', \qquad (4.5)$$

where the zeroth approximation f^0 is the local equilibrium distribution function and the first-order correction f' is a linear homogeneous function of the gradients. f^0 is determined so as to satisfy the equation

$$\lim_{\epsilon \to +0} \Psi(0, i\epsilon) f^0(\mathbf{r}, \mathbf{p}, t) = 0 \qquad (4.6)$$

and the condition that the currents (2.17) and (2.18) should vanish, if we calculate them by making use of f^0. We construct such an f^0 as follows. First, let us notice that the averaged equilibrium distribution function $\langle f^{eq}(\mathbf{r}, \mathbf{p}) \rangle$ is a stationary solution of Eq. (3.13):

$$i\mathbf{k} \cdot \mathbf{v} \langle f_\mathbf{k}^{eq}(\mathbf{p}) \rangle = -\lim_{\epsilon \to +0} \Psi(\mathbf{k}, i\epsilon) \langle f_\mathbf{k}^{eq}(\mathbf{p}) \rangle, \qquad (4.7)$$

so that $\langle f_0^{eq}(\mathbf{p}) \rangle$ is a solution of Eq. (4.6). We can take any functional of our one-particle Hamiltonian (2.1) as $f^{eq}(\mathbf{r}, \mathbf{p})$, but we select the Maxwell–Boltzmann function

$$f^{eq}(\mathbf{r}, \mathbf{p}) = \exp\left[\{\zeta - \mathcal{K}(\mathbf{r}, \mathbf{p})\}/kT \right], \qquad (4.8)$$

by taking into account the fact that in the real conductor the particle system interacts, for instance, with the crystal lattice vibrations. Since the component with $\mathbf{k} = 0$ is nothing else than the spatial average, $f_0^{eq}(\mathbf{p})$ is essentially the Maxwellian distribution function. The averaged function $\langle f^{eq}(\mathbf{r}, \mathbf{p}) \rangle$ is obviously independent of $\mathbf{r}$, so that

$$\langle f_0^{eq}(\mathbf{p}) \rangle = \langle f^{eq}(\mathbf{r}, \mathbf{p}) \rangle. \qquad (4.9)$$

On the other hand, the operator $\Psi(0, i\epsilon)$ contains neither $\mathbf{r}$ nor ∇. Thus the general form of the local equilibrium function f^0 is given by

$$f^0(\mathbf{r}, \mathbf{p}) = \langle \exp\left[\{\zeta(\mathbf{r}, t) - \mathcal{K}(\mathbf{r}, \mathbf{p})\}/kT(\mathbf{r}, t) \right] \rangle, \quad (4.10)$$

where $\zeta(\mathbf{r}, t)$ and $T(\mathbf{r}, t)$ denote the local chemical potential and the local temperature, respectively. Keeping only terms linear in the deviations from the true equilibrium value

$$\Delta\zeta(\mathbf{r}, t) = \zeta(\mathbf{r}, t) - \zeta, \quad \Delta T(\mathbf{r}, t) = T(\mathbf{r}, t) - T, \quad (4.11)$$

we may use the expression

$$f^0(\mathbf{r}, \mathbf{p}, t) = \langle f^{eq}(\mathbf{r}, \mathbf{p}) \rangle \left\{ 1 + \frac{\Delta\zeta(\mathbf{r}, t)}{kT} \right\}$$
$$+ \left\langle \frac{\mathcal{K}(\mathbf{r}, \mathbf{p}) - \zeta}{kT} f^{eq}(\mathbf{r}, \mathbf{p}) \right\rangle \frac{\Delta T(\mathbf{r}, t)}{T}, \quad (4.12)$$

or

$$f_\mathbf{k}^0(\mathbf{p}, t) = \langle f_0^{eq}(\mathbf{p}) \rangle \left\{ \delta_{\mathbf{k},0} + \frac{[\Delta\zeta(t)]_\mathbf{k}}{kT} \right\}$$
$$+ \left\langle \left[\frac{\mathcal{K}(\mathbf{r}, \mathbf{p}) - \zeta}{kT} f^{eq}(\mathbf{r}, \mathbf{p}) \right]_0 \right\rangle \frac{[\Delta T(t)]_\mathbf{k}}{T}. \quad (4.13)$$

We see that the $\mathbf{k}$ dependence of $f_\mathbf{k}^0$ appears only through the $\mathbf{k}$ dependence of $[\Delta\zeta]_\mathbf{k}$ and $[\Delta T]_\mathbf{k}$, so that $f_\mathbf{k}^0$ has its appreciable value only for small $\mathbf{k}$, provided that $\Delta\zeta(\mathbf{r}, t)$ and $\Delta T(\mathbf{r}, t)$ are assumed to be slowly varying with respect to $\mathbf{r}$. Since the $\mathbf{p}$ dependence of f^0 is Maxwellian, f^0 is actually an even function of $\mathbf{p}$.

Now let us determine the first-order correction f' from Eq. (3.12). According to our assumption that the time dependence of $\langle f \rangle$ is only through the hydrodynamical variables T and ζ, which is assumed to have the same functional dependence on n and T as the equilibrium one, $\partial_t \langle f \rangle$ can be expressed in terms of $\partial_t n$ and $\partial_t T$. $\partial_t n$ and $\partial_t T$ are in turn expressed by the divergence of the currents $\mathbf{j}$ and $\mathbf{q}$ by means of the hydrodynamical equations (4.1) and (4.2). Keeping terms linear in the gradient operator ∇, we may use the zeroth approximation for the currents, which vanish. Thus we may put

$$\partial_t \langle f \rangle = 0 \qquad (4.14)$$

in our kinetic equation (3.12). This means that our method of solution is a quasistatic formalism at least for our model, and that the non-Markoffian effect completely drops, because it comes out from the z dependence of $\Psi(\mathbf{k}, z)$. Remembering Eqs. (4.5) and (4.6), we obtain at once

$$f_{\mathbf{k}}{}'(\mathbf{p},\,t) = \lim_{\epsilon \to +0} \frac{1}{\Psi(0,\,i\epsilon)} \left\{ \mathbf{v} - \lim_{\mathbf{k}' \to 0} i\frac{\partial \Psi(\mathbf{k}',\,i\epsilon)}{\partial \mathbf{k}'} \right\}$$

$$\cdot (-i\mathbf{k})f_{\mathbf{k}}^{0}(\mathbf{p},\,t)$$

$$= \lim_{\epsilon \to +0} \frac{1}{\Psi(0,\,i\epsilon)} \left\{ \mathbf{v} - \lim_{\mathbf{k}' \to 0} i\frac{\partial \Psi(\mathbf{k}',\,i\epsilon)}{\partial \mathbf{k}'} \right\}$$

$$\cdot \left\{ \langle f_0^{\mathrm{eq}}(\mathbf{p}) \rangle \frac{[-\nabla \zeta(t)]_{\mathbf{k}}}{kT} \right.$$

$$+ \left. \left\langle \left[\frac{\mathfrak{K}(\mathbf{r},\,\mathbf{p}) - \zeta}{kT} f^{\mathrm{eq}}(\mathbf{r},\,\mathbf{p}) \right]_0 \right\rangle \right.$$

$$\times [-\nabla \ln T(t)]_{\mathbf{k}} \bigg\}. \qquad (4.15)$$

The $\mathbf{k}$ dependence appears again only through the gradients.

In the same way from Eq. (3.16), by keeping only linear terms with respect to the gradients, we obtain

$$g_{1,\mathbf{k}}(\mathbf{p},\,t) = \lim_{\epsilon \to +0} \left\{ \mathfrak{C}(1,\,0,\,i\epsilon)f_{\mathbf{k}}'(\mathbf{p},\,t) \right.$$

$$+ \lim_{\mathbf{k}' \to 0} i\frac{\partial \mathfrak{C}(1,\,\mathbf{k}',\,i\epsilon)}{\partial \mathbf{k}'} \cdot (-i\mathbf{k})f_{\mathbf{k}}^{0}(\mathbf{p},\,t) \bigg\}. \qquad (4.16)$$

5. THERMAL TRANSPORT COEFFICIENTS

For the actual evaluation of the thermal transport coefficients from the viewpoint of kinetic approach, the expressions (4.15) and (4.16) could be more convenient. However, for our purpose of expressing transport coefficients in terms of correlation functions it is convenient to rewrite them in the following way.

First, let us rewrite the $\mathbf{k}'$ derivative in the first curly brackets of Eq. (4.15). Since the matrix elements of $-\lambda\delta\mathcal{L}$, Eq. (2.13), depend only on the difference of wave vectors, they become independent of $\mathbf{k}'$, if we measure the wave vector appearing in a "intermediate state" of the expression (3.3) from $\mathbf{k}'$. We have only to differentiate the matrix elements of intermediate propagators

$$\left(\mathbf{k}' + 1 \left| \frac{1}{\mathcal{L}_0 - z} \right| \mathbf{k}' + 1' \right) = \frac{\delta_{1',1}}{(\mathbf{k}' + 1)\cdot\mathbf{v} - z}. \qquad (5.1)$$

We see that the differentiation with respect to $\mathbf{k}'$ is equivalent to the replacement

$$\frac{1}{\mathcal{L}_0 - z} \to \frac{1}{\mathcal{L}_0 - z}(-\mathbf{v})\frac{1}{\mathcal{L}_0 - z}, \qquad (5.2)$$

and we obtain

$$-\lim_{\mathbf{k}' \to 0} i\frac{\partial \Psi(\mathbf{k}',\,i\epsilon)}{\partial \mathbf{k}'} = \sum_{n=1}^{\infty} \sum_{m=1}^{n} \left\langle\!\left\langle 0 \left| \left(-\lambda\delta\mathcal{L}\frac{1}{\mathcal{L}_0 - z} \right)^{m} \mathbf{v} \right.\right.\right.$$

$$\times \left. \left. \left\{ \frac{1}{\mathcal{L}_0 - z}(-\lambda\delta\mathcal{L}) \right\}^{n+1-m} \right| 0 \right\rangle\!\right\rangle^{\mathrm{conn}}. \qquad (5.3)$$

In Eq. (4.15), this operator acts on either the Maxwellian $\langle f_0^{\mathrm{eq}}(\mathbf{p}) \rangle$ or the same function multiplied by a linear combination of the unperturbed energy $\mathbf{p}^2/2m$ and a constant, $\langle [(\mathfrak{K} - \zeta)f^{\mathrm{eq}}(\mathbf{r},\,\mathbf{p})]_0 \rangle/kT$. Let us write these operands simply $\varphi(\mathbf{p}^2/2m)$, and introduce

$$F_{\mathbf{k}}(\mathbf{p},\,z)$$

$$= \sum_{n=0}^{\infty} \left(\mathbf{k} \left| \left\{ \frac{1}{\mathcal{L}_0 - z}(-\lambda\delta\mathcal{L}) \right\}^{n} \right| 0 \right)\varphi\!\left(\frac{\mathbf{p}^2}{2m} \right). \qquad (5.4)$$

This quantity is a solution of the equation

$$F_{\mathbf{k}}(\mathbf{p},\,z)$$

$$= \varphi\!\left(\frac{\mathbf{p}^2}{2m} \right)\delta_{\mathbf{k},0} - \sum_{\mathbf{k}'} \frac{(\mathbf{k}\,|\lambda\delta\mathcal{L}|\,\mathbf{k}')}{\mathbf{k}\cdot\mathbf{v} - z}F_{\mathbf{k}'}(\mathbf{p},\,z), \qquad (5.5)$$

and hence $\lim_{\epsilon \to +0} F_{\mathbf{k}}(\mathbf{p},\,i\epsilon)$ is a solution of the stationary Liouville equation

$$\sum_{\mathbf{k}'} (\mathbf{k}\,|\mathcal{L}|\,\mathbf{k}') \lim_{\epsilon \to +0} F_{\mathbf{k}'}(\mathbf{p},\,i\epsilon) = 0 \qquad (5.6)$$

such that it reduces to $\varphi(\mathbf{p}^2/2m)$ as $\lambda \to 0$ and satisfies the normalization condition

$$\int d\mathbf{p} \lim_{\epsilon \to +0} F_0(\mathbf{p},\,i\epsilon) = \int d\mathbf{p}\,\varphi\!\left(\frac{\mathbf{p}^2}{2m} \right). \qquad (5.7)$$

We know that such a solution is the equilibrium function

$$\lim_{\epsilon \to +0} F_{\mathbf{k}}(\mathbf{p},\,i\epsilon)$$

$$= f_{\mathbf{k}}^{\mathrm{eq}}(\mathbf{p}) \quad \text{or} \quad \left[\frac{\mathfrak{K}(\mathbf{r},\,\mathbf{p}) - \zeta}{kT}f^{\mathrm{eq}}(\mathbf{r},\,\mathbf{p}) \right]_{\mathbf{k}}. \qquad (5.8)$$

The right-hand side of Eq. (5.7) is the average of the left-hand side, and to take the Fourier zero-component is equivalent to taking the spatial average. Thus the left-hand side of Eq. (5.7) is either the total number of particles divided by Ω or the total energy (minus the total Gibbs free energy) divided by Ω (and kT). These quantities should be independent of the arrangement of scattering centers and equal to their averaged value. Thus Eq. (5.7) is satisfied. If our system is assumed to be ergodic, our solution (5.8) is unique.

Inserting the results (5.8) and (5.3) into Eq. (4.15), we obtain

$$f_k'(\mathbf{p}, t) = \lim_{\epsilon \to +0} \frac{1}{\Psi(0, i\epsilon)} \sum_{k'} \sum_{n=0}^{\infty} \left\{ \left\langle \left(0 \left| \left(-\lambda \delta \mathcal{L} \frac{1}{\mathcal{L}_0 - z} \right)^n \right| k' \right) v f_k^{eq}(p) \right\rangle^{conn} \cdot \frac{[-\nabla \zeta]_k}{kT} \right.$$

$$\left. + \left\langle \left(0 \left| \left(-\lambda \delta \mathcal{L} \frac{1}{\mathcal{L}_0 - z} \right)^n \right| k' \right) v \left[\frac{\mathcal{H}(\mathbf{r}, \mathbf{p}) - \zeta}{kT} f^{eq}(\mathbf{r}, p) \right]_{k'} \right\rangle^{conn} \cdot [-\nabla \ln T]_k \right\}$$

$$= \lim_{\epsilon \to +0} \sum_{k'} \frac{1}{\Psi(0, i\epsilon)} \sum_{n=0}^{\infty} \left\langle \left(0 \left| \left(-\lambda \delta \mathcal{L} \frac{1}{\mathcal{L}_0 - z} \right)^n \right| k' \right) \left[\left\{ \mathbf{v} \cdot \frac{[-\nabla \zeta]_k}{kT} + \frac{\mathbf{w} - \zeta \mathbf{v}}{kT} \cdot [-\nabla \ln T]_k \right\} f^{eq} \right]_{k'} \right\rangle^{con}$$

$$= \lim_{\epsilon \to +0} \sum_{k'} (-i) \left\langle \left(0 \left| \frac{1}{\mathcal{L} - z} \right| k' \right) \left[\left\{ \mathbf{v} \cdot \frac{[-\nabla \zeta]_k}{kT} + \frac{\mathbf{w} - \zeta \mathbf{v}}{kT} \cdot [-\nabla \ln T]_k \right\} f^{eq} \right]_{k'} \right\rangle$$

$$= \lim_{\epsilon \to +0} \int_0^{\infty} dt' e^{-\epsilon t'} \sum_{k'} \left\langle (0 \left| e^{-i\mathcal{L}t'} \right| k') \left[\left\{ \mathbf{v} \cdot \frac{[-\nabla \zeta]_k}{kT} + \frac{\mathbf{w} - \zeta \mathbf{v}}{kT} \cdot [-\nabla \ln T]_k \right\} f^{eq} \right]_{k'} \right\rangle. \tag{5.9}$$

Here we have made use of the theorem that, if the limit of a product of two functions exists and if the limit of one factor exists, we may take the limit for both factors separately.

In the same way, we can carry out the k' differentiation in Eq. (4.19), and arrive at

$$g_{1,k}(\mathbf{p}, t)$$

$$= \lim_{\epsilon \to +0} \int_0^{\infty} dt' e^{-\epsilon t'} \sum_{k'} \sum_{\alpha} \left\langle e^{i1 \cdot \mathbf{R}\alpha} (1 \left| e^{-i\mathcal{L}t'} \right| k') \right.$$

$$\times \left. \left[\left\{ \mathbf{v} \cdot \frac{[-\nabla \zeta]_k}{kT} + \frac{\mathbf{w} - \zeta \mathbf{v}}{kT} \cdot [-\nabla \ln T]_k \right\} f^{eq} \right]_{k'} \right\rangle. \tag{5.10}$$

By making use of Eqs. (5.9) and (5.10), we can easily write down the expression (2.19) for the Fourier component of energy flow vector in the form

$$\mathbf{q}_k(t) = \lim_{\epsilon \to +0} \int_0^{\infty} dt' e^{-\epsilon t'} \int d\mathbf{p} \frac{1}{\Omega} \int_{\Omega} d\mathbf{r} \left\langle f^{eq} \mathbf{w} e^{-i\mathcal{L}t'} \right.$$

$$\times \left. \left\{ \mathbf{v} \cdot \frac{[-\nabla \zeta]_k}{kT} + \frac{\mathbf{w} - \zeta \mathbf{v}}{kT} \cdot [-\nabla \ln T]_k \right\} \right\rangle \tag{5.11}$$

and for the Fourier component of diffusion current

$$\mathbf{j}_k(t) = \frac{1}{\Omega} \int_{\Omega} d\mathbf{r} e^{-i\mathbf{k} \cdot \mathbf{r}} \mathbf{j}(\mathbf{r}, t)$$

$$= \lim_{\epsilon \to +0} \int_0^{\infty} dt' e^{-\epsilon t'} \int d\mathbf{p} \frac{1}{\Omega} \int_{\Omega} d\mathbf{r} \left\langle f^{eq} \mathbf{v} e^{-i\mathcal{L}t'} \right.$$

$$\times \left. \left\{ \mathbf{v} \cdot \frac{[-\nabla \zeta]_k}{kT} + \frac{\mathbf{w} - \zeta \mathbf{v}}{kT} \cdot [-\nabla \ln T]_k \right\} \right\rangle. \tag{5.12}$$

In these expressions (5.11) and (5.12), the $\mathbf{k}$ dependence on the right-hand sides comes only from the gradients. Therefore, the linear relations

$$\mathbf{j}(\mathbf{r}, t) = -\mathbf{L}_{DD} \cdot \nabla \zeta(\mathbf{r}, t) - \mathbf{L}_{DQ} \cdot \nabla \ln T(\mathbf{r}, t), \tag{5.13}$$

$$\mathbf{q}(\mathbf{r}, t) - \zeta \mathbf{j}(\mathbf{r}, t) = -\mathbf{L}_{QD} \cdot \nabla \zeta(\mathbf{r}, t)$$

$$- \mathbf{L}_{QQ} \cdot \nabla \ln T(\mathbf{r}, t),$$

hold with constant transport coefficients

$$\mathbf{L}_{\mu\nu} = \lim_{\epsilon \to +0} \int_0^{\infty} dt e^{-\epsilon t} \frac{1}{kT\Omega}$$

$$\cdot \iint d\mathbf{p} \, d\mathbf{r} \langle f^{eq} \mathbf{J}_\mu e^{-i\mathcal{L}t} \mathbf{J}_\nu \rangle, \tag{5.14}$$

where μ and ν are either D or Q, and

$$\mathbf{J}_\mu = \begin{cases} \mathbf{v}, & \text{if } \mu = D, \\ \mathbf{w} - \zeta \mathbf{v}, & \text{if } \mu = Q. \end{cases} \tag{5.15}$$

Our result (5.14) has just the classical form similar to the Kubo formula for the electric conductivity, and hence satisfies the Onsager relation.

ACKNOWLEDGMENT

Perturbation Variation Methods for a Quantum Boltzmann Equation

R. F. SNIDER

Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada
(Received 12 June 1964)

For molecules with degenerate internal states, the single-particle distribution function must be replaced by a density matrix, or better, if the translational motion is treated classically, by a Wigner distribution-function density matrix. The modified Boltzmann integro-differential equation for this quantity has been previously derived but so far only limited solutions of the resulting equation have been obtained. Methods are herein discussed which enable the standard methods for the solution of the classical Boltzmann equation to be applied to the solution of this equation. Complications involving commutation properties are resolved.

I. INTRODUCTION

WHEN degenerate internal states are present, it is no longer sufficient to describe a free particle in a gas by a probability density. One must rather use a singlet density matrix. This has the advantage that it correctly accounts for the phase relations between the degenerate states. For a dilute gas the equation of change for the singlet density matrix, or better for the Wigner distribution function-density matrix is a generalization of the classical Boltzmann integro-differential equation.[1] This equation has been derived by Waldmann[2] and independently using a different method by the present author.[3]

Only for the case of spin-$\frac{1}{2}$ particles has a solution of the full matrix equation been considered.[4] In this case, Waldmann[4] found it convenient to express the distribution function-density matrix as a linear combination of the Pauli spin matrices and the identity matrix. For higher spin values this expansion will likely be cumbersome. Moreover, this form of expansion may be contrasted to the usual classical perturbation method that considers firstly a deviation from local equilibrium and secondly an expansion in the independent vectors or tensors of the system. In Sec. III methods are discussed which allow this latter order to be utilized. It is the purpose of Secs. IV and V to adapt the formulation of the linearized Boltzmann equation so that the recently reviewed variational methods

for solving the classical Boltzmann equation[5] may be applied to the quantum case.

The mathematical problems involved are due to the noncommutation of the local equilibrium distribution function-density matrix and the perturbation. For this purpose the first and second derivatives of the exponential function in a noncommutative algebra are discussed in Appendix B. The thermal conductivity of a gas with internal rotational states[6] has recently been considered and is based on the above method.

II. THE MODIFIED BOLTZMANN EQUATION AND ENTROPY PRODUCTION

Quantum mechanically, the probability of a state of one molecule in a gas is given by a one-particle density matrix. However, for the purposes of the kinetic theory of gases it is more appropriate to utilize a position-momentum distribution function for the translational motion much like the classical case and this is most easily accomplished by using a Wigner distribution function-density matrix. This is a distribution function of position $\mathbf{r}$ and linear momentum $\mathbf{p}$ for the translational states of the molecule while retaining its behavior as a density matrix for the internal states of the molecule. Since the intermolecular interactions in a gas are usually short range (the only case treated here) the collisions may be considered (for a dilute gas) to be localized and thus to occur at one point. In this manner the Boltzmann equation that describes the change of the singlet Wigner distribution function-density matrix involves only one position. Thus an alternate interpretation of the Wigner distribution function-density matrix is that it is a density matrix in internal states and a

[1] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), Chap. 7.
[2] L. Waldmann, Z. Naturforsch. **12a**, 661 (1957); **13a**, 609 (1958); *Handbuch der Physik*, edited by S. Flugge (Springer-Verlag, Berlin, 1958), Vol. 12.
[3] R. F. Snider, J. Chem. Phys. **32**, 1051 (1960). Some minor errors of this paper are corrected in Appendix A.
[4] L. Waldmann, Nuovo Cimento **14**, 898 (1959); Z. Naturforsch. **15a**, 19 (1960); **18a**, 86 (1963); H. D. Kupatt, *ibid.* **19a**, 301 (1964).

[5] R. F. Snider, J. Chem. Phys. **41**, 591 (1964). Eq. (47) of this article should read $X_{TI} = -X$.
[6] F. R. McCourt and R. F. Snider (to be published).

diagonal density matrix in momentum space while being *parameterized* by a position $\mathbf{r}$. It is this latter interpretation which will be used exclusively in the succeeding arguments. This will enable one to make use of operator formalism and to designate the trace over internal states and integration[7] over momentum by the one symbol, trace (tr). In particular the singlet Wigner distribution function-density matrix f will be normalized as

$$\operatorname{tr} f(\mathbf{r}, t) = n(\mathbf{r}, t), \tag{1}$$

where $n(\mathbf{r}, t)$ is the local number density at position $\mathbf{r}$ and time t. An immediate generalization of Eq. (1) when the system under consideration is a ν-component mixture may be treated by thinking of f and n as ν-component vectors in a chemical species space besides f having all the properties already mentioned in this paragraph. In this last case, wherever the mass m or relative mass μ appears, these must be considered as diagonal operators in the chemical species space.

On assuming that f is always diagonal in the energy, the Boltzmann equation for f when there are degenerate internal states may be written[2,3]

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} + i(H_{\mathrm{int}} f - f H_{\mathrm{int}})$$

$$= (2\pi)^4 \operatorname{tr}_1 \left[t f f_1 \delta(\mathbf{P} - \mathbf{P}) \delta(E - K) t^\dagger \right.$$

$$\left. + \frac{1}{2\pi i} (t f f_1 - f f_1 t^\dagger) \right], \tag{2}$$

where Planck's constant $\hbar$ is set equal to 1 and Boltzmann statistics have been assumed. Tr_1 refers to a trace over the states of the 2nd molecule, which is labelled by subscript 1 with f_1 also referring to this molecule. The external force $\mathbf{F}$ changes the linear momentum of the molecules while H_{int} is the one particle hamiltonian for internal states which may include external forces acting on the internal states (e.g., a magnetic field when magnetic dipoles are present). The t operator is the *relative coordinate t* operator which may be defined by

$$t = V + \lim_{\epsilon \to 0} V(E - K + i\epsilon)^{-1} t, \tag{3}$$

where V is the intermolecular potential and K is the kinetic energy (including internal energy) all in relative coordinates. The reason for stressing

relative coordinates is so that no δ function for conservation of momentum will appear in t or its quantum mechanical adjoint $t^\dagger$. Another way of expressing this would be to say that t, V, and K are equal to the corresponding quantum mechanical operators in pair-space divided by a δ function for momentum conservation. Lastly, E is the eigenvalue of K on which t (or $t^\dagger$) operates while the total linear momentum operator for a pair of molecules is $\mathbf{P}$ with eigenvalue $\mathbf{P}$ on which t^+ operates. The two δ functions in Eq. (2) thus signify the conservation of linear momentum and energy. On the energy-momentum shell, that is, for matrix elements between states with the same relative energy and total linear momentum, the following operator equations are valid:

$$t^\dagger - t = 2\pi i t^\dagger \delta(E - K) \delta(\mathbf{P} - \mathbf{P}) t$$

$$= 2\pi i t \delta(E - K) \delta(\mathbf{P} - \mathbf{P}) t^\dagger. \tag{4}$$

The standard expression for the entropy density ρs in a one-particle picture is

$$\rho s = -k \operatorname{tr} f \ln \left(\frac{h^3}{e} f \right), \tag{5}$$

in which ρ is the mass density and s is the entropy per unit mass. From the Boltzmann equation, it is then easily shown that the equation of entropy change is

$$\frac{\partial \rho s}{\partial t} = -\frac{\partial}{\partial \mathbf{r}} \cdot (\rho s \mathbf{v}_0 + \mathbf{J}_s) + \sigma, \tag{6}$$

where $\mathbf{v}_0$ is the local mass stream velocity, $\mathbf{J}_s$ is the entropy flux, and σ is the entropy production given by

$$\sigma = -(2\pi)^4 k \operatorname{tr} \ln f \left[t f f_1 \delta(\mathbf{P} - \mathbf{P}) \delta(E - K) t^\dagger \right.$$

$$\left. + \frac{1}{2\pi i} (t f f_1 - f f_1 t^\dagger) \right]. \tag{7}$$

A mathematical property of a density matrix is that it is an Hermitian (self-adjoint), positive definite operator on the Hilbert space $\mathfrak{H}$ of quantum mechanical state vectors. Furthermore it has a finite trace which together with its positive definiteness and self-adjointness implies that it is in the trace-class[8,9] of operators on a Hilbert space $\mathfrak{H}$ which in turn requires that it belong to the Schmidt-class[8] (σc) of operators with finite Schmidt norm

$$(A \mid A)^{\frac{1}{2}} \equiv (\operatorname{tr} A^\dagger A)^{\frac{1}{2}} < \infty, \tag{8}$$

[7] More correctly, this should be the appropriately normalized summation over momentum states or rather over the wave numbers for a box of unit volume, see the normalization of Eq. (1). As is usual in the kinetic theory of dilute gases, this unit volume must be of macroscopic size.

[8] R. Schatten, *Norm Ideals of Completely Continuous Operators*, Vol. 27 of *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (Springer-Verlag, Berlin, 1960), Chaps. 2 and 3.
[9] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).

$(A \, \epsilon(\sigma c))$. In particular, the Schmidt norm of f will be less than $n(\mathbf{r}, t)$ and thus an element of the Hilbert space (σc). Again, this has the consequence that f is a completely continuous operator on $\mathfrak{H}$ and thus has a discrete (real and positive) spectrum with spectral representation[8,9]

$$f = \sum_i p_i P_i' = \sum_i p_i \, |i\rangle\langle i|, \qquad (9)$$

in which p_i is an eigenvalue and $P_i = |i\rangle\langle i|$ is the corresponding projection operator onto the eigenfunction $|i\rangle$ (element of $\mathfrak{H}$) of f. (It is hoped that no confusion will arise between the linear momentum variables $\mathbf{p}$ and eigenvalues p_i nor between the total linear momentum $\mathbf{P}$ and the projection operator P_i since two of these are vectors written in bold face while the others are in italics and no magnitudes of the vectors will appear.) The eigenvalues p_i must satisfy the equations

$$\sum_i p_i = n \quad \text{and} \quad \sum_i p_i^2 \le n^2. \qquad (10)$$

Boltzmann's H theorem is to show that the entropy production is always greater than or equal to zero. This was proven by Waldmann[2] by essentially the following method. Define the superoperator[10] $\mathfrak{I}$ which is a linear operator on the Hilbert space[11] $(\sigma c) \times (\sigma c)_1$ as

$$\mathfrak{I}A \equiv tA \, \delta(E - K)\delta(\mathbf{P} - \mathbf{P})t^\dagger. \qquad (11)$$

Then on symmetrizing the $\ln f$ term in Eq. (7) and making use of Eq. (4) and the scalar product analogous to the metric defined in Eq. (8), the entropy production σ can be written as

$$\sigma = -\tfrac{1}{2}(2\pi)^4 k[(\ln f \mid \mathfrak{I}f) - (1 \mid \mathfrak{I}(f \ln f))], \qquad (12)$$

where 1 is the identity operator, f is the pair density matrix ff_1 (both in pair space) and although not explicitly expressed, the scalar product is again in pair space. With the help of the second identity of Eq. (4) it also follows that

$$(f \mid \mathfrak{I}(1)) = (1 \mid \mathfrak{I}(f)). \qquad (13)$$

The spectral representation of f will be denoted by

$$f = \sum_i \mathfrak{p}_i \mathfrak{P}_i. \qquad (14)$$

The matrix elements of $\mathfrak{I}$ with respect to the pro-

jection operators $\mathfrak{P}_i$ are then

$$\mathfrak{I}_{ii} = (\mathfrak{P}_i \mid \mathfrak{I}(\mathfrak{P}_i))$$
$$= \delta(E_i - E_j)\delta(\mathbf{P}_i - \mathbf{P}_j) \, |\langle i| \, t \, |j\rangle|^2 \ge 0, \qquad (15)$$

which follows from the definition of $\mathfrak{I}$, Eq. (11) and the fact that f is diagonal in the one-particle energy so that f and $\mathfrak{P}_i$ are diagonal in the K representation. In fact, the collision cross section $\sigma_{j \to i}$ for the scattering of a pair of molecules from state j (pair-state) to state i is given by

$$\delta(E_i - E_j)\delta(\mathbf{P}_i - \mathbf{P}_j)\sigma_{j \to i}$$
$$= (2\pi)^4 \mu_i \mu_j (k_i / k_j) \, |\langle i| \, t \, |j\rangle|^2 \, \delta(E_i - E_j)\delta(\mathbf{P}_i - \mathbf{P}_j),$$
$$= (2\pi)^4 \mu_i \mu_j (k_i / k_j) \, \mathfrak{I}_{ii}, \qquad (16)$$

where k_i, k_j are the corresponding wave numbers in relative coordinates. It then follows from Eqs. (12)–(15) that

$$\sigma = -\tfrac{1}{2}(2\pi)^4 k \sum_{ij} \mathfrak{I}_{ij}(\mathfrak{p}_i \ln \mathfrak{p}_i - \mathfrak{p}_i \ln \mathfrak{p}_j + \mathfrak{p}_j - \mathfrak{p}_i), \qquad (17)$$

which is greater than or equal to zero because the quantity in parentheses is negative for all real and positive $\mathfrak{p}_i$ and $\mathfrak{p}_j$. The above proof is much simpler if detailed balance holds ($\mathfrak{I}_{ij} = \mathfrak{I}_{ji}$ or $t^\dagger = t$), in which case the positiveness of σ arises by symmetrization between i and j in the first two terms of Eq. (17) and no use is required of Eq. (13).

A state of local equilibrium is said to exist when the entropy production σ is zero. It is easily seen, Eq. (17), that this occurs only if[12] $\mathfrak{p}_i^{(0)} = \mathfrak{p}_j^{(0)}$ for values of i and j such that $\mathfrak{I}_{ij} \ne 0$. From the product property of $f^{(0)}(f^{(0)} = f^{(0)}f_1^{(0)})$ it follows that $g = -\ln f^{(0)}$ must be a summational invariant,[1] in other words a linear combination of mass, linear momentum, angular momentum, and energy operators.

III. LINEARIZATION OF THE BOLTZMANN EQUATION

The local equilibrium Wigner distribution function-density matrix $f^{(0)}$ satisfies the Boltzmann equation (2) with no left hand side. In the next approximation,[13] the Boltzmann equation becomes

$$f^{(0)}X = -(2\pi)^4 \, \mathrm{tr}_1 \bigg[\mathfrak{I}(f^{(0)}f_1^{(1)} + f^{(1)}f_1^{(0)})$$
$$+ \frac{1}{2\pi i} \{ t(f^{(0)}f_1^{(1)} + f^{(1)}f_1^{(0)}) - (f^{(0)}f_1^{(1)} + f^{(1)}f_1^{(0)})t^\dagger \} \bigg], \qquad (18)$$

---

[10] A superoperator is an operator which transforms an operator on $\mathfrak{H}$ to a new operator on $\mathfrak{H}$. This terminology was introduced by J. A. Crawford, Nuovo Cimento 10, 698 (1958). For a survey of some of the properties of superoperators and bibliography see H. Primas, Mol. Phys. 6, 225 (1963); Rev. Mod. Phys. 35, 710 (1963). See also Appendix B.

[11] $\mathfrak{I}$ operates on operators in pair-space (pairs of molecules) rather than on (σc) so is an operator on the Cartesian product of (σc) with $(\sigma c)_1$, $(\sigma c) \times (\sigma c)_1$.

[12] Superscript (0) stands for local equilibrium value or function.

[13] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, London, 1952), Chap. 7.

where X is obtained from $f^{(0)}$ by

$$X = -\frac{1}{f^{(0)}}\left[\frac{\partial f^{(0)}}{\partial t} + \frac{\mathbf{p}}{m}\cdot\frac{\partial f^{(0)}}{\partial \mathbf{r}}\right.$$

$$\left. + \mathbf{F}\cdot\frac{\partial f^{(0)}}{\partial \mathbf{p}} + i(H_{int}f^{(0)} - f^{(0)}H_{int})\right] \quad (19)$$

with the time derivative $\partial f^{(0)}/\partial t$ eliminated by means of the hydrodynamic equations evaluated in an approximation consistant with $f^{(0)}$. In this way, $f = f^{(0)} + f^{(1)}$ is a steady-state solution of the Boltzmann equation to terms linear in the macroscopic gradients. It is customary[1,13] to write $f^{(1)} = f^{(0)}\phi$, but in this case both $f^{(0)}$ and ϕ are operators on $\mathfrak{H}$ and hence, in general, will not commute. Is this then the best way of writing the perturbation? The following considerations are aimed at answering this question.

There are three obvious forms in which one can write f or $f^{(1)}$, namely:

(a) $f^{(1)} = f^{(0)}\phi,$ (20)

(b) $f^{(1)} = \frac{1}{2}(\phi^{\dagger}f^{(0)} + f^{(0)}\phi),$ (21)

or

(c) $f^{(1)} = f - f^{(0)} = e^{-g+\phi} - e^{-g}.$ (22)

Form (a) looks much the simpler but since f, $f^{(0)}$, and $f^{(1)}$ must be Hermitian, then ϕ must be such that its adjoint $\phi^{\dagger}$ is given by

$$\phi^{\dagger} = f^{(0)}\phi f^{(0)-1} = e^{-g}\phi e^{g} = e^{-\Delta}\phi. \quad (23)$$

In this expression use has been made of the inverse $f^{(0)-1}$ of $f^{(0)}$ which trivially exists from its definition and the superoperator Δ is the commutation superoperator which takes the commutator of g with the operator upon which it acts, i.e.,

$$\Delta\phi = g\phi - \phi g. \quad (24)$$

On the other hand, the Hermitian property of $f^{(1)}$ and $f^{(0)}$ are built into the form (b) and (c); in form (b) by the $\phi^{\dagger}$ term while in form (c), by a Hermitian ϕ.

It is easily shown that ϕ in form (b) may be restricted to being Hermitian by first assuming ϕ to be Hermitian and then showing that it is determined uniquely by $f^{(1)}$. By simple algebra, ϕ is shown to be given by

$$\phi = \frac{2e^{g}}{1 + e^{\Delta}}f^{(1)}, \quad (25)$$

which always has a solution for any $f^{(1)}$ as long as the superoperator $1 + e^{\Delta}$ has an inverse. This is true since g is Hermitian and thus the spectrum of Δ is real.

Since physical observables are Hermitian it might be felt that forms (b) and (c) may have a more physical interpretation because of their hermiticity of ϕ. This should be contrasted with the apparent simplicity of Eq. (20).

The linearized Boltzmann equation (18) may be written as

$$X = \mathfrak{R}\phi \quad (26)$$

with the superoperator $\mathfrak{R}$ given as follows for form (a):

$$\mathfrak{R}\phi = \mathfrak{R}^{(a)}\phi = -(2\pi)^{4}\,\mathrm{tr}_{1}\,f_{1}^{(0)}\left[\Im(\phi + \phi_{1})\right.$$

$$\left. + \frac{1}{2\pi i}\{t(\phi + \phi_{1}) - (\phi + \phi_{1})t^{\dagger}\}\right], \quad (27)$$

in which use has been made of the fact that $\mathfrak{f}^{(0)} = f^{(0)}f_{1}^{(0)}$ commutes with the collision operator t. For the choice (b), Eq. (21), it follows that

$$\mathfrak{R}\phi = \mathfrak{R}^{(b)}\phi = \frac{1}{2}\mathfrak{R}^{(a)}(\phi + e^{\Delta}\phi^{\dagger}), \quad (28)$$

while for the choice (c), linearization of $f^{(1)}$ with respect to ϕ requires an expansion of the exponential which is accomplished in Appendix B. Thus to first order in ϕ,

$$f = e^{-g+\phi} = f^{(0)} + f^{(0)}\frac{e^{\Delta} - 1}{\Delta}\phi + \cdots, \quad (29)$$

and, consequently,

$$\mathfrak{R}\phi = \mathfrak{R}^{(c)}\phi = \mathfrak{R}^{(a)}\frac{e^{\Delta} - 1}{\Delta}\phi. \quad (30)$$

In order to find expressions for the entropy density ρs and entropy production σ, it is necessary to have expressions for both the first and second order terms in the expansion of exp and ln for a noncommutative algebra; these expressions are derived in Appendix B. As usual in perturbing the Boltzmann equation, the expectation value of all of the summational invariants are required to be determined entirely by $f^{(0)}$. Thus $f^{(1)}$ must be orthogonal to all the summational invariants in the scalar product of Eq. (8). These auxiliary conditions on $f^{(1)}$ are useful here in that they simplify the expressions for ρs and σ. Form (c) will be considered first:

$$\rho s^{(c)} = -k\,\mathrm{tr}\,f\ln\left[\frac{h^{3}}{e}f\right]$$

$$= k\,\mathrm{tr}\left[g - \ln\frac{h^{3}}{e} - \phi\right]e^{-g+\phi}$$

$$= k\,\mathrm{tr}\left[g - \ln\frac{h^{3}}{e}\right]e^{-g} - k\,\mathrm{tr}\,\phi e^{-g+\phi}, \quad (31)$$

in which use has been made of the orthogonality of g (a summational invariant) to $f^{(1)}$. With the help of expansion (29) and keeping only terms quadratic in ϕ,

$$\rho s^{(c)} = \rho s^{(0)} - k \operatorname{tr} \phi f^{(0)} \frac{e^{\Delta} - 1}{\Delta} \phi, \qquad (32)$$

where $\rho s^{(0)}$ is the local equilibrium value of the entropy density. The corresponding entropy production is

$$\sigma^{(c)} = k \operatorname{tr} \phi f^{(0)} \mathfrak{R}^{(c)} \phi. \qquad (33)$$

By a similar though more complicated argument (see Appendix B), form (a) yields an expression

$$\rho s^{(a)} = \rho s^{(0)} - \tfrac{1}{2} k \operatorname{tr} f^{(0)} \phi \frac{\Delta}{e^{\Delta} - 1} \phi, \qquad (34)$$

for the entropy density and for the entropy production

$$\sigma^{(a)} = k \operatorname{tr} f^{(0)} \phi \frac{\Delta}{e^{\Delta} - 1} \mathfrak{R}^{(a)} \phi. \qquad (35)$$

Finally, since form (b) with ϕ Hermitian can be written in form (a) with $\phi^{(a)} = \tfrac{1}{2}(1 + e^{\Delta})\phi^{(b)}$, substitution of this into Eqs. (34–5) gives

$$\rho s^{(b)} = \rho s^{(0)} - \tfrac{1}{8} k \operatorname{tr} \phi f^{(0)} \frac{\Delta(e^{\Delta} + e^{-\Delta} + 2)}{1 - e^{-\Delta}} \phi \qquad (36)$$

and

$$\sigma^{(b)} = \tfrac{1}{2} k \operatorname{tr} \phi f^{(0)} \frac{\Delta(1 + e^{-\Delta})}{1 - e^{-\Delta}} \mathfrak{R}^{(b)} \phi \qquad (37)$$

for the corresponding entropy density and entropy production in this case.

The above formulas are quite complicated and the relations between the respective entropy densities and entropy productions will be considered in the next section on variational methods. Blount[14] has also worried about the commutation problems in expressing the entropy production in general quantum mechanical systems but has not given any explicit formulas for this quantity.

It may be noticed that in the classical limit (no commutation problems), the following reductions hold: firstly, $\mathfrak{R}^{(a)} = \mathfrak{R}^{(b)} = \mathfrak{R}^{(c)}$ together with all the entropy production terms being of the form (33); secondly, the entropy density for forms (a) and (b) reduce to

$$\rho s = \rho s^{(0)} - \tfrac{1}{2} k \operatorname{tr} f^{(0)} \phi^2, \qquad (38)$$

whereas Eq. (32)

$$\rho s^{(c)} = \rho s^{(0)} - k \operatorname{tr} f^{(0)} \phi^2. \qquad (39)$$

_____
[14] E. I. Blount, Phys. Rev. **131**, 2354 (1963), Appendix.

The difference between these last two equations is due to the different orthogonality properties of ϕ. In the classical case the perturbation is usually written in form (a) so that Eq. (38) results.

IV. VARIATIONAL METHODS

The variational methods for solving the Boltzmann equation in the classical case have been recently reviewed.[5] Consequently, the development presented herein will be restricted to introducing a suitable notation so that the classical results can be applied to the quantum case.

The essential feature of the classical case is that there is an operator $\mathfrak{R}$ in a (real) Hilbert space which is positive definite but not necessarily self-adjoint (symmetric), the variational methods being then designed to solve Eq. (26) for such an operator. In this section appropriate Hilbert spaces are considered for each form of the perturbation, the positive-definiteness of $\mathfrak{R}$ demonstrated in each case and appropriate definitions made so that Sec. II of Ref. 5 may be taken over as far as possible to the quantum case.

That the Hilbert space is over the real field in the quantum mechanical case follows immediately from the fact that $f^{(1)}$ must be Hermitian. However the appropriate scalar product is different for each of the forms of the perturbation. Classically, the norm of the Hilbert space is related to the expression for the entropy density, so that classically ϕ must satisfy

$$\operatorname{tr} f^{(0)} \phi^2 < \infty, \qquad (40)$$

which by Eqs. (38)–(39) yields a finite entropy density. On the basis of the entropy density, for form (a), the appropriate scalar product is

$$\begin{aligned}
\langle \phi \mid \psi \rangle &= \operatorname{tr} f^{(0)} \phi \frac{\Delta}{e^{\Delta} - 1} \psi, \\
&= \operatorname{tr} (u f^{(0)-\frac{1}{2}} f^{(0)} \phi f^{(0)-\frac{1}{2}})(u f^{(0)-\frac{1}{2}} f^{(0)} \psi f^{(0)-\frac{1}{2}}),
\end{aligned} \qquad (41)$$

which is real since $f^{(0)}\phi$, $f^{(0)}\psi$ and $[\Delta/(e^{\Delta} - 1)]\psi$ are Hermitian by Eqs. (20, 23) and positive definite by the last form of Eq. (41). Here u is the superoperator which is the positive root of the superoperator equation

$$u^2 = \frac{\Delta}{2 \sinh \tfrac{1}{2} \Delta}. \qquad (42)$$

Thus uA is Hermitian if A is and also u has the property that

$$\operatorname{tr} AuB = \operatorname{tr} (uA)B = \operatorname{tr} BuA, \qquad (43)$$

which has been used in the last form of Eq. (41).

In order to apply the variational methods of Ref. 5 to the present case, it is necessary to show that $\mathcal{R}^{(a)}$ is a positive definite superoperator relative to the above scalar product and that the entropy production is given by

$$\sigma = k\langle\phi|\; \mathcal{R}\;|\phi\rangle. \qquad (44)$$

The last statement follows directly from the definition Eq. (35) while the former can be proven by an adaptation of an argument of Waldmann's,[15] namely that with the definition,

$$\Psi = u\mathfrak{f}^{(0)\frac{1}{2}}\phi + u_1\mathfrak{f}^{(0)\frac{1}{2}}\phi_1, \qquad (45)$$

one has

$$\langle\phi|\; \mathcal{R}^{(a)}\;|\phi\rangle = \mathrm{tr}\; (u\mathfrak{f}^{(0)\frac{1}{2}}\phi)\mathfrak{f}^{(0)\frac{1}{2}}\mathcal{R}^{(a)}(u\phi),$$

$$= -\tfrac{1}{2}(2\pi)^4\; \mathrm{tr}\; \mathrm{tr}_1\; \Psi\left[3\Psi + \frac{1}{2\pi i}\;(t\Psi - \Psi t^\dagger),\right]$$

$$= -\tfrac{1}{2}(2\pi)^4\; \mathrm{tr}\; \mathrm{tr}_1\; [\Psi t\Psi - \Psi\Psi t]\delta(E - K)\,\delta(\mathbf{P} - \mathbf{P})t^\dagger,$$

$$= \tfrac{1}{4}(2\pi)^4\; \mathrm{tr}\; \mathrm{tr}_1\; A\,\delta(E - K)\,\delta(\mathbf{P} - \mathbf{P})A^\dagger \geq 0, \qquad (46)$$

where

$$A = \mathfrak{f}^{(0)\frac{1}{4}}(\Psi t - t\Psi)\mathfrak{f}^{(0)-\frac{1}{4}}. \qquad (47)$$

Lastly, it is easily shown that the superoperator $\mathcal{R}^{(a)\mp}$ which is adjoint to $\mathcal{R}^{(a)}$ relative to the scalar product (41) is given by

$$R^{(a)\mp}\phi = -(2\pi)^4\; \mathrm{tr}_1\; \mathfrak{f}_1^{(0)}\left[t^\dagger(\phi + \phi_1)\,\delta(E - K)\,\delta(\mathbf{P} - \mathbf{P})t\right.$$

$$\left. + \frac{1}{2\pi i}\;\{(\phi + \phi_1)t - t^\dagger(\phi + \phi_1)\}\right]. \qquad (48)$$

The above equations are quite complicated to apply because of the superoperator u which appears in the scalar product. A simpler variational method can be obtained from the above by setting $u = 1$, in which case the appropriate scalar product is

$$\langle\phi\;|\;\psi\rangle = \mathrm{tr}\; f^{(0)\frac{1}{2}}\phi f^{(0)\frac{1}{2}}\psi, \qquad (49)$$

and all the previous statements are correct except for the identification of the quadratic form (44) with the entropy production.

Form (b), Eq. (21), for the perturbation is very similar to form (a), in fact, one can be obtained from the other as has been noted previously since $\phi^{(a)} = \tfrac{1}{2}(1 + e^\Delta)\phi^{(b)}$. Making this substitution in Eqs. (41–48), the same results are obtained as in form (a) above with the proper interpretations of entropy density and production as given in Eqs. (36)–(37). A simpler variational method can be based

on the scalar product (49) provided slight modifications are made in the linearized Boltzmann equation (26), namely by defining $X^{(d)}$ and $\mathcal{R}^{(d)}$ by the following

$$X^{(d)} \equiv f^{(0)\frac{1}{2}}Xf^{(0)-\frac{1}{2}} = \mathcal{R}^{(d)}\phi \equiv f^{(0)\frac{1}{2}}(\mathcal{R}^{(b)}\phi)f^{(0)-\frac{1}{2}}$$

$$= \cosh\;(\tfrac{1}{2}\Delta)\mathcal{R}^{(a)}\phi. \qquad (50)$$

With these definitions, it is easily shown by an argument similar to Eq. (46) that $\mathcal{R}^{(d)}$ is positive definite with respect to the scalar product (49) and that its adjoint is given by

$$\mathcal{R}^{(d)\mp}\phi = \frac{e^{\frac{1}{2}\Delta} + e^{-\frac{1}{2}\Delta}}{2}\;\mathcal{R}^{(a)\mp}\phi = \cosh\;(\tfrac{1}{2}\Delta)\mathcal{R}^{(a)\mp}\phi. \qquad (51)$$

Form (c), Eq. (22), is somewhat different because of its normalization. Thus using as a scalar product

$$\langle\phi\;|\;\psi\rangle = \mathrm{tr}\; \phi f^{(0)}\;\frac{e^\Delta - 1}{\Delta}\;\psi$$

$$= \mathrm{tr}\; (u^{-1}f^{(0)\frac{1}{4}}\phi f^{(0)\frac{1}{4}})(u^{-1}f^{(0)\frac{1}{4}}\psi f^{(0)\frac{1}{4}}), \qquad (52)$$

and modifying the linearized Boltzmann equation to be

$$X^{(c)} \equiv \frac{\Delta}{e^\Delta - 1}\;X = \mathcal{R}^{(a)}\phi, \qquad (53)$$

it follows that $\mathcal{R}^{(a)}$ is positive definite relative to the scalar product (52), has the same adjoint as before, Eq. (48), and gives the correct entropy production by Eq. (44), namely reduces to Eq. (33). Again, for simplicity, u can be set equal to 1 in Eq. (52) but then there is no correspondence with the entropy density or entropy production.

With the respective scalar products and superoperators, the variational methods discussed in Sec. III of Ref. 5 may be taken over to solve the linearized Boltzmann equation.

V. TIME-REVERSAL AND SPACE-INVERSION SYMMETRY

The variational methods referred to in the previous section make use of the relation between the Boltzmann superoperator $\mathcal{R}$ and its adjoint $\mathcal{R}^\mp$. In the classical case[5] the relation between these two (super) operators is physically related to time reversal and space inversion. That this is also true in the quantum mechanical case will be shown in this section.

Quantum mechanically, the time-reversal operator θ on state vectors in the Hilbert space $\mathfrak{H}$ is antiunitary.[16] Consequently, the time-reversal

[15] L. Waldmann, Z. Naturforsch. 15a, 19 (1960), sec. 2.

[16] E. P. Wigner, Group Theory (Academic Press, Inc., New York, 1959).

superoperator Θ defined by

$$\Theta A = \theta A \theta^{-1} \qquad (54)$$

is also antiunitary. In the absence of a magnetic field $\mathbf{H}$ the Hamiltonians K and V are time-reversal invariant, that is

$$\Theta K = K, \qquad \Theta V = V, \qquad (55)$$

so that the t operator Eq. (3) satisfies

$$\Theta t = t^{\dagger}. \qquad (56)$$

Thus for example, the collision cross section, Eq. (16), satisfies

$$k_i^2 \sigma_{j \to i} = k_{-i}^2 \sigma_{-i \to -j} \qquad (57)$$

in which $-i$ is the state obtained by time-reversing state i.

In analogy with the classical case $f^{(0)}$ is not time-reversal invariant if there is a local stream velocity or a local angular momentum density but this difficulty can be overcome by defining the operator Θ' which acts on the parameters in $f^{(0)}$ to reverse the stream velocity and angular momentum density parameters. It is convenient to require Θ' to also change the sign of t and $\mathbf{H}$. The combined operator $\Theta_T = \Theta \Theta'$ then will leave $f^{(0)}$ invariant, i.e., $\Theta_T f^{(0)} = f^{(0)}$. It then follows from Eqs. (48) and (56) that

$$\mathcal{R}_T^{(a)} \equiv \Theta_T \mathcal{R}^{(a)} \Theta_T^{-1} = \mathcal{R}^{(a)*}, \qquad (58)$$

which is exactly Eq. (40) of Ref. 5.

Space-inversion symmetry is also the same quantum mechanically as classically. The space-inversion operator Π on state vectors in $\mathfrak{H}$ is unitary as is the corresponding superoperator Π given by

$$\Pi A = \Pi A \Pi. \qquad (59)$$

Thus if the Hamiltonians K and V are space-inversion invariant, then so is t, i.e.,

$$\Pi K = K, \qquad \Pi V = V, \qquad \Pi t = t, \qquad (60)$$

so that the cross section, Eq. (16) satisfies

$$\sigma_{j \to i}(\mathbf{p}_j \to \mathbf{p}_i) = \sigma_{j \to i}(-\mathbf{p}_j \to -\mathbf{p}_i). \qquad (61)$$

The inversion of the relative linear momenta of the initial and final states has been explicitly indicated. However, if the states have a spatial configuration as is the case for optical isomers, then the spatial configuration must also be reversed.

Since $f^{(0)}$ is not invariant to space inversion, one again defines an operator Π' which inverts the stream velocity $\mathbf{v}_0$ so that $f^{(0)}$ is now space-

inversion invariant. It must be remembered that $f^{(0)}$ is *parameterized* by the position $\mathbf{r}$ but that the origin of the space inversion is to be considered the position $\mathbf{r}$ so that $\mathbf{r}$ is unchanged by the above inversion process. The combined space-inversion operator $\Pi_I = \Pi' \Pi$ will leave $f^{(0)}$ invariant so that

$$\mathcal{R}_I^{(a)} \equiv \Pi_I \mathcal{R}^{(a)} \Pi_I = \mathcal{R}^{(a)}. \qquad (62)$$

Thus for form (a) and the scalar product (41), all the statements of Secs. III and IV of Ref. 5 may be applied to the quantum case. This also holds for the other two forms and their respective scalar products. For the simpler scalar products, everything will again be applicable except for the identification of entropy density and entropy production as is discussed in Sec. IV.

ACKNOWLEDGMENTS

APPENDIX A

Some errors in Ref. 3 are corrected:

Eq. (33): replace μ^2 by $\mu_a \mu_b$;
Eq. (37): replace μ_a^2 by $\mu_a \mu_c$;
Eqs. (44), (51), (58), (59): replace i by $-i$;
Eq. (69): insert $1/\mu_{LL_1}$ inside the integral.

APPENDIX B

In linearizing the Boltzmann equation according to form (c) an expansion of the exponential $e^{-\theta + \phi}$ is required. Similarly, in evaluating the entropy density and entropy production by forms (a) and (b) the linear and quadratic terms in ϕ of $\ln [f^{(0)}(1 + \phi)]$ have been used. These formulas will be developed here for a noncommutative algebra.

The classic Baker–Hausdorff formula for the expansion of $z = \ln (e^x e^y)$ in powers of y has recently been discussed by several authors.[17] This expansion may be written as follows:

$$z = x + z_1 + z_2 + \cdots, \qquad (B1)$$

where

$$z_1 = \frac{\Delta_x}{1 - e^{-\Delta_x}} y \qquad (B2)$$

---

[17] J. Wei, J. Math. Phys. **4**, 575, 1337 (1963); G. Weiss and A. A. Maradudin, *ibid.* **3**, 771 (1962).

and

$$z_m = \frac{1}{m}\left(\frac{\partial}{\partial x} z_{m-1}\right) z_1. \tag{B3}$$

In these formulas, Δ_z is the commutator superoperator

$$\Delta_z y = xy - yx \tag{B4}$$

and the derivative is the Fréchet derivative[18] (which is again a superoperator). The first term Eq. (B2) is in convenient form for application, though it was found difficult to find an explicit expression for z_2 in terms of x and y from Eq. (B3). Consequently, a different formulation of this result will be obtained.

Let $z(\lambda)$ be considered as an analytic function of the complex variable λ; then by a Taylor series expansion of the exponential and identification of the results, the following derivative is obtained:

$$\frac{\partial}{\partial \lambda} e^{z(\lambda)} = \sum_{n=0}^{\infty} \sum_{m=0}^{n-1} \frac{1}{n!} z^m \frac{\partial z}{\partial \lambda} z^{n-m-1},$$

$$= \sum_{n=0}^{\infty} \sum_{p=0}^{\infty} \frac{m!p!}{(m+p+1)!} \frac{z^m}{m!} \frac{\partial z}{\partial \lambda} \frac{z^p}{p!},$$

$$= \sum_{m,p} \int_0^1 (1-\alpha)^m \alpha^p \, d\alpha \, \frac{z^m}{m!} \frac{\partial z}{\partial \lambda} \frac{z^p}{p!},$$

$$= \int_0^1 e^{(1-\alpha)z} \frac{\partial z}{\partial \lambda} e^{\alpha z} \, d\alpha, \tag{B5}$$

wherein $p = n - m - 1$ and the integral formula for the beta function $m!p!/(m+p+1)!$ has been used.[19] The last integral may be evaluated using the commutator superoperator Δ_z defined analogous to Δ_z in Eq. (B4). Thus,

$$\int_0^1 e^{(1-\alpha)z} \frac{\partial z}{\partial \lambda} e^{\alpha z} \, d\alpha = e^z \int_0^1 e^{-\alpha \Delta_z} \, d\alpha \, \frac{\partial z}{\partial \lambda}$$

$$= \frac{e^z(1 - e^{-\Delta_z})}{\Delta_z} \frac{\partial z}{\partial \lambda}. \tag{B6}$$

For $z(\lambda) = \ln(e^x e^{\lambda y})$, Eq. (B6) after inversion of the superoperator and setting $\lambda = 0$ gives the first Gâteaux differential[18] of $z(\lambda)$ which is just Eq. (B2).

[18] E. Hille and R. S. Phillips, *Functional Analysis and Semi-groups* (American Mathematical Society, Providence, Rhode Island, 1957).
[19] See, for example, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 1, Chap. 1.

The second Gâteaux differential of $z(\lambda)$ is obtained in the same way, namely by differentiating Eq. (B5) with respect to λ, setting $z(\lambda) = \ln e^x e^{\lambda y}$, taking the limit $\lambda \to 0$, and solving for $\partial^2 z/\partial \lambda^2$ using superoperators. Thus,

$$\frac{\partial^2}{\partial \lambda^2} e^{z(\lambda)} = \int_0^1 \int_0^1 \left[e^{(1-\alpha)z} \frac{\partial^2 z}{\partial \lambda^2} e^{\alpha z} \right.$$

$$+ e^{(1-\alpha)(1-\beta)z}(1-\alpha) \frac{\partial z}{\partial \lambda} e^{\beta(1-\alpha)z} \frac{\partial z}{\partial \lambda} e^{\alpha z}$$

$$\left. + e^{(1-\alpha)z} \frac{\partial z}{\partial \lambda} e^{(1-\beta)\alpha z} \alpha \frac{\partial z}{\partial \lambda} e^{\beta \alpha z} \right] d\alpha \, d\beta,$$

$$= e^z \frac{1 - e^{-\Delta_z}}{\Delta_z} \frac{\partial^2 z}{\partial \lambda^2} + e^z \frac{1 - e^{-\Delta_z}}{\Delta_z} \left[\left(\frac{1}{\Delta_z} \frac{\partial z}{\partial \lambda} \right) \frac{\partial z}{\partial \lambda} \right]$$

$$- \left[\frac{1}{\Delta_z} \frac{\partial z}{\partial \lambda} \right] e^z \frac{1 - e^{-\Delta_z}}{\Delta_z} \frac{\partial z}{\partial \lambda} + e^z \left(\frac{1 - e^{-\Delta_z}}{\Delta_z} \frac{\partial z}{\partial \lambda} \right) \frac{1}{\Delta_z} \frac{\partial z}{\partial \lambda}$$

$$- e^z \frac{1 - e^{-\Delta_z}}{\Delta_z} \left[\frac{\partial z}{\partial \lambda} \frac{1}{\Delta_z} \frac{\partial z}{\partial \lambda} \right], \tag{B7}$$

so that for $\lambda = 0$ and $z(\lambda) = \ln e^x e^{\lambda y}$ there results

$$2z_2 = \left. \frac{\partial^2 z}{\partial \lambda^2} \right|_{\lambda=0}$$

$$= \frac{\Delta_z}{1 - e^{-\Delta_z}} y^2 - \left(\frac{1}{1 - e^{-\Delta_z}} y \right) \frac{\Delta_z}{1 - e^{-\Delta_z}} y$$

$$+ \frac{\Delta_z}{1 - e^{-\Delta_z}} \left[\left(\frac{1}{e^{\Delta_z} - 1} y \right) y - y \frac{1}{1 - e^{-\Delta_z}} y \right]$$

$$+ \left(\frac{\Delta_z}{1 - e^{-\Delta_z}} y \right) \frac{1}{1 - e^{-\Delta_z}} y. \tag{B8}$$

To apply these equations to the main text, one sees first of all that the second term in Eq. (29) is just Eq. (B6) in the limit $\lambda = 0$ with the proper replacement of x by $-g$ and $\partial z/\partial \lambda$ by ϕ. Equations (34)–(35) are derived from the expansion (B1) for $z(\lambda) = \ln e^x e^{\lambda y}$ with the replacement of x by $-g$ and y by $\phi - \frac{1}{2}\phi^2$, the latter being necessary if e^y is to be equal to $1 + \phi$ to second order in ϕ. A small amount of algebraic manipulation and the identity

$$\phi f^{(0)} = f^{(0)} e^\Delta \phi \tag{B9}$$

[with Δ as in Eq. (24)] then yields the required results.

Observables in Relativistic Quantum Mechanics*

W. C. DAVIDON

Haverford College, Haverford, Pennsylvania

AND

H. EKSTEIN

Argonne National Laboratory, Argonne, Illinois

(Received 3 March 1964; final manuscript received 2 July 1964)

The conventional statement of statistical determinism is that "the expectation values of all (Heisenberg) observables are determined by the expectation values of the observables *at one time.*" This requires that a full algebra of self-adjoint operators be in one-to-one correspondence with measurement procedures performed *at one time.* For instance, it requires that if two noncommuting observables p and q are defined at $t = 0$, there should exist a measurement procedure at $t = 0$ corresponding to $p + q$. No such procedure is known. The contrast between the positive assertion of the existence of certain laboratory procedures and the inability to describe them constitutes perhaps the weakest point of quantum mechanics. However, the conventional statement of statistical causality is shown to be untenable in a relativistic theory. This paper proposes a weaker form of causality which (1) uses measurements made within a truncated light cone rather than at one time for predictive purposes, and (2) which involves only strictly localized states, i.e., states which are vacuumlike outside a finite volume. Failure of the conventional causality statement implies that the set of quasilocal observables is not necessarily linear, i.e., if A and B are in a set, $A + B$ is not necessarily in it. This remark may open the way to a systematic inquiry into the problems of associating laboratory procedures to self-adjoint operators.

I. INTRODUCTION

THE fact that quantum mechanics is an incomplete theory is generally acknowledged and deplored by those who are interested in fundamental problems. Quantum mechanics asserts[1,2] that measurement procedures at one time are in one-to-one correspondence with an algebra of self-adjoint operators on Hilbert space, but it does not specify the procedures. As an example, assume that procedures for measuring the position q and the momentum p at the time $t = 0$ are known. Quantum mechanics asserts that there exist procedures performed at $t = 0$ which correspond to $p + q$. The assertion does not mean only that it is possible to design a procedure by which the sum of the expectation values $\langle p \rangle_\Psi + \langle q \rangle_\Psi$ is obtained for every state Ψ. If this were the whole assertion, the procedure could be trivially specified as an arithmetic addition of numbers obtained from many individual measurements of p and of q on samples of the ensemble Ψ. The assertion is that the same procedure should also yield the expectation values of $(p + q)^2$ and of other real-valued functions of the operator $p + q$. For this purpose, results of the measurement of $(p + q)$ on individual samples may be squared and averaged. One could, for instance, measure q and

then p in rapid succession and consider the sum of the observed values as the value of $(p + q)$. However, the more accurately q is measured, the wider the statistical dispersion of subsequent values of p, until, in the limit, the measured value of p becomes entirely independent of the original state.

Also, two measurements of $(p + q)$ performed in rapid succession should give the same or almost the same value. These requirements, imposed by the theory on the apparatus, cannot be met by any known device. On the other hand, the sum of two commuting observables A and B may be defined simply as the arithmetic addition operation on the two procedures. Operationally, the test for commutativity is to determine if the expectation value $\langle A + B \rangle$ is independent of the order in which the measurements are performed.

Why is it necessary to maintain the stringent postulate in the face of obvious difficulties? What would the theory lose in predictive power if the postulate were dropped or weakened? It is shown that the usual assumption about the correspondence between operators and procedures is indispensable for the commonly accepted form of statistical causality (or determinism).[3] The assertion of causality is

(A) "The expectation values of the observables measured at one time (on a spacelike hypersurface) de-

---

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, England 1947), p. 26.

[2] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Dover Publications, Inc., New York, 1943), p. 167.

[3] No precise distinction between the two words seems to enjoy universal acceptance.

termine the expectation values of all observables at later times."

This seems to be a minimal substitute for classical causality, and it is understandable that one goes far to save it.

We shall see that in the context of the general principles of quantum mechanics the causality (A) requires, in effect, that the observables on one spacelike hypersurface $t = 0$ form a linear set. In particular, for a complete set of dynamical variables $p(0)$, $q(0)$, the linear combination $p(0) + q(0)$ must also be an observable at $t = 0$. The dilemma apparently is this: either we must find procedures for measuring such quantities as $(p + q)$ at $t = 0$, or we must abandon what seems to be a reasonably minimal form of causality. Yet, as we shall show, the statement (A) conflicts with the combination of (1) the relativistic principle of signal propagation with a finite velocity and (2) well-established non-classical effects such as measurability of parity. Therefore, one must accept a weaker form of statistical causality which does not refer to such all-inclusive categories as "observables at time t anywhere in the universe" but, more modestly and realistically, to quasilocal observables [Sec. IV, Statement (C)].

The weaker form of causality does not demand that observables at one time form a linear set, and hence relieves us of the burden of trying to design extraordinary experimental procedures to satisfy the requirements of a theory. This result opens the way to a systematic investigation of the relation between laboratory procedures and self-adjoint operators on Hilbert space.

II. CONSEQUENCES OF CONVENTIONAL CAUSALITY

A measurement procedure in a space–time volume V or spacelike hyperplane S is a set of instructions and apparatus for an operation carried out within V or S; that is, all interaction between the apparatus and the system takes place within V or S.

The assumption that such procedures exist clearly requires some extrapolative idealization. If a measuring instrument begins to interact with the system at the time t in a space volume v, the instrument must have been brought there previously, thus disturbing the system. To justify this assumption, it must be asserted that the interaction previous to t can be minimized to any desired degree.

In Secs. II–IV we are not interested in correlation measurements, i.e., subsequent measurements on the same sample of an ensemble. We may assume that each sample of the ensemble is destroyed or discarded after the measurement. However, we do not assume that *each* measurement is instantaneous, and we classify observables by the time interval of measurement, i.e., the interval beginning with the interaction between apparatus and object and ending at the moment when the necessary information is stored.

Many measurement procedures are equivalent in that they give identical results for all ensembles. An equivalence class of measurement procedures in V (or S) are called an observable in V (or S). Different observables may have identical expectation values for all ensembles, e.g., the momentum of a free particle, measured at different times. This defines an equivalence class of observables which, following Dirac, we call a "dynamical variable." Self-adjoint operators on Hilbert space may be considered as images of observables in a many-to-one mapping, or as images of dynamical variables in a one-to-one mapping.

We follow the conventional assumption to the extent that the set E of dynamical variables is assumed to form a normed linear space so that the set of all observables is closed under addition. For example, if $p(0)$ and $q(0)$ are observables, then $q(0) + p(0)$ may not be an observable at $t = 0$; but it is *an* observable. This is a much weaker assumption than isomorphism between dynamical variables and observables at one time. For instance, for a free particle [$q(t) = q(0) + pt$], the Heisenberg operator q measured at the time $t = 1$ is equal to $p(0) + q(0)$. In other words, the equivalence class of the dynamical variable $q(0) + p(0)$ may not include an observable at $t = 0$, but it does include one at $t = 1$. In the remainder of this section, we consider only observables at one time (or on a spacelike hyperplane).

Let G_t be the set of dynamical variables observable at t, i.e., G_t consists of those dynamical variables whose equivalence class includes an observable at t. To an ensemble ρ, one associates expectation values of dynamical variables $\langle A \rangle_\rho$ $(A \in E)$. They form a positive linear functional $f_\rho(A)$ on the dynamical variables. According to Statement (A), the expectation values of the particular dynamical variables B $(B \in G_t)$ determine all expectation values. In other words, if two ensembles have identical expectation values for all dynamical variables $B \in G_t$, then they also have identical expectation values for all dynamical variables. That is,

$$f_\rho(B) = f_{\rho'}(B) \qquad (B \in G_t) \qquad (2.1)$$
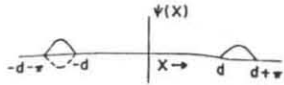
implies

FIG. 1. Wavefunction of a one-dimensional particle with positive (full line) and negative (dotted line) reflection parity.

$$f_\rho(A) = f_{\rho'}(A) \qquad (A \in E). \qquad (2.2)$$

The vanishing of a linear functional $f_\rho - f_{\rho'}$ in G_t implies its vanishing on the whole set E; such a subset G_t is called *total*. *Note added in proof*: It is assumed that every linear functional can be represented as the difference between two positive linear functionals. This assumption is justified only for certain topologies of the space of dynamical variables.

We need two definitions in order to state the consequences of this postulate.

Definition 1. A subset G is dense in E if, for each $y \in E$, there exists a Cauchy sequence of elements $X_n \in G$ so that $X_n \to y$.

Definition 2. A subset G is fundamental if the set of all linear combinations of elements of G is dense in E.

The condition for G_t being total [i.e., the condition for the postulate (A)] is then given by the theorem:[4]

Theorem. A subset G is total if and only if it is fundamental.

For the purpose of designing measurement procedures, we can go farther. The knowledge of the expectation values of observables A_i is equivalent to the knowledge of the expectation values of all linear combinations of the A_i. Also, there is no physical distinction between a procedure for obtaining a mean value and one which allows approximating it to any desired degree. Hence:

Physically, a fundamental set of dynamical variables is equivalent to the whole set.

To summarize, the postulate (A), together with the general principles of quantum mechanics, requires a one-to-one correspondence between the set of observables at one time and the set of all dynamical variables.

In an attempt to avoid the unpleasant consequences, one might weaken the statement of causality in an obvious way by requiring knowledge of expectation values of observables in a spacelike *slab* of finite thickness in the time dimension [Statement (B)].[5]

[4] S. Banach, *Théorie des Opérations Linéaires* (Hafner Publishing Company, New York, 1932), p. 58.
[5] R. Haag and B. Schroer, J. Math. Phys. **3**, 249 (1963).

While this weakening constitutes a further departure from the idea that the present determines the future, it does not seem unreasonable as long as the thickness of the slab is small. This idea will not be pursued in the present paper since the next section will show that neither Statement (A) nor Statement (B) is tenable in a relativistic theory.

III. THE FAILURE OF STATISTICAL CAUSALITY IN RELATIVITY

The finite velocity of signal propagation imposes severe restrictions on the possibilities of the measuring apparatus. Since the measuring instruments are macroscopic, it is sufficient to apply the basic principles of classical relativity to their operation.

Consider a space volume v at a time t, and let S_v be the set of all observables that can be measured in v at time t. If v' is another nonintersecting volume, a measurement in v cannot influence one in v' at the time t. That is, any instantaneous measurement by an instrument which occupies both v and v' supplies no more information than could be obtained by simultaneous separate measurements in v and in v'. The same conclusion obviously holds if V and V' are space–time volumes which are spacelike with respect to each other, i.e., if V includes only points that are separated by spacelike intervals from all points of V'.

Consider a state that is vacuumlike everywhere except in two congruent disjoint volumes v and v', i.e., the expectation values of all quasilocal observables at $t = 0$ outside of v and v' are those of the vacuum state. The remaining information is supplied by quasilocal observables in the space–time volumes V and V' which include v and v'. We may assume V and V' to be spacelike, with respect to each other. As an example, consider a one-dimensional one-particle system with two states described by the wavefunctions $\psi(x)$ (Fig. 1):

$$\psi = \begin{cases} 0 & \text{except for } d < |x| < d + \pi, \\ \tfrac{1}{2}(2)^{\frac{1}{2}} \sin(x-d) & \text{for } d < x < d + \pi, \\ \pm\tfrac{1}{2}(2)^{\frac{1}{2}} \sin(-x+d) & \text{for } -d-\pi < x < -d. \end{cases}$$
$$(3.1)$$

Instantaneous observations in the two segments $d < |x| < d + \pi$ cannot distinguish between the two signs. By the principle of finite signal propagation, the time necessary to obtain additional information cannot be made arbitrarily small. If a photon is used for the purpose of *comparing* the physical situations in the two segments, the minimal time for obtaining information would be $2(d + \pi)/c$.

On the other hand, we know that the reflection operator, defined by

$$R\psi(x) = \psi(-x) \qquad (3.2)$$

corresponds to an observable. The two functions in Eq. (3.1) are eigenfunctions of R with parity (eigenvalue) ± 1, and there are known methods for determination of parity.

If, more generally, observations in a finite time interval are admitted, the same conclusions hold if the space–time volumes

$$-d - \pi < x < -d, \qquad |t| < \Delta t$$

and

$$d < x < d + \pi, \qquad |t| < \Delta t$$

are spacelike with respect to each other. For any finite timelike thickness Δt, there are states (characterized by ψ in our example) whose observable properties cannot be determined by an observation in the timelike slice.

We conclude that the strong causality [Statement (A)] as well as the slightly weakened form (B) are untenable in relativistic quantum mechanics.

IV. WEAK CAUSALITY

A strictly localized ensemble ρ_v has the property that at $t = 0$ the expectation values of all observables are vacuumlike outside the space volume v. More precisely, if w is a space volume entirely outside v, and A_w a quasilocal observable at $t = 0$ in w, then the expectation value $\langle A_w \rangle_{\rho_v}$ is equal to the vacuum expectation value $\langle A_w \rangle_\Omega$ of this observable. According to Sec. III, there exist observables whose expectation values are not functions of the instantaneous expectation values $\langle A \rangle_{\rho_v}$. Consider, however, a four-dimensional cone defined as follows. Let R be the radius of the smallest sphere that contains v. Then this sphere and the hypersurface consisting of all light rays from the surface of the sphere to its center defines a space–time cone $C(v)$, shown in Fig. 2, such that observations in C can ascertain any "phase relation" between parts of the physical system in v. Without contradicting either the relativistic principle of finite signal velocity or well-established results of quantum mechanics, we can state a weaker form of causality:

(C) *For a strictly localized ensemble ρ_v in the space volume v, the expectation values of the observables A_C in the corresponding space–time cone $C(v)$ determine all expectation values.*

It might appear, at first, that Statement (C) is
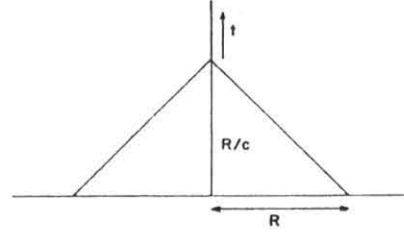
FIG. 2. Weak causality: For a state strictly localized within a sphere of radius R_i, a determining set of observations must include durations of R/c.

too weak to be useful as a substitute for (A) since the strictly localized ensembles are a very special class of ensembles or states. However, the concept of a physical system is meaningful only to the extent that it is not influenced by other parts of the universe which are left out in considering the system. If, nevertheless, the system is idealized so that it extends everywhere, then we must make the assumption that nothing else exists, i.e., that the expectation values of quasilocal observables are vacuumlike at sufficiently large distances. Statement (C) seems to be adequate not because most ensembles are strictly localized but because the only way to deal with actual ensembles is to approximate them by strictly localized ones. In contrast to Statements (A) and (B), (C) evidently does not require that the set of observables associated with a precise instant should be closed under addition, and thereby relieves the theorist of a heavy burden.

Another consequence of the principle of finite signal velocity is that a collection of strictly localized ensembles $\{\rho_v\}$ for a fixed volume v is invariant under operators that are images of the corresponding set of quasilocal observables $\{A_v\}$. Indeed, according to the principles of quantum mechanics the vector $A_v \Psi_v / \|A_v \Psi_v\|$ is the state created immediately after an instantaneous measurement A_v. If this state differed from the vacuum state outside of v, a signal would be transmitted instantaneously from v to other space points.

This remark can serve to confirm the impossibility of determining the phase δ in a state of the type considered in Sec. III, viz.

$$\Phi = \Psi_v + e^{i\delta}\Psi_{v'}$$

by instantaneous measurements if v and v' are disjoint simultaneous space volumes. Clearly,

$$(\Psi_{v'}, \Psi_v) = 0,$$

and according to our previous remark

$$(\Psi_{v'}, A\Psi_{v'}) = 0$$

whether A is in $\{A_v\}$ or $\{A_{v'}\}$. Hence, for any observable A,

$$(\Phi, A\Phi) = (\Psi_v, A\Psi_v) + (\Psi_{v'}, A\Psi_{v'})$$

and the cross term always vanishes, so that no instantaneous information about the phase is available.

V. THE PROJECTION AXIOM

The weakening of classical determinism in quantum mechanics is of two kinds: either the statements refer to all observables and all states but to the ensemble rather than the individual sample, or they refer to some observations on some states and successive observations on one sample. The latter cases are realized by a special kind of measurement procedure called a "procedure of the first kind,"[6] which is aptly described as *filtering*. A filter selects a subset of an original ensemble, and some unambiguous predictions can be made with respect to each sample of such a subset. Let us consider the restrictions that relativity imposes on these predictions.

One of von Neumann's postulates is the projection axiom (M)[7]: "If the observable R is measured on a system twice in succession, both observations yield the same value." Clearly, this form of the statement must be taken with a grain of salt. Margenau[8] has pointed out that in the overwhelming majority of measurements the system under observation is destroyed; it or its parts become permanently attached to the measuring apparatus. In the spirit of Pauli,[6] a more literal version would preface the sentence by "In every equivalence class of procedures belonging to the observable R, there exists one such that . . . ."

Is Axiom M necessary at all? It is argued here that at least in some modified form "Axiom M" is both physically desirable and indispensable. In classical physics the immediate repetition of an observation confirms the first result. This fact is tacitly accepted as the basis of any science. If it were not so, could one speak of objectively true events at all? Since quantum mechanics must agree with classical physics in some limit, quantum mechanics must surely include some statement with predictive claim on successive measurements of individual samples. What could the statement be?

Von Neumann points out that there are, *a priori*, three possible forms of causality or acausality in relation to the repetition test. (The words "con-

firmability" or "objectivity" would perhaps be more felicitous than "causality.") Given a repetition, (1) the first and second results could be statistically independent, (2) the first result could have a statistical dispersal, but the second be each time identical with the first, or (3) both results could be uniquely determined by the initial state.

The third case is that of classical mechanics; the first would come close to denying the existence of any objective observation, and hence of natural science. There remains the second case which is embodied in Axiom M—and perhaps a fourth possibility, namely that the results of the second measurement could be statistically correlated to the first. The principle of simplicity impels us to choose the second rather than the fourth possibility unless there is definite evidence against the former.

The point in which von Neumann's axiom needs revision (in addition to the minor restriction made above) is the time after which a confirmatory repetition can be made. As we have seen, in relativistic quantum mechanics some observables that are indispensable for prediction cannot be measured instantly, i.e., there is an inevitable delay between the beginning of the interaction and the recording of the information. It is now shown that there is equally an inevitable delay before the second measurement can confirm the first result.

In discussing the time sequence of measurements, it is convenient to think of a retrospective analysis of measurements completed in the distant past, rather than of a theory to be applied to experiments in actual progress. The first advantage of this view is that the use of probability in the sense of a rational judgment on the basis of existing and, ordinarily, incomplete evidence never arises; the only kind of probability involved is the relative frequency of past events. The second advantage of the retrospective view is that the question of signal velocity between recording devices never arises. It must be remembered that, literally speaking, a *prediction* is not possible even in classical relativistic physics, since the time necessary to communicate information from local observing devices to a central predictor would be precisely as long as the time for which the theoretical predictive ability claims validity, viz. $t = \Delta/c$, where Δ is the distance between the most distant of the simultaneous recording devices. Instead, by prediction we mean the establishment of a functional relation between observations recorded at different times, all of them in the distant past with respect to the time at which the verification is made.

---

[6] W. Pauli, *Handbuch der Physik* (Springer-Verlag, Berlin, 1933), p. 152.
[7] Reference 2, p. 177.
[8] H. Margenau, Phys. Rev. 49, 240 (1936).

We consider the procedures of the first kind which measure the parity of a sample at two different times in such a way that the first and second measurements have identical results.

Resonance scattering of light provides such a procedure for some systems. Let there be two energy "ground" eigenstates with opposite parity, and let the state be a coherent superposition of the two (which are assumed to be nearly degenerate). If there is an excited energy eigenstate of known parity (say $+1$), then resonance scattering for sufficiently long wavelength is possible only with parity change of the system. If the energy spread of the photon covers the energy difference between the two ground states and the excited states then the system is certain to be in the state with parity -1 after resonance of the photon has been observed. The question is now: What is the smallest time between the beginning of the interaction between system and photon and its cessation? To simplify the question, think of a system which is initially in the negative-parity state, and ask for the time at which the wavefunction of the combination (system and photon) becomes a product function with the negative-parity eigenfunction as one factor.

Consider an electrodynamic system (such as a positronium) consisting of two particles localized approximately in small volumes v and v' with a large distance between them. Intuitively, the answer to our question is then the following. In order to be sensitive to the parity of the state, the photon has to be scattered by one particle (say, in v) and run to v' to be rescattered—or vice versa. The smallest time for such a process is evidently d/c.

A more quantitative estimate may be derived by elementary perturbation theory. Dyson's operator[9] $U(t, t_0)$ for finite times can be expanded and the terms transformed in the usual manner. The result of the contractions can be represented by the usual diagrams. The relevant fourth-order diagrams are shown in Fig. 3; it is the interference of these two terms that is sensitive to parity. The rules for the evaluation of the diagrams differ from the usual ones only in that the integration over the coordinates of points 1 and 4 is omitted. The result is the probability amplitude for a process in which a photon reaches the system at t_1 and leaves it at t_4. In order to obtain the probability for a resonance scattering from a ground state, the resulting matrix element would have to be integrated with respect to the

⁹ See, for example, S. S. Schweber, H. A. Bethe, and F. De Hoffmann, *Mesons and Fields* (Row, Peterson and Co., Evanston, Illinois, 1955), Vol. I.
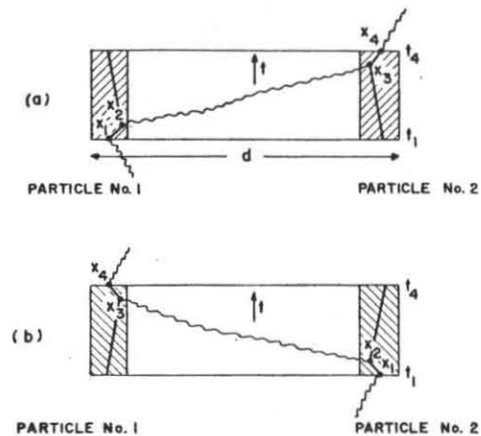
FIG. 3. Feynman diagrams for resonance scattering. The photon is absorbed and re-emitted by one particle, then absorbed and re-emitted by the other particle.

initial and final coordinates of the particles. Similarly, an integration over some localized wave packet of photons would have to be performed on the final and initial coordinates of the photon.

The usual evaluation of the diagrams exhibits the function $D_F(x_3 - x_2)$, where the time coordinates of x_3 and x_2 cannot differ by more than $t_4 - t_1$ while the space coordinates differ approximately by d. Since the function D_F decreases rapidly outside the light cone, the matrix element is negligible unless the time $t_4 - t_1$ is larger than d/c. The measurement begins when the photon interacts with particle No. 1 (or 2) and is repeatable after it has interacted with particle 2 (or 1) in diagram (a) [or (b)]. In the intermediate period, one of the particles is in an excited state and the total system is clearly not in the ground state, so that an additional photon would not be scattered in the same manner.

Only a particular class of fourth-order diagrams has been considered, and one may ask why others should not contribute to the measurement. Physically, the reason is that the photon energy has been chosen for resonance (i.e., so that Thompson scattering, Compton scattering, etc., are negligible), but mathematically this cannot be shown from perturbation since the excited intermediate state is not obtainable by perturbation.

It has thus been shown that relativity imposes a time delay between a measurement of the first kind and the subsequent confirmatory observation. Therefore, the term "immediate repetition" in Axiom M must be replaced by the phrase "repetition after the time Δt that characterizes the space–time

volume V assigned to the observables." This leads to the modified axiom:

(M') *In every equivalence class of procedures belonging to an observable A_V, there exists a procedure such that for all systems whose Hamiltonian commutes with A_V, its repetition after the time Δt gives the same* result as the first measurement, where the interval Δt is the largest timelike interval in V.

Relativistic Coulomb Scattering of Electrons*

R. L. GLUCKSTERN† AND SHIN-R LIN†
Physics Department, Yale University, New Haven, Connecticut
(Received 28 May 1964)

A simple and useful relation between the Coulomb amplitudes F and G (in Mott's notation) is derived and F and G are evaluated analytically up to α^4 terms for arbitrary $q = \alpha/\beta$. These results are valid for all angles, but are particularly useful at small angles. The general analytic behavior of F and G in the variable $x = \sin \frac{1}{2}\theta$ is discussed. The method is applicable to higher-order terms (α^6 and up). A double integral representation of F is also derived by using the Sommerfeld–Watson transformation. This integral representation exhibits the dependence on α, q, and θ separately.

1. INTRODUCTION

THE solutions for the relativistic scattering of electrons in a Coulomb field were first obtained by Mott[1] in the form of partial wave amplitudes. These amplitudes were expressed as functions of the two parameters α and q, where $\alpha = Z/137$, $q = \alpha/\beta$, and $\beta = v/c$. Attempts to sum the partial wave series analytically were successful only in powers of α (with β considered to be of order 1).[2,3] The most recent of these attempts[3] led to expressions for the Coulomb amplitudes F and G (in Mott's notation) accurate to order α^2 and α^3 respectively, with extremely complicated coefficients which were functions of β and $x = \sin \frac{1}{2}\theta$.

We have obtained a simple and useful relation between the Coulomb amplitudes F and G, and have succeeded in summing the partial wave series in powers of α^2 for arbitrary q, up to and including the terms in α^4. This organization of the expansion appears to be simpler and more natural than that in simultaneous power of α and q,[2,3] since the major complexity of the latter comes from expansion of the Coulomb phase factor, exp $(2iq \ln x)$, in powers of q. In our expansion the result is separated into two terms, one of which contains the phase factor exp $(2iq \ln x)$, the other of which does not. These results are then analytic in the variable x, apart from the Coulomb phase factor. This separation is similar to that given by Drell and Pratt[4] for $\beta = 1$.

Our results are related to those of Rosen,[5] and of Fradkin, Weber, and Hammer.[6] Rosen derived a double-integral representation of the coefficients of powers of α^2. Fradkin et al. derived a similar expansion in terms of a two-parameter function $T(\theta, q)$ up to α^2. We have evaluated these coefficients as convergent expansions in powers of x, which are most useful in the small-angle region (near $x = 0$). In addition, the method is applicable for the α^6 and higher terms, although the algebra is tedious and has not been carried out.

For completeness we have also obtained a double-integral representation of the Coulomb amplitudes, in which the dependence on α, q, and θ is exhibited in separate factors.

Applications of the considerations in the present paper to physical problems have been considered

[1] N. F. Mott, Proc. Roy. Soc. (London) A124, 425 (1929); 135, 429 (1932).
[2] W. A. McKinley and H. Feshbach, Phys. Rev. 74, 1759 (1948); R. H. Dalitz, Proc. Roy. Soc. (London) A206, 509 (1951).
[3] W. R. Johnson, T. A. Weber, and C. J. Mullin, Phys. Rev. 121, 933 (1962).
[4] S. D. Drell and R. H. Pratt, Phys. Rev. 125, 1394 (1962).
[5] B. Rosen, J. Math. Phys. 4, 392 (1963).
[6] D. M. Fradkin, T. A. Weber, and C. L. Hammer, Ann. Phys. (N. Y.) 27, 338 (1964).

volume V assigned to the observables." This leads to the modified axiom:

(M′) *In every equivalence class of procedures belonging to an observable A_V, there exists a procedure such that for all systems whose Hamiltonian commutes with A_V, its repetition after the time Δt gives the same* result *as the first measurement, where the interval Δt is the largest timelike interval in V.*

Relativistic Coulomb Scattering of Electrons*

R. L. GLUCKSTERN† AND SHIN-R LIN†

Physics Department, Yale University, New Haven, Connecticut

(Received 28 May 1964)

A simple and useful relation between the Coulomb amplitudes F and G (in Mott's notation) is derived and F and G are evaluated analytically up to α^4 terms for arbitrary $q = \alpha/\beta$. These results are valid for all angles, but are particularly useful at small angles. The general analytic behavior of F and G in the variable $x = \sin \frac{1}{2}\theta$ is discussed. The method is applicable to higher-order terms (α^6 and up). A double integral representation of F is also derived by using the Sommerfeld–Watson transformation. This integral representation exhibits the dependence on α, q, and θ separately.

1. INTRODUCTION

THE solutions for the relativistic scattering of electrons in a Coulomb field were first obtained by Mott[1] in the form of partial wave amplitudes. These amplitudes were expressed as functions of the two parameters α and q, where $\alpha = Z/137$, $q = \alpha/\beta$, and $\beta = v/c$. Attempts to sum the partial wave series analytically were successful only in powers of α (with β considered to be of order 1).[2,3] The most recent of these attempts[3] led to expressions for the Coulomb amplitudes F and G (in Mott's notation) accurate to order α^2 and α^3 respectively, with extremely complicated coefficients which were functions of β and $x = \sin \frac{1}{2}\theta$.

We have obtained a simple and useful relation between the Coulomb amplitudes F and G, and have succeeded in summing the partial wave series in powers of α^2 for arbitrary q, up to and including the terms in α^4. This organization of the expansion appears to be simpler and more natural than that in simultaneous power of α and q,[2,3] since the major complexity of the latter comes from expansion of the Coulomb phase factor, exp ($2iq \ln x$), in powers of q. In our expansion the result is separated into two terms, one of which contains the phase factor exp ($2iq \ln x$), the other of which does not. These results are then analytic in the variable x, apart from the Coulomb phase factor. This separation is similar to that given by Drell and Pratt[4] for $\beta = 1$.

Our results are related to those of Rosen,[5] and of Fradkin, Weber, and Hammer.[6] Rosen derived a double-integral representation of the coefficients of powers of α^2. Fradkin et al. derived a similar expansion in terms of a two-parameter function $T(\theta, q)$ up to α^2. We have evaluated these coefficients as convergent expansions in powers of x, which are most useful in the small-angle region (near $x = 0$). In addition, the method is applicable for the α^6 and higher terms, although the algebra is tedious and has not been carried out.

For completeness we have also obtained a double-integral representation of the Coulomb amplitudes, in which the dependence on α, q, and θ is exhibited in separate factors.

Applications of the considerations in the present paper to physical problems have been considered

[1] N. F. Mott, Proc. Roy. Soc. (London) **A124**, 425 (1929); **135**, 429 (1932).
[2] W. A. McKinley and H. Feshbach, Phys. Rev. **74**, 1759 (1948); R. H. Dalitz, Proc. Roy. Soc. (London) **A206**, 509 (1951).
[3] W. R. Johnson, T. A. Weber, and C. J. Mullin, Phys. Rev. **121**, 933 (1962).
[4] S. D. Drell and R. H. Pratt, Phys. Rev. **125**, 1394 (1962).
[5] B. Rosen, J. Math. Phys. **4**, 392 (1963).
[6] D. M. Fradkin, T. A. Weber, and C. L. Hammer, Ann. Phys. (N. Y.) **27**, 338 (1964).

in a separate paper.[7] Among these applications are:

(1) Accurate evaluation of the poorly convergent partial wave series for the Coulomb amplitudes at forward angles.

(2) Comparison of attractive and repulsive Coulomb scattering, particularly in the nonrelativistic limit of large $|q|$ but finite α.

(3) Behavior of the cross section and asymmetry function for small angles.

(4) Behavior of the cross section and asymmetry function at backward angles, where magnetic scattering effects will be important.

(5) Modification of the Coulomb amplitudes because of screening by atomic electrons.

2. RELATION BETWEEN F AND G FOR RELATIVISTIC COULOMB SCATTERING

a. Partial Wave Expansion

The relativistic amplitudes for Coulomb scattering have been obtained in partial wave expansion by Mott,[1] and his results are reproduced below. The amplitudes of the scattered wave in the two spin states are related to the functions $f(\theta)$ and $g(\theta)$, given in the general case by

$$2ikf(\theta) = \sum_{n=0}^{\infty} [(n+1)(e^{2i\eta_n} - 1) + n(e^{2i\eta_{-n-1}} - 1)]P_n(\cos \theta),$$ (2.1)

$$2ikg(\theta) = \sum_{n=0}^{\infty} (-e^{2i\eta_n} + e^{2i\eta_{-n-1}})P_n^1(\cos \theta).$$

For Coulomb scattering the phase shifts in (2.1) are determined from the asymptotic form of the wavefunction; the prescription here is

$$e^{2i\eta_{-n-1}} = -[n - iq(1 - \beta^2)^{\frac{1}{2}}]C_n,$$
$$e^{2i\eta_{n-1}} = -[n + iq(1 - \beta^2)^{\frac{1}{2}}]C_n,$$ (2.2)

where[8]

$$C_n = -e^{i\pi(n-\rho_n)}\Gamma(\rho_n - iq)/\Gamma(\rho_n + 1 + iq)$$ (2.3)

and

$$\beta = v/c, \quad \rho_n = (n^2 - \alpha^2)^{\frac{1}{2}}, \quad \alpha = Z/137, \quad q = \alpha/\beta.$$ (2.4)

Using

$$kf(\theta) = -iq(1 - \beta^2)^{\frac{1}{2}}F(\theta) + G(\theta),$$
$$kg(\theta) = iq(1 - \beta^2)^{\frac{1}{2}}F(\theta) \cot \tfrac{1}{2}\theta + G(\theta) \tan \tfrac{1}{2}\theta,$$ (2.5)

Mott obtains

---

[7] R. L. Gluckstern and S.-R Lin, Phys. Rev. **136**, B 859 (1964).
[8] The notation is identical with that of Mott (see Ref. 1) except that our C_n differs from his by a factor $(-1)^n$.

$$F(\theta) = \frac{i}{2} \sum_{n=1}^{\infty} nC_n[P_n(\cos \theta) - P_{n-1}(\cos \theta)],$$

$$G(\theta) = \frac{i}{2} \sum_{n=1}^{\infty} n^2C_n[P_n(\cos \theta) + P_{n-1}(\cos \theta)].$$ (2.6)

b. Convergence of Partial Wave Expansion

The series as they stand in (2.6) do not converge for all θ. For large n, one has

$$C_n \sim n^{-1-2iq},$$

and for $\theta \neq 0$, $P_n(\cos \theta) \sim n^{-\frac{1}{2}}$. Therefore in the strict sense, $F(\theta)$ is conditionally convergent and $G(\theta)$ is divergent. Nevertheless, the expression for $G(\theta)$ in (2.6) is to be understood as the limit of a corresponding sum which includes a suitable convergence factor.

A more convenient approach is to evaluate the sums in (2.6) for $\alpha = 0$, obtaining the usual Born approximation (F_0 and G_0)and then expand the difference between F, G and F_0, G_0 in partial wave series which converge properly. Mott has performed the sums for $\alpha = 0$, obtaining

$$F_0(\theta) = \tfrac{1}{2}i[\Gamma(1 - iq)/\Gamma(1 + iq)](\sin \tfrac{1}{2}\theta)^{2iq},$$ (2.7)
$$G_0(\theta) = -iq \cot^2 \tfrac{1}{2}\theta F_0(\theta),$$

and the convergent partial wave expressions for $F_1(\theta) = F(\theta) - F_0(\theta)$ and $G_1(\theta) = G(\theta) - G_0(\theta)$ are then given by

$$F_1(\theta) = \frac{i}{2} \sum_{n=1}^{\infty} nD_n[P_n(\cos \theta) - P_{n-1}(\cos \theta)],$$

$$G_1(\theta) = \frac{i}{2} \sum_{n=1}^{\infty} n^2D_n[P_n(\cos \theta) + P_{n-1}(\cos \theta)],$$ (2.8)

where

$$D_n = C_n - C_n(\alpha = 0).$$ (2.9)

The convergence of the series in (2.8) is now assured since n^2D_n is bounded as $n \to \infty$.

c. Relation between F and G[9]

The recurrence relations satisfied by the Legendre polynomials can be put in the form

$$(1 - \cos \theta)d[P_n(\cos \theta) + P_{n-1}(\cos \theta)]/d \cos \theta$$
$$= -n[P_n(\cos \theta) - P_{n-1}(\cos \theta)],$$ (2.10)

$$(1 + \cos \theta)d[P_n(\cos \theta) - P_{n-1}(\cos \theta)]/d \cos \theta$$
$$= n[P_n(\cos \theta) + P_{n-1}(\cos \theta)].$$ (2.11)

Equations (2.11) and (2.8) lead immediately to the

---

[9] The existence of such a relation was first pointed out by G. Rawitscher.

relation

$$G_1(\theta) = (1 + \cos \theta)dF_1(\theta)/d \cos \theta$$

$$= - \cot \tfrac{1}{2}\theta dF_1(\theta)/d\theta. \qquad (2.12)$$

From (2.7) one can easily demonstrate that

$$G_0(\theta) = (1 + \cos \theta)dF_0(\theta)/d \cos \theta$$

$$= - \cot \tfrac{1}{2}\theta dF_0(\theta)/d\theta. \qquad (2.13)$$

Therefore F and G satisfy the same relation, namely

$$G(\theta) = (1 + \cos \theta)dF(\theta)/d \cos \theta$$

$$= - \cot \tfrac{1}{2}\theta dF(\theta)/d\theta. \qquad (2.14)$$

The significance of this relation is twofold: In any computation or evaluation of Coulomb amplitudes, it is only necessary to consider $F(\theta)$. The other amplitude $G(\theta)$ can be obtained by differentiation once $F(\theta)$ is available. Moreover, since the series for $F_1(\theta)$ in (2.8) converges more rapidly than that for $G_1(\theta)$, numerical results obtained using (2.12) will be more accurate than those obtained with (2.8).

d. Non-Coulomb Scattering

The relation (2.14) suggests that it may be useful to cast the results (2.1) for general relativistic scattering into a more convergent form. If one defines the complex quantities A_n and B_n in terms of the phase shifts by

$$e^{2i\eta_{n-1}} = A_n,$$

$$e^{2i\eta_{-n-1}} = B_n, \qquad (2.15)$$

the amplitudes f and g can be written, for all directions except $\theta = 0$, as

$$2ikf(\theta) = \sum_{n=1}^{\infty} n(A_n P_n + B_n P_{n-1}),$$

$$\qquad (2.16)$$

$$2ikg(\theta) = \frac{d}{d\theta} \sum_{n=1}^{\infty} (-A_n P_{n-1} + B_n P_n).$$

Defining

$$\dot{E}(\theta) = \frac{1}{4i} \sum_{n=1}^{\infty} n(A_n + B_n)(P_n + P_{n-1}),$$

$$\qquad (2.17)$$

$$H(\theta) = \frac{1}{4i} \sum_{n=1}^{\infty} n(A_n - B_n)(P_n - P_{n-1}),$$

one finds

$$kf(\theta) = E(\theta) - H(\theta)$$

$$\qquad (2.18)$$

$$kg(\theta) = E(\theta) \tan \tfrac{1}{2}\theta + H(\theta) \cot \tfrac{1}{2}\theta,$$

where the recurrence relations (2.10) and (2.11) have been used. Comparison with (2.5) shows that $E(\theta)$ and $H(\theta)$ now play the role of $iq(1 - \beta^2)^{\frac{1}{2}}F(\theta)$ and

$G(\theta)$. However, the relation (2.14) no longer exists between $E(\theta)$ and $H(\theta)$.

The rate of convergence of the expansions(2.17) may be increased by repeated applications of (2.10) and (2.11). For example, one may define the sums

$$\bar{E}(\theta) = \frac{1}{4i} \sum_{n=1}^{\infty} (A_n + B_n)(P_n - P_{n-1}),$$

$$\qquad (2.19)$$

$$\bar{H}(\theta) = \frac{1}{4i} \sum_{n=1}^{\infty} (A_n - B_n)(P_n + P_{n-1}),$$

leading to

$$E(\theta) = -\tan \tfrac{1}{2}\theta \, d\bar{E}(\theta)/d\theta$$

$$\qquad (2.20)$$

$$H(\theta) = \cot \tfrac{1}{2}\theta \, d\bar{H}(\theta)/d\theta.$$

This prescription must be used cautiously in numerical calculations, since the more rapid convergence of (2.19) is to be balanced against the greater accuracy required for the numerical differentiation required in (2.20).

e. Limits for $F(\theta)$, $G(\theta)$ at $\theta = 0, \pi$

The convergent expression for $F_1(\theta)$ in (2.8) leads immediately to the results

$$F_1(0) = 0,$$

$$\qquad (2.21)$$

$$F_1(\pi) = i \sum_{n=1}^{\infty} (-1)^n n D_n.$$

Explicit results for coefficients of $F_1(\pi)$ expanded in powers of α are given later. It can similarly be shown from (2.8) or (2.14) with some care regarding convergence, that

$$G_1(\pi) = 0. \qquad (2.22)$$

The quantity $G_1(0)$ is infinite as is also shown later, although this may be inferred directly from (2.8).

3. SMALL-ANGLE EXPANSION FOR F AND G

It is obvious from (2.8) that the main contributions to the behavior of $F_1(\theta)$ and $G_1(\theta)$ near $\theta = 0$ come from large n. In fact, an expansion of D_n for large n allows one to perform sums of the terms in this expansion for small θ, leading to the form

$$F_1(\theta) \simeq \sum_{i=1}^{\infty} a_i \theta^{i+2iq} + \sum_{i=2}^{\infty} b_i \theta^i,$$

$$\qquad (3.1)$$

$$G_1(\theta) \simeq \sum_{i=-1}^{\infty} c_i \theta^{i+2iq} + \sum_{i=0}^{\infty} d_i \theta^i.$$

After much labor, the first few values of a_i and c_i can be written down for arbitrary α, q. However b_i and d_i cannot be obtained explicitly this way.

It can be shown that (3.1) correctly describes the analytic property of $F_1(\theta)$, $G_1(\theta)$ in the variable θ.

Specifically, each function separates into two parts; one is regular in the variable θ and the other has a branch point at $\theta = 0$, just as $F_0(\theta)$ and $G_0(\theta)$ in (2.7). We now obtain the series in (3.1) by expanding in powers of α^2 *for fixed* q. This modified Born series correctly exhibits the analytic behavior in θ, although succeeding powers in α^2 are more difficult to obtain.

The starting point of this expansion is the integral representation for D_n:

$$D_n = \frac{1}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq} t^{n-iq-1}$$
$$\times [1 - e^{(n-\rho_n)(i\pi - \ln t)}]. \qquad (3.2)$$

Writing

$$n - \rho_n \simeq \alpha^2/2n + \alpha^4/8n^3,$$

one has

$$nD_n \simeq \frac{1}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq} t^{n-iq-1}$$
$$\times [\tfrac{1}{2}\alpha^2(-i\pi + \ln t) - \tfrac{1}{8}\alpha^4(-i\pi + \ln t)^2$$
$$+ (\alpha^4/8n^2)(-i\pi + \ln t)]$$
$$\equiv \tfrac{1}{2}\alpha^2 d_n^{(1)} - \tfrac{1}{8}\alpha^4 d_n^{(2)} + \tfrac{1}{8}\alpha^4 d_n^{(3)}. \qquad (3.3)$$

The most convenient form to use in evaluating the sums in (2.8) is the one which has the appropriate power of n to cancel that of the $d_n^{(i)}$ in (3.3). Specifically, one can write

$$f^{(1)} = \sum d_n^{(1)}[P_n(\cos\theta) - P_{n-1}(\cos\theta)],$$
$$g^{(2)} = \sum nd_n^{(2)}[P_n(\cos\theta) + P_{n-1}(\cos\theta)], \qquad (3.4)$$
$$h^{(3)} = \sum n^2 d_n^{(3)}[P_n(\cos\theta) - P_{n-1}(\cos\theta)].$$

Use of the recurrence relations (2.10) gives

$$g^{(2)} = -\cot\tfrac{1}{2}\theta\, df^{(2)}/d\theta, \quad g^{(1)} = -\cot\tfrac{1}{2}\theta\, df^{(1)}/d\theta,$$
$$h^{(3)} = \tan\tfrac{1}{2}\theta\, dg^{(3)}/d\theta, \quad g^{(3)} = -\cot\tfrac{1}{2}\theta\, df^{(3)}/d\theta. \qquad (3.5)$$

One therefore evaluates $f^{(1)}$, $g^{(2)}$, $h^{(3)}$ from (3.4), obtains $g^{(3)}$, $f^{(3)}$, $f^{(2)}$ from (3.5), and finally arrives at

$$F_1(\theta) = \tfrac{1}{2}i[\tfrac{1}{2}\alpha^2 f^{(1)} - \tfrac{1}{8}\alpha^4 f^{(2)} + \tfrac{1}{8}\alpha^4 f^{(3)}],$$
$$G_1(\theta) = \tfrac{1}{2}i[\tfrac{1}{2}\alpha^2 g^{(1)} - \tfrac{1}{8}\alpha^4 g^{(2)} + \tfrac{1}{8}\alpha^4 g^{(3)}]. \qquad (3.6)$$

Let us write the sums explicitly, replacing the factor $(-i\pi + \ln t)$ by $(-i\pi + \partial/\partial\epsilon)t^\epsilon |_{\epsilon=0}$. This leads to

$$f^{(1)} = \frac{D_\epsilon}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq} t^{\epsilon-iq-1}$$
$$\times \sum_{n=1} t^n(P_n - P_{n-1}),$$
$$\qquad (3.7)$$
$$g^{(2)} = \frac{D_\epsilon}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq} t^{\epsilon-iq-1}$$
$$\times \sum_{n=1} t^n(P_n + P_{n-1}),$$
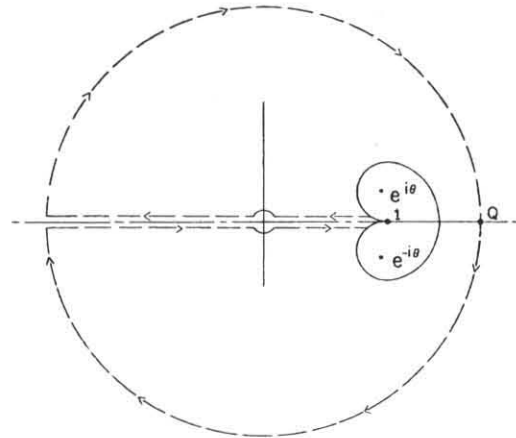
FIG. 1. The contour C for $P(x)$.

and

$$h^{(3)} = f^{(1)},$$

where

$$D_\epsilon \equiv (-i\pi + \partial/\partial\epsilon)_{\epsilon=0}.$$

The sums over n lead to

$$f^{(1)} = \frac{D_\epsilon}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq} t^{\epsilon-iq-1}$$
$$\times \left[\frac{1 - t}{(1 - 2t\cos\theta + t^2)^{\frac{1}{2}}} - 1\right], \qquad (3.8)$$
$$g^{(2)} = \frac{D_\epsilon^2}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq} t^{\epsilon-iq-1}$$
$$\times \left[\frac{1 + t}{(1 - 2t\cos\theta + t^2)^{\frac{1}{2}}} - 1\right]. \qquad (3.9)$$

Since

$$1 - 2t\cos\theta + t^2 = (1 - te^{i\theta})(1 - te^{-i\theta}),$$

the integrals can be written as double hypergeometric functions of the arguments $e^{\pm i\theta}$. We are interested in the behavior for small θ, and our task is therefore to obtain a suitable analytic continuation of the double hypergeometric function. We do this directly from (3.8) and (3.9), illustrating our method with (3.8).

a. Term Proportional to α^2 for Arbitrary q

Let us start with the integral representation

$$p(x) = \int_0^1 dt \frac{(1 - t)^{2iq+1} t^{\epsilon-iq-1}}{[(1 - t)^2 + 4tx^2]^{\frac{1}{2}}}, \qquad (3.10)$$

where $x = \sin\tfrac{1}{2}\theta$. From (3.8) one has

$$f^{(1)} = [D_\epsilon/\Gamma(1 + 2iq)][p(x) - p(0)]. \qquad (3.11)$$
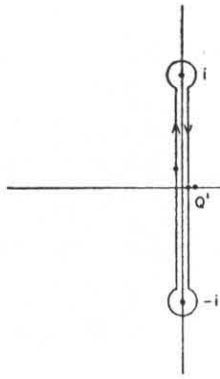
FIG. 2. The contour C' for $P(x)$.

Let us now consider the contour integral

$$P(x) = \int_c dt \frac{(t-1)^{2iq+1} t^{\epsilon-iq+1}}{[(t-e^{i\theta})(t-e^{-i\theta})]^{\frac{1}{2}}}, \quad (3.12)$$

where the contour C is shown as the solid curve in Fig. 1. The singularities in the t plane are at $t = 0, 1, e^{i\theta}$, and $e^{-i\theta}$, and the phase of each factor is chosen to be zero at the point Q. If one deforms the contour to the dotted curve shown, one can neglect the contributions around $t = 0$ and $t = \infty$, obtaining

$$P(x) = 2p(x) \sinh 2\pi q - 2r(x) \sinh (\pi q - i\pi\epsilon), \quad (3.13)$$

where

$$r(x) = \int_{-\infty}^{0} dt \frac{(1-t)^{2iq+1}(-t)^{\epsilon-iq-1}}{[(1-t)^2 + 4tx^2]^{\frac{1}{2}}}. \quad (3.14)$$

It is clear from Fig. 1 that the integral representation $p(x)$ does not allow one to approach the limit $\theta = 0$, since the branch points $t = e^{\pm i\theta}$ merge with one of the endpoints of integration. However $p(x)$ may be expressed in terms of $P(x)$ and $r(x)$ from (3.13), each of which is now evaluated as $\theta \to 0$.

In the expression for $r(x)$, $|4tx^2|$ is always less than $(1-t)^2$ and one may expand the denominator in

powers of x, obtaining

$$r(x) = \sum_{j=0}^{\infty} \frac{(2x)^{2j} \Gamma(\frac{1}{2})(-1)^j}{j! \ \Gamma(\frac{1}{2} - j)}$$

$$\times \int_{-\infty}^{0} dt(1-t)^{2iq-2j}(-t)^{\epsilon-iq+j-1}. \quad (3.15)$$

Setting $u = (1-t)^{-1}$, one finds

$$r(x)$$

$$= \sum_{j=0}^{\infty} \frac{(-1)^j (2x)^{2j} \Gamma(\frac{1}{2}) \Gamma(j-iq-\epsilon) \Gamma(j-iq+\epsilon)}{j! \ \Gamma(\frac{1}{2} - j) \Gamma(2j-2iq)}. \quad (3.16)$$

In the integral representation for $P(x)$ one may change variables from t to v according to $t - 1 = 2vx$. In this way one writes

$$P(x) = (2x)^{2iq+1} \int_c dv \ v^{2iq+1}(1 + 2vx)^{\epsilon-iq-1}$$

$$\times (v^2 + 2vx + 1)^{-\frac{1}{2}}, \quad (3.17)$$

where the contour is the same as Fig. 1, displaced to $v = 0$. For small x, one can expand both the factors $(1 + 2vx)^a$ and $(v^2 + 1 + 2vx)^b$ in powers of x, obtaining

$$P(x) = (2x)^{2iq+1}$$

$$\times \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{(2x)^{j+k} \Gamma(k+iq+1-\epsilon) \Gamma(\frac{1}{2})(-1)^k}{k! \ \Gamma(iq+1-\epsilon)j! \ \Gamma(\frac{1}{2} - j)}$$

$$\times \int_{c'} dv \frac{v^{2iq+1+j+k}}{(v^2+1)^{\frac{1}{2}+j}}, \quad (3.18)$$

where the contour C' is shown in Fig. 2.

The phases of v and $v^2 + 1$ are taken to be zero at Q'. The integral over C' can be separated into two parts, one in the upper and the other in the lower half-plane. Each is a beta function, leading to the following result for $P(x)$:

$$P(x) = (2x)^{2iq+1} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \frac{(2x)^{j+k} \Gamma(k+iq+1-\epsilon) \Gamma(\frac{1}{2})(-1)^k(-2\pi i)}{k! \ \Gamma(iq+1-\epsilon)j! \ \Gamma[\frac{3}{2} + iq + \frac{1}{2}(k-j)] \Gamma[-iq - \frac{1}{2}(j+k)]}. \quad (3.19)$$

Setting $j = l - k$, one can write

$$P(x) = \sum_{l=1}^{\infty} \frac{(2x)^{l+1+2iq} \Gamma(\frac{1}{2})(-2\pi i)}{\Gamma(-iq - l/2) \Gamma(iq + 1 - \epsilon)}$$

$$\times \sum_{k=0}^{l} \frac{(-1)^k \Gamma(k+iq+1-\epsilon)}{k! \ (l-k)! \ \Gamma(\frac{3}{2} + iq - \frac{1}{2}l + k)}. \quad (3.20)$$

The sum over k may be performed by using the beta function integral representation of $\Gamma(k+a)/\Gamma(k+b)$. One finally obtains

$$P(x) = -2i\Gamma(\frac{1}{2})(2x)^{1+2iq}.$$

$$\times \sum_{l=0}^{\infty} \frac{(2x)^l \Gamma(\frac{1}{2} + \frac{1}{2}l - \epsilon) \Gamma(\frac{1}{2} + \epsilon + \frac{1}{2}l)}{\Gamma(-iq - \frac{1}{2}l) \Gamma(\frac{3}{2} + iq - \frac{1}{2}l)l!}$$

$$\times \cos (\frac{1}{2}\pi l - \pi\epsilon). \quad (3.21)$$

Combining (3.13), (3.16), and (3.21), one finds

$$\frac{p(x) - p(0)}{\Gamma(1 + 2iq)} = \frac{1}{\pi} \left[-x^{2iq+1} \sum_{l=0}^{\infty} \frac{\Gamma(-iq)}{\Gamma(-iq - \frac{1}{2}l)} \right.$$

$$\times \frac{\Gamma(-iq + \frac{1}{2})}{\Gamma(\frac{3}{2} + iq + \frac{1}{2}l)} \frac{(2x)^l}{l!} \Gamma(\frac{1}{2} + \frac{1}{2}l - \epsilon)$$

$$\times \Gamma(\tfrac{1}{2} + \tfrac{1}{2}l + \epsilon) \cos(\tfrac{1}{2}\pi l - \epsilon\pi)$$

$$- i \sum_{j=1}^{\infty} (2x)^{2i} \frac{\Gamma(\tfrac{1}{2})(-)^i \Gamma(1 + 2iq - 2j)}{j! \, \Gamma(\tfrac{1}{2} - j)\Gamma(1 + 2iq)}$$

$$\times \Gamma(j - iq - \epsilon)\Gamma(j - iq + \epsilon) \sinh(\pi q - i\pi\epsilon) \Big].$$
$$(3.22)$$

The operator D_ϵ may now be applied in order to obtain $f^{(1)}$. Since $\Gamma(a - \epsilon)\Gamma(a + \epsilon)$ is even in ϵ, the differentiation need only be applied to the trigonometric factors and ϵ can be set equal to zero in the gamma functions. Thus

$$f^{(1)} = ix^{2iq+1} \sum_{i=0}^{\infty} \frac{\Gamma(-iq)}{\Gamma(-iq - \tfrac{1}{2}l)} \frac{\Gamma(-iq + \tfrac{1}{2})}{\Gamma(\tfrac{3}{2} + iq + \tfrac{1}{2}l)}$$

$$\times \frac{(2x)^l \Gamma^2(\tfrac{1}{2} + \tfrac{1}{2}l)i^l}{l!} - \frac{\pi e^{\pi q}}{\sinh \pi q} \sum_{j=1}^{\infty} \frac{(2x)^{2i}\Gamma(\tfrac{1}{2})}{j! \, \Gamma(\tfrac{1}{2} - j)}$$

$$\times \frac{\Gamma(1 + 2iq - 2j)}{\Gamma(1 + 2iq)} \frac{\Gamma(j - iq)}{\Gamma(1 - j + iq)}. \qquad (3.23)$$

One can therefore write for the ratio F_1/F_0, accurate to order α^2,

$$\frac{F_1}{F_0} \simeq \frac{\alpha^2}{2} f^{(1)} x^{-2iq} \frac{\Gamma(1 + iq)}{\Gamma(1 - iq)} = xe^{i\psi} \sum_{i=0}^{\infty} a_i x^{2i}$$

$$+ x^2 \sum_{j=0}^{\infty} b_i x^{2i} + \frac{e^{\pi q}}{\sinh \pi q} x^{2-2iq} \sum_{i=0}^{\infty} c_i x^{2i}, \qquad (3.24)$$

where

$$a_i^{(2)} = \frac{\alpha^2 i\pi}{1 + 2iq} \frac{\Gamma(j + \tfrac{1}{2})}{\Gamma(\tfrac{1}{2})j!}$$

$$\times \frac{\Gamma(\tfrac{3}{2} + iq)}{\Gamma(\tfrac{3}{2} + iq + j)} \frac{\Gamma(j + 1 + iq)}{\Gamma(1 + iq)}, \qquad (3.25)$$

$$b_i^{(2)} = -\alpha^2 \frac{(1 + 2iq)}{2iq(1 + iq)} \frac{\Gamma(\tfrac{3}{2})\Gamma(j + 1)}{\Gamma(j + \tfrac{3}{2})}$$

$$\times \frac{\Gamma(2 + iq)}{\Gamma(2 + iq + j)} \frac{\Gamma(\tfrac{3}{2} + iq + j)}{\Gamma(\tfrac{3}{2} + iq)}, \qquad (3.26)$$

$$c_i^{(2)} = \alpha^2 \frac{\pi}{2q} \frac{\Gamma(\tfrac{1}{2} + j)}{\Gamma(\tfrac{1}{2})j!}$$

$$\times \frac{\Gamma(-iq + \tfrac{1}{2})}{\Gamma(-iq + \tfrac{1}{2} + j)} \frac{\Gamma(j - iq)}{\Gamma(-iq)}, \qquad (3.27)$$

and

$$\psi = 2 \arg \Gamma(1 + iq) - 2 \arg \Gamma(\tfrac{1}{2} - iq)$$

$$= 4 \arg \Gamma(1 + iq) - 2 \arg \Gamma(1 + 2iq)$$

$$+ 2q \ln 2. \qquad (3.28)$$

The superscript (2) stands for the term in α^2.

It is obvious from the form of the coefficients a_i, b_i, c_i that F_1/F_0 may be written in terms of hypergeometric functions as

$$\frac{F_1}{F_0} = \alpha^2 \Big\{ \frac{i\pi xe^{i\psi}}{(1 + 2iq)} \, {}_2F_1(\tfrac{1}{2}, 1 + iq, \tfrac{3}{2} + iq; x^2)$$

$$- \frac{(1 + 2iq)x^2}{2iq(1 + iq)} \, {}_3F_2(1, 1, \tfrac{3}{2} + iq, \tfrac{3}{2}, 2 + iq; x^2)$$

$$+ \frac{x^{-2iq}\pi e^{\pi q}}{2q \sinh \pi q} [{}_2F_1(\tfrac{1}{2}, -iq, \tfrac{1}{2} - iq; x^2) - 1] \Big\},$$
$$(3.29)$$

where we have used the Pochhammer notation for the hypergeometric function,[10]

$$_mF_n(a_1, \cdots, a_m, b_1, \cdots, b_n, z)$$

$$= \sum_{i=0}^{\infty} \frac{\Gamma(a_1 + j)}{\Gamma(a_1)} \cdots \frac{\Gamma(a_m + j)}{\Gamma(a_m)}$$

$$\times \frac{\Gamma(b_1)}{\Gamma(b_1 + j)} \cdots \frac{\Gamma(b_n)}{\Gamma(b_n + j)} \frac{z^j}{j!}. \qquad (3.30)$$

To the same approximation one can write

$$\frac{G_1}{G_0} = xe^{i\psi} \sum_{i=0}^{\infty} \bar{a}_i x^{2i}$$

$$+ x^2 \sum_{j=0}^{\infty} \bar{b}_i x^{2i} + \frac{e^{\pi q}}{\sinh \pi q} x^{2-2iq} \sum_{i=0}^{\infty} \bar{c}_i x^{2i}, \qquad (3.31)$$

where

$$\bar{a}_i = [1 + (2j + 1)/2iq]a_i, \qquad (3.32)$$

$$\bar{b}_i = [1 + (j + 1)/iq]b_i, \qquad (3.33)$$

$$\bar{c}_i = [(j + 1)/iq]c_i. \qquad (3.34)$$

Equations (3.32)–(3.34) are obtained from (2.14) rewritten in the form

$$\frac{G_1}{G_0} = \Big[1 + \frac{x}{2iq} \frac{d}{dx}\Big] \frac{F_1}{F_0}. \qquad (3.35)$$

In terms of the hypergeometric function, one finds, to order α^2,

$$G_1/G_0 = \alpha^2 \{(\tfrac{1}{2}\pi/q)xe^{i\psi} \, {}_2F_1(\tfrac{1}{2}, 1 + iq, \tfrac{1}{2} + iq; x^2)$$

$$+ \frac{1 + 2iq}{2q^2} x^2 \, {}_3F_2(1, 1, \tfrac{3}{2} + iq, \tfrac{3}{2}, 1 + iq; x^2)$$

$$- \frac{x^{2-2iq}}{q(1 - 2iq)} \frac{\pi e^{\pi q}}{\sinh \pi q} \, {}_2F_1(\tfrac{3}{2}, 1 - iq, \tfrac{3}{2} - iq; x^2) \Big\}.$$
$$(3.36)$$

b. Term Proportional to α^4 for Arbitrary q

The above procedure works as well for the term proportional to α^4, although the algebra is lengthier and the final result more complicated. A brief account of the analysis is given.

The term with superscript (3) in (3.6) comes directly from $h^{(3)} = f^{(1)}$, evaluated in (3.23) and (3.24).

---

[10] See for example, A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, p. 182.

From (3.5) one finds

$$g^{(3)}(\theta) = \int_0^x 2\frac{dx}{x} h^{(3)} + A, \qquad (3.37)$$

where the constant A is chosen so that $g^{(3)}(\pi) = 0$, i.e.,

$$A = -\int_0^1 2\frac{dx}{x} h^{(3)} = g^{(3)}(0). \qquad (3.38)$$

One then has

$$g^{(3)} = 2i\left[x^{2iq+1} \sum_{l=0}^{\infty} \frac{\Gamma(-iq)}{\Gamma(-iq - \frac{1}{2}l)}\right.$$

$$\times \frac{\Gamma(-iq + \frac{1}{2})}{\Gamma(\frac{3}{2} + iq + \frac{1}{2}l)} \frac{(2x)^l}{l!} \frac{\Gamma^2(\frac{1}{2} + \frac{1}{2}l)}{2iq + l + 1} i^l$$

$$- \frac{\pi e^{\pi q}}{\sinh \pi q} \sum_{j=1}^{\infty} \frac{(2x)^{2j}}{2j} \frac{\Gamma(\frac{1}{2})}{j! \; \Gamma(\frac{1}{2} - j)}$$

$$\times \left. \frac{\Gamma(1 + 2iq - 2j)}{\Gamma(1 + 2iq)} \frac{\Gamma(j - iq)}{\Gamma(1 - j + iq)} \right] + A, \qquad (3.39)$$

and $f^{(3)}$ is then obtained from (3.5).

The term with superscript (2) in (3.6) comes from $g^{(2)}$ which is given in (3.9) and can be rewritten as

$$g^{(2)}(\theta) = \frac{D_\epsilon^2}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq-1} t^{\epsilon - iq - 1}$$

$$\times (1 + t)\left[\frac{1 - t}{[(1 - t)^2 + 4tx^2]^{\frac{1}{2}}} - 1\right] + B, \qquad (3.40)$$

where the constant

$$B = \frac{2D_\epsilon^2}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq-1} t^{\epsilon - iq} \qquad (3.41)$$

is equivalent to the value $g^{(2)}(0)$.

The integral over x in the expression for A in (3.38) can be performed using (3.8), and leads to

$$A = -2\frac{D_\epsilon d_\mu}{\Gamma(1 + 2iq)} \int_0^1 dt(1 - t)^{2iq+\mu} t^{\epsilon - iq}, \qquad (3.42)$$

where the operator d_μ is defined as $\partial/\partial\mu \mid_{\mu=0}$. According to (3.6), $g^{(3)}$ and $g^{(2)}$ occur in the combination $g^{(3)} - g^{(2)}$. One can then write the terms in α^4 in the form (3.24) and (3.31). After much algebra, one finds

$$\bar{a}_j^{(4)} = \begin{cases} 0, & j = 0 \\ \dfrac{i\alpha^4}{4q^2} \sum_{n=1}^{i} \dfrac{\pi^{\frac{1}{2}}\Gamma(n - \frac{1}{2})(-1)^n(2iq + 3n - \frac{3}{2})\Gamma(1 - iq)\Gamma(\frac{1}{2} + iq)}{\Gamma(n)(n - \frac{1}{2})(2iq + 2n - 1)\Gamma(1 - n - iq)\Gamma(\frac{1}{2} + n + iq)}, & j \neq 0 \end{cases} \qquad (3.43)$$

$$\bar{b}_j^{(4)} = \begin{cases} -\dfrac{\pi^2\alpha^4}{8q^2}, & j = 0 \\ \dfrac{\alpha^4}{8q^2}\left(-\pi^2 + \sum_{n=1}^{i} \dfrac{\pi^{\frac{1}{2}}\Gamma(n)(2iq + 3n)\Gamma(1 + iq)\Gamma(\frac{1}{2} - iq)}{n\Gamma(n + \frac{1}{2})(iq + n)\Gamma(1 + iq + n)\Gamma(\frac{1}{2} - iq - n)}\right), & j \neq 0 \end{cases} \qquad (3.44)$$

$$\bar{c}_j^{(4)} = \begin{cases} -\dfrac{i\alpha^4\pi}{4q^2}\left(\dfrac{\Gamma'(1 + 2iq)}{\Gamma(1 + 2iq)} - \dfrac{\Gamma'(1 - iq)}{\Gamma(1 - iq)}\right), [11] & j = 0 \\ -\dfrac{\alpha^4}{8q} \sum_{n=0}^{i} \dfrac{\pi^{\frac{1}{2}}(-1)^n\Gamma(\frac{3}{2} + n)(iq - 3n - 3)\Gamma(iq - n - \frac{1}{2})\Gamma(n + 1 - iq)}{\Gamma(n + 2)(n + 1)(iq - n - 1)\Gamma(iq + \frac{1}{2})\Gamma(1 - iq)}, & j \neq 0. \end{cases} \qquad (3.45)$$

The quantities $a_j^{(4)}$, $b_j^{(4)}$, and $c_j^{(4)}$ are obtained from (3.43)–(3.45) by using (3.32)–(3.34).

A study of the form of the coefficients leads to the following conclusions:

(1) The form of F_1/F_0 and G_1/G_0 to all orders of α^2 is given by (3.24) and (3.31).

(2) The terms not containing the factor x^{-2iq} start with successively higher powers of x as one goes to higher order in α^2. For example,

$$a_0^{(4)} = 0, \quad a_0^{(6)} = b_0^{(6)} = 0, \quad a_0^{(8)} = a_1^{(8)} = b_0^{(8)} = 0.$$

(3) The expressions given are convergent for all values of x except $x = 1$. For this value of $x(\theta = \pi)$ alternate expressions obtained in the following sub-section are valid.

(4) The terms containing the factor x^{-2iq} all contain an additional factor $e^{\pi q}/\sinh \pi q$. Since the factor x^{-2iq} is rapidly oscillating near $\theta = 0$, the character of the Coulomb amplitudes for large posi-

---

[11] The quantity

$$\frac{\Gamma'(1 + 2iq)}{\Gamma(1 + 2iq)} - \frac{\Gamma'(1 - iq)}{\Gamma(1 - iq)}$$

may also be written as

$$\frac{1}{2}\frac{d}{dq} \arg \Gamma(1 + 2iq) - \frac{d}{dq} \arg \Gamma(1 + iq)$$

$$+ \frac{1}{2}\left\{ \frac{3}{2iq} + \frac{i\pi}{2\tanh \pi q}(3 + \tanh^2 \pi q)\right\},$$

and can also be expressed in terms of $d\psi/dq$ by means of (3.28).

tive and negative q will be substantially different. This has been pointed out by Fradkin, Weber, and Hammer,[6] and also by Rawitscher,[12] who note the absence of oscillations in the nonrelativistic limit (large $|q|$) for the repulsive Coulomb case ($q < 0$).

c. Values at $\theta = \pi$

To complete the expressions given in (3.24) and (3.31) the value of F_1/F_0 at $\theta = \pi$ accurate to order α^2 will be obtained. This term is given by

$$F_1(\pi) \simeq \frac{i\alpha^2}{4} f^{(1)}(\pi)$$

$$= \frac{-i\alpha^2}{2} \frac{D_\epsilon}{\Gamma(1 + 2iq)} \int_0^1 dt \frac{(1 - t)^{2iq} t^{\epsilon - iq}}{1 + t}$$

$$= -\frac{i\alpha^2}{4} \sum_{j=0}^\infty \frac{\Gamma(2iq + j + 1)}{\Gamma(2iq + 1)} \frac{1}{2^j}$$

$$\times D_\epsilon \frac{\Gamma(\epsilon - iq + 1)}{\Gamma(j + iq + \epsilon + 2)}. \qquad (3.46)$$

One then has

$$\frac{F_1(\pi)}{F_0(\pi)} = -\frac{\alpha^2}{2} \sum_{j=0}^\infty \frac{\Gamma(2iq + j + 1)\Gamma(1 + iq)}{\Gamma(2iq + 1)\Gamma(2 + iq + j)} \frac{1}{2^j}$$

$$\times \left[-i\pi + \frac{\Gamma'(1 - iq)}{\Gamma(1 - iq)} - \frac{\Gamma'(2 + iq + j)}{\Gamma(2 + iq + j)} \right]. \qquad (3.47)$$

4. INTEGRAL REPRESENTATION FOR $F(\theta)$

In order to complete the more-formal aspects of our work we have obtained an integral representation of the function $F(\theta)$ which separates the dependence on the parameters α and q and which exhibits the analytic properties in the $x = \sin \frac{1}{2}\theta$ plane. The technique used to obtain the result parallels that of Rosen,[5] who obtains representations as double integrals for the individual terms in an expansion in power of α^2. We obtain a double integral for $F(\theta)$ without expansion in α or q.[13]

For this purpose we write $F(\theta)$ as

$$F(\theta) = \frac{i}{2} \sum_{n=1}^\infty n C_n (P_n - P_{n-1})$$

$$= -\frac{i}{2} (1 - z) \frac{d}{dz} \sum_{n=1}^\infty C_n (P_n + P_{n-1}), \qquad (4.1)$$

where $z = \cos \theta$ and C_n is given in (2.3). Applying the Sommerfeld–Watson transformation one has

$$F(\theta) = \frac{1}{4}(1 - z)$$

$$\times \frac{d}{dz} \int_C \left[\frac{P_\nu(-z) - P_{\nu-1}(-z)}{\sin \nu\pi} \right] C(\nu) \, d\nu, \qquad (4.2)$$

---

[12] G. Rawitscher, Phys. Letters **9**, 337 (1964).
[13] The possibility of such an integral representation was pointed out by L. Brown (private communication) who obtained the corresponding result for the Klein–Gordon equation.
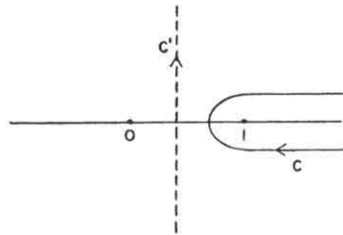
FIG. 3. The contour C and C' for $F(\theta)$.

where the contour C is shown as the solid curve in Fig. 3. The quantity $C(\nu)$ is given by

$$C(\nu) = -e^{-i(\rho - \nu)\pi}$$

$$\times \Gamma(\rho(\nu) - iq)/\Gamma(\rho(\nu) + iq + 1), \qquad (4.3)$$

where $\rho(\nu) = (\nu^2 - \alpha^2)^{\frac{1}{2}}$. It can be shown that both $P_\nu(z)$ and $C(\nu)$ are bounded for large ν in the region of interest and that the contour can be deformed to the straight-line contour C'.

We now use the integral representation

$$\frac{1}{\sin \nu\pi} \frac{d}{dz} P_\nu(-z) = -\frac{1}{\pi} \int_0^\infty \frac{ds \, s^{-\nu}}{(s^2 + 1 - 2sz)^{\frac{3}{2}}}, \qquad (4.4)$$

which converges for $-2 < \text{Re } \nu < 1$. This leads to

$$F(\theta) = \frac{1}{4\pi} (1 - z) \int_0^\infty \frac{ds(1 + s)}{(s^2 + 1 - 2sz)^{\frac{3}{2}}}$$

$$\times \int_{C'} d\nu C(\nu) s^{-\nu}. \qquad (4.5)$$

The poles in the ν plane are located at

$$\nu^2 = \alpha^2 + (m + 1 + iq)^2, \quad m = 0, 1, 2 \cdots$$

which are all to the left of the line $\text{Re } \nu = 1$. The contour C' may therefore be closed to the right giving a vanishing contribution for $s > 1$. If one then uses the beta-function integral representation for $C(\nu)$, one has

$$F(\theta) = -\frac{1}{4\pi} (1 - z) \frac{1}{\Gamma(2iq + 1)} \int_0^1 \frac{ds(1 + s)}{(s^2 + 1 - 2sz)^{\frac{3}{2}}}$$

$$\times \int_0^1 dt \frac{d}{dt} \lfloor (1 - t)^{2iq} t^{-iq} \rfloor$$

$$\times \int_{C'} \frac{d\nu}{\rho} e^{-i(\rho - \nu)\pi} e^{-\nu \ln s + \rho \ln t}, \qquad (4.6)$$

where an integration by parts has been performed in the variable t. The behavior of the integrand for large ν is governed by the exponent

$$e^{-\nu \ln s + \rho \ln t}$$

with $\rho \simeq \nu$. In the right half-plane, the integrand tends to zero if $s > t$. Contributions to the integral therefore come only for $t > s$, and these may be evaluated by closing the contour to the left to form

a loop around the branch points at $\nu = \pm \alpha$ which are the only singularities remaining in the ν plane. Setting

$$\nu = \alpha \cos \phi, \qquad \rho = i\alpha \sin \phi,$$

the integral over the loop becomes

$$i \int_{-\pi}^{\pi} d\phi \, e^{(-i\pi + \ln t) i\alpha \sin \phi + (i\pi - \ln s) \alpha \cos \phi}$$

$$= 2\pi i I_0 \{\alpha [(i\pi - \ln s)^2 - (i\pi - \ln t)^2]^{\frac{1}{2}}\}. \quad (4.7)$$

The final integral representation for $F(\theta)$ is therefore

$$F(\theta) = -\frac{i}{2} (1-z) \frac{1}{\Gamma(2iq+1)} \int_0^1 \frac{ds(1+s)}{(s^2+1-2sz)^{\frac{3}{2}}}$$

$$\times \int_s^1 dt \frac{d}{dt} [(1-t)^{2iq} t^{-iq}]$$

$$\times I_0 \{\alpha [(i\pi - \ln s)^2 - (i\pi - \ln t)^2]^{\frac{1}{2}}\}. \quad (4.8)$$

This integral representation clearly separates the α and q dependence and suggests that the analytic behavior of $F(\theta)$ in the variables α and q will be different.

The analytic form of $F(\theta)$ in the variable $x = \sin \frac{1}{2}\theta$ will be the same as that given in the α^2 and α^4 terms already obtained. In fact the same contour deformation as in Sec. 3 can be used in the s plane to demonstrate that $F(\theta)$ can be separated into the form in (3.24) for arbitrary α. The coefficients of successive powers of α^2 can be obtained by using the power series expansion for I_0. The individual terms present themselves as double integrals in this formulation, as in the work of Rosen.[5] The terms in α^2 and α^4 have been reduced after much labor to those obtained in (3.24)–(3.34) and (3.43)–(3.45). Higher terms can similarly be obtained, but the results do not appear to be simple, nor do they seem to be easier to calculate in this way than with the method of Sec. 3.

5. RELATION WITH BORN EXPANSIONS

At this point it is useful to discuss the relationship of our present results with the previous Born expansion[2,3] results obtained by simultaneous expansion in powers of α and $q = \alpha/\beta$. The coefficients obtained in this way have logarithmic dependence on the angle, a result which is easy to understand since the expansion of x^{2iq} in powers of α is

$$x^{2iq} = \sum_{m=0}^{\infty} \frac{(2i)^m}{m!} \left(\frac{\alpha}{\beta}\right)^m (\ln x)^m. \quad (5.1)$$

A useful prescription for Coulomb-type problems is therefore to extract the phase $2iq \ln (\sin \frac{1}{2}\theta)$ wherever the logarithmic factors occur. If this procedure is applied to the Born expansions, agreement is obtained with our results to all orders available.

It is also clear at this point that the analytic behavior obtained here for the Coulomb amplitudes of the Dirac equation must be quite similar to that for the Klein–Gordon equation, where the phase shifts are given by

$$e^{2i\eta_l} = e^{-i\pi(\rho' - n')}$$

$$\times \Gamma(\rho' + \frac{1}{2} - iq)[\Gamma(\rho' + \frac{1}{2} + iq)]^{-1} \quad (5.2)$$

with

$$n' = l + \frac{1}{2}, \qquad \rho' = (n'^2 - \alpha^2)^{\frac{1}{2}}.$$

These results are discussed for large $|q|$ by Fradkin, Weber and Hammer,[6] and by Rawitscher.[12]

ACKNOWLEDGMENTS

Possible Relationship between Electric Charge and Dual Charge*

Elihu Lubkin

Department of Physics, Brown University, Providence, Rhode Island
(Received 21 April 1964)

Electric charge has no direct meaning for strong interactions, yet is involved in the octet model in such a way that the size of the elementary unit of electric charge becomes tied to the topology of the group of strong interaction symmetries. This tie may indicate a relationship of electric charge to another kind of charge got directly from topology of the group.

1. INTRODUCTION

IT has been noted by Gell-Mann[1] that the theory of $SU(3)$ symmetry and the identification of the operator

$$Z = \tfrac{1}{2}Y + I_3 \tag{1}$$

as the charge in units of that of the positron $Z = Q/e$, suggest the existence of particles of charge $\pm\tfrac{1}{3}e$, "subelectrons." This comes from the possibility that arbitrary representations $D(p, q)$ of $SU(3)$ are realized by multiplets of particles.

On the other hand, there is an "octet model" which rules out subelectron representations. The octet model may be posited in several ways. Briefest is the rule[2]

$$p \equiv q \bmod 3 \tag{2}$$

for a physical representation. Equivalently, that all physical or "octet" representations be obtained from $D(1, 1)$, the octet, by tensor product and reduction. Justifications for (2) in the form of physical models arise because tensor product and reduction appear in the discussion of bound states. One viewpoint is that the basic particles belong to $D(1, 1)$ octets. Another[3] is that the subelectron multiplets are missed experimentally, perhaps because they are associated with high mass, and all ordinary particles are got from $D(1, 0) \otimes D(0, 1) = D(0, 0) \oplus D(1, 1)$ binding. If the distinction between basic and compound particles is removed, however, these models leave Rule (2) an undigested empirical fact.

The subelectron matrix representations are also characterized by being triple-valued representations of $pSU(3)$, the factor group of $SU(3)$ by its center C, whereas octet representations are single-valued. This,

though generally known,[4] will be explained in Sec. 2, because it is closely related to the point raised here. Namely, this distinction between subelectron and octet representations is a direct manifestation of the fact that $pSU(3)$ is not simply connected. A similar manifestation is the notion of "dual charge."[5] The first of these is a mathematical fact, the second is related to a conserved quantity and is therefore more of a physical hypothesis. It is unfortunately not completely physical, as dual charge also entails topological complexity of ordinary space–time, whereas there is no scheme of calculation wedding quantum mechanics to topological complexity of space–time. The point of this note is to suggest a relationship between subelectrons and dual charge. It will be seen that a difference exists in a classical theory between fields belonging to subelectron representations and those belonging to others, if it is postulated that dual charge and electric charge are essentially the same. Whether this distinction is reflected in mass differences, in a law against subelectrons, or in some other way, is hard to say.

2. REVIEW OF THE DISTINCTION BETWEEN SUBELECTRON AND OCTET REPRESENTATIONS OF $SU(3)$

$SU(3) = D(1, 0)$ is the group of all 3 by 3 complex unitary matrices of determinant 1. Corresponding infinitesimal operators, defined with an "i" factored out, are the Hermitian matrices of trace 0, $su(3)$. By leaving one "third" complex dimension in the representation space alone, one finds an $SU(2)$ subgroup of isotopic spin $\tfrac{1}{2}$;

$$I_3 = \operatorname{diag}\left(\tfrac{1}{2}, -\tfrac{1}{2}, 0\right), \tag{3}$$

[1] Lectures of T. D. Lee and S. L. Glashow (unpublished).
[2] J. J. de Swart, Rev. Mod. Phys. 35, 916 (1963), Eq. (7.4). The argument and matrices at the beginning of Sec. 2 are adapted from de Swart.
[3] This idea of M. Gell-Mann is said to have motivated his "Eightfold Way."

[4] The author gratefully acknowledges information to this effect from S. L. Glashow. The referee has provided the following references: M. Gourdin, Nuovo Cimento 30, 587, (1963); L. C. Biedenharn and E. C. Fowler, "Fractional Charges in the $SU3$ Scheme", Duke University preprint, 1963; and C. R. Hagen and A. J. MacFarlane, "Quarks, Triality, and Unitary Symmetry Schemes" (Rochester University preprint, 1964).
[5] E. Lubkin, Ann. Phys. (N. Y.) 23, 233 (1963), especially pp. 263–273.

and the $su(3)$ matrices which commute with the isotopic spin matrices are all proportional to

$$Y = \text{diag} \left(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}\right). \quad (4)$$

The requirement that the hypercharge operator Y commute with I_1, I_2, I_3, and belong to $su(3)$ fixes (4) up to a scale factor, which is then adjusted to fit Eq. (1) with Z's which match nuclear physics and the successful assignments of particles to $D(1, 1)$ and $D(3, 0)$ multiplets. From (1), (3), (4),

$$Z = \text{diag} \left(\tfrac{2}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}\right), \quad (5)$$

and in particular, an eigenvalue $Z' \equiv -\tfrac{1}{3}$ mod 1.

The representation $D(p, q)$ is obtained from the tensor product of p factors $D(1, 0)$ with q of the contragredient representation $D(0, 1)$ of transposed inverse matrices, equivalently complex-conjugate matrices, by subjecting the representation space of tensors to symmetry conditions. Since complex conjugation corresponds to reversing sign in $su(3)$, the matrices in $D(0, 1)$ for I_3, Y, Z are the negatives of (3)–(5). In particular, $Z' \equiv \tfrac{1}{3}$ mod 1 in $D(0, 1)$. Eigenvalues of infinitesimal operators add when tensor products are taken, so

$$Z' \equiv -\tfrac{1}{3}(p - q) \text{ mod 1}, \quad (6)$$

generally, whence the rule (2) against subelectrons.

The center C of $SU(3)$ consists of its constant matrices, by Schur's lemma. The condition that the determinant be 1 shows that these are only $M(\lambda) = \text{diag} (\lambda, \lambda, \lambda)$, where $\lambda = 1$, ω, or ω^2, and $\omega = \exp(-i\tfrac{2}{3}\pi)$. Observe that

$$M(\omega) = \exp(-i2\pi Z). \quad (7)$$

Topology of the group is brought into the discussion through the following circumstance. Most physical $SU(3)$ arguments really concern only $su(3)$, so that it is of interest to consider replacing $SU(3)$ by another connected compact Lie group G with an infinitesimal algebra isomorphic to $su(3)$.

The only other such group is $pSU(3)$: $SU(3)$ is simply connected.[6] All G are obtained from the simply connected one by congruence modulo a discrete central subgroup.[7] But the only central subgroups are the identity alone, which gives back $SU(3)$, and the whole three-element center C, which gives $pSU(3)$.[8]

[6] C. C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), p. 60.
[7] Reference 6, p. 49.
[8] $pSU(3) \cong D(1, 1)$ is also obtained as the adjoint representation, and by asking for the action of $SU(3)$ [or even of $U(3)$ or $GL(3)$] on rays instead of vectors, whence the notation "p" for "projective", suggested by Dr. Bruno Harris of the Department of Mathematics.

Topology enters because the discrete central subgroup employed in the construction is also the first homology group of the group G: In $pSU(3)$, there are three types of closed paths, corresponding to a curve beginning at $M(1)$ and ending at $M(1)$, $M(\omega)$, or $M(\omega^2)$ in $SU(3)$ itself; only the first of these types may be shrunk to a point. A matrix representation R of $su(3)$ if extended over $pSU(3)$ is single-valued if and only if the matrix $R[M(\lambda)]$ representing $M(\lambda)$ is independent of λ; i.e., if $R[M(\omega)]$ is the appropriate unit matrix. The other matrix representations are triple valued.

In $D(p, q)$, $M(\omega)$ is represented by $\omega^p(\omega^*)^q = \omega^{p-q}$ times a unit matrix; this also from (6) and (7). If $\omega^{p-q} = \omega^{-3Z'}$ is the "type" of the representation, then Type 1 corresponds to single-valued representations of $pSU(3)$ and rule (2), whereas the triple-valued representations are those of type ω and ω^2, $p \equiv q \pm 1$ mod 3, $Z' \equiv \mp\tfrac{1}{3}$ mod 1, respectively Subelectron representations correspond to $pSU(3)$-multivalued types.

3. $pSU(3)$

It would be premature to take single-valuedness of a $D(p, q)$ considered as a $pSU(3)$ representation as a law for nonexistence of subelectrons. An analogous law for the group of rotations in ordinary space, $O(3) \cong pSU(2)$, would forbid half-integral spin. Furthermore, the use of the representations as ray representations or projective representations in quantum mechanics collapses their multivaluedness: the three representative matrices for an element of $pSU(3)$ in a multivalued ordinary matrix representation are obtained from one of them by multiplying by the phase factors 1, ω, and ω^2; and are therefore identical in their transformation of rays. In fact, because of the requirement of unitarity and the irrelevance of a uniform phase factor, the reduction[8] $GL(n) \to pSU(n)$ together with the loss of multivaluedness in virtue of the replacement of matrix representations by ray representations is the largest symmetry group conceivable belonging to an n-level quantum mechanical system, together with the correct description of the representations as they are used.

The only remaining general feature known to the author and associated with the three-element group that might have physical meaning is dual charge.

4. REVIEW OF DUAL CHARGE[5]

Suppose $su(3)$ symmetry is taken seriously, in spite of the fact of broken symmetry, so that a choice of a reference frame in an attached $D(p, q)$

representation space of "tensors" at a space–time point is subject to a $D(p, q)$ ambiguity, and that following the philosophy of general relativity as explained by Yang and Mills,[9] a parallel displacement field (gauge field) must be introduced to define an absolute (covariant) derivative, and that this absolute derivative, not the literal derivative, must appear in some unnamed Lagrangian which is to tie the theory to physics. The parallel displacement assigns a $D(p, q)$ matrix to each loop in ordinary space, with the ambiguity of an unimportant conjugation which will not be mentioned further. That $D(p, q)$ matrix explains how a tensor is altered when displaced parallelly around the loop. Consider a spherical bag in ordinary space as a sequence of parallels of latitude, each one of which is a loop, beginning with a degenerate or point loop (north pole) and ending with another (south pole). Each parallel of latitude corresponds to a $D(p, q)$ matrix, the poles to the unit matrix. Therefore, the bag as a whole corresponds to a loop L of $D(p, q)$ matrices, beginning and ending at the unit matrix. This loop L can be broken into a succession of many small transformations, each one, roughly speaking, associated with an element of $su(3)$. By putting these small transformations together in $SU(3)$, the loop L is imaged back in $SU(3)$, so that it appears as a path P from the identity $M(1)$ to either $M(1)$, $M(\omega)$, or $M(\omega^2)$. In the first case, the dual charge D contained in the bag is defined to be 0, in the second case it is -1, in the third case, 1.

If the bag can be continuously deformed in ordinary space to a point, without running through singularities of the displacement or tensor fields, then $D = 0$, so that speculations on $D \neq 0$ entail singularities or wormholes in ordinary space.

The dual charge of a system is the sum of the dual charges of its components, reduced modulo 3.

5. POSSIBLE ARGUMENT AGAINST SUBELECTRONS

If $p \not\equiv q \bmod 3$, then the path P in $SU(3)$ must terminate at $M(1)$, and $D = 0$. This is because, if the path P is not a closed loop, then L is not, as the representation map $SU(3) \to D(p, q)$ is an isomorphism. Now suppose space should be imagined as a froth of wormholes,[10] with D's having all possible values. Then $D(p, q)$ tensors can't be supported unless $p \equiv q \bmod 3$.

However, if the picture of a tensor field is replaced by a "triad" field of unordered triads of tensors, mutually differing by factors of ω, ω^2, $D = 0$ corresponds to the three sheets being separate; $D = \pm 1$ to their being united in one "Riemann surface." This possibility should be considered, because the multivaluedness collapses on going to ray representations, and is therefore not excluded in the sense of quantum mechanics even by a law of a single-valuedness for quantum fields.

The surface of a bag is simply connected. Therefore, it cannot support three attached sheets which are separate over each point.[11] $D \neq 0$ triad fields therefore entail at least one branch point on a bag. To avoid singularity, the magnitude of the triad field must vanish there. We are reminded of Dirac's nodal lines.[12]

6. POSSIBLE IDENTIFICATION OF DUAL CHARGE WITH ELECTRIC CHARGE

If

$$\tfrac{1}{3}D \equiv Z' \bmod 1. \tag{8}$$

or if $\tfrac{1}{3}D \equiv -Z' \bmod 1$, so that electric charge is a refinement of the crude topological notion of dual charge, then all subelectrons possess $D = \pm 1$. Even without requiring the vacuum to be a froth of wormholes, the option of using a triad field would then be necessary to realize a single-valued quantum field for a subelectron, even if the problem of writing such a field is restricted to the surface of a sphere of large radius, which contains a single subelectron. If a triad field with its three branches united is for some reason impossible, the subelectron would become impossible under (8) at the same time that the basis (8) of the argument becomes vacuous!

7. RELATION TO GRAVITY

It is hard to imagine (8) without a wormhole model for charge in general. Nevertheless, the inability to shrink a bag to a point is definitely argued only for $D \neq 0$, so that under the hypothesis (8), only subelectrons would necessarily be associated with wormholes. If the reason for the unshrinkability is singularities, then the metric of space–time would be expected to depart greatly from flat-space values. If the reason is that nonsingular fields are distributed over a space–time manifold with an actual "neck," again a large distortion would be expected in the ordinary metric. If such distortion is pictured physically as a strong gravitational field, and if the mass tensor is the only source for the development of this field, then even if the phenom-

[9] C. N. Yang and R. L. Mills, Phys. Rev. 96, 191 (1954).
[10] J. A. Wheeler, *Geometrodynamics* (Academic Press Inc., New York, 1962).

[11] Reference 6, p. 46, Theorem 2.
[12] P. A. M. Dirac, Proc. Roy. Soc. (London) 133, 60 (1931).

enon has no essential roots in gravity theory, there is an argument here for large concentration of mass. This argument cannot be converted into any fantastically large mass, however, because the large concentration of mass may be confined to an arbitrarily small region.

In greater detail, a mass m yields a Compton wavelength m^{-1} and a gravitational radius proportional to m. These are equal[13] for $m = 2.2 \times 10^{-5}\,\mathrm{g} \sim 10^{19}\,\mathrm{BeV}$. But this fantastic mass drops out of the picture if the Compton wavelength of the particle is allowed to greatly exceed the characteristic dimension of the wormhole; if the wormhole is quantum mechanically smeared in its mean position.[14]

8. DUAL CHARGE FOR THE ROTATION GROUP

Conserved quantum numbers not incorporated in any group-theoretical scheme are atomic number, electron number, and muon number. These numbers are all associated with fermions, which belong to double-valued representations of the rotation group, $O(3)$. The use of spinor doublets in place of spinor fields changes the situation from that of no dual charge for ordinary fermion fields[15] to possible dual charge for all fields, and then by sheer analogy to (8), perhaps to dual charge only for fermion fields, or only for some fermion fields. For $O(3)$, dual charge is of course an integer modulo 2, but may be a crude representation of one of the unclassified free integer conserved quantities. Of course, the conservation of fermion number modulo 2 follows simply from conservation of angular momentum, and in itself needs no explanation; the above indicates merely the possibility of a parallel in $O(3)$ of the previous discussion. The relation of the rotation group to the 10-parameter Lorentz group which ties Fermi statistics to half-integral spin, has not been explored in relation to the above.

[13] Reference 10, p. 77.
[14] Compare Reference 10, p. 82.

[15] Reference 5, p. 271.

Coulomb Green's Function

JULIAN SCHWINGER
Harvard University, Cambridge, Massachusetts
(Received 19 June 1964)

A one-parameter integral representation is given for the momentum space Green's function of the nonrelativistic Coulomb problem.

IT has long been known that the degeneracy of the bound states in the nonrelativistic Coulomb problem can be described by a four-dimensional Euclidean rotation group, and that the momentum representation is most convenient for realizing the connection. It seems not to have been recognized, however, that the same approach can be used to obtain an explicit construction for the Green's function of this problem. The derivation[1] is given here.

The momentum representation equation for the

[1] It was worked out to present at a Harvard quantum mechanics course given in the late 1940's. I have been stimulated to rescue it from the quiet death of lecture notes by recent publications in this Journal, which give alternative forms of the Green's function: E. H. Wichmann and C. H. Woo, J. Math. Phys. **2**, 178 (1961); L. Hostler, *ibid.* **5**, 591 (1964).

Green's function is ($\hbar = 1$)

$$\left(E - \frac{p^2}{2m}\right)G(\mathbf{p}, \mathbf{p}') + \frac{Ze^2}{2\pi^2} \int (d\mathbf{p}'') \frac{1}{(\mathbf{p} - \mathbf{p}'')^2}$$
$$\times G(\mathbf{p}'', \mathbf{p}') = \delta(\mathbf{p} - \mathbf{p}').$$

We shall solve this equation by assuming, at first, that

$$E = -(p_0^2/2m)$$

is real and negative. The general result is inferred by analytic continuation.

The parameters

$$\xi = -\frac{2p_0\mathbf{p}}{p_0^2 + p^2}, \qquad \xi_0 = \frac{p_0^2 - p^2}{p_0^2 + p^2}$$

define the surface of a unit four-dimensional Eu-

enon has no essential roots in gravity theory, there is an argument here for large concentration of mass. This argument cannot be converted into any fantastically large mass, however, because the large concentration of mass may be confined to an arbitrarily small region.

In greater detail, a mass m yields a Compton wavelength m^{-1} and a gravitational radius proportional to m. These are equal[13] for $m = 2.2 \times 10^{-5} \, \text{g} \sim 10^{19} \, \text{BeV}$. But this fantastic mass drops out of the picture if the Compton wavelength of the particle is allowed to greatly exceed the characteristic dimension of the wormhole; if the wormhole is quantum mechanically smeared in its mean position.[14]

8. DUAL CHARGE FOR THE ROTATION GROUP

Conserved quantum numbers not incorporated in any group-theoretical scheme are atomic number, electron number, and muon number. These numbers are all associated with fermions, which belong to double-valued representations of the rotation group, $O(3)$. The use of spinor doublets in place of spinor fields changes the situation from that of no dual charge for ordinary fermion fields[15] to possible dual charge for all fields, and then by sheer analogy to (8), perhaps to dual charge only for fermion fields, or only for some fermion fields. For $O(3)$, dual charge is of course an integer modulo 2, but may be a crude representation of one of the unclassified free integer conserved quantities. Of course, the conservation of fermion number modulo 2 follows simply from conservation of angular momentum, and in itself needs no explanation; the above indicates merely the possibility of a parallel in $O(3)$ of the previous discussion. The relation of the rotation group to the 10-parameter Lorentz group which ties Fermi statistics to half-integral spin, has not been explored in relation to the above.

---

[13] Reference 10, p. 77.
[14] Compare Reference 10, p. 82.

[15] Reference 5, p. 271.

Coulomb Green's Function

JULIAN SCHWINGER
Harvard University, Cambridge, Massachusetts
(Received 19 June 1964)

A one-parameter integral representation is given for the momentum space Green's function of the nonrelativistic Coulomb problem.

IT has long been known that the degeneracy of the bound states in the nonrelativistic Coulomb problem can be described by a four-dimensional Euclidean rotation group, and that the momentum representation is most convenient for realizing the connection. It seems not to have been recognized, however, that the same approach can be used to obtain an explicit construction for the Green's function of this problem. The derivation[1] is given here.

The momentum representation equation for the

---

[1] It was worked out to present at a Harvard quantum mechanics course given in the late 1940's. I have been stimulated to rescue it from the quiet death of lecture notes by recent publications in this Journal, which give alternative forms of the Green's function: E. H. Wichmann and C. H. Woo, J. Math. Phys. **2**, 178 (1961); L. Hostler, *ibid.* **5**, 591 (1964).

Green's function is ($\hbar = 1$)

$$\left(E - \frac{p^2}{2m}\right)G(\mathbf{p}, \mathbf{p}') + \frac{Ze^2}{2\pi^2} \int (d\mathbf{p}'') \frac{1}{(\mathbf{p} - \mathbf{p}'')^2}$$
$$\times G(\mathbf{p}'', \mathbf{p}') = \delta(\mathbf{p} - \mathbf{p}').$$

We shall solve this equation by assuming, at first, that

$$E = -(p_0^2/2m)$$

is real and negative. The general result is inferred by analytic continuation.

The parameters

$$\xi = -\frac{2p_0\mathbf{p}}{p_0^2 + p^2}, \qquad \xi_0 = \frac{p_0^2 - p^2}{p_0^2 + p^2}$$

define the surface of a unit four-dimensional Eu-

clidean sphere,

$$\xi_0^2 + \xi^2 = 1,$$

the points of which are in one to one correspondence with the momentum space. The element of area on the sphere is

$$d\Omega = \frac{(d\xi)}{|\xi_0|} = \left(\frac{2p_0}{p_0^2 + p^2}\right)^3 (d\mathbf{p}),$$

if one keeps in mind that $p \gtrless p_0$ corresponds to the two semispheres $\xi_0 = \mp(1 - \xi^2)^{\frac{1}{2}}$. As another form of this relation, we write the delta function connecting two points on the unit sphere as

$$\delta(\Omega - \Omega') = \left(\frac{p_0^2 + p^2}{2p_0}\right)^3 \delta(\mathbf{p} - \mathbf{p}').$$

Next, observe that

$$(\xi - \xi')^2 = (\xi_0 - \xi_0')^2 + (\xi - \xi')^2$$

$$= \frac{4p_0^2}{(p_0^2 + p^2)(p_0^2 + p'^2)} (\mathbf{p} - \mathbf{p}')^2.$$

Then, if we define

$$\Gamma(\Omega, \Omega') = -\frac{1}{16mp_0^3} (p_0^2 + p^2)^2 G(\mathbf{p}, \mathbf{p}')(p_0^2 + p'^2)^2,$$

that function obeys a four-dimensional Euclidean surface integral equation,

$$\Gamma(\Omega, \Omega') - 2\nu \int d\Omega'' \, D(\xi - \xi'')\Gamma(\Omega'', \Omega')$$

$$= \delta(\Omega - \Omega'),$$

where

$$D(\xi - \xi') = \frac{1}{4\pi^2} \frac{1}{(\xi - \xi')^2} ,$$

and

$$\nu = Ze^2m/p_0.$$

The function D that is defined similarly throughout the Euclidean space is the Green's function of the four-dimensional Poisson equation,

$$-\partial^2 D(\xi - \xi') = \delta(\xi - \xi').$$

It can be constructed in terms of a complete set of four-dimensional solid harmonics. In the spherical coordinates indicated by ρ, Ω, these are

$$(\rho^{n-1}, \rho^{-n-1}) Y_{nlm}(\Omega), \qquad n = 1, 2, \cdots,$$

where the quantum numbers l, m provide a three-dimensional harmonic classification of the four-dimensional harmonics. The largest value of l contained in the homogeneous polynomial $\rho^{n-1}Y_{nlm}(\Omega)$ is the degree of the polynomial, $n - 1$. Thus,

$$-l \le m \le l, \qquad 0 \le l \le n - 1$$

label the n^2 distinct harmonics that have a common value of n.

The Green's function D is exhibited as

$$D(\xi - \xi') = \sum_{n=1}^{\infty} \frac{\rho_<^{n-1}}{\rho_>^{n+1}} \frac{1}{2n} \sum_{lm} Y_{nlm}(\Omega) Y_{nlm}(\Omega')^*,$$

where

$$\delta(\Omega - \Omega') = \sum_{nlm} Y_{nlm}(\Omega) Y_{nlm}(\Omega')^*$$

conveys the normalization and completeness of the surface harmonics. One can verify that D has the radial discontinuity implied by the delta function inhomogeneity of the differential equation,

$$-\rho^3 \frac{\partial}{\partial \rho} D(\xi - \xi') \Big]_{\rho'-0}^{\rho'+0} = \delta(\Omega - \Omega').$$

The function D is used in the integral equation for Γ with $\rho = \rho' = 1$. The equation is solved by

$$\Gamma(\Omega, \Omega') = \sum_{nlm} \frac{Y_{nlm}(\Omega) Y_{nlm}(\Omega')^*}{1 - (\nu/n)}.$$

The singularities of this function at $\nu = n = 1, 2, \cdots$ give the expected negative energy eigenvalues. The residues of G at the corresponding poles in the E plane provide the normalized wavefunctions, which are

$$\Psi_{nlm}(p) = \frac{4p_0^{5/2}}{(p_0^2 + p^2)^2} Y_{nlm}(\Omega),$$

$$p_0 = Ze^2m/n.$$

One can exhibit $\Gamma(\Omega, \Omega')$ in essentially closed form with the end of the expansion for D. We use the following version of this expansion:

$$\frac{1}{2\pi^2} \frac{1}{(1 - \rho)^2 + \rho(\xi - \xi')^2}$$

$$= \sum_{n=1}^{\infty} \rho^{n-1} \frac{1}{n} \sum_{lm} Y_{nlm}(\Omega) Y_{nlm}(\Omega')^*,$$

where ξ and ξ' are of unit length and $0 < \rho < 1$. Note, incidentally, that if we set $\xi = \xi'$ and integrate over the unit sphere, of area $2\pi^2$, we get

$$\frac{1}{(1 - \rho)^2} = \sum_{n=1}^{\infty} \rho^{n-1} \frac{1}{n} m_n,$$

where m_n is the multiplicity of the quantum number n. This confirms that $m_n = n^2$.

The identity

$$\frac{1}{1 - (\nu/n)} = 1 + \frac{\nu}{n} + \nu^2 \frac{1}{n(n - \nu)},$$

together with the integral representation

$$\frac{1}{n - \nu} = \int_0^1 d\rho \rho^{-\nu} \rho^{n-1},$$

valid for $\nu < 1$, gives

$$\Gamma(\Omega, \Omega') = \delta(\Omega - \Omega') + \frac{\nu}{2\pi^2} \frac{1}{(\xi - \xi')^2}$$

$$+ \frac{\nu^2}{2\pi^2} \int_0^1 d\rho \rho^{-\nu} \frac{1}{(1 - \rho)^2 + \rho(\xi - \xi')^2}. \quad (1)$$

Equivalent forms, produced by partial integrations, are

$$\Gamma(\Omega, \Omega') = \delta(\Omega - \Omega')$$

$$+ \frac{\nu}{2\pi^2} \int_0^1 d\rho \rho^{-\nu} \frac{d}{d\rho} \frac{\rho}{(1 - \rho)^2 + \rho(\xi - \xi')^2}, \quad (2)$$

and

$$\Gamma(\Omega, \Omega') = \frac{1}{2\pi^2} \int_0^1 d\rho \rho^{-\nu} \frac{d}{d\rho} \frac{\rho(1 - \rho^2)}{[(1 - \rho)^2 + \rho(\xi - \xi')^2]^2}, \quad (3)$$

which uses the limiting relation

$$\delta(\Omega - \Omega') = \lim_{\rho \to 1} \frac{1}{2\pi^2} \frac{1 - \rho^2}{[(1 - \rho)^2 + \rho(\xi - \xi')^2]^2}.$$

Note that Γ is a function of a single variable, $(\xi - \xi')^2$.

The restriction $\nu < 1$ can be removed by replacing the real integrals with contour integrals,

$$\int_0^1 d\rho \rho^{-\nu}(\ ) \to \frac{i}{2 \sin \pi\nu} e^{\pi i \nu} \int_C d\rho \rho^{-\nu}(\ ).$$

The path C begins at $\rho = 1 + 0i$, where the phase of ρ is zero and terminates at $\rho = 1 - 0i$, after encircling the origin within the unit circle.

The Green's function expressions implied by (1), (2), and (3) are

$$G(\mathbf{p}, \mathbf{p}') = \frac{\delta(\mathbf{p} - \mathbf{p}')}{E - T} - \frac{Ze^2}{2\pi^2} \frac{1}{E - T} \frac{1}{(\mathbf{p} - \mathbf{p}')^2} \frac{1}{E - T'}$$

$$- \frac{Ze^2}{2\pi^2} \frac{1}{E - T} \left[i\eta \int_0^1 d\rho \rho^{-i\eta} \right.$$

$$\left. \times \frac{1}{(\mathbf{p} - \mathbf{p}')^2 \rho - (m/2E)(E - T)(E - T')(1 - \rho)^2} \right] \frac{1}{E - T'}, \quad (1')$$

where

$$T = p^2/2m, \qquad \eta = -i\nu = Ze^2 m/k;$$

$$G(\mathbf{p}, \mathbf{p}') = \frac{\delta(\mathbf{p} - \mathbf{p}')}{E - T} - \frac{Ze^2}{2\pi^2} \frac{1}{E - T} \left[\int_0^1 d\rho \rho^{-i\eta} \frac{d}{d\rho} \right.$$

$$\left. \times \frac{\rho}{(\mathbf{p} - \mathbf{p}')^2 \rho - (m/2E)(E - T)(E - T')(1 - \rho)^2} \right] \frac{1}{E - T'}; \quad (2')$$

and

$$G(\mathbf{p}, \mathbf{p}') = -\frac{i}{4\pi^2} \frac{k}{E} \int_0^1 d\rho \rho^{-i\eta} \frac{d}{d\rho}$$

$$\times \frac{\rho(1 - \rho^2)}{[(\mathbf{p} - \mathbf{p}')^2 \rho - (m/2E)(E - T)(E - T')(1 - \rho)^2]^2}. \quad (3')$$

The Green's function is regular everywhere in the complex E plane with the exception of the physical energy spectrum. This consists of the negative-energy eigenvalues already identified and the positive-energy continuum. The integral representations $(1')$, $(2')$, and $(3')$ are not completely general since it is required that

$$\mathrm{Re}\, i\eta = -\mathrm{Im}\, \eta < 1.$$

As we have indicated, this restriction can be removed. It is not necessary to do so, however, if one is interested in the limit of real k. These representations can therefore be applied directly to the physical scattering problem.

The asymptotic conditions that characterize finite angle deflections are

$$E - T \sim 0, \qquad E - T' \sim 0, \qquad (\mathbf{p} - \mathbf{p}')^2 > 0.$$

The second of the three forms given for G is most convenient here. The asymptotic behavior is dominated by small ρ values, and one immediately obtains

$$G(\mathbf{p}, \mathbf{p}') \sim G^0(\mathbf{p})(-1/4\pi^2 m)f(\mathbf{p}, \mathbf{p}')G^0(\mathbf{p}'),$$

where

$$G^0(\mathbf{p}) = \frac{1}{E - T} \exp\left[-i\eta \log \frac{E - T}{4E} \right]\left(\frac{2\pi\eta}{e^{2\pi\eta} - 1} \right)^{\frac{1}{2}}$$

and

$$f(\mathbf{p}, \mathbf{p}') = \frac{2mZe^2}{(\mathbf{p} - \mathbf{p}')^2} \exp\left[-i\eta \log \frac{4k^2}{(\mathbf{p} - \mathbf{p}')^2} \right],$$

$$p^2 = p'^2 = k^2.$$

One would have found the same asymptotic form for any potential that decreases more rapidly than the Coulomb potential at large distances, but with $G^0(\mathbf{p}) = (E - T)^{-1}$. The factors $G^0(\mathbf{p}')$ and $G^0(\mathbf{p})$ describe the propagation of the particle before and after the collision, respectively, and f is identified as the scattering amplitude. The same interpretation is applicable here since the modified G^0 just incorporates the long-range effect of the Coulomb potential. This is most evident from the asymptotic behavior of the corresponding spatial function, which is a distorted spherical wave,

$$\int \frac{(d\mathbf{p})}{(2\pi)^3} e^{i\mathbf{p}\cdot\mathbf{r}} G(\mathbf{p})$$

$$\sim (-m/2\pi r) \exp\left[i(kr + \eta \log 2kr + \zeta) \right],$$

$$\zeta = \arg \Gamma(1 - i\eta).$$

The scattering amplitude obtained in this way coincides with the known result,

$$f(\vartheta) = (Ze^2/4E) \csc^2 \tfrac{1}{2}\vartheta \exp\left[-i\eta \log \csc^2 \tfrac{1}{2}\vartheta \right].$$

On the Construction of a Unitary Matrix with Elements of Given Moduli

MATTS ROOS
NORDITA, Copenhagen, Denmark
(Received 12 June 1964)

Given n^2 nonnegative real numbers u_{jk} which form a matrix with row and column vectors of unit magnitude, it is shown under what conditions there exists a unitary matrix (U_{jk}), such that $|U_{jk}| = u_{jk}$. The results may be shown to contain a theorem on unitary matrices.

THE formulation of a physical test of the quantal superposition principle[1] is dependent on the solution of the following mathematical problem: Given n^2 nonnegative real numbers u_{jk} which obey the $2n - 1$ independent relations

$$\sum_{j=1}^{n} u_{jk}^2 = \sum_{k=1}^{n} u_{jk}^2 = 1, \tag{1}$$

when is it possible to construct a unitary matrix (U_{jk}), such that

$$|U_{jk}| = u_{jk}? \tag{2}$$

Alternatively, the question may be formulated: When do there exist real solutions v_{jk} to the set of equations

$$\sum_{j=1}^{n} u_{jk}u_{jl} \exp i(v_{jl} - v_{jk}) = 0, \tag{3}$$

where $k = 1, \cdots, n$ and $l = k + 1, \cdots, n$, and the v_{jk} are the arguments of the matrix elements

$$U_{jk} = u_{jk} \exp iv_{jk}? \tag{4}$$

Rather than attempting to consider the difficult nonlinear equations (3), we try to solve the problem by making use of a convenient parameterization of unitary matrices. Murnaghan[2] has shown that an n-dimensional unitary matrix may be factorized into a diagonal unitary matrix and $\frac{1}{2}n(n - 1)$ unitary unimodular matrices of a particularly simple structure. Thus, if U^n denotes the unitary matrix and $D(\alpha_1, \cdots, \alpha_n)$ denotes the diagonal matrix with elements

$$D_{jk} = \delta_{jk} \exp i\alpha_j, \tag{5}$$

then

$$U^n = D(\alpha_1, \cdots, \alpha_n)\mathfrak{I}^n, \tag{6}$$

where $\mathfrak{I}^n$ denotes the product

$$\mathfrak{I}^n = \prod_{k=2}^{n} \prod_{j=1}^{k-1} \mathfrak{U}_{jk}^n(\theta_{jk}, \sigma_{jk}). \tag{7}$$

The unitary unimodular matrices $\mathfrak{U}_{jk}^n(\theta_{jk}, \sigma_{jk})$ are functions of only two parameters each: θ_{jk} and σ_{jk}. Only four elements in $\mathfrak{U}_{jk}^n$ are different from δ_{jk}: the (jj) and (kk) elements have the value $c_{jk} \equiv \cos \theta_{jk}$, the (jk) element has the value $-s_{jk} \exp (-i\sigma_{jk})$, where $s_{jk} \equiv \sin \theta_{jk}$, and the (kj) element has the value $s_{jk} \exp i\sigma_{jk}$.

The n-dimensional unitary group is an n^2-parameter group. When the parameters are chosen, as in (6) and (7), to be the n parameters α_j, the $\frac{1}{2}n(n - 1)$ parameters θ_{jk} and the $\frac{1}{2}n(n - 1)$ parameters σ_{jk}, the parametric space is

$$-\pi < \alpha_j \leq \pi, \quad -\pi < \sigma_{jk} \leq \pi, \quad 0 \leq \theta_{jk} \leq \frac{1}{2}\pi.$$

To solve our problem, as formulated in (2), we first observe that

$$|U_{jk}^n| = |\mathfrak{I}_{jk}^n|,$$

because the unitary diagonal matrix D in (6) does not alter the values of the moduli of the elements of $\mathfrak{I}^n$ when multiplied into $\mathfrak{I}^n$. Thus the parameters α_j need not enter into our problem.

Let us identify the moduli of the elements of $\mathfrak{I}^4$ with 16 given nonnegative real numbers u_{jk} which obey (1), so that nine of them are independent, and try to solve for the θ's and σ's. We have

$$\mathfrak{I}^4 = \mathfrak{U}_{12}^4(\theta_{12}, \sigma_{12}) \cdot \mathfrak{U}_{13}^4(\theta_{13}, \sigma_{13}) \cdot \mathfrak{U}_{23}^4(\theta_{23}, \sigma_{23})$$
$$\cdot \mathfrak{U}_{14}^4(\theta_{14}, \sigma_{14}) \cdot \mathfrak{U}_{24}^4(\theta_{24}, \sigma_{24}) \cdot \mathfrak{U}_{34}^4(\theta_{34}, \sigma_{34}).$$

The product of the last three matrices becomes explicitly

$$\begin{bmatrix} c_{14} & -s_{14}s_{24}e^{i(-\sigma_{14}+\sigma_{24})} & -s_{14}c_{24}s_{34}e^{i(-\sigma_{14}+\sigma_{34})} & -s_{14}c_{24}c_{34}e^{-i\sigma_{14}} \\ 0 & c_{24} & -s_{24}s_{34}e^{i(-\sigma_{24}+\sigma_{34})} & -s_{24}c_{34}e^{-i\sigma_{24}} \\ 0 & 0 & c_{34} & -s_{34}e^{-i\sigma_{34}} \\ s_{14}e^{i\sigma_{14}} & c_{14}s_{24}e^{i\sigma_{24}} & c_{14}c_{24}s_{34}e^{i\sigma_{34}} & c_{14}c_{24}c_{34} \end{bmatrix}. \tag{8}$$

[1] M. Roos and B. E. Laurent (to be published).
[2] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan Books, Washington, D. C., 1962).

Multiplying $\mathfrak{u}_{23}^4$ into (8), the second and third rows mix in such a way that it is convenient to introduce a new parameter

$$\gamma_{23} \equiv \sigma_{23} - \sigma_{24} + \sigma_{34}.$$

Then (8) becomes explicitly, omitting the unaltered first and fourth rows,

$$\begin{bmatrix} 0 & c_{23}c_{24} & -(c_{23}s_{24}s_{34}e^{i\gamma_{23}} + s_{23}c_{34})e^{-i\sigma_{23}} & (-c_{23}s_{24}c_{34} + s_{23}s_{34}e^{-i\gamma_{23}})e^{-i\sigma_{24}} \\ 0 & s_{23}c_{24}e^{i\sigma_{23}} & -s_{23}s_{24}s_{34}e^{i\gamma_{23}} + c_{23}c_{34} & (-s_{23}s_{24}c_{34}e^{i\gamma_{23}} - c_{23}s_{34})e^{-i\sigma_{24}} \end{bmatrix}.$$

Next, multiplying $\mathfrak{u}_{13}^4$ into this, the first and third rows mix in such a way that it is convenient to introduce a second new parameter

$$\gamma_{13} \equiv \sigma_{13} - \sigma_{14} - \sigma_{23} + \sigma_{24}.$$

Multiplying in the final factor $\mathfrak{u}_{12}^4$ and introducing a third new parameter

$$\gamma_{12} \equiv \sigma_{12} - \sigma_{14} + \sigma_{24},$$

we obtain the product matrix $\mathfrak{J}^4$:

$$\mathfrak{J}^4 = \begin{bmatrix} c_{12}c_{13}c_{14} & z_{12}(\gamma_{12}, \gamma_{13}, \theta_{23})e^{-i\sigma_{12}} & z_{13}(\gamma_{12}, \gamma_{13}, \gamma_{23}, \theta_{23})e^{-i\sigma_{13}} & z_{14}(\gamma_{12}, \gamma_{13}, \gamma_{23}, \theta_{23})e^{-i\sigma_{14}} \\ s_{12}c_{13}c_{14}e^{i\sigma_{12}} & z_{22}(\gamma_{12}, \gamma_{13}, \theta_{23}) & z_{23}(\gamma_{12}, \gamma_{13}, \gamma_{23}, \theta_{23})e^{-i\sigma_{23}} & z_{24}(\gamma_{12}, \gamma_{13}, \gamma_{23}, \theta_{23})e^{-i\sigma_{24}} \\ s_{13}c_{14}e^{i\sigma_{13}} & z_{32}(\gamma_{13}, \theta_{23})e^{i\sigma_{23}} & z_{33}(\gamma_{13}, \gamma_{23}, \theta_{23}) & z_{34}(\gamma_{13}, \gamma_{23}, \theta_{23})e^{-i\sigma_{34}} \\ s_{14}e^{i\sigma_{14}} & c_{14}s_{24}e^{i\sigma_{24}} & c_{14}c_{24}s_{34}e^{i\sigma_{34}} & c_{14}c_{24}c_{34} \end{bmatrix}. \tag{9}$$

The z_{ik} are complex polynomials, in which the argument of each term is a linear sum of γ's or zero.

It is evident from (9) that the moduli $|\mathfrak{J}_{ik}^4|$ are functions of nine parameters only: the six parameters θ_{ik} and the three parameters γ_{ik}, but not of the remaining three independent parameters σ_{ik}, which enter only in the arguments of the matrix elements.

We now want to express the parameters θ_{ik} and γ_{ik} in terms of nine independent u_{ik}'s. From

$$u_{21} = s_{12}c_{13}c_{14}, \qquad u_{31} = s_{13}c_{14}, \qquad u_{41} = s_{14}, \qquad u_{42} = c_{14}s_{24}, \qquad u_{43} = c_{14}c_{24}s_{34},$$

we can immediately solve for five of the θ_{ik}'s, which become

$$\theta_{12} = \arcsin u_{21}(1 - u_{31}^2 - u_{41}^2)^{-\frac{1}{2}}; \qquad \theta_{13} = \arcsin u_{31}(1 - u_{41}^2)^{-\frac{1}{2}}; \qquad \theta_{14} = \arcsin u_{41};$$

$$\theta_{24} = \arcsin u_{42}(1 - u_{41}^2)^{-\frac{1}{2}}; \qquad \theta_{34} = \arcsin u_{43}(1 - u_{41}^2 - u_{42}^2)^{-\frac{1}{2}}.$$

There remain four equations containing the parameters

$\theta_{23}, \gamma_{12}, \gamma_{13}$ and γ_{23}:

$$u_{32} = |z_{32}(\gamma_{13}, \theta_{23})|; \quad u_{23} = |z_{23}(\gamma_{12}, \gamma_{13}, \gamma_{23}, \theta_{23})|; \quad u_{22} = |z_{22}(\gamma_{12}, \gamma_{13}, \theta_{23})|; \quad u_{33} = |z_{33}(\gamma_{13}, \gamma_{23}, \theta_{23})|. \tag{10}$$

From the first of Eqs. (10), $\cos\gamma_{13}$ may be solved in terms of θ_{23} and inserted into the other three equations. Then from the last two equations (10), $\cos\gamma_{12}$ and $\cos\gamma_{23}$, respectively, may be solved in terms of θ_{23} and inserted into the u_{23} equation, which thus becomes

$$u_{23} = f(\theta_{23}). \tag{11}$$

The answer given by this procedure to our initially formulated question is then as follows: it is possible to construct a unitary matrix (U_{ik}) with $|U_{ik}| = u_{ik}$ only when

(1) there exists a solution $0 \leq \theta_{23} \leq \frac{1}{2}\pi$ to Eq. (11), and when this solution, inserted into (10), yields

(2)

$$|\cos\gamma_{13}| \leq 1; \qquad |\cos\gamma_{23}| \leq 1; \qquad |\cos\gamma_{12}| \leq 1. \tag{12}$$

The conditions (12) are restrictions on the allowed volume in the nine-dimensional u_{ik}-space. If these conditions all hold with the inequality sign, the unitary matrix U^4 is completely determined by the nine given u_{ik}'s and seven phases: the three independent σ_{ik}'s and the four α_i's.

If one or several of the conditions (12) hold with the equality sign, the number of independent u_{ik}'s decreases correspondingly and the arbitrariness of the unitary matrix exceeds that of the seven phases.

It is straightforward to generalize the four-dimensional case to n dimensions. $\mathfrak{I}^n$ then consists of $\frac{1}{2}n(n-1)$ factors; the product of the last $n-1$ factors gives the generalization of (8).

The first γ_{ik} appears when we multiply this product with the nth matrix

$$\mathfrak{U}^n_{n-2,n-1}(\theta_{n-2,n-1}, \sigma_{n-2,n-1}),$$

which mixes the rows of number $n-2$ and $n-1$. This γ_{ik} will be

$$\gamma_{n-2,n-1} \equiv \sigma_{n-2,n-1} - \sigma_{n-2,n} + \sigma_{n-1,n}.$$

Each time we multiply the next factor matrix into the previous product, it becomes convenient to introduce a new parameter γ_{ik}, which is a sum of σ_{ik}'s with weights $+1$ or -1. The final product $\mathfrak{I}^n$ will thus contain $\frac{1}{2}n(n-1)$ parameters θ_{ik}, $\frac{1}{2}(n-1)(n-2)$ parameters γ_{ik} and $n-1$ parameters σ_{ik}. The moduli $|\mathfrak{I}^n_{jk}|$ of the matrix elements will be functions of the parameters θ_{ik} and γ_{ik}, totaling $(n-1)^2$, and the parameters σ_{ik} enter only in the arguments of $\mathfrak{I}^n_{jk}$.

Next, we write down the $(n-1)^2$ requirements

$$|\mathfrak{I}^n_{ik}| = u_{ik}, \tag{13}$$

where, for instance, $j = 2, \cdots, n$ and $k = 1, \cdots, n-1$. We can immediately solve for the $2n-3$ parameters θ_{1j} and θ_{kn} $(k \neq 1)$, and get expressions similar to those of the four-dimensional case. The number of remaining equations (13), corresponding to Eqs. (10), is $(n-2)^2$.

The structure of the remaining equations allows each $\cos \gamma_{ik}$ to be solved as a function of the θ_{ik}'s. Carrying through this process of elimination, we are finally left with $\frac{1}{2}(n-2)(n-3)$ equations, corresponding to (11) and containing the remaining $\frac{1}{2}(n-2)(n-3)$ θ_{ik}'s.

The problem (13) or (2) thus has a solution when (1) the set of $\frac{1}{2}(n-2)(n-3)$ equations have a solution with each θ_{ik} within the interval

$$0 \leq \theta_{ik} \leq \frac{1}{2}\pi$$

and, using this solution in the expressions for the $\cos \gamma_{ik}$,

$$(2) \qquad |\cos \gamma_{ik}| \leq 1 \tag{14}$$

for each γ_{ik}.

The conditions (14) are restrictions on the allowed volume in the $(n-1)^2$-dimensional u_{ik}-space. If these conditions all hold with the inequality sign, the unitary matrix U^n is completely determined by the $(n-1)^2$ given u_{ik}'s and $2n-1$ phases: the $n-1$ independent σ_{ik}'s and the n α_i's. Let us call such a unitary matrix *regular*, whereas if (14) holds with one or more equality signs we call the unitary matrix *irregular*.

Let us further define a *phase transformation* by $D_1 M D_2$, where D_1 and D_2 are diagonal unitary matrices. Such a transformation has $2n-1$ parameters; it further has the property of leaving the moduli $|M_{jk}|$ of the elements of an arbitrary matrix M unchanged.

Part of our results may then be stated as the following

Theorem. A regular unitary matrix is determined, up to a phase transformation, by the moduli of its elements.

The irregular unitary matrices are clearly not completely determined by the moduli of their elements and a phase transformation, because there are relations between the moduli and thus the number of independent moduli is smaller than $(n-1)^2$.

On the Generation of Anisotropic Tensors*

G. F. SMITH

Yale University, New Haven, Connecticut
(Received 21 May 1964)

The problem of generating a complete set of linearly independent nth-order tensors which are invariant under a crystallographic group is considered. A number of methods for the solution of this problem such as the use of tensor bases, the addition of tensors of lower symmetry, and the method of polynomial invariants are discussed. The limitations of these methods are outlined.

I. INTRODUCTION

AN nth-order tensor $C_{i_1 i_2 \cdots i_n}$ which satisfies the equations

$$C_{i_1 i_2 \cdots i_n} = t_{i_1 j_1} t_{i_2 j_2} \cdots t_{i_n j_n} C_{j_1 j_2 \cdots j_n} \qquad (1)$$

for all transformations $T = \|t_{ij}\|$ belonging to a group G is said to be invariant under G. A tensor which is invariant under the orthogonal group is said to be an isotropic tensor. A tensor which is invariant under a proper subgroup of the orthogonal group, for example, one of the crystallographic groups, is said to be an anisotropic tensor. We are concerned with the problem of generating a complete set of linearly independent nth-order tensors invariant under the crystallographic groups and would like to comment on the efficiency of certain of the available methods when applied to the generation of invariant tensors of high order.

II. TENSOR BASES

A tensor basis for the group G is a set of tensors each of which is invariant under G such that any tensor which is invariant under G is expressible as a linear combination of outer products of the basic tensors. We list in Table I the tensor bases associated with those crystallographic groups G for which the three-dimensional representation of the group G furnished by the symmetry transformation matrices $T_1, \ldots, T_m$ is reducible to the sum of three one-dimensional representations. We also list the number $P_n(G)$ of linearly independent nth-order tensors which are invariant under G. The quantity $P_n(G)$ is readily obtained from group-theoretic considerations. The crystallographic groups are identified by their Schonflies and Shubnikov symbols and the vectors e_1, e_2, e_3, B, C appearing in Table I are defined by

$$e_{1i} = (1, 0, 0), \quad e_{2i} = (0, 1, 0), \quad e_{3i} = (0, 0, 1), \qquad (2)$$
$$B_i = (1, i, 0), \quad C_i = (1, -i, 0), \quad i^2 = -1.$$

The first eight tensor bases are those given by Sirotin.[1] The remaining eight tensor bases are obtained immediately from the "selection rules" given by Sirotin, although the tensor bases listed by Sirotin for these cases differ from those in Table I. Tensor bases for the crystallographic groups D_{2d}, C_{4v}, D_4, D_{4h}, C_{3v}, D_3, D_{3d}, D_{3h}, C_{6v}, D_6, D_{6h}, T, T_h, T_d, and O_h have been recently determined by Smith and Rivlin.[2] While it is a simple matter to generate the complete set of $P_n(G)$ linearly independent nth-order tensors invariant under G from the tensor basis elements for those groups G listed in Table I, we maintain that for the remaining crystallographic groups the use of tensor bases is not the appropriate way to generate the set of nth-order linearly independent invariant tensors if n is large. This statement also applies to the crystallographic groups C_4, S_4, C_{4h}, C_3, C_{3h}, C_6, C_{6h}, C_{3i} if the tensor bases listed by Sirotin[1] or Smith and Rivlin[2] are used instead of those given in Table I. This contrasts with the statement by Sirotin[1] that if the tensor basis is known one may readily construct the set of linearly independent invariant tensors. In support of this point of view, we note that it may be readily verified that every nth-order (n is even) two-dimensional isotropic tensor is expressible as a linear combination of the $\frac{1}{2}\binom{n}{\frac{1}{2}n}$ distinct linearly independent isotropic tensors obtained from

$$B_{\alpha_1} \cdots B_{\alpha_m} C_{\alpha_{m+1}} \cdots C_{\alpha_n}$$
$$+ C_{\alpha_1} \cdots C_{\alpha_m} B_{\alpha_{m+1}} \cdots B_{\alpha_n} \quad (m = \tfrac{1}{2}n), \qquad (3)$$
$$B_\alpha = (1, i), \quad C_\alpha = (1, -i), \quad i^2 = -1$$

by the $n!$ permutations of the subscripts $\alpha_1, \cdots, \alpha_n$. The usual statement concerning two-dimensional isotropic tensors is that $\delta_{\alpha\beta}$ forms a tensor basis for the two-dimensional orthogonal group, i.e., that any nth-order two-dimensional isotropic tensor is expressible as a linear combination of the $n!/2^{n/2} \cdot$

[1] Iu. I. Sirotin, Dokl. Akad. Nauk SSSR **133**, 321 (1960) [English transl.: Soviet Phys.—Doklady **5**, 774 (1961)].
[2] G. F. Smith and R. S. Rivlin, Arch. Ratl. Mech. Anal. **15**, 170 (1964).

TABLE I. Tensor bases.

Point groups	Tensor bases	P_n
$C_1, 1$	e_{1i}, e_{2i}, e_{3i}	$P_n = 3^n$
$C_i, \bar{2}$	$e_{1i}e_{1j}, e_{1i}e_{2j}, e_{1i}e_{3j}, e_{2i}e_{2j}, e_{2i}e_{3j}, e_{3i}e_{3j}$	$2P_n = 3^n + (-3)^n$
C_s, m	$e_{1i}, e_{2i}, e_{3i}e_{3j}$	$2P_n = 3^n + 1$
$C_2, 2$	$e_{1i}e_{1j}, e_{2i}e_{2j}, e_{3i}, e_{1i}e_{2j}$	$2P_n = 3^n + (-1)^n$
$C_{2h}, 2:m$	$e_{1i}e_{1j}, e_{2i}e_{2j}, e_{3i}e_{3j}, e_{1i}e_{2j}$	$4P_n = 3^n + (-3)^n + 1 + (-1)^n$
$C_{2v}, 2 \cdot m$	$e_{1i}e_{1j}, e_{2i}e_{2j}, e_{3i}$	$4P_n = 3^n + 2 + (-1)^n$
$D_2, 2:2$	$e_{1i}e_{1j}, e_{2i}e_{2j}, e_{3i}e_{3j}, e_{1i}e_{2j}e_{3k}$	$4P_n = 3^n + 3 \cdot (-1)^n$
$D_{2h}, m \cdot 2 : m$	$e_{1i}e_{1j}, e_{2i}e_{2j}, e_{3i}e_{3j}$	$8P_n = 3^n + (-3)^n + 3 + 3 \cdot (-1)^n$
$C_4, 4$	$e_{3i}, B_iC_j, B_iB_jB_kB_l, C_iC_jC_kC_l$	$4P_n = 3^n + 2 + (-1)^n$
$S_4, \bar{4}$	$e_{3i}e_{3j}, B_iC_j, B_iB_jB_kB_l, C_iC_jC_kC_l, e_{3i}B_jB_k, e_{3i}C_jC_k$	$4P_n = 3^n + 3 \cdot (-1)^n$
$C_{4h}, 4:m$	$e_{3i}e_{3j}, B_iC_j, B_iB_jB_kB_l, C_iC_jC_kC_l$	$8P_n = 3^n + (-3)^n + 3 + 3 \cdot (-1)^n$
$C_3, 3$	$e_{3i}, B_iC_j, B_iB_jB_k, C_iC_jC_k$	$3P_n = 3^n$
$C_{3h}, 3:m$	$e_{3i}e_{3j}, B_iC_j, B_iB_jB_k, C_iC_jC_k$	$6P_n = 3^n + 2 \cdot (-2)^n + 1$
$C_6, 6$	$e_{3i}, B_iC_j, B_iB_jB_kB_lB_mB_n, C_iC_jC_kC_lC_mC_n$	$6P_n = 3^n + 2 \cdot (-2)^n + (-1)^n$
$C_{6h}, 6:m$	$e_{3i}e_{3j}, B_iC_j, B_iB_jB_kB_lB_mB_n, C_iC_jC_kC_lC_mC_n$	$12P_n = 3^n + (-3)^n + 2 \cdot 2^n + 2 \cdot (-2)^n + 1 + (-1)^n$
$C_{3i}, \bar{6}$	$e_{3i}e_{3j}, B_iC_j, B_iB_jB_kB_lB_mB_n, C_iC_jC_kC_lC_mC_n, e_{3i}B_jB_kB_l, e_{3i}C_jC_kC_l$	$6P_n = 3^n + (-3)^n$

$(n/2)$! distinct tensors formed from

$$\delta_{\alpha_1 \alpha_2} \delta_{\alpha_3 \alpha_4} \cdots \delta_{\alpha_{n-1} \alpha_n} \qquad (4)$$

by the $n!$ permutations of the subscripts. We observe that for $n = 12$ we obtain $12!/2^6 \cdot 6! = 10395$ distinct tensors from (4). However from (3) we see that these are expressible in terms of $\frac{1}{2}\binom{12}{6} = 462$ isotropic tensors. Thus the generation of the set of invariant tensors from the tensor basis elements leads in this case to the introduction of a large number of redundant tensors. The generation of three dimensional isotropic tensors from the tensor basis δ_{ij} also leads to the introduction of redundant tensors for (even) n greater than 6. Since either $\delta_{\alpha\beta}$ or δ_{ij} is an element of the tensor bases given by Smith and Rivlin[2] for the crystallographic groups not listed in Table I, it is clear that for large values of n the use of tensor bases would lead in these cases to the generation of redundant tensors. A procedure which is readily applicable for the crystallographic groups $D_{2d}, \cdots, O_h$ is described by Sirotin.[1]

III. ADDITION OF TENSORS OF LOWER SYMMETRY

In a recent paper,[3] Lokhin and Sedov have proposed an alternate method for generating the set of $P_n(G)$ linearly independent nth-order tensors invariant under a crystallographic group G. Consider the sequence of groups $G \subset G_1 \subset G_2 \subset \cdots \subset G_p$ where each group is a subgroup of the groups to the right of it in the sequence. It is clear that any tensor

invariant under a group G_r is also invariant under the groups preceding G_r. The method then proceeds by first listing the set of nth-order tensors invariant under G_p, then augmenting this set by the tensors which are invariant under G_{p-1} but are not invariant under $G_p, \cdots$, and finally listing the tensors which are invariant under G but which are not invariant under G_1. This method has been previously discussed by Sirotin[4,5] and has considerable merit in special cases.[5] However when applied to the general case the method as outlined by Lokhin and Sedov[3] has some objectionable features. For example, the group G_p is in many instances taken to be the three-dimensional orthogonal group and it is then required to list the set of three-dimensional isotropic tensors of given order. This is readily accomplished for $n = 6$ or less but for larger values of n (n even) it is a matter of considerable difficulty. The set of linearly independent isotropic tensors of order 10 (say) may be generated from $\delta_{i_1 i_2} \delta_{i_3 i_4} \cdots \delta_{i_9 i_{10}}$ by applying Young's symmetry[6-8] operators to this tensor. However, this process is tedious and the resulting tensors are sufficiently cumbersome so as to render the procedure impractical for even moderate values of n. It is also to be noted that while it may not be difficult to generate $P_n(G_{r-1}) - P_n(G_r)$ tensors which are invariant under G_{r-1} but which are not invariant

[3] V. V. Lokhin and L. I. Sedov, Prikl. Math. Mech. 27, 393 (1963).

[4] Iu. I. Sirotin, Kristallografica 5, 171 (1960) [English transl.: Soviet Phys.— Cryst. 5, 157 (1960)].

[5] Iu. I. Sirotin, Kristallografica 6, 331 (1961) [English transl.: Soviet Phys.—Cryst. 6, 263 (1961)].

[6] T. L. Wade, Am. J. Math. 63, 645 (1941).

[7] M. Hammermesh, Group Theory (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1962), p. 244.

[8] H. Weyl, The Classical Groups (Princeton University Press, Princeton, New Jersey, 1939), pp. 96–136.

under G_r, it may prove to be very troublesome to verify that they are linearly independent.

IV. POLYNOMIAL INVARIANTS

A polynomial function W of the symmetric tensor S is said to be invariant under the group G if

$$W(S) = W(TST^{-1}) \tag{5}$$

for every transformation T belonging to the group G. Consider the problem of determining the form of the eighth-order tensor $C_{i_1 \cdots i_8}$ associated with the fourth-order elastic constants for the group O_h, i.e., the tensor $C_{i_1 \cdots i_8}$ appearing in the expression

$$W_4 = C_{i_1 \cdots i_8} G_{i_1 i_2} G_{i_3 i_4} G_{i_5 i_6} G_{i_7 i_8}, \quad G_{ij} = G_{ji}, \tag{6}$$

where W_4 is a scalar function of degree four in the finite strain tensor G_{ij} which is invariant under the group O_h. The tensor $C_{i_1 \cdots i_8}$ appearing in (6) is required to be invariant under O_h and is also required to be symmetric in i_1 and i_2, i_3 and i_4, etc. The form of the general polynomial function of G_{ij} which is invariant under O_h is known[9] and we note that this information immediately yields the form of the fourth-order elastic constant tensor $C_{i_1 \cdots i_8}$ and also the form of the tensor $C_{i_1 \cdots i_{2n}}$ associated with the elastic constants of arbitrary order n. Thus, it has been shown that any polynomial function of G_{ij} which is invariant under O_h is uniquely expressible in the form

$$S_0 + S_1 L_1 + S_2 L_2 + S_3 L_3 + S_4 L_1^2 + S_5 L_2 L_3, \tag{7}$$

where $S_0, \cdots, S_5$ are polynomials in the quantities $K_1, \cdots, K_6$ defined by

$$(K_1, \cdots, K_6) = (\sum G_{11}, \sum G_{11} G_{22}, G_{11} G_{22} G_{33},$$
$$\sum G_{23}^2, \sum G_{23}^2 G_{31}^2, G_{23} G_{31} G_{12}), \tag{8}$$

and where

$$L_1 = \sum G_{11}(G_{31}^2 + G_{12}^2), \quad L_2 = \sum G_{11} G_{31}^2 G_{12}^2,$$
$$L_3 = \sum G_{23}^2 G_{22} G_{33}. \tag{9}$$

In (8) and (9), $\sum G_{11} \cdots G_{12}$ denotes the sum of the three quantities obtained by permuting the subscripts in the summand cyclically. Hence the function W_4 given by (6) is expressible as a linear com-

bination of the eleven quantities

$$K_1^4, \; K_1^2 K_2, \; K_1^2 K_4, \; K_1 K_3, \; K_2^2, \; K_4^2,$$
$$K_2 K_4, \; K_5, \; K_1 K_6, \; K_1 L_1, \; L_3. \tag{10}$$

We may associate a tensor with each of the invariants (10). Thus, since

$$K_1^4 = \delta_{i_1 i_2} \delta_{i_3 i_4} \delta_{i_5 i_6} \delta_{i_7 i_8} G_{i_1 i_2} G_{i_3 i_4} G_{i_5 i_6} G_{i_7 i_8}, \tag{11}$$

we associate $\delta_{i_1 i_2} \delta_{i_3 i_4} \delta_{i_5 i_6} \delta_{i_7 i_8}$ with K_1^4. The tensor $C_{i_1 \cdots i_8}$ in (6) is then expressible as a linear combination of the eleven tensors associated with the invariants (10). It is clear that this analysis may be extended so as to obtain the elastic constants of arbitrary order.

The fourth-order elastic constants for O_h have been obtained recently by Ghate[10] and are in agreement with the results listed above. We feel that obtaining the form of the elastic constant tensors from consideration of the form of polynomial invariants of a single symmetric tensor G_{ij} is preferable to the method adopted by Ghate[10] and note that elastic constant tensors of arbitrary order and for all crystallographic groups may be read off almost immediately from the results given by Smith.[9]

V. CONCLUSION

For the crystallographic groups $C_1, C_i, C_s, C_2, C_{2h}, C_{2v}, D_2, D_{2h}, C_4, C_{4h}, C_3, C_{3h}, C_6, C_{6h}, C_{3i}$, it is preferable to generate the set of invariant tensors from the tensor bases listed in Table I. For the crystallographic groups $D_{2d}, C_{4v}, D_4, D_{4h}, C_{3v}, D_3, D_{3d}, D_{3h}, C_{6v}, D_6, D_{6h}, T, T_h, T_d, O, O_h$, the set of nth-order invariant tensors may be generated efficiently by use of tensor bases only for small values of n (say 5 or less) but for larger values of n it is preferable to use the method discussed by Sirotin.[1] The method of addition of tensors of lower symmetry may prove to be highly efficient in special cases but in general it will suffer by comparison with the use of tensor bases for the low-symmetry crystal classes and with the method of Sirotin[1] for the high-symmetry crystal classes. For special cases such as the determination of the elastic constant tensor, the use of polynomial invariants is highly efficient.

[9] G. F. Smith, Arch. Ratl. Mech. Anal. 10, 108 (1962).

[10] P. B. Ghate, J. Appl. Phys. 35, 337 (1964).

Lagrangian Formulation of the Phonon Field Equations

ARNOLD D. LEVINE

West Virginia University, Morgantown, West Virginia

(Received 7 July 1964)

The equations of motion of lattice vibrations are formulated with the action principle as a starting point. As a result, one obtains, in addition to the equations of motion, the conservation laws for energy and momentum. The latter are contained in a set of finite difference equations. Boundary conditions on the field variables must be specified over a region equal to one lattice spacing in order for the entire procedure to be meaningful. The quantized version of the theory can be constructed in a conventional way, and the commutators of the field variables exhibit a set of periodically spaced singularities. In this way we construct a field which is nonlocal with respect to its dependence on space variables, but is local with respect to its time dependence.

THE conventional theory of lattice vibrations is formulated in terms of a set of coupled finite difference equations. Because of the periodicity of the lattice, one may transform to normal modes where it is possible, in a completely straightforward manner, to quantize the decoupled equations.[1] Such a theory is strictly one of particle dynamics, and the wave character of the lattice vibrations can be discussed only by going to the continuum limit.

Now there is a fundamental difference between a particle theory as compared to a wave theory of lattice vibrations; for the latter requires the specification of a function throughout all points in space, while the former allows field functions to be specified only at the lattice sites. Nevertheless, the interacting phonon field can be studied in terms of a point interaction with, say, the electron field. A conventional field theoretic description of this electron phonon system can then be developed, but this is convenient only if we consider the continuum limit for the phonon field.[2]

At present, to the knowledge of the author, no formulation has been given in the literature of a theory involving field variables which are continuous functions of the space variables and which satisfy the finite difference equations of the particle theory at the lattice points. Such a theory is outlined below in the following sections.[3] The development is presented with the action integral as a starting point. This allows a straightforward deduction of the conservation laws, and is particularly convenient when one quantizes the theory. We consider the one-dimensional case only, since all of the salient features

manifest themselves here, and the only new result to be obtained in the three-dimensional problem is the angular momentum conservation law, as well as the derivation of the spin of the phonon.

I

We consider a one-dimensional linear chain of length L, consisting of N particles separated by a distance a. In the particle theory of lattice vibrations we obtain the set of coupled finite difference equations for the displacements of the lattice points

$$\ddot{\eta}_l = (c^2/a^2)(\eta_{l+1} + \eta_{l-1} - 2\eta_l), \tag{1}$$

where η_l is the displacement of the lattice site. We introduce a field, defined at all points in space, which can be made identical to the lattice displacements at the lattice sites. Thus, we want an equation of the form

$$\ddot{\phi}(x, t) = [c^2/a^2][\phi(x + a, t) + \phi(x - a, t) - 2\phi(x, t)]. \tag{1'}$$

For the sake of brevity, we introduce the following notation:

$$f(x \pm a) = T_{\pm}f(x)$$

and

$$\int_{x_1}^{x_1+a} dx f(x) - \int_{x_1}^{x_1+a} dx f(x) = \left\{\int_{x_1} - \int_{x_1}\right\} f(x).$$

Let us now introduce the following Lagrangian density[4]

---

[1] J. Ziman, *Electrons and Phonons* (Oxford University Press, London, 1960), Chap. 1.
[2] Y. Nambu, Phys. Rev. **117**, 648 (1960).
[3] A preliminary report of this work is presented in the following reference: A. D. Levine and A. N. Vaidya, Bull. Am. Phys. Soc. **7**, 546 (1962).

[4] The method of taking finite differences for the purposes of obtaining the equations of motion is not unique. For example, we could employ finite differences of the field operators evaluated at $(x \pm a/2)$ with equally satisfactory results. The only modifications that appear in the final results correspond to the manner of specification of the variations at the boundaries. We employ the present convention on the finite differences as a matter of convenience.

$$L = [\dot{\phi}(x, t)]^2/2 - [c^2/2a^2][(T_+ - 1)\phi(x, t)]^2 \quad (2)$$

with the action integral

$$I = \int_{x_1}^{x_2} dx \int_{t_1}^{t_2} dt\, L\{\phi, T_+\phi, \dot{\phi}\}. \quad (3)$$

The variation of the action integral will then be

$$\delta I = \iint dx\, dt\Big[(\partial L/\partial\phi)\delta_0\phi + (\partial L/\partial T_+\phi)\delta_0 T_+\phi$$
$$+ (\partial L/\partial\dot{\phi})\delta_0\dot{\phi} + \int_{x_1}^{x_2} dx\, L\delta t|_{t_1}^{t_2}$$
$$+ \int_{t_1}^{t_2} dt\, L\delta x|_{x_1}^{x_2}\Big], \quad (4)$$

where

$$\delta_0\phi = -\dot{\phi}\delta t - \phi_x\delta x \quad (5)$$

is the intrinsic variation of ϕ at the boundaries.[5] Now the third term in the square-bracketed expression of (4) may be transformed, using integration by parts in a conventional manner. The second term in the same bracketed expression is transformed in the following way:

$$\int_{x_1}^{x_2} dx\, f\{\phi(x)\}\,\delta_0 T_+\phi(x)$$
$$= \int_{x_1}^{x_2} dx\, T_+([T_-f\{\phi(x)\}]\delta_0\phi(x)).$$

Now, the latter expression may be rewritten as

$$\int_{x_1}^{x_2} dx(T_+ - 1)([T_-f\{\phi(x)\}]\delta_0\phi(x))$$
$$+ \int_{x_1}^{x_2} dx[T_-f\{\phi(x)\}]\delta_0\phi(x).$$

But the first integral of this last expression may be rewritten as a surfacelike integral:

$$\Big\{\int_{x_2} - \int_{x_1}\Big\}[T_-f\{\phi(x)\}]\delta_0\phi(x).$$

Thus, combining all of our results, the variation of the action integral may be written as

$$\delta I = \iint dx\, dt[\partial L/\partial\phi$$
$$+ T_-(\partial L/\partial T_+\phi) - \partial_t(\partial L/\partial\dot{\phi})]\delta_0\phi$$
$$- \int_{t_1}^{t_2} dt(1/a)\Big\{\int_{x_2} - \int_{x_1}\Big\}S_1\,\delta x$$
$$- \int_{t_1}^{t_2} dt\, S_2\,\delta x|_{x_1}^{x_2} - \int_{x_1}^{x_2} dx\, \Pi\,\delta x|_{t_1}^{t_2}$$
$$- \int_{t_1}^{t_2}(1/a)\Big\{\int_{x_2} - \int_{x_1}\Big\}G\,\delta t - \int_{x_1}^{x_2} H\, dx\,\delta t|_{t_1}^{t_2}, \quad (6)$$

[5] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), Chaps. 1 and 2.

where we have written

$$S = S_1 + S_2 \quad S_1 = aT_-\{\partial L/\partial T_+\phi\}\phi x, \quad S_2 = -L,$$
$$\Pi = (\partial L/\partial\dot{\phi})\phi_x,$$
$$G = aT_-(\partial L/\partial T_+\phi)\dot{\phi},$$
$$H = (\partial L/\partial\dot{\phi})\dot{\phi} - L. \quad (7)$$

The symbols introduced are as follows: S will be the stress density, Π the momentum density, G the power flow density, and H the energy density. We have employed (5) to represent the variation of the field function at the boundary layers in order to obtain (6).

We must impose certain conditions on the system and the allowed variations before we can use the action principle to obtain the desired results. For one thing, as is already evident if one observes the surfacelike integrals of (6), we must continue the definition of the system beyond the spatial limits x_1 and x_2, at least to within one lattice spacing. This means that our results will be restricted to the interiors of large or infinite lattices. Another condition which we must impose, in order to obtain meaningful conservation laws is that the variations of x and t are uniform at the boundaries x_1 and x_2 within a distance of one lattice spacing. Thus, we cannot specify variations at the boundaries which involve relative displacements of the end lattice points as a function of the distance between the lattice points. This seems to be a reasonable assumption of the physical situation that one encounters. Of course, it must be strictly understood that we are omitting all effects of surface vibrations from our considerations.

We may now obtain all of the desired information about the system using (6). The application of the action integral at the interior of the space–time domain gives us the equation of motion:

$$(\partial L/\partial\phi) + T_-(\partial L/\partial T_+\phi) - \partial_t(\partial L/\partial\dot{\phi}) = 0. \quad (8)$$

If we insert the Lagrangian (2) into (8) we obtain (1') as we had set out to do. Application of the action principle to the coefficients of δx and δt in (6) gives us the conservation laws for momentum and energy, respectively:

$$\int_{t_1}^{t_2} dt(1/a)\Big\{\int_{x_2} - \int_{x_1}\Big\}S_1$$
$$+ \int_{t_1}^{t_2} dt\, S_2|_{x_1}^{x_2} + \int_{x_1}^{x_2} dx\, \Pi|_{t_1}^{t_2} = 0, \quad (A9)$$

$$\int_{t_1}^{t_2} dt(1/a)\Big\{\int_{x_2} - \int_{x_1}\Big\}G + \int_{x_1}^{x_2} dx\, H|_{t_1}^{t_2} = 0. \quad (9B)$$

These conservation laws appear in integral form.

The corresponding differential forms for these laws are

$$(1/a)(T_+ - 1)S_1 + \partial S_2/\partial x + \partial \Pi/\partial t = 0, \qquad (9A')$$

$$(1/a)(T_+ - 1)G + \partial H/\partial t = 0. \qquad (9B')$$

Note that these latter equations occur as finite difference equations. The reader may verify readily that the equation of motion (8) is consistent with the conservation laws (9).

One may consider the continuum limit of the equations which we have used, and show that it is identical to the continuum theory of the one-dimensional wave-equation problem. For example, the Lagrangian (2) becomes

$$L = \tfrac{1}{2}\dot{\phi}^2 - (\tfrac{1}{2}c^2)\phi_x^2.$$

The equation of motion (1') is $\ddot{\phi} = c^2\phi_{xx}$ while the conservation equations (9) become ordinary differential equations. We might also note that the stress density S_1, the power flow density G, and the energy density H, become identical to the corresponding quantities of the continuum theory. (The momentum density Π already is of the form which we encounter in a conventional continuum theory.) If we introduce a plane-wave representation

$$e^{i(kx - \omega t)},$$

the equation of motion (1') yields the dispersion law

$$\omega = (2c/a)\sin\tfrac{1}{2}ka \qquad (10)$$

which, as we may note approaches the result for the continuum limit when ka is much less than unity.

It is well to bear in mind the fact that some of the results of the theory we have outlined above are a consequence of the assumption of nearest-neighbor interactions. If this restriction is removed, the basic structure of the theory remains unaltered, although the individual equations become more involved. The most important difference occurs in the assumptions pertaining to boundary conditions and variations at the boundaries. Thus, for example, if we were to include second nearest neighbor interactions, it would be necessary to continue the domain of the lattice to a length equal to two lattice spacings beyond x_2. At the same time, it would be necessary to demand that the variations of x and t at the boundaries x_1 and x_2 be uniform in a spacelike region equal to two lattice spacings beyond these two limits. With these alterations the entire procedure goes through as before. If now, one were to increase the number of neighbors whose interaction must be accounted for, it would become necessary to maintain uniform

variations of x and t throughout a substantial portion of the lattice domain, and the physical interpretation of such a theory would become quite obscure. However, when we begin to take all of the neighboring interactions into account, we go over to a limit where the lattice model is no longer adequate, and we must consider the system as a many-body problem. Since we confine our considerations to those cases where the lattice model is adequate, the physical interpretation of our theory does not present any difficulty.

II

We may now proceed to the quantization of the classical field equations which we have developed. Starting with the canonical momentum[6]

$$P(t) = \int_{x_1}^{x_2} dx'\Pi(x', t), \qquad (11)$$

we have

$$[P(t), \phi(x, t)] = (\hbar/i)\phi_x(x, t). \qquad (11')$$

Combining these relations with (7) and (2) we find

$$[\phi(x, t), \phi(x', t')] = D(x - x', t - t') \qquad (12)$$

with

$$D(x - x', 0) = 0, \qquad (13a)$$

$$D_t(x - x', t - t')|_{t=t'} = (\hbar/i)\delta(x - x'). \qquad (13b)$$

The explicit form of these commutator functions can be obtained by using a plane-wave representation

$$\phi(x, t) = (1/L)^{-\frac{1}{2}} \sum_k (\tfrac{1}{2}\hbar\omega_k)^{\frac{1}{2}}$$

$$\times \{b_k e^{-i\omega_k t} + b_k^* e^{i\omega_k t}\}e^{ikx}; \qquad (14a)$$

we obtain (12) and (13) if we set

$$[b_k, b_{k'}] = 0, \qquad [b_k, b_{k'}^*] = \delta_{kk'}. \qquad (14b)$$

Here the operators b^* and b are the conventional creation and annihilation operators for the plane-wave states. The singularities of the commutator function are determined by (13) and the dispersion law (10). The explicit form of the commutator function can be deduced quite readily using a plane-wave expansion of the form (14), and the procedure is sufficiently well known, so that we may limit ourselves to quoting the result[7]:

[6] J. M. Jauch and F. Rohrlich, *The Theory of Electrons and Photons* (Addison-Wesley Publishing Company, Inc., Cambridge, Massachusetts, 1955), p. 27.

[7] We make use of the Fourier decomposition for Bessel functions in arriving at this result. See: G. H. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, New York, 1944), p. 20.

$$D(\xi, \tau) = \sum_m \delta(\xi - ma) \int_0^\tau J_{2m}(2ct'/a) \, dt', \qquad (15)$$

where the sum over m runs from zero to infinity and the J are the ordinary Bessel functions. Now the commutator function is also the Green's function for the solution of the homogeneous equation of motion (1′). If, at some given time t, we may specify the field operator and its time derivative, then at any subsequent time the field operator will be given by

$$\phi(x, t) = i \int dx' [D\dot{t}(x - x', t - t_1)\phi(x', t_1)$$
$$- D(x - x', t - t_1)\dot{\phi}(x', t_1)]. \qquad (15')$$

The derivation of this last result may be obtained, starting with (1′) and proceeding as one does in the continuum theory. If we have some disturbance, then from (15) it is clear that this disturbance will be propagated from one lattice cell to the next, as if the system consisted of a set of percussion centers.

We write down explicitly the (normal product) plane-wave expansion of the energy and momentum. This, of course, could be obtained by substituting (14) into (7),

$$P = \sum_k \hbar k b_k^* b_k,$$

$$E = \sum_k \hbar\omega_k b_k^* b_k = \int dxH. \qquad (16)$$

There are two points worthy of note in connection with these last relationships. In the particle theory of lattice vibrations, there is a degeneracy in the system due to the nature of the dispersion law (10) so that it is necessary to divide up the domain of wavenumber space into zones,[8] and sum all dynamical effects over the zones. If this is not done, dynamical variables such as energy and momentum will contain coherent sums over the degenerate states in the separate zones, and these will give rise to unphysical results. In a wave theory, such as we have considered in this paper, such degeneracies do not occur, and there need be no restriction in the summations of relation (16). Of course, in practical applications of the theory a sum over zones may be very helpful, the point being that here it is no longer needed. We should also keep in mind the fact that we are dealing with a nonlocal theory. The difficulties which one normally encounters for such a theory[9]

do not appear because the theory which we have developed is not covariant, and is local in its dependence on time, even though it is nonlocal in its dependence on space variables.

The three-dimensional problem may be treated in the same manner as the one-dimensional problem described here. A realistic three-dimensional model must include second nearest neighbor interactions.[10] Thus, the relations become quite involved. However, the procedure and the results are very similar to the ones obtained here. The main difference arises from the variations corresponding to rotations of the system from which one may deduce the angular momentum conservation laws. The properties involving the spin of the phonon that enter into these considerations have been discussed elsewhere in terms of a continuum theory, and are very similar to the ones given in that reference.[11] The dispersion law and the commutator function do not have the simple forms given in this paper, but one can readily discuss the singularities, and demonstrate explicitly how the wave propogates in such a system.

In this article, we have demonstrated the possibility of quantizing a system of finite difference equations and constructing a nonlocal field that corresponds to these equations. Apart from the purely heuristic value, it is interesting to look at the interactions of such a field, and study it to determine what physical effects manifest themselves when one considers the renormalized interacting systems. It is also interesting to consider electromagnetic interactions of such systems and attempt to gain an insight into the gauge condition that plays such an important role in the electromagnetic properties of these interacting systems. Such studies are now under way.

[8] L. Brillouin, *Wave Propogation and Periodic Structures* (McGraw-Hill Book Company, Inc., New York, 1946), pp. 102–107, 148–164.
[9] A. Pais and G. Uhlenbeck, Phys. Rev. **82**, 914 (1951).

[10] C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York, 1960), pp. 96–101.
[11] A. D. Levine, Nuovo Cimento **26**, 190 (1962).

Multiple Scattering of Electromagnetic Waves by Random Scatterers of Finite Size

N. C. MATHUR

University of Roorkee, Roorkee, India

AND

K. C. YEH

University of Illinois, Urbana, Illinois

(Received 9 June, 1964)

The problem of multiple scattering of waves by randomly positioned objects has been treated by several authors, for example, Foldy, Lax, Twersky, Waterman, and Truell. The present work extends the theory to electromagnetic vector fields and to scatterers of arbitrary size and properties. A general formulation has been made for scattering by any type of discrete and identical scatterers which are similarly oriented. The case of spherical scatterers has been treated by using the rigorous Mie theory both for sparse and dense concentration. Results indicate that in case of sparse concentration, the statistical expectation of the total field has a polarization similar to that of the normally incident wave and the distribution of scatterers is equivalent to a homogeneous medium with a modified refractive index. In case of dense concentration the medium can sustain a number of plane-wave modes. A dispersion relation for the modified medium has been obtained. When the special cases of small spheres is considered, the well-known results obtained by other authors are recovered.

1. INTRODUCTION

THE study of wave propagation in a random medium is interesting both theoretically and from the experimental point of view due to its numerous practical applications. Consequently, it has received considerable attention in the literature. Various approaches are used for a theoretical investigation of the subject, depending upon the statistical model chosen to describe the medium. In this paper we consider a random distribution of distinct obstacles. As the wave propagates in such a medium, it is scattered by the obstacles and the problem is, therefore, formulated in terms of multiply scattered waves.

The problem of scattering by distributions of objects dates back to 1881 when the Lorentz–Lorenz formula for the refractive index of a gas was developed. This was followed by Lord Rayleigh's classical work in 1899 on the scattering of waves by random distributions which explained the color of the sky. However, the first systematic treatment of multiple scattering of waves was given in a paper by Foldy in 1495.[1] Much work has since been done on the subject with valuable contributions from Lax,[2] Twersky,[3] and Waterman and Truell.[4]

The formulation used in this paper follows closely the work of Waterman and Truell. The new features considered here are the vector nature of the electromagnetic waves and the finite size of the scatterers. We consider the incidence of an electromagnetic wave on a semi-infinite medium containing a random distribution of identical, similarly oriented scattering objects. The statistical expectation of the field for an ensemble of configurations of the scatterers is obtained using the joint probability density distribution of scatterers. The special case of spherical scatterers is considered in detail. When the concentration of spheres is sparse, an expression is obtained for the refractive index of the synthetic medium by considering the first-order scattering only. The polarization of a normally incident wave is found to remain unchanged. For dense concentrations, multiple scattering effects have to be taken into account and a dispersion equation for the refractive index of the synthetic medium is obtained. For finite-size spheres, this equation is higher than a quadratic and shows that more than one mode can propagate in the synthetic medium.

2. FORMULATION OF THE PROBLEM

Let us consider a random distribution of m identical, similarly oriented scatterers of arbitrary size, shape, and scattering properties. Let the various configurations of scatterers be governed by the probability density distribution $p(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m)$. Here $p(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m) \, dv_1 \, dv_2 \cdots dv_m$ is the joint probability of finding one scatterer in the volume dv_1 centered at $\mathbf{r}_1$, another scatterer in the volume

[1] L. L. Foldy, Phys. Rev. **67**, 107 (1945).
[2] M. Lax, Rev. Mod. Phys. **23**, 287 (1951); Phys. Rev. **85**, 621 (1952).
[3] V. Twersky, J. Math. Phys. **3**, 700 (1962).
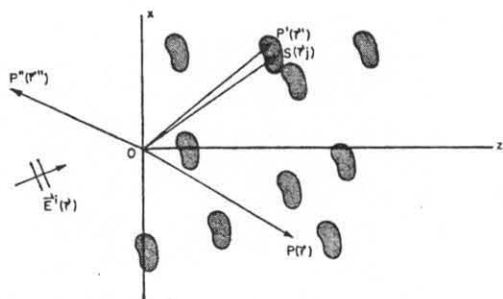[4] P. C. Waterman and R. Truell, J. Math. Phys. **2**, 512 (1961).

FIG. 1. The geometry of the problem.

dv_2 centered at $\mathbf{r}_2$ and so on. Since all scatterers are identical and similarly oriented, a configuration of scatterers is specified by the scatterer positions alone. To facilitate mathematical formulation, let us introduce the following notation.

Let S be the domain of all points in the configurational space. Let S_1 be the domain of all points lying in the right half-space $z \geq 0$, and S_2 the domain of all points lying in the left half-space $z < 0$, so that $S_1 \cup S_2 = S$. Let S_{r_i} be the domain of all points lying inside the scatterer centered at $\mathbf{r}_i$. To be specific we shall restrict the centers of all scatterers to the right half-space $z \geq 0$ and shall not permit interpenetration of scatterers. This is expressed by imposing the following conditions on the density function:

$$p(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m) = 0 \quad \text{for} \quad \mathbf{r}_i \in S_2, \quad \forall j,$$

$$p(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m) = 0 \quad \text{for} \quad S_{r_i} \cap S_{r_k} \neq 0,$$

$$\forall j, k; \quad j \neq k.$$

We shall consider only elastic scattering and assume that the scatterers are in no way affected by the incident field and that the motion of scatterers, if any, is too slow to be of significance.

Let an electromagnetic wave $\mathbf{E}^i(\mathbf{r}, t)$ be incident from the left. We shall consider only the forced oscillation case with time dependence $e^{-i\omega t}$ and shall suppress the time dependence for convenience. For a configuration $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m$ of scatterers, the total field at a point $\mathbf{r}$ is denoted by $\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m)$. Clearly, if $\mathbf{r} \in S_1$, it may lie outside all scatterers (as at P in Fig. 1) or it may lie within some scatterer centered at $\mathbf{r}_i$ (as at P'). However, if $\mathbf{r} \in S_2$ (as at P'') and far from the boundary $z = 0$, it must lie outside all scatterers. For convenience, we shall not consider the case when $\mathbf{r} \in S_2$ but is so close to the boundary that it is within some scatterer whose center is in S_1. This "edge" effect seems to be very complicated and in the following we shall ignore this small region and only consider the fields in S_1 and S_2 away from the edge. The incident

field as well as the field outside of all scatterers in a given configuration is supposed to satisfy the wave equation

$$(\nabla^2 + k^2)\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m) = 0, \quad \mathbf{r} \in S - \bigcup_{j=1}^{m} S_{r_j}.$$

If $\mathbf{r}$ is inside any scatterer, say at $\mathbf{r}_i$, then the field satisfies

$$(\nabla^2 + k_s^2)\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m) = 0, \quad \mathbf{r} \in S_{r_i}, \quad \forall j.$$

All these fields are also supposed to satisfy certain radiation and boundary conditions.

2.1 Derivation of Exact Equations

Let $\mathbf{E}^s(\mathbf{r}, \mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m)$ denote the scattered field at $\mathbf{r}$ from a scatterer at $\mathbf{r}_i$ for the configuration $\mathbf{r}_1, \cdots, \mathbf{r}_m$ of scatterers. We shall write this scattered field in terms of the exciting field $\mathbf{E}^E(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m)$ at the scatterer at $\mathbf{r}_i$ and a scattering operator $T(\mathbf{r}, \mathbf{r}_i)$. Thus,

$$\mathbf{E}^s(\mathbf{r}, \mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m) = T(\mathbf{r}, \mathbf{r}_i)\mathbf{E}^E(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m),$$

$$\mathbf{r} \notin \bigcup_{j=1}^{m} S_{r_j}.$$

If $\mathbf{r}$ lies inside some scatterer, say one centered at $\mathbf{r}_i$, then we shall make use of the interior scattering operator $T^I(\mathbf{r}, \mathbf{r}_i)$ and write

$$\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m) = T^I(\mathbf{r}, \mathbf{r}_i)\mathbf{E}^E(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m),$$

$$\mathbf{r} \in S_{r_i}.$$

The scattering operators T and T^I are merely formal and are introduced purely to stress the functional dependences. For convenience we shall let

$$T(\mathbf{r}, \mathbf{r}_i) = 0, \quad \text{for} \quad \mathbf{r} \in S_{r_i},$$

$$T^I(\mathbf{r}, \mathbf{r}_i) = 0, \quad \text{for} \quad \mathbf{r} \notin S_{r_i}.$$

The total field at a point lying in the scattering medium, i.e., $\mathbf{r} \in S_1$ can, therefore, be written in a self-consistent manner as

$$\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m)$$

$$= \begin{cases} \mathbf{E}^i(\mathbf{r}) + \sum_{j=1}^{m} T(\mathbf{r}, \mathbf{r}_i)\mathbf{E}^E(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m), \\ \qquad\qquad\qquad\qquad\qquad \mathbf{r} \notin \bigcup_{j=1}^{m} S_{r_j}, \\ T^I(\mathbf{r}, \mathbf{r}_i)\mathbf{E}^E(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m), \quad \mathbf{r} \in S_{r_i}. \end{cases}$$

Following Waterman and Truell, these equations can be combined into one by using the symbol $\alpha(\mathbf{r}, \mathbf{r}_k)$ defined as follows

$$\alpha(\mathbf{r}, \mathbf{r}_k) = \begin{cases} 0, & \mathbf{r} \in S_{r_k}, \\ 1, & \mathbf{r} \notin S_{r_k}. \end{cases}$$

Using this symbol, the total field for a given configuration can be written as

$$\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m) = \left[\prod_{k=1}^{m} \alpha(\mathbf{r}, \mathbf{r}_k) \right]$$

$$\times \left[\mathbf{E}^{\mathrm{i}}(\mathbf{r}) + \sum_{j=1}^{m} T(\mathbf{r}, \mathbf{r}_j) \mathbf{E}^{\mathrm{E}}(\mathbf{r}_j : \mathbf{r}_1, \cdots, \mathbf{r}_m) \right]$$

$$+ \sum_{k=1}^{m} [1 - \alpha(\mathbf{r}, \mathbf{r}_k)][T^{\mathrm{i}}(\mathbf{r}, \mathbf{r}_k) \mathbf{E}^{\mathrm{E}}(\mathbf{r}_k : \mathbf{r}_1, \cdots, \mathbf{r}_m)],$$

$$\mathbf{r} \in S_1. \quad (1)$$

It has been found by various authors that the field given by Eq. (1) cannot be evaluated explicitly except in simplest cases. The scheme here is to take the ensemble average of (1) as it stands and make approximations later.

The statistical expectation, or the average value, of the total field is defined by

$$\langle \mathbf{E}(\mathbf{r}) \rangle = \int dv_1 \int dv_2 \cdots \int dv_m$$

$$\times p(\mathbf{r}_1, \cdots, \mathbf{r}_m) \mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m).$$

In this, the volume of integration for each scatterer is the entire volume accessible to scatterers. If the average is taken with one or more scatterers held fixed, we get the first, second, etc., partial average. Thus

$$\langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}_1 : \mathbf{r}_1) \rangle = \text{first partial average of the exciting field at}$$

$$\mathbf{r}_1 \text{ with the scatterer at } \mathbf{r}_1 \text{ held fixed,}$$

$$= \int dv_2 \cdots \int dv_m p(\mathbf{r}_2, \cdots, \mathbf{r}_m : \mathbf{r}_1)$$

$$\times \mathbf{E}^{\mathrm{E}}(\mathbf{r}_1 : \mathbf{r}_1, \cdots, \mathbf{r}_m),$$

where $p(\mathbf{r}_2, \cdots, \mathbf{r}_m : \mathbf{r}_1)$ is the conditional probability density function when the scatterer at $\mathbf{r}_1$ is given. It has been shown by Waterman and Truell that due to exclusion of interpenetration, we can write

$$p(\mathbf{r}_1, \cdots, \mathbf{r}_m) \prod_{k=1}^{m} \alpha(\mathbf{r}, \mathbf{r}_k)$$

$$= p(\mathbf{r}_1, \cdots, \mathbf{r}_m) \left[1 - \sum_{k=1}^{m} \{1 - \alpha(\mathbf{r}, \mathbf{r}_k)\} \right].$$

Using this relation, Eq. (1) can be averaged to get the result

$$\langle \mathbf{E}(\mathbf{r}) \rangle = \mathbf{E}^{\mathrm{i}}(\mathbf{r}) \left[1 - \int_{\mathbf{r} \in S_{\mathbf{r}'}} dv' \rho(\mathbf{r}') \right]$$

$$+ \int_{\mathbf{r} \notin S_{\mathbf{r}'}} dv' \rho(\mathbf{r}') T(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}' : \mathbf{r}') \rangle - \int_{\mathbf{r} \in S_{\mathbf{r}'}} dv' \rho(\mathbf{r}')$$

$$\times \int_{\mathbf{r} \in S_{\mathbf{r}''}, S_{\mathbf{r}'} \cap S_{\mathbf{r}''} = 0} dv'' \rho(\mathbf{r}'' : \mathbf{r}') T(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}' : \mathbf{r}', \mathbf{r}'') \rangle$$

$$+ \int_{\mathbf{r} \in S_{\mathbf{r}'}} dv' \rho(\mathbf{r}') T^{\mathrm{i}}(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}' : \mathbf{r}') \rangle, \quad \mathbf{r} \in S_1. \quad (2)$$

Here, $\rho(\mathbf{r}')$ is the density of scatterers at $\mathbf{r}'$ and is related to the single scatterer probability through $p(\mathbf{r}') = \rho(\mathbf{r}')/m$. The conditional density $\rho(\mathbf{r}'' : \mathbf{r}')$ is related to the conditional probability through $p(\mathbf{r}'' : \mathbf{r}') = \rho(\mathbf{r}'' : \mathbf{r}')/(m - 1)$. The domain of integration $\mathbf{r} \in S_{\mathbf{r}'}$ indicates that the $\mathbf{r}'$ integration is to be carried out over all points $\mathbf{r}'$ such that $\mathbf{r}$ is inside the scatterer at $\mathbf{r}'$. Similarly, the domain $\mathbf{r} \notin S_{\mathbf{r}'}$ indicates that $\mathbf{r}$ is to be outside the scatterer at $\mathbf{r}'$ and $S_{\mathbf{r}'} \cap S_{\mathbf{r}''} = 0$ indicates that the $\mathbf{r}''$ integration is to be carried out over all points $\mathbf{r}''$ such that the scatterer at $\mathbf{r}''$ does not penetrate the scatterer at $\mathbf{r}'$.

If the point of observation lies in the left halfspace, i.e., $\mathbf{r} \in S_2$, then the total field is given by

$$\mathbf{E}(\mathbf{r} : \mathbf{r}_1, \cdots, \mathbf{r}_m)$$

$$= \mathbf{E}^{\mathrm{i}}(\mathbf{r}) + \sum_{j=1}^{m} T(\mathbf{r}, \mathbf{r}_j) \mathbf{E}^{\mathrm{E}}(\mathbf{r}_j : \mathbf{r}_1, \cdots, \mathbf{r}_m).$$

The average total field then becomes

$$\langle \mathbf{E}(\mathbf{r}) \rangle = \mathbf{E}^{\mathrm{i}}(\mathbf{r}) + \int_{\mathbf{r} \notin S_{\mathbf{r}'}} dv' \rho(\mathbf{r}') T(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}' : \mathbf{r}') \rangle,$$

$$\mathbf{r} \in S_2. \quad (3)$$

The partial averages of the exciting field that occur in Eqs. (2) and (3) can be obtained from the self-consistent equation

$$\mathbf{E}^{\mathrm{E}}(\mathbf{r}_j : \mathbf{r}_1, \cdots, \mathbf{r}_m)$$

$$= \mathbf{E}^{\mathrm{i}}(\mathbf{r}_j) + \sum_{\substack{k=1 \\ \neq j}}^{m} T(\mathbf{r}_j : \mathbf{r}_k) \mathbf{E}^{\mathrm{E}}(\mathbf{r}_k : \mathbf{r}_1, \cdots, \mathbf{r}_m). \quad (4)$$

The ensemble average of this equation taken with the scatterer at $\mathbf{r}_j$ held fixed gives the first partial average

$$\langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}_j : \mathbf{r}_j) \rangle = \mathbf{E}^{\mathrm{i}}(\mathbf{r}_j)$$

$$+ \int_{S_{\mathbf{r}_j} \cap S_{\mathbf{r}'} = 0} dv' \rho(\mathbf{r}' : \mathbf{r}_j) T(\mathbf{r}_j, \mathbf{r}') \langle \mathbf{E}^{\mathrm{E}}(\mathbf{r}' : \mathbf{r}', \mathbf{r}_j) \rangle. \quad (5)$$

This equation involves the first and second partial averages of the exciting field. An equation for the second partial average can be similarly obtained and will involve the third partial average. Similarly, each equation for a partial average will involve a partial average of one higher order. In order to break this hierarchy of equations and get an inte-

gral equation in the closed form, some approximations have to be made.

2.2 Approximations

There are several types of approximations possible. In a weakly random medium, we can use different orders of iteration of Eq. (4) for the exciting field. In the first iteration the exciting field is replaced by the incident field alone. In the second iteration we consider the exciting field as made up of the incident field plus the once scattered fields and so on. At each stage of iteration in such an approach, we are confronted with an essentially new problem of integration when any specific type of scatterers are considered.

Another approach is to consider the exciting field at a scatterer at r_i in a given configuration as an expansion in which the first term is the total field at r_i when this scatterer is not there [that is, in a configuration of $(m - 1)$ scatterers]. The second and higher terms then include the rescattering of the field scattered from this scatterer when it is put back in the configuration. Thus, we have

$$\mathbf{E}^{E}(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m) = \mathbf{E}(\mathbf{r}_i : \mathbf{r}_1, \cdot' , \mathbf{r}_m)$$

$$+ \sum_{\substack{k=1 \\ \neq j}}^{m} T(\mathbf{r}_j, \mathbf{r}_k) T(\mathbf{r}_k, \mathbf{r}_j) \mathbf{E}(\mathbf{r}_i : \mathbf{r}_1, \cdot', \mathbf{r}_m) + \cdots . \quad (6)$$

The prime on the configuration $\mathbf{r}_1, \cdots, \mathbf{r}_m$ indicates that the scatterer at $\mathbf{r}_i$ has been removed. The approximation then consists of neglecting the second and higher terms in the right-hand side of Eq. (6). This approximation has been considered in detail by Waterman and Truell[4] and they have developed a criterion of its validity. According to this criterion, the second and higher terms are much smaller than the first if

$$\rho_0 Q_s / k \ll 1,$$

where ρ_0 is the number density of scatterers (assumed constant), Q_s is the scattering cross section of a single scatterer and k is the propagation constant of the medium in which the scatterers are located. Although this criterion has been developed using point scatterers and scalar waves, it is shown to be quite generally valid.

A third approach is to consider the hierarchy of equations for partial averages of the exciting field of which Eq. (5) is the first. The approximation consists of breaking the hierarchy at some point, that is, taking

$$\langle \mathbf{E}^{E}(\mathbf{r}_1 : \mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_i, \mathbf{r}_j) \rangle \approx \langle \mathbf{E}^{E}(\mathbf{r}_1 : \mathbf{r}_1, \cdots, \mathbf{r}_i) \rangle,$$

for some i and j. If we break the hierarchy at the first equation itself, then we have

$$\langle \mathbf{E}^{E}(\mathbf{r}_1 : \mathbf{r}_1, \mathbf{r}_2) \rangle \approx \langle \mathbf{E}^{E}(\mathbf{r}_1 : \mathbf{r}_1) \rangle. \quad (7)$$

This approximation has been discussed by Lax[2] and is designated as the "quasicrystalline" approximation by him. He has shown that this is a very good approximation in the case of dense systems when multiple scattering effects are most important. For the case of statistically independent distributions, it can be shown to be equivalent to the second approximation considered.

We shall simplify the equation by using the second approximation and considering the case of statistically independent distributions only. Thus we put

$$\mathbf{E}^{E}(\mathbf{r}_i : \mathbf{r}_1, \cdots, \mathbf{r}_m) \approx \mathbf{E}(\mathbf{r}_i : \mathbf{r}_1, \cdot' , \mathbf{r}_m)$$

and

$$p(\mathbf{r}_1, \cdots, \mathbf{r}_m) = p(\mathbf{r}_1) p(\mathbf{r}_2) \cdots p(\mathbf{r}_m).$$

Straightforward simplification then leads to the equations

$$\langle \mathbf{E}(\mathbf{r}) \rangle = \left[1 - \int_{\mathbf{r} \in S_{r'}} dv' \rho(\mathbf{r}') \right]$$

$$\times \left[\mathbf{E}^{i}(\mathbf{r}) + \int_{\mathbf{r} \notin S_{r'}} dv' \rho(\mathbf{r}') T(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{E}(\mathbf{r}' : \mathbf{r}') \rangle \right]$$

$$+ \int_{\mathbf{r} \in S_{r'}} dv' \rho(\mathbf{r}') T^{I}(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{E}(\mathbf{r}' : \mathbf{r}') \rangle, \quad \mathbf{r} \in S_1, \quad (8)$$

$$\langle \mathbf{E}(\mathbf{r}) \rangle = \mathbf{E}^{i}(\mathbf{r}) + \int_{\mathbf{r} \notin S_{r'}} dv' \rho(\mathbf{r}') T(\mathbf{r}, \mathbf{r}') \langle \mathbf{E}^{E}(\mathbf{r}' : \mathbf{r}') \rangle,$$

$$\mathbf{r} \in S_2, \quad (9)$$

$$\langle \mathbf{E}^{E}(\mathbf{r} : \mathbf{r}) \rangle = \mathbf{E}^{i}(\mathbf{r}) + \int_{S_r \cap S_{r'} = 0} dv' \rho(\mathbf{r}') T(\mathbf{r}, \mathbf{r}')$$

$$\times \langle \mathbf{E}^{E}(\mathbf{r}' : \mathbf{r}') \rangle. \quad (10)$$

These equations are quite generally valid for scatterers of arbitrary size and shape. We shall now consider the special case when the scatterers are spheres of arbitrary radius a and electromagnetic properties μ_s, ϵ_s, k_s with constant number density ρ_0 and when the incident wave is a linearly polarized plane wave and is incident normally. We shall consider both the sparse concentration case when single scattering theory is good enough as well as the dense concentration case when multiple scattering has to be taken into account.

3. SINGLE SCATTERING BY SPHERICAL SCATTERERS

When the scatterers are sparsely packed, only first-order scattering (Born approximation) need

be considered. In this case we replace the exciting field by the incident field and neglect the second term in Eq. (10). Considering the incident field of the form $\hat{\imath}_x e^{ikz}$, Eqs. (8) and (9) now involve the scattered and "transmitted" fields $[T(\mathbf{r}, \mathbf{r}')\hat{\imath}_x e^{ikz'}]$ and $[T^I(\mathbf{r}, \mathbf{r}')\hat{\imath}_x e^{ikz'}]$ at the point $\mathbf{r}$ for a sphere centered at $\mathbf{r}'$. The problem of scattering of a linearly polarized plane wave by a sphere was first solved rigorously by Gustav Mie in 1908 and the results are expressed in terms of an infinite series involving spherical vector wavefunctions.[5] Using the coordinate systems shown in Fig. 2 and letting $c_n = i^n (2n + 1)/n(n + 1)$, the various fields can be expressed as follows:

$$\mathbf{E}^i(\mathbf{r}) = \hat{\imath}_x e^{ikz} = \hat{\imath}_x e^{ik(z'+z_1)}$$

$$= e^{ikz'} \sum_{n=1}^{\infty} c_n[\mathbf{m}_{01n}^1(\mathbf{r}_1, k) - i\mathbf{n}_{e1n}^1(\mathbf{r}_1, k)],$$

$$T(\mathbf{r}, \mathbf{r}')\mathbf{E}^i(\mathbf{r}')$$

$$= e^{ikz'} \sum_{n=1}^{\infty} c_n[a_n^s\mathbf{m}_{01n}^3(\mathbf{r}_1, k) - ib_n^s\mathbf{n}_{e1n}^3(\mathbf{r}_1, k)],$$

$$T^I(\mathbf{r}, \mathbf{r}')\mathbf{E}^i(\mathbf{r}')$$

$$= e^{ikz'} \sum_{n=1}^{\infty} c_n[a_n^t\mathbf{m}_{01n}^1(\mathbf{r}_1, k_s) - ib_n^t\mathbf{n}_{e1n}^1(\mathbf{r}_1, k_s)].$$

The coefficients a_n^s, b_n^s, a_n^t, and b_n^t are obtained from boundary conditions and are functions of the radius and properties of the spheres. These fields are now expressed in terms of a coordinate system centered at the center of the scattering sphere. The position of the scatterer is taken care of by the phase factor $e^{ikz'}$. However, since the center of the scatterer, $\mathbf{r}'$, is the variable of integration, the integrand must be expressed in terms of some other coordinate system which is fixed. The translation addition theorems for spherical vector wavefunctions are very involved, in general, and, therefore, we choose the fixed coordinate system to be centered at the point of observation P. We now use the simple coordinate transformation $\mathbf{r}_2 = -\mathbf{r}_1$. It can be easily shown that

$$\mathbf{m}_{01n}^{1,3}(\mathbf{r}_1, k) = (-1)^n\mathbf{m}_{01n}^{1,3}(\mathbf{r}_2, k),$$

$$\mathbf{n}_{e1n}^{1,3}(\mathbf{r}_1, k) = (-1)^{n+1}\mathbf{n}_{e1n}^{1,3}(\mathbf{r}_2, k).$$

[5] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York 1941), p. 564. In most cases we are using his notations with slight and obvious modifications. It was pointed out by the reviewer that similar techniques as that used in this section have also been applied to the periodic case by N. Kasterin, Koning. Akd. Wentens. 4, 460 (1897) and to waves in a lattice of spherical particles based on angular momentum theorems by P. H. Morse, Proc. Acad. Sci. 42, 276 (1956).
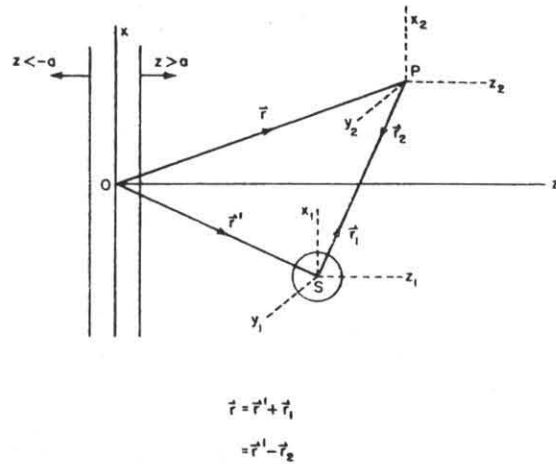
FIG. 2. The coordinate systems.

Equations (8) and (9) can, therefore, be written as

$$\langle\mathbf{E}(\mathbf{r})\rangle = (1 - v_s)\hat{\imath}_x e^{ikz}$$

$$+ (1 - v_s)\rho_0 e^{ikz} \int_{r_2>a, z_2 \geq -z} dv_2 e^{ikz_2}$$

$$\times \left[\sum_{n=1}^{\infty} (-1)^n c_n\{a_n^s\mathbf{m}_{01n}^3(\mathbf{r}_2, k) + ib_n^s\mathbf{n}_{e1n}^3(\mathbf{r}_2, k)\} \right]$$

$$+ \rho_0 e^{ikz} \int_{r_2<a} dv_2 e^{ikz_2}$$

$$\times \left[\sum_{n=1}^{\infty} (-1)^n c_n\{a_n^t\mathbf{m}_{01n}^1(\mathbf{r}_2, k_s) + ib_n^t\mathbf{n}_{e1n}^1(\mathbf{r}_2, k_s)\} \right],$$

$$(11)$$

where $\mathbf{r} \in S_1$, and

$$\langle\mathbf{E}(\mathbf{r})\rangle = \hat{\imath}_x e^{ikz} + \rho_0 e^{ikz} \int dv_2 e^{ikz_2}$$

$$\times \left[\sum_{n=1}^{\infty} (-1)^n c_n\{a_n^s\mathbf{m}_{01n}^3(\mathbf{r}_2, k) + ib_n^s\mathbf{n}_{e1n}^3(\mathbf{r}_2, k)\} \right], \quad (12)$$

where $\mathbf{r} \in S_2$. Here $v_s = \int_{r \in S_{r'}} \rho_0 \, dv' = (\frac{4}{3})\pi a^3 \rho_0$, is the fractional volume occupied by the scatterers. The domains of integration used here are shown in Fig. 3. For convenience, we restrict the point of observation P to lie outside the slab region $-a < z < a$. The integration in Eqs. (11) and (12) can be carried out exactly by expanding the vector wavefunctions $\mathbf{m}$ and $\mathbf{n}$ in terms of their Cartesian components.[6] The summation and integration can be interchanged. The integrands involved are of the form

$$e^{ikz_2}P_n^m(\cos\theta_2)z_n(kr_2)_{\sin}^{\cos}(m\phi_2),$$

where P_n^m is the associated Legendre polynomial

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Part II, p. 1899.

(a) $r_2 > a$, $z_2 \gtreqless -z$; $z > a$

(b) $r_2 < a$

(c) $z_2 \gtreqless -z$; $z < -a$

FIG. 3. The domains of integration (shaded regions).

and z_n stands for either the spherical Bessel function or spherical Hankel function. Since the domains of integration have a ϕ_2 symmetry the terms of the type

$$\int_0^{2\pi} d\phi_2 \, {\cos \atop \sin}(m\phi_2), \qquad m \neq 0,$$

go to zero. Thus the only terms which do not go to zero are those which are independent of ϕ_2. This reduces Eqs. (11) and (12) to the following:
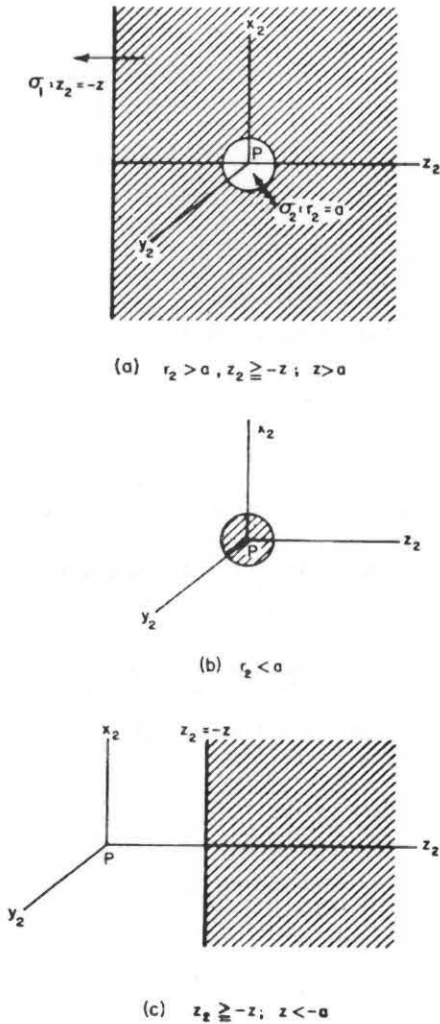
$$\langle E(\mathbf{r})\rangle_x = (1 - v_s)e^{ikz} + (1 - v_s)e^{ikz}\rho_0 \sum_{n=1}^{\infty} \frac{(-i)^n}{2}$$
$$\times \int_{r_2 > a,\, z_2 \geq -z} dv_2 e^{ikz_2}[(2n+1)a_n^s P_n(\cos\theta_2)h_n(kr_2)$$
$$+ ib_n^s\{(n+1)P_{n-1}(\cos\theta_2)h_{n-1}(kr_2)$$
$$- nP_{n+1}(\cos\theta_2)h_{n+1}(kr_2)\}] + e^{ikz}\rho_0 \sum_{n=1}^{\infty} \frac{(-i)^n}{2}$$

$$\times \int_{r_2 < a} dv_2 e^{ikz_2}[(2n+1)a_n^t P_n(\cos\theta_2)j_n(k_s r_2)$$
$$+ ib_n^t\{(n+1)P_{n-1}(\cos\theta_2)j_{n-1}(k_s r_2)$$
$$- nP_{n+1}(\cos\theta_2)j_{n+1}(k_s r_2)\}], \qquad (13)$$

when $r \in S_1$, and

$$\langle E(\mathbf{r})\rangle_x = e^{ikz} + e^{ikz}\rho_0 \sum_{n=1}^{\infty} \frac{(-i)^n}{2}\int_{z_2 \geq -z} dv_2 e^{ikz_2}$$
$$\times [(2n+1)a_n^s P_n(\cos\theta_2)h_n(kr_2)$$
$$+ ib_n^s\{(n+1)P_{n-1}(\cos\theta_2)h_{n-1}(kr_2)$$
$$- nP_{n+1}(\cos\theta_2)h_{n+1}(kr_2)\}], \qquad (14)$$

when $\mathbf{r} \in S_2$. It is thus seen that the average total field is also linearly polarized in the x direction, like the incident field. This was also found by Twersky.[3] It is to be noticed that there is no singularity in the integrands in any of the domains of integration. Previous work on this subject is restricted to the case of small spheres and the fields interior to the spheres are ignored. The problem is thus reduced to that of point scatterers. This introduces a singularity in the kernel of the integral equation and there is some ambiguity in treating such integrals since the results depend upon the shape of the volume excluded in the Cauchy principal value technique (for instance, see Waterman and Truell[4]). In our treatment, where the fields interior to the scatterers are properly accounted for, such ambiguity no longer exists. The spherical Hankel function is regular at infinity and the spherical Bessel function is regular at the origin. Typical integrals are now of the form

$$e^{ikz_2}P_n(\cos\theta_2)h_n(kr_2) \quad \text{and} \quad e^{ikz_2}P_n(\cos\theta_2)j_n(kr_2).$$

In the domain shown in Fig. 3(b), the integration can be carried out in a straightforward manner and in the domain 3(c) we make use of cylindrical coordinates and the relation[7]

$$P_n(\cos\theta_2)h_n(kr_2) = (-i)^n P_n\left(\frac{1}{ik}\frac{\partial}{\partial z_2}\right)\frac{e^{ikr_2}}{ikr_2}.$$

These integrations can then be carried out following the work of Waterman and Truell. For the domain shown in Fig. 3(a), we convert the volume integral to a surface integral by the formula

$$\int_v dv_2 e^{ikz_2} P_n(\cos\theta_2)h_n(kr_2)$$
$$= \int_\sigma \left[P_n(\cos\theta_2)h_n(kr_2)\nabla\left\{e^{ikz_2}\left(\frac{1}{4k^2} - \frac{iz_2}{2k}\right)\right\}\right.$$
$$\left. - e^{ikz_2}\left(\frac{1}{4k^2} - \frac{iz_2}{2k}\right)\nabla\{P_n(\cos\theta_2)h_n(kr_2)\}\right]\cdot d\mathbf{S}. \qquad (15)$$

[7] B. van der Pol, Physica **3**, 393 (1936).

The surface σ encloses the volume V and $d\mathbf{S}$ is in the direction of the outward normal. This formula has been developed in the report. After some lengthy computations using the above relations, the average total field can be shown to be given by

$$\langle E(\mathbf{r})\rangle_x = (1 - v_s)e^{ikz}$$

$$+ (1 - v_s)e^{ikz}\rho_0\left[\frac{\pi z}{k^2}\sum_{n=1}^{\infty}(2n+1)(a_n^s + b_n^s)\right.$$

$$+ \frac{\pi}{2k^3}\sum_{n=1}^{\infty}\{(2n+1)a_n^s\alpha_n + (n+1)b_n^s\alpha_{n-1} + nb_n^s\alpha_{n+1}\}\right]$$

$$+ \rho_0 e^{ikz}\left[\frac{\pi}{2k^3}\sum_{n=1}^{\infty}\{(2n+1)a_n^t\beta_n + (n+1)b_n^t\beta_{n-1}\right.$$

$$\left.+ nb_n^t\beta_{n+1}\}\right], \qquad \mathbf{r} \in S_1. \tag{16}$$

The symbols α_n and β_n are defined as follows

$$\alpha_n = \zeta^2[h_n(\zeta)j_n'(\zeta) + h_n'(\zeta)j_n(\zeta)$$

$$+ 2\zeta\{h_n(\zeta)j_n''(\zeta) - h_n'(\zeta)j_n'(\zeta)\}],$$

$$\beta_n = \frac{4\zeta^2}{1 - N_s^2}[N_s j_n(\zeta)j_{n-1}(N_s\zeta) - j_{n-1}(\zeta)j_n(N_s\zeta)],$$

where $\zeta = ka$, $N_s = k_s/k$ and the primes indicate differentiation with respect to the argument of the function concerned. Equation (16) can be written in the form

$$\langle E(\mathbf{r})\rangle_x = E_1^t e^{ikz}(1 + i\delta kz),$$

with obvious definitions for E_1^t and δ. If δ is small, as it will be in cases where the Born approximation is sufficiently good, then this can also be written as

$$\langle E(\mathbf{r})\rangle_x = E_1^t e^{iN_B kz}, \qquad \mathbf{r} \in S_1, \tag{17}$$

where $N_B = 1 + \delta$. Thus, the wave propagates with a "transmission" coefficient E_1^t in a medium of refractive index N_B.

In the left half-plane, Eq. (12) leads to

$$\langle E(\mathbf{r})\rangle_x = e^{ikz}$$

$$+ e^{-ikz}\left[\frac{\rho_0\pi i}{2k^3}\sum_{n=1}^{\infty}(-1)^n(2n+1)(a_n^s - b_n^s)\right]. \tag{18}$$

This equation is of the form

$$\langle E(\mathbf{r})\rangle_x = e^{ikz} + E_1^r e^{-ikz}, \qquad \mathbf{r} \in S_2. \tag{19}$$

This also shows that the right half-space containing the scatterers acts like a modified medium which reflects part of the incident field with a "reflection" coefficients E_1^r. The reflection coefficient, transmission coefficient and refractive index of the equivalent medium are functions of the size, density and electromagnetic properties of the scatterers

and of the wavelength considered. Some of their properties are discussed in Sec. 5.

4. MULTIPLE SCATTERING BY SPHERICAL SCATTERERS

4.1 The Exciting Field

In a dense distribution, the effects of multiple scattering can not be neglected. The various orders of scattering can be considered by successive iteration of the exciting field equation. However in this method, as was pointed out earlier, a new problem in integration is confronted at each stage. The complexity of the integrals involved increases very rapidly even for such a simple shape of scatterer as a sphere. We shall, therefore, direct our attention to solving Eq. (10) for the exciting field.

Most of the earlier work on multiple scattering has shown that a distribution of scatterers can be replaced by a modified homogeneous medium. Thus Foldy has obtained an expression for the refractive index of such a modified medium for the case of isotropic point scatterers. A similar result for anisotropic point scatterers has been obtained by Waterman and Truell for the case of scalar waves. The single scattering approach of Sec. 3 gives the refractive index of the modified medium when vector waves are considered and no restriction is placed on the size of the scatterers. On the basis of these results we shall assume that the exciting field can be represented by a collection of uniform plane-wave modes when multiple scattering effects are taken into account. From the geometry of the problem and the results of Born approximation it is clear that these plane waves will all travel in the positive z direction and will be linearly polarized like the incident wave. Therefore, we assume the following form for the exciting field as a trial solution

$$\langle \mathbf{E}^E(\mathbf{r} : \mathbf{r})\rangle = \sum_{l=1}^{\infty} \hat{i}_x E_l e^{ik_l z},$$

where all the k_l's are assumed to be distinct, i.e., $k_l \neq k_{l'}$ for $l = l'$. Substituting this in Eq. (10) we get

$$\sum_{l=1}^{\infty} \hat{i}_x E_l e^{ik_l z} = \hat{i}_x e^{ikz}$$

$$+ \rho_0\int_{S_r\cap S_{r'}=0} dv'\left[T(\mathbf{r},\mathbf{r}')\sum_{l=1}^{\infty}\hat{i}_x E_l e^{ik_l z'}\right]. \tag{20}$$

In order to carry out the integration, we need to

know $[T(\mathbf{r}, \mathbf{r}') \sum_{l=1}^{\infty} \hat{\imath}_z E_l^{i k_l z'}]$, which is the scattered field at $\mathbf{r}$ from a scatterer at $\mathbf{r}'$ excited by a collection of plane waves of the type $\hat{\imath}_x E_l \exp i k_l z$. Each of these plane waves gives rise to a scattered field which travels in a medium of propagation constant k, and a "transmitted" field inside the scatterer where the propagation constant is k_s. Thus using the coordinate system of Fig. 2, we can write the incident, scattered and transmitted fields of each plane-wave mode as follows:

$$\hat{\imath}_x E_l e^{i k_l z} = E_l e^{i k_l z'} \sum_{n=1}^{\infty} c_n [\mathbf{m}_{01n}^1(\mathbf{r}_1, k_l) - i \mathbf{n}_{e1n}^1(\mathbf{r}_1, k_l)],$$

(21a)

$$T(\mathbf{r}, \mathbf{r}') \hat{\imath}_x E_l e^{i k_l z'}$$

$$= E_l e^{i k_l z'} \sum_{n=1}^{\infty} c_n [A_{ln}^s \mathbf{m}_{01n}^3(\mathbf{r}_1, k) - i B_{ln}^s \mathbf{n}_{e1n}^3(\mathbf{r}_1, k)],$$

(21b)

$$T^t(\mathbf{r}, \mathbf{r}') \hat{\imath}_x E_l e^{i k_l z'}$$

$$= E_l e^{i k_l z'} \sum_{n=1}^{\infty} c_n [A_{ln}^t \mathbf{m}_{01n}^1(\mathbf{r}_1, k_s) - i B_{ln}^t \mathbf{n}_{e1n}^1(\mathbf{r}_1, k_s)].$$

(21c)

The media corresponding to the incident, scattered and transmitted fields are characterized by (μ_l, ϵ_l, k_l), (μ, ϵ, k), and (μ_s, ϵ_s, k_s), respectively. This is the so-called "two-exterior" formalism of Twersky,[3] indicating that the incident and scattered fields travel in two different media. The coefficients are found from boundary conditions and are given by

$$A_{ln}^s = \frac{\mu}{\mu_l} \frac{\mu_s j_n(N_s \zeta)[N_l \zeta j_n(N_l \zeta)]' - \mu_l j_n(N_l \zeta)[N_s \zeta j_n(N_s \zeta)]'}{\mu h_n(\zeta)[N_s \zeta j_n(N_s \zeta)]' - \mu_s j_n(N_s \zeta)[\zeta h_n(\zeta)]'},$$

$$A_{ln}^t = \frac{\mu_s}{\mu_l} \frac{\mu h_n(\zeta)[N_l \zeta j_n(N_l \zeta)]' - \mu_l j_n(N_l \zeta)[\zeta h_n(\zeta)]'}{\mu h_n(\zeta)[N_s \zeta j_n(N_s \zeta)]' - \mu_s j_n(N_s \zeta)[\zeta h_n(\zeta)]'},$$

(22)

$$B_{ln}^s = \frac{\epsilon_l}{\epsilon N_l} A_{ln}^s(\mu \to \epsilon); \qquad B_{ln}^t = \frac{\epsilon_l}{\epsilon_s} \frac{N_s}{N_l} A_{ln}^t(\mu \to \epsilon).$$

The notation $A_{ln}^{s,t}(\mu \to \epsilon)$ means that μ, μ_s, μ_l are to be replaced by ϵ, ϵ_s, ϵ_l in the expressions for $A_{ln}^{s,t}$. The relations

$$k^2 = \omega^2 \mu \epsilon, \qquad k_l^2 = \omega^2 \mu_l \epsilon_l, \qquad k_s^2 = \omega^2 \mu_s \epsilon_s,$$

$$k_s = N_s k, \qquad k_l = N_l k$$

have also been used.

We can now substitute Eq. (21b) in Eq. (20) and carry out the integration by referring the whole integrand to a coordinate system centered at $\mathbf{r}$. The domain of integration in this case is the right half-space excluding a sphere of radius $2a$ centered at $\mathbf{r}$, since the sphere the sphere at $\mathbf{r}'$ must not penetrate the sphere at $\mathbf{r}$. The integration is carried out by converting to a surface integral using Green's theorem. Some lengthy but straightforward computations lead to the equation

$$\sum_{l=1}^{\infty} E_l e^{i k_l z} = e^{i k z}$$

$$+ \sum_{l=1}^{\infty} E_l e^{i k z} \frac{\pi i \rho_0}{k^3 (N_l - 1)} \left[\sum_{n=1}^{\infty} (2n + 1)(A_{ln}^s + B_{ln}^s) \right]$$

$$+ \sum_{l=1}^{\infty} E_l e^{i k_l z} \frac{2 \pi \rho_0}{k^3 (N_l^2 - 1)} \left[\sum_{n=1}^{\infty} \{(2n + 1) A_{ln}^s \gamma_{ln} \right.$$

$$\left. + (n + 1) B_{ln}^s \gamma_{l,n-1} + n B_{ln}^s \gamma_{l,n+1} \} \right],$$

(23)

where

$$\gamma_{ln} = (2\zeta)^2 [N_l j_{n-1}(2 N_l \zeta) h_n(2\zeta) - j_n(2 N_l \zeta) h_{n-1}(2\zeta)].$$

Since this equation is true for all $z > 2a$ in the right half-space, we can equate the coefficients of $e^{i k_l z}$, for all l, and of $e^{i k z}$ and get the following equations

$$\sum_{n=1}^{\infty} [(2n + 1) A_{ln}^s \gamma_{ln} + (n + 1) B_{ln}^s \gamma_{l,n-1} + n B_{ln}^s \gamma_{l,n+1}]$$

$$= (2\zeta^3 / 3 v_s)(N_l^2 - 1), \qquad l = 1, 2, \cdots, \quad (24)$$

and

$$\sum_{l=1}^{\infty} \rho_0 E_l \frac{3 i v_s}{4 \zeta^3 (N_l - 1)}$$

$$\times \left[\sum_{n=1}^{\infty} (2n + 1)(A_{ln}^s + B_{ln}^s) \right] + 1 = 0. \quad (25)$$

Equation (24) is the dispersion relation governing the refractive index of the medium. Its roots are the different modes which the medium can sustain. However, since the coefficients A_{ln}^s and B_{ln}^s involve both N_l and μ_l, one more equation is needed. This can be derived by considering an incident $\mathbf{H}$ field so that $\mathbf{H}^i(\mathbf{r}) = \hat{\imath}_x e^{i k z}$. A similar analysis leads to the equation

$$\sum_{n=1}^{\infty} [(2n + 1) C_{ln}^s \gamma_{ln} + (n + 1) D_{ln}^s \gamma_{l,n-1} + n D_{ln}^s \gamma_{l,n+1}]$$

$$= (2\zeta^3 / 3 v_s)(N_l^2 - 1), \quad (26)$$

where

$$C_{ln}^s = \mu_l B_{ln}^s / N_l \mu, \quad \text{and} \quad D_{ln}^s = N_l \mu_l A_{ln}^s / \mu.$$

Between Eqs. (24) and (26) we can get a transcendental equation in which the only unknown is N_l. The different modes will be governed by this equation.

4.2 The Average Total Field

The average total field can be derived in a straightforward manner by substituting Eqs. (21a), (21b), (21c) into Eqs. (8) and (9). The techniques of integration are similar to those outlined in Sec. 3. Equation (8) leads to the following equation giving the average total field in the right half-space ($z > a$)

$$\langle E(\mathbf{r}) \rangle_z = (1 - v_s) e^{ikz}$$

$$+ \sum_{l=1}^{\infty} (1 - v_s) E_l \rho_0 \frac{\pi i}{k^3 (N_l - 1)} e^{ikz}$$

$$\times \left[\sum_{n=1}^{\infty} (2n + 1)(A_{ln}^s + B_{ln}^s) \right]$$

$$+ \sum_{l=1}^{\infty} E_l \rho_0 \frac{2\pi}{k^3} e^{ik_l z} \left[\frac{1 - v_s}{N_l^2 - 1} \sum_{n=1}^{\infty} \{(2n + 1) A_{ln}^s \delta_{ln} \right.$$

$$+ (n + 1) B_{ln}^s \delta_{l,n-1} + n B_{ln}^s \delta_{l,n+1}\}$$

$$+ \frac{1}{N_l^2 - N_s^2} \sum_{n=1}^{\infty} \{(2n + 1) A_{ln}^t \epsilon_{ln} $$

$$+ (n + 1) B_{ln}^t \epsilon_{l,n-1} + n B_{ln}^t \epsilon_{l,n+1}\} \Bigg],$$

where

$$\delta_{ln} = \zeta^2 [N_l j_{n-1}(N_l \zeta) h_n(\zeta) - j_n(N_l \zeta) h_{n-1}(\zeta)]$$

and

$$\epsilon_{ln} = \zeta^2 [N_s j_n(N_l \zeta) j_{n-1}(N_s \zeta) - N_l j_{n-1}(N_l \zeta) j_n(N_s \zeta)].$$

By virtue of Eq. (25), the first two terms of this equation add up to zero. The equation, therefore, reduces to the form

$$\langle E(\mathbf{r}) \rangle_z = \sum_{l=1}^{\infty} E_l^t e^{ik_l z}, \quad z > a. \quad (27)$$

Thus, the average total field propagates in the

right half-space as a collection of plane-wave modes. The extinction theorem is verified since there is no e^{ikz} component in the field. The transmission coefficients E_l^t are given by the equation

$$E_l^t = \frac{3v_s(1 - v_s)}{2\zeta^3(N_l^2 - 1)} E_l \sum_{n=1}^{\infty} [(2n + 1) A_{ln}^s \delta_{ln}$$

$$+ (n + 1) B_{ln}^s \delta_{l,n-1} + n B_{ln}^s \delta_{l,n+1}]$$

$$+ \frac{3v_s}{2\zeta^3(N_l^2 - N_s^2)} E_l \sum_{n=1}^{\infty} [(2n + 1) A_{ln}^t \epsilon_{ln}$$

$$+ (n + 1) B_{ln}^t \epsilon_{l,n-1} + n B_{ln}^t \epsilon_{l,n+1}]. \quad (28)$$

The average total field in the left half-space is similarly obtained from Eq. (9) and is given by

$$\langle E(\mathbf{r}) \rangle_z = e^{ikz} + E^r e^{-ikz}, \quad z < -a, \quad (29)$$

where

$$E^r = \sum_{l=1}^{\infty} \frac{3iv_s}{4\zeta^3(N_l + 1)}$$

$$\times E_l \left[\sum_{n=1}^{\infty} (-1)^n (2n + 1)(A_{ln}^s - B_{ln}^s) \right]. \quad (30)$$

This treatment has given a fairly good picture of the multiply scattered field. There is not enough information to determine uniquely the amplitudes E_l of the plane-wave modes. Because of the complexity of integrals the treatment has excluded from consideration the region $-a < z < a$. However, sufficient information has been obtained to determine the refractive index of the modified medium from the dispersion relation.

5. SPECIAL CASES

The results obtained in Secs. 3 and 4 for single and multiple scattering reduce to simple and well known results when special cases are considered. Thus at low frequencies, when the radius of the sphere is small compared to the wavelength, the parameter $\zeta(=ka)$ is very small. In this case we can take the small-argument approximation of spherical Bessel and Hankel functions. In the case of single scattering the refractive index for this case is given by

$$N_B = 1 + \frac{3}{2} v_s \left[\frac{\mu_s - \mu}{\mu_s + 2\mu} - \frac{\mu_s - \mu N_s^2}{2\mu_s + \mu N_s^2} \right] \Big/ \left\{ 1 + \frac{v_s}{4} \left[\frac{\mu_s - \mu}{\mu_s + 2\mu} - \frac{4}{5} \frac{\mu_s - \mu N_s^2}{2\mu_s + \mu N_s^2} \right] - \frac{3v_s}{(1 - v_s)} \frac{\mu_s}{(2\mu_s - \mu N_s^2)} \right\},$$

$$(31)$$

where ζ, $N_s \zeta \ll 1$. If the permeability μ_s of the spheres is nearly equal to that of the background medium and the case of sparse concentration of spheres is considered, then $\mu_s \approx \mu$ and $v_s \ll 1$.

In this case, to the first power of v_s we get

$$N_B = 1 + \frac{3}{2} v_s \frac{N_s^2 - 1}{N_s^2 + 2} = 1 + 2\pi\rho_0\alpha, \quad (32)$$

where the polarizability of spheres is $\alpha = a^3(N_s^2 - 1)/(N_s^2 + 2)$. This is the well known refractive index of Rayleigh scattering given in the Clausius–Mossotti form. The small-sphere approximation is equivalent to neglecting all orders of multipoles, except the first-order electric dipole in the Mie expansion.

For the case of small perfectly conducting spheres, we cannot let $N_s \rightarrow \infty$ directly in the above equation, since this was derived for $N_s\zeta \ll 1$. However, taking asymptotic expansions for a_n^s and b_n^s for $N_s \rightarrow \infty$ and $\zeta \ll 1$, the total field and refractive index in this case are given by

$$\langle E(\mathbf{r}) \rangle_x = \begin{cases} (1 - \tfrac{9}{8}v_s)e^{iN_Bkz}, & z > a, \\ e^{ikz} - \tfrac{9}{8}v_s e^{-ikz}, & z < -a, \end{cases} \quad (33)$$

and

$$N_B = 1 + \tfrac{3}{4}v_s. \quad (34)$$

When the sphere size is comparable to wavelength the contribution of the higher order multipoles can no longer be neglected. In this case the refractive index N_B for the Born approximation will, in general, have an imaginary part also, indicating attenuation in the medium. Numerical values of N_B for perfectly conducting spheres for values of ζ from 0.1 to 5 and for v_s from 0.001 to 0.1 have been computed and are given in the report.

For the case of multiple scattering, when ζ, $N_s\zeta$, and $N_l\zeta$ are much less than unity, Eqs. (24) and (26) reduce to

$$3\frac{\mu}{\mu_l}\left[\frac{\mu_s - \mu_l}{\mu_s + 2\mu}\right]N_l^2 + \left[\frac{\epsilon_s - \epsilon_l}{\epsilon_s + 2\epsilon}\right](2 + N_l^2)$$
$$= (N_l^2 - 1)/v_s,$$

$$3\frac{\epsilon}{\epsilon_l}\left[\frac{\epsilon_s - \epsilon_l}{\epsilon_s + 2\epsilon}\right]N_l^2 + \left[\frac{\mu_s - \mu_l}{\mu_s + 2\mu}\right](2 + N_l^2)$$
$$= (N_l^2 - 1)/v_s.$$

The constants μ_l and ϵ_l can be determined from these equations for any type of spheres. Suitable combination of these two equations yields a quartic equation in $\mu_l\epsilon_l$. To the lowest order in v_s the refractive index N_l is 1. To the next order in v_s we obtain

$$N_l = 1 + \tfrac{3}{2}v_s\left(\frac{\epsilon_s/\epsilon - 1}{\epsilon_s/\epsilon + 2} + \frac{\mu_s/\mu - 1}{\mu_s/\mu + 2}\right), \quad (35)$$

which is just the generalization of the Clausius–Mossotti relation given in Eq. (32).

When the spheres are perfectly conducting, the dispersion relation reduces to the following

$$\left(\frac{7}{18}\zeta^2\right)N_l^4 - \left[\frac{1}{3}\left(1 + \frac{2}{v_s}\right) + \frac{23}{10}\zeta^2 + \left(\frac{1}{9}i\right)\zeta^3\right]N_l^2$$
$$+ \left[\frac{2}{3}\left(2 + \frac{1}{v_s}\right) + \frac{10}{3}\zeta^2 - \left(\frac{1}{9}i\right)\zeta^3\right] = 0.$$

For very small values of ζ, we have only one mode with refractive index given by

$$N_l^2 = (1 + 2v_s)/(1 + \tfrac{1}{2}v_s).$$

This expression has been derived earlier by Twersky.[3] For sparse concentrations, we get back the refractive index and transmission and reflection coefficients obtained in the case of Born approximation. By comparing the results with the standard expressions for the transmission and reflection coefficients, we get

$$\mu_l = \mu(1 - \tfrac{3}{2}v_s), \qquad \epsilon_l = \epsilon(1 + 3v_s).$$

These expressions also agree with Twersky's.[3] As ζ increases we get more than one mode and their refractives indices have to be computed numerically.

Dyadic Analysis of Space–Time Congruences*

Frank B. Estabrook and Hugo D. Wahlquist

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received 9 June 1964)

A physical 3-vector and dyadic formalism for the treatment of general relativistic problems is derived, by systematic introduction of a proper tetrad field. The method is especially appropriate when there exists a physically or geometrically preferred timelike congruence; all quantities in the formalism are then shown to have immediate physical interpretation as proper local observables. A complete and nonredundant set of equations for the analysis of timelike congruences is developed in this operational language. Application is made to some simple examples involving local observations, and the direct measurement of the Riemann tensor discussed.

A. INTRODUCTION

THE spinor analysis and the tetrad (or vierbein) formalism were both employed in the 1930's, in connection with attempts to generalize general relativity and to formulate a unified theory of electricity and gravitation. The lack of success in this particular endeavor, however, led to a subsequent lack of interest in the formal techniques thus opened up. Now, in just the last few years, greatly renewed interest in the spinor analysis has followed upon its successful application to cases of gravitational radiation, *within* the now-classical theory of Einsteinian general relativity.[1] Such cases are characterized by having preferred null congruences. The tetrad formalism, we believe, can be of equally great service, within Einstein theory, when appropriately applied to situations having preferred timelike congruences. When a tetrad formalism is based on a preferred congruence it then naturally leads to a three-dimensional dyadic and vector formulation which explicitly depends on (and expresses) the dimensionality and signature of physical space–time. For the many important results that depend on this dimensionality and signature for their validity, the usual tensor calculus is rather an imperfect instrument, tending to prove easily only more general results, valid in n dimensions with arbitrary signatures.

The dyadic formalism we present in the present paper has the advantages of physical *interpretability*, mathematical *completeness*, and wide *applicability*. We are at considerable pains in several sections of the paper to give the physical interpretation of all dyadic quantities arising from the formalism—in almost all cases this is rather easily done, for indeed the naturally occurring dyadic quantities are found to be those already familiar either from classical mechanics or from quite simple geometric considerations. The result is a much more understandable set of relations, than in the more customary 4-tensor formulation of general relativity, especially when a physically distinguished congruence is present. The second advantage is in the completeness of the dyadic partial differential equations. The more usual tensorial techniques for discussing congruences in curved (3 + 1)-dimensional manifolds are quite *ad hoc*, and although the literature is replete with many elegant results for special cases, a systematic mathematical approach or algorithm which overlooks no such results, writes no redundant equations, and yet is completely general, seems not to be available. Although this technical point is difficult to express in an introduction, it should become clear in the body of the paper. Finally, there promise to be many areas of application of the dyadic formalism: a timelike congruence which is in some way distinguished or preferred occurs in such varied situations as space–times supporting matter–energy distributions, cosmological models with preferred galactic distributions, and space–times having symmetries and isometries described by congruences. The possibility of generating new exact solutions of the field equations should also be mentioned, especially since the dyadic formalism is not wedded to a choice of (holonomic) coordinates. The applicability of the dyadic formalism to the explicit prediction of experimental results is noteworthy: the dyadic quantities are world scalars, proper components everywhere resolved along the orthogonal space and time axes of local Lorentz tetrads; they are, that is, precisely the raw material of observational physics. We demonstrate this last point by presenting equations for the differential absolute acceleration and precession between adjacent inertially oriented test particles, which show in principle how 14 components of the Riemann tensor are locally measurable.

[1] E. Newman and R. Penrose, J. Math Phys. **3**, 566 (1962).

The differential precession equation in particular seems not to have been given previously in a form involving strictly local, proper, observations, and uniting the differential Thomas precession of accelerating particles with the general relativistic Fokker precession.

In Sec. B of this paper we discuss tetrad fields and the formulation of general relativity in terms of such anholonomic reference systems. Section C introduces the 3-dyadic formalism, based on a tetrad field attached to a preferred timelike congruence, and elucidates the physical significance of the dyadic quantities. In Sec. D we transcribe the tetrad equations into this physical dyadic language.

In a succeeding paper[2] we will demonstrate the utility of the dyadic formalism in a consideration of the Herglotz–Nöther theorem on the motion of Born-rigid bodies, which assumes a simplicity otherwise entirely concealed. In addition we will there derive new results for Born congruences in curved space–times. In future papers, we intend to present the dyadic method applied in several other situations having, again, preferred timelike congruences.

B. TETRAD FORMALISM

1. Tetrad Fields

The use of auxiliary ennuples in differential geometry is of course not new, going back at least to the work of Ricci. To introduce the 3-dyadic treatment of Secs. C and D, we nevertheless must briefly recapitulate in a uniform notation much of the formalism expounded, for example, in Schouten[3] and Eisenhart.[4]

The method of analysis follows upon systematic introduction of a tetrad field based on a given timelike congruence; we will in fact use four orthonormal reference vector fields $_r\lambda^\mu$, where $r = 0$ labels a timelike vector, and $r = 1, 2, 3$ are three spacelike vectors. The label r is a "Lorentz index" in the terminology of Synge,[5] and we will reserve Latin indices for this purpose. These unit vector fields $_r\lambda^\mu$ will trace out four congruences not, in general, 3-surface orthogonal. The method is thus equivalent to the introduction of convenient, everywhere orthogonal, but anholonomic coordinates, in the terminology of Schouten.[3]

By transvection with the contravariant tetrad vectors $_r\lambda^\mu$ or their covariant duals, $^s\lambda_\nu$, we will systematically "strangle" all tensor indices of fields of interest, thus replacing these indices by Lorentz indices, labeling the resulting arrays of world scalars. This formalism in many ways bridges the conventional approach in which tensors are considered as arrays of components, and that of the school of Cartan, with its perhaps more physical emphasis on algebraic quantities in tangent vector spaces.[6]

At any point of space–time, the given timelike congruence, and in particular the orthonormal vector tetrad there, defines a preferred local Minkowskian frame, with respect to which Lorentz indices take meaning as labeling proper components, spacelike, timelike, and mixed. We will use the special relativistic Minkowski metric form $\eta^{rs} = \eta_{rs} = \mathrm{diag}(-1, 1, 1, 1)$ to raise and lower Lorentz indices, and so to express the tetrad orthonormality relations

$$^r\lambda^\mu \, ^s\lambda_\mu = \eta^{rs}, \quad _r\lambda^\mu \, _s\lambda_\mu = \eta_{rs}. \tag{B.1}$$

The metric tensor components are, as in the Cartan formalism, simply given by quadratic forms in the unit vectors:

$$g_{\mu\nu} = {}^r\lambda_\mu \, _r\lambda_\nu, \quad g^{\mu\nu} = {}^r\lambda^\mu \, _r\lambda^\nu,$$
$$g^\mu_\nu \equiv \delta^\mu_\nu = {}^r\lambda^\mu \, _r\lambda_\nu. \tag{B.2}$$

In general, it appears that results which are valid *only* for a certain dimensionality and signature of a space, are much more easily and directly demonstrated with such a tetrad formalism. The main algebraic inconvenience which will arise is due to the lack of commutivity in the process of successive "intrinsic" differentiation of scalars (i.e., absolute differentiation along the unit vector fields); we derive the necessary commutation formulas in Sec. B3.

2. The Object of Anholonomy

The variation of the tetrad field is described by the set of strangled intrinsic derivatives of the unit vectors:

$$\Gamma_{rst} \equiv {}_t\lambda^\mu \, _s\lambda_{\mu;\nu} \, _r\lambda^\nu. \tag{B.3}$$

These are essentially the "rotation coefficients" introduced by Ricci. It is shown in Sec. B3 that the set of scalars Γ_{rst} can properly be regarded as the anholonomic components of the affinity in our 3 + 1 metric space. From Eq. (B.1) it immediately follows that $\Gamma_{rst} = -\Gamma_{rts}$ and indeed there are here exactly 24 scalar fields. A more elegant set of 24 scalars,

[2] H. D. Wahlquist and F. B. Estabrook, unpublished.
[3] J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.
[4] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1926).
[5] J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960).

[6] See, for example, A. Lichnerowicz, *Elements of Tensor Calculus* (Methuen and Company, Ltd., London, 1962).

however, may be defined using only simple curls of the vector fields:

$$\Omega^r_{.st} \equiv \tfrac{1}{2}(^r\lambda_{\mu,\nu} - {}^r\lambda_{\nu,\mu})\lambda^\mu_{.s}\lambda^\nu_{.t}. \qquad (B.4)$$

In our metric space this is equivalent to

$$\Omega_{rst} = \tfrac{1}{2}(\Gamma_{tsr} - \Gamma_{str}). \qquad (B.5)$$

The Ω_{tsr} are again antisymmetric on the last pair of indices: $\Omega_{tsr} = -\Omega_{trs}$. For the present case of orthonormal unit vectors the Eq. (B.5) can be readily solved for the anholonomic affinity components, which demonstrates the equivalence of the two sets:

$$\Gamma_{tsr} = \Omega_{str} + \Omega_{rst} + \Omega_{tsr}. \qquad (B.6)$$

It is thus clear that the curls of the tetrad fields carry all the metric information, and so knowledge of the 40 Christoffel symbols is not now required. This is an advantage of an orthonormal tetrad formulation, also met with in the spinor calculus, where there are just 24 components in the spin connection. In the following we give explicit expressions for the Riemann tensor components in terms of the Γ_{tsr} fields.

The components $\Omega^r_{.st}$ defined as in Eq. (B.4) are termed the "object of anholonomity" by Schouten,[3] who introduces them in general, non-Riemannian, spaces. The vanishing of the $\Omega^r_{.st}$ everywhere is the integrability condition for the unit vectors to be gradients of four families of hypersurfaces—hence, derivable from ordinary or holonomic coordinates. In our present case, the vanishing of $\Omega^r_{.st}$ would imply the existence of four everywhere orthogonal, equally spaced (hence, Cartesian) coordinate families, which is to say, the flatness of space–time.

Intrinsic differentiation of Eq. (B.4) with respect to $_p\lambda^\sigma$ and subsequent complete antisymmetrization with respect to Lorentz indices s, t, and p, results in a set of 16 first-order differential identities:

$$\Omega^r_{.[st,p]} = 2\Omega^q_{.[ps}\Omega^r_{.t]q}. \qquad (B.7)$$

Here the brackets denote complete antisymmetrization—in the case of three indices, this involves adding six terms with appropriate signs according to the even or odd permutation of the indices, and multiplication by $\tfrac{1}{6}$. These equations are to be found in Ref. 3, p. 101; they are in fact *integrability* conditions on the 24 world scalar fields $\Omega^r_{.st}$, allowing them to be derivable from four congruences or vector fields $_s\lambda^\mu$ in the manner given.

The 16 integrability conditions are especially noteworthy, in that the metric properties of space–time nowhere enter in their derivation. There are twenty other equations implied in a metric space–

time by the form of Eq. (B.5); when second covariant derivatives are eliminated by antisymmetrization (this time on two indices only) components of the strangled Riemann tensor R^{rstp} are introduced. If the 16 relations already written are systematically eliminated, by imposing the algebraic symmetries of the Riemann tensor in metric 4-space, one finally obtains the further independent set:

$$\begin{aligned}
\Omega^{(rp)(s.t)} + \Omega^{(st)(r.p)} &= 2\Omega^{(rp)q}\Omega^{(st)}_{..q} - \Omega^{(sp)q}\Omega^{(rt)}_{..q} \\
&\quad - \Omega^{(rs)q}\Omega^{(tp)}_{..q} - \tfrac{3}{2}\Omega^{q(t(p}\Omega^{r)s)}_{..q} \\
&\quad + \Omega^{q(p(t}\Omega^{s)r)}_{..q} + \Omega^{q(t(p}\Omega^{r)s)}_{..q} - \tfrac{3}{4}S^{strp}. \quad (B.8)
\end{aligned}$$

Here we have used parentheses to denote total symmetrization—in the case of two indices, for example, this means summation of two terms with indices transposed, and multiplication by $\tfrac{1}{2}$. In addition, it has proved algebraically convenient to use the *symmetrized* Riemann tensor (Ref. 5, p. 54),

$$S^{strp} = -\tfrac{1}{3}(R^{srtp} + R^{sptr}). \qquad (B.9)$$

It is clear that all of Eq. (B.8) has the same symmetries as S^{strp}: viz., symmetry on the first pair of indices, symmetry on the second pair, symmetry on the two pairs of indices, and a cyclic symmetry on, say, the last three indices. Hence there are precisely 20 independent relations in Eq. (B.8). The complete set of 36 differential relations for the tetrad field, consists of Eqs. (B.7) and (B.8).

Although their separate origins are obscured by the process, it is often convenient to have Eqs. (B.7) and (B.8) written together in one set of 36 equations involving the usual Riemann tensor, the anholonomic affinity components, and their intrinsic derivatives (Ref. 4, p. 98):

$$\begin{aligned}
\Gamma^{[t|sr|,p]} &= \tfrac{1}{2}\Gamma^{prq}\Gamma^{ts}_{..q} - \tfrac{1}{2}\Gamma^{trq}\Gamma^{ps}_{..q} \\
&\quad + \Gamma^{[p|t|q}\Gamma^{sr}_q + \tfrac{1}{2}R^{srtp}, \qquad (B.10)
\end{aligned}$$

where indices enclosed between bars are excluded from the antisymmetrization brackets. Equation (B.10) is, of course, also the promised direct expression of the components of the Riemann tensor in terms of the tetrad field.

3. Further Relations

In Sec. D the dyadic forms of Eqs. (B.7) and (B.8) are presented as a general tool for the analysis of space–time congruences. We must, however, first supplement these equations by commutation formulas, and by the Bianchi Identities.

Because of the anholonomy, two successive intrinsic derivative indices do not commute—even though they are derivatives of world scalars. This is easily seen from the definition of intrinsic deriva-

tive; it is perhaps more illuminating, however, to derive the important resulting commutation formula from the general formula for strangulation of covariant derivatives. Consider a tensor $T^{r\cdots}_{\cdots\mu\cdots;\sigma}$ with a single covariant differentiation index; strangle by multiplication with ${}^t\lambda_\tau \cdots {}_u\lambda^\mu \cdots {}_s\lambda^\sigma$; using the orthonormal properties of the tetrad, the resulting expression can be written in terms of intrinsic derivatives:

$$(T^{r\cdots}_{\cdots\mu\cdots;\sigma})^t\lambda_\tau \cdots {}_u\lambda^\mu \cdots {}_s\lambda^\sigma = T^{t\cdots}_{\cdots u,s}$$
$$+ \; \Gamma^{\cdot\cdot t}_{sq}T^{q\cdots}_{\cdots u\cdots} + \cdots - \Gamma^{\cdot\cdot q}_{su}T^{t\cdots}_{\cdots q\cdots} - \cdots . \quad \text{(B.11)}$$

In this scalar expression the set of $\Gamma^{\cdots r}_{si}$ plays exactly the formal role of an affinity. We emphasize, however, that whereas with ordinary holonomic coordinates an affinity in a Riemann space is symmetric on the first two indices (and so in four dimensions has 40 components), as a result of the orthonormal nature of our present anholonomic reference frame $\Gamma^{\cdots r}_{si}$ is antisymmetric on the last two indices and in four dimensions has 24 components.[7]

Since we may commute covariant derivatives of any scalar, $T^{t\cdots}_{\cdots u\cdots;[\nu\sigma]} = 0$, it then follows immediately upon strangulation according to Eq. (B.11) that the commutation formula for intrinsic differentiation is (suppressing all nonderivative Lorentz indices)

$$T_{,[rs]} = \Gamma^{\cdots q}_{[sr]}T_{,q} = \Omega^q_{\cdot rs}T_{,q}. \quad \text{(B.12)}$$

We conclude this section by recording the integrability conditions for the (20) components of the Riemann tensor field, allowing them to be derivable as in Eq. (B.8). If we are *given* a Riemannian metric form, these conditions are of course identically satisfied: they are indeed the Identities of Bianchi. In our tetrad notation, they follow readily upon intrinsic differentiation of Eq. (B.8), antisymmetrization, and use of the commutation relation Eq. (B.12) to eliminate second derivatives. The Bianchi Identities may be most easily written in terms of the strangled double-dual of the Riemann tensor; they are

$$*R^{prst}_{\cdots\cdots,t} + 2*R^{sqt[r}\Gamma^{\cdot\cdot p]}_{qt} + 2*R^{prt[q}\Gamma^{\cdot\cdot s]}_{qt} = 0, \quad \text{(B.13)}$$

where

$$*R^{prst} = \tfrac{1}{4}\epsilon^{prqn}\epsilon^{stvu}R_{qnvu}, \quad \text{(B.14)}$$

and ϵ^{prqn} is the usual four-dimensional permutation symbol. As is immediately obvious in the dyadic notation, there are exactly 20 independent conditions in Eq. (B.13); these include the four contracted Bianchi Identities. These 20 equations are of great importance and utility when deriving the

consequences of special assumptions and symmetries imposed on the gravitational field, and on the stress-energy tensor; both of these, in Einstein's theory, are comprised in the geometrical Riemann tensor.

C. 3-VECTOR AND 3-DYADIC ALGEBRAIC FORMALISM AND INTERPRETATION

1. Introduction

In the general tetrad formalism the associated congruences are geometrical reference objects more or less devoid of intrinsic physical significance. If, however, we identify the timelike congruence with the world lines of a material continuum, described by the velocity 4-vector field ${}_0\lambda^\mu$, this is no longer the case. The ${}_0\lambda^\mu$ congruence might represent, in various instances, the motion of a relativistic fluid, a rigid body as defined by Born's constraint condition, a proper frame of reference for the performance of experiments, or a privileged cosmological matter distribution. But regardless of the particulars, it is the attitude of considering the timelike congruence to be a physically given object that provides the rationale for the 3-dyadic formalism to be presented here. A region of space–time in which such a congruence exists is endowed with a unique time direction at each point, and it becomes physically reasonable then to dissolve the 4-dimensional union of space and time with respect to the congruence. Of course, such a decoupling is almost always done at some point in any physical problem in relativity theory, by selection of a "convenient" set of coordinates. With the tetrad and dyadic formalisms this is done at the outset before further specification of the particular system at hand, and without prejudice as to the admissibility or desirability of any holonomic coordinate system.

In Sec. C 2 we introduce a representation of the anholonomic affinity, $\Gamma^{\cdots i}_{rs}$, by splitting its components into independent three-dimensional arrays having spacelike Lorentz indices only. The three spacelike tetrad vectors used to generate these components are not, of course, unique. In Sec. C3 it is shown that certain restricted transformations between sets of these auxiliary vectors are the analogs of the familiar orthogonal rotations of Cartesian axes in 3-space, and that the arrays of proper components will transform precisely as conventional 3-vector or 3-dyadic fields under such spatial rotations. A detailed discussion of the kinematical and geometrical significance of the quantities thus introduced is given in Secs. C4 and C5.

---

[7] It is mnemonically most convenient to write *all* the "correction" terms in Eq. (B.11) with plus signs, summing always on the *second* index of the anholonomic affinity.

2. Three-Dimensional Representation of Γ_{rst}

The splitting apart of the components of the anholonomic affinity into independent 3-dimensional arrays proceeds by segregating those components which differ in the number and location of timelike indices, here denoted by 0. It should be noted that raising or lowering the 0 index changes the sign of a quantity. We shall henceforth reserve the letters from the first half of the Latin alphabet $(a \cdots m)$ to indicate spacelike indices. These take on the values 1, 2, 3 only, and *the summation convention for such indices is limited* to this range. Since the local spacelike metric $\eta^{ab} = \delta_b^a$, the vertical position of these indices does not matter. Parentheses and brackets around indices have the same significance as in Sec. B, and ϵ_{abc} denotes the usual three-dimensional permutation symbol.

With these conventions, the components of Γ_{rst} having at least one timelike index may be written:

$$\Gamma_{00a} = -\Gamma_{0a0} \equiv a_a, \tag{C.1}$$

$$-\Gamma_{ab0} = \Gamma_{a0b} \equiv S_{ab} + \epsilon_{abc}\Omega^c, \tag{C.2}$$

$$\Gamma_{0ab} = -\Gamma_{0ba} \equiv \epsilon_{abc}\omega^c, \tag{C.3}$$

where the quantities on the right constitute a three-dimensional representation consistent with the antisymmetry of Γ_{rst} on its last two indices. The array of scalars, S_{ab}, is defined to be symmetric to the interchange of a and b; from Eq. (C.2) it follows that

$$S_{ab} \equiv -\Gamma_{(ab)0}. \tag{C.4}$$

These definitions provide a representation for 15 of the 24 independent components of the affinity. The remaining nine, comprised in the wholly spacelike Γ_{abc}, describe characteristics of the nonunique auxiliary congruences. Again by virtue of the antisymmetry on b and c, we may represent six of these quantities by a symmetric array, N_{ad}, and the final three by L_b as follows:

$$\tfrac{1}{2}\epsilon_{dcb}\Gamma_a^{.bc} \equiv N_{ad} - \tfrac{1}{2}N_{.b}^b\delta_{ad} + \epsilon_{ad.}^{..b}L_b, \tag{C.5}$$

where the contraction, $N_{.b}^b$, has been explicitly subtracted for reasons of formal simplicity later. From this equation we further have:

$$N_{ad} - \tfrac{1}{2}N_{.b}^b\delta_{ad} \equiv \tfrac{1}{2}\epsilon_{(d}^{.cb}\Gamma_{a)bc}, \tag{C.6}$$

$$N_{.b}^b \equiv \epsilon_{abc}\Gamma^{abc}, \tag{C.7}$$

and

$$L_b \equiv \tfrac{1}{2}\Gamma_{.ab}^a. \tag{C.8}$$

For future reference it is convenient also to catalog the components of the object of anholonomy, Ω_{rst}, in terms of this representation, *viz.*,

$$\Omega_{00a} = -\Omega_{0a0} = \tfrac{1}{2}a_a, \tag{C.9}$$

$$\Omega_{0ab} = -\Omega_{0ba} = \epsilon_{abc}\Omega^c, \tag{C.10}$$

$$\Omega_{ab0} = -\Omega_{a0b} = \tfrac{1}{2}[-S_{ab} + \epsilon_{abc}(\Omega^c - \omega^c)], \tag{C.11}$$

$$\Omega_{abc} = -\Omega_{acb} = \tfrac{1}{2}(\epsilon_{.cb}^d N_{ad} + 2L_{[c}\delta_{b]a}). \tag{C.12}$$

3. Vector-Dyadic Notation

In the representation just developed, the set of 24 components of either Γ_{rst} or Ω_{rst} clearly falls into natural three-dimensional subarrays for which a vector and dyadic notation would be convenient. In such notation the equations involving these quantities would preserve the familiar formalism of 3-space rotation covariance which here corresponds to the arbitrariness remaining in the selection of the auxiliary spacelike tetrad vectors, even when $_0\lambda^\mu$ is physically given. Since the quantities in question are defined in terms of the tetrad vectors themselves and their derivatives, it is not obvious that this program must succeed at all; especially if we insist that the vector or dyadic character shall hold not just at a single event, but throughout space–time.

Accordingly, we now perform an analysis of the transformation properties of the arrays under a general, four-dimensional, proper orthogonal transformation of the tetrad fields which leaves $_0\lambda^\mu$ fixed. We determine the widest group of such transformations under which the arrays will have the 3-vector and dyadic character at every point. Not surprisingly, the set of acceptable transformations is quite restricted, in the sense that the parameters of the transformation at one event determine the transformation throughout space–time. For such transformations, however, we show that the arrays a_a and L_a are polar 3-vectors, say $\mathbf{a}$ and $\mathbf{L}$; while Ω_a and ω_a form axial vectors, $\boldsymbol{\Omega}$ and $\boldsymbol{\omega}$. The symmetric arrays S_{ab} and N_{ab} transform as dyadics, S and N, although the latter has a pseudocharacter under inversions of the spatial tetrad vectors.

Consider then an orthogonal transformation of the three auxiliary spacelike vector fields. We may write such a transformation most generally as

$$_a\bar\lambda^\mu = A_{.\nu}^\mu\,_a\lambda^\nu, \tag{C.13}$$

where $A_{.\nu}^\mu$ is a tensor field satisfying

$$A_{\mu\sigma}\,g^{\sigma\tau}A_{\nu\tau} = g_{\mu\nu}. \tag{C.14}$$

In the present case we require that the orthogonal tensor field be proper, and that it leave unchanged the $_0\lambda^\mu$ congruence; it follows that it has an unmoved 2-flat and can be written in the canonical form[8]

[8] F. B. Estabrook, California Institute of Technology, Pasadena, California, Jet Propulsion Laboratory, Research Summary No. 36-14, p. 119 (1962).

$$A_{\mu\nu} = \cos\theta\, g_{\mu\nu} + \sin\theta\, (-g)^{-\frac{1}{2}} \epsilon_{\mu\nu\sigma\tau}\, _0\lambda^\sigma \zeta^\tau$$
$$+ 2\sin^2(\tfrac{1}{2}\theta)\,(\zeta_\mu \zeta_\nu - {}_0\lambda_\mu\, _0\lambda_\nu). \qquad (C.15)$$

ζ^ν is a unit spacelike four-vector orthogonal to $_0\lambda^\mu$; together they define the unmoved 2-flat. In the local tetrad frame, we see a simple spatial rotation by angle θ about the ζ^ν direction. Equation (C.14) and the invariance of $_0\lambda^\mu$ and ζ^ν can be verified immediately by direct computation. Strangling Eq. (C.15) we get the familiar 3-space proper rotation matrix

$$O_{ab} = \cos\theta\, \delta_{ab} + \sin\theta\, \epsilon_{abc}\zeta^c + 2\sin^2(\tfrac{1}{2}\theta)\, \zeta_a \zeta_b. \quad (C.16)$$

$\hat{\zeta}$ is the unit 3-vector with strangled components $\zeta_a \equiv \zeta_\mu\, _a\lambda^\mu$; it points along the axis of the rotation. Noting that $O_{0a} = 0$, $O_{00} = -1$, we can also write

$$A_{\mu\nu} = O_{\sigma\tau}\, {}^s\lambda_\mu\, {}^\tau\lambda_\nu = O_{ab}\, {}^a\lambda_\mu\, {}^b\lambda_\nu - {}^0\lambda_\mu\, {}^0\lambda_\nu. \quad (C.17)$$

Any vector V^μ orthogonal to $_0\lambda^\mu$ may be expanded in either auxiliary tetrad system,

$$V^\mu = {}^a V\, _a\lambda^\mu = {}^b \bar{V}\, _b\bar{\lambda}^\mu, \qquad (C.18)$$

and substituting from Eq. (C.13) we can see that the components ^aV transform contragradiently to the unit vectors:

$$^c\bar{V} = {}^a V O_a^{\cdot c}. \qquad (C.19)$$

The arrays of components a_a, Ω_a and S_{ab} can be immediately shown, from their definitions Eqs. (C.1) and (C.2), to transform according to Eq. (C.19) (or its dyadic generalization, in the case of S_{ab}), and so this justifies our use of 3-vector and dyadic notation for them: a, $\boldsymbol{\Omega}$, and S.

We now consider the change of ω_c, defined in Eq. (C.3), under the transformation of Eq. (C.13). From the definition,

$$\Gamma_{0ba} = -\epsilon_{abc}\omega^c. \qquad (C.20)$$

If we similarly set

$$\bar{\Gamma}_{0ba} = {}_a\bar{\lambda}^\mu\, _b\bar{\lambda}_{\mu;\nu}\, _0\lambda^\nu = -\epsilon_{abc}\bar{\omega}^c, \qquad (C.21)$$

substitution from Eqs. (C.13) and (C.16) leads finally to the transformation law

$$\bar{\omega}^a = \omega^d O_d^{\cdot a} + \tfrac{1}{2}\epsilon^{abc}\dot{O}_{\cdot b}^d O_{dc}. \qquad (C.22)$$

Equivalent to this is

$$\bar{\omega}^a = \omega^d O_d^{\cdot a} - \dot{\theta}\zeta^a - \sin\theta\, \dot{\zeta}^a$$
$$- (1 - \cos\theta)\,(\hat{\zeta} \times \dot{\hat{\zeta}})^a, \qquad (C.23)$$

where the superimposed dot means the intrinsic derivative in the $_0\lambda^\mu$ direction, e.g., $\dot{\theta} = \theta_{,\mu}\, _0\lambda^\mu$. If (and only if) we restrict the orthogonal transformation tensor $A^\mu_{\cdot\nu}$ to one for which $\dot{\theta}$ and $\dot{\hat{\zeta}}$ everywhere vanish, which is to say $\dot{O}_{ab} = 0$, we

arrange that the quantities ω^a transform precisely like a 3-vector, and so justify our choice of notation for this set of three components. The restriction $\dot{O}_{ab} = 0$ amounts to correlating the rotation induced by O_{ab} of the three spacelike unit vectors of a fundamental tetrad at a given event, to the rotations of all other such tetrads induced at all other events along the world line of the $_0\lambda^\mu$ congruence through the given event, so that $\boldsymbol{\omega}$ is not intrinsically changed, but only locally projected on a different anholonomic coordinate mesh.

The remaining components, those of the symmetric 3-dyadic N_{ab} or N, and the 3-vector L_a, or L, will also transform precisely as the notation suggests only under special forms of O_{ab}. In fact, one finds

$$\bar{N}^{ab} = N^{cd}O_c^{\cdot a}O_d^{\cdot b} + \tfrac{1}{2}\epsilon^{hfg}O_{fd,g}O_h^{\cdot d}\delta^{ab}$$
$$+ \tfrac{1}{2}O_{fd,g}O^{g(a}\epsilon^{b)cd}O_{\cdot c}^f \qquad (C.24)$$

and

$$\bar{L}^a = L^d O_d^{\cdot a} - \tfrac{1}{2}O_{\cdot\cdot\cdot c}^{ca}. \qquad (C.25)$$

Upon substitution of the explicit form of O_{ab} from Eq. (C.16), it is found from equations analogous to Eq. (C.23) that the extraneous terms in Eqs. (C.24) and (C.25) can vanish in general if and only if $O_{ab,c} = 0$. Combining this with our previous result, we can state: $\boldsymbol{\omega}$, N, and L transform properly as three-dimensional vector and dyadic fields, for those orthogonal transformations having the array O_{ab} constant everywhere.

We have then the following situation: *given* $_0\lambda^\mu$, a further orthonormal set $_a\lambda^\mu$ may be chosen at every event. Three quite arbitrary auxiliary spacelike congruences are thus determined. From this auxiliary set, however, we usually allow only transformations to other sets derived from it by choosing an arbitrary unit spacelike 3-vector $\hat{\zeta}$, *whose components with respect to the spacelike unit vectors are the same at every event*, and rotating the spacelike set at every event by the *same* angle θ about the direction $\hat{\zeta}$. Any such transformation thus derives a new set of three auxiliary spacelike congruences from the first. We call such a new set of auxiliary orthogonal congruences *3-space rotated* with respect to the original set. Under such 3-space rotation, a, $\boldsymbol{\Omega}$, S, $\boldsymbol{\omega}$, N, and L transform in familiar three-dimensional orthogonal fashion, and form-invariant equations between these quantities can be written in the familiar language of the Gibbsian vector analysis.

In such equations, I denotes the unit dyadic, with components $I_{ab} = \delta_{ab}$. By (tr S) we mean the contraction or trace, $S_{\cdot a}^a$. The dot notation for inner products is used, and a double dot product

of two dyadics is equivalent to the trace of their inner product. The cross product is defined in the usual right-handed way. When a $\times$ operates on a dyadic, it operates on the nearest index when expressed in terms of components; e.g.,

$$(\Omega \times S)_{ab} \equiv \epsilon_{acd}\Omega^c S^d_{.b}. \tag{C.26}$$

The double cross product of two dyadics often provides a convenient brevity of notation. It is used only between symmetric dyadics so that no ambiguity of ordering can arise in its definition; viz.,

$$(Q \,\substack{\times \\ \times}\, S)_{ab} \equiv \epsilon_{acd}\epsilon_{bfg}Q^{fc} S^{dg}. \tag{C.27}$$

The result is again a symmetric dyadic having the expansion

$$Q \,\substack{\times \\ \times}\, S \equiv Q \cdot S + S \cdot Q - (\text{tr } S)Q - (\text{tr } Q)S$$
$$+ [(\text{tr } S)(\text{tr } Q) - Q : S]1. \tag{C.28}$$

We use the 3-vector symbol $\mathbf{D}$ for spatial intrinsic derivation: thus $\phi_{,a}$ becomes $\mathbf{D}\phi$, a gradient; $V^a_{.,a}$ becomes $\mathbf{D}\cdot\mathbf{V}$, a divergence; $\frac{1}{2}(V_{c,b} - V_{b,c})$ when multiplied by ϵ^{abc} becomes the curl, $\mathbf{D}\times\mathbf{V}$; etc. Another spatial differential operator, linearly related to $\mathbf{D}$, is introduced in Sec. C5; denoted ∇, this operator is convenient in many of our equations, and is the triad-strangled operation of covariant differentiation in spatial subspaces (when such exist). The operations of gradient, divergence and curl with the ∇ operator are defined in Sec. C5.

4. Physical Interpretation of the Dyadic Quantities

The identification of $_0\lambda^\mu$ with a physical motion imbues many of the components of the anholonomic affinity with immediate physical or kinematical significance. We first develop the interpretations by recalling some definitions met with in the usual tensorial description of the kinematics of a relativistic continuum. In a sense this procedure is logically inverted, but it has the advantage of quickly connecting quantities in the present notation with the familiar tensor quantities. A more basic approach will follow.

Let a fluid motion be described by a velocity 4-vector field λ^μ, with $\lambda_\mu\lambda^\mu = -1$. From the derivatives $\lambda^\mu_{.;\nu}$ one resolves canonical sets of first-order differential quantities:[9] the acceleration vector $a_\mu \equiv \lambda_{\mu;\nu}\lambda^\nu$; the (antisymmetric) angular velocity tensor $\Omega_{\mu\nu} \equiv \lambda_{[\mu;\nu]} + a_{[\mu}\lambda_{\nu]}$; and the (symmetric) rate-of-strain tensor $\sigma_{\mu\nu} \equiv \lambda_{(\mu;\nu)} + a_{(\mu}\lambda_{\nu)}$. From the angular-velocity tensor can be defined an equivalent

local angular-velocity vector, Ω^μ, by setting

$$\Omega^\mu = \frac{1}{2}(-g)^{-\frac{1}{2}}\epsilon^{\mu\nu\sigma\tau}\Omega_{\nu\sigma}\lambda_\tau. \tag{C.29}$$

This can be solved for $\Omega_{\mu\nu}$,

$$\Omega_{\mu\nu} = (-g)^{-\frac{1}{2}}\epsilon_{\mu\nu\sigma\tau}\Omega^\sigma\lambda^\tau, \tag{C.30}$$

which demonstrates the equivalence. All these canonical tensor quantities are projected into the local proper frame; that is,

$$a_\mu\lambda^\mu = \Omega_{\mu\nu}\lambda^\nu = \Omega_\mu\lambda^\mu = \sigma_{\mu\nu}\lambda^\nu \equiv 0. \tag{C.31}$$

Now, we identify $\lambda^\mu \equiv _0\lambda^\mu$ and take the proper components of the canonical tensors with respect to the local tetrad, using Eq. (B.3) to introduce affinity components. Clearly, by Eq. (C.31), transvection with $_0\lambda^\mu$ itself will always give a zero result. For the acceleration vector, then, we have using Eq. (B.3) and Eq. (C.1),

$$a_\mu \,_a\lambda^\mu = _0\lambda_{\mu;\nu} \,_0\lambda^\nu \,_a\lambda^\mu = \Gamma_{00a} = a_a, \tag{C.32}$$

so that our 3-vector $\mathbf{a}$ is precisely the local proper acceleration of the $_0\lambda^\mu$ congruence. Likewise from (C.2)

$$\Omega_\mu \,_a\lambda^\mu = \frac{1}{2}\epsilon_a^{.bc}\Gamma_{cb0} = \Omega_a, \tag{C.33}$$

and from (C.10)

$$\Omega_{\mu\nu} \,_a\lambda^\mu \,_b\lambda^\nu = -\Omega_{0ab} = \epsilon_{bac}\Omega^c, \tag{C.34}$$

which identifies the 3-vector $\mathbf{\Omega}$ as the local angular velocity of the medium. Analyses by Synge,[5] Pirani,[10] and others have made it clear that, like $\mathbf{a}$, this $\mathbf{\Omega}$ is an *absolute* entity: the angular velocity of the material medium with respect to Weyl's "compass of inertia."

The rate-of-strain tensor $\sigma_{\mu\nu}$ gives six proper components, all spacelike, and using (C.4),

$$\sigma_{\mu\nu} \,_a\lambda^\mu \,_b\lambda^\nu = -\Gamma_{(ab)0} = S_{ab}, \tag{C.35}$$

so that S is the local, three-dimensional, rate-of-strain dyadic. With this, we have found transcriptions for all the canonical tensors and will turn to the interpretation of ω, Eq. (C.3).

Projecting the local time derivatives (i.e., the intrinsic derivatives in the $_0\lambda^\mu$ direction, for which we use the superimposed dot notation throughout) of the spacelike tetrad vectors themselves, one has for the timelike components

$$_a\dot{\lambda}_\mu \,_0\lambda^\mu = -_a\lambda_\mu \,_0\dot{\lambda}^\mu = -_a\lambda_\mu a^\mu = -a_a \tag{C.36}$$

from the orthogonality relations alone. And in the spatial directions, from Eq. (C.3) we have

$$_a\dot{\lambda}_\mu \,_b\lambda^\mu = \Gamma_{0ab} = \epsilon_{abc}\omega^c. \tag{C.37}$$

[9] See, for example, J. Ehlers and W. Kundt, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962).

[10] F. A. E. Pirani, Helv. Phys. Acta Suppl. IV, 198 (1956); Acta Phys. Polon. 15, 389 (1956).

The vector ω thus describes the orthogonal propagation of the spacelike auxiliary triads *along* the $_0\lambda^\mu$ congruence; kinematically, ω is the local angular velocity of the auxiliary orthonormal triad with respect to the compass of inertia. Conversely, $(-\omega)$ is the angular velocity of a "stable-platform" relative to the triad.

From Eqs. (C.36) and (C.37), we may express the condition for Fermi–Walker transport of the spacelike triad along a given line of the $_0\lambda^\mu$ congruence simply by setting $\omega = 0$ on that line. Putting $\omega = 0$ everywhere would prescribe the introduction of a tetrad field such that the spacelike triad attached to each material point represents a local, inertially nonrotating reference frame. It is an advantage of the dyadic notation that this condition is a 3-vector equation, form invariant under 3-space rotation. We show in Sec. C5 that it is always possible initially to introduce the tetrad field according to any such prescription for ω.

Elucidation of the kinematical significance of the quantities Ω, ω, and S is alternatively obtained by considering an equation for the proper orthogonal separation, say ρ^μ, of two closely adjacent members of the $_0\lambda^\mu$ congruence. In the local, proper frame ρ^μ will appear as the displacement vector between two proximate material particles. Its rate of change with local time is given by[10]

$$\dot{\rho}^\mu = (_0\lambda^\mu_{.;\nu} + {}_0\lambda^\mu a_\nu)\rho^\nu. \qquad (C.38)$$

Projecting ρ^μ onto the tetrad defines locally Cartesian spatial coordinates r_a, or components of a local displacement 3-vector r, where

$$r_a = \rho^\mu \, _a\lambda_\mu. \qquad (C.39)$$

The local time derivative of these is found with the help of Eq. (C.38) and (C.11) to be given by

$$\dot{r}_a = 2\Omega_{a0b}r^b = [S_{ab} + \epsilon_{acb}(\Omega^c - \omega^c)]r^b. \qquad (C.40)$$

Equation (C.40) is valid to first order in the displacements r_a. These displacement components are a *Cartesian* vector, in the (flat) tangent space at the origin $r_a = 0$: the S_{ab}, Ω_c and ω_c are Cartesian components evaluated at $r_a = 0$. Remembering these limitations, we may still use dyadic notation:

$$\dot{r} + \omega \times r = S \cdot r + \Omega \times r \qquad (C.41)$$

from which immediately

$$\tfrac{1}{2}D \times \dot{r} = \Omega - \omega \qquad (C.42)$$

and

$$\tfrac{1}{2}(D\dot{r} + \dot{r}D) = S. \qquad (C.43)$$

These equations manifest the local kinematical

significance of Ω, ω, and S and basically provide interpretations for the canonical tensors as well. Since Fermi–Walker transport of the basis vectors is accomplished by setting $\omega = 0$, the interpretation of Ω as the local angular velocity of the material relative to the compass of inertia is clear. In the general dyadic equations to be written later it is evident that a particularly convenient choice for ω is rather to propagate the tetrads so that $\Omega - \omega = 0$. This alternative is called corotating transport, or "body-fixed axes," since as Eq. (C.42) shows, the local reference frame is thereby rotated with respect to the compass of inertia so as to follow the physical rotation of the neighboring members of the $_0\lambda^\mu$ congruence. Again, the condition for body-fixed axes is form invariant under 3-space rotation.

Interpretation of the quantities L and N, which express characteristics of the auxiliary congruences, is somewhat less evident. In fact their significance, being more geometrical than physical, emerges most clearly in the special circumstance when the given timelike congruence comprises the orthogonal trajectories of a family of 3-surfaces immersed in space–time. This is discussed in some detail in the next section. First, however, the relationship of L and N to the properties of the spacelike congruences is obtained.

The first curvature vector of a curve of the congruence generated by $_a\lambda^\mu$ is defined by $_a\lambda_{\mu;\nu} \, _a\lambda^\nu$ (a not summed), and its components in the local tetrad basis are

$$(_a\lambda_{\mu;\nu} \, _a\lambda^\nu)_r\lambda^\mu = \Gamma_{aar} \quad (a \text{ not summed}). \qquad (C.44)$$

Referring to Eq. (C.4) we see that the timelike component is given by the diagonal element of S,

$$\Gamma_{a0a} = S_{aa} \quad (a \text{ not summed}), \qquad (C.45)$$

which determines the rate of convergence in the $_a\lambda^\mu$ direction of the timelike congruence curves $_0\lambda^\mu$. For $r = b \neq a$ we have

$$\Gamma_{aab} = \epsilon_{dba}N^d_{.a} + L_b \quad (a \text{ not summed}), \qquad (C.46)$$

which involves L and only the off-diagonal elements of N. If we were to perform a 3-space rotation to diagonalize N at a given event, the spacelike components at that point of the first curvature vectors of the new set of auxiliary congruences thus obtained would be expressed by L alone. In general, of course, such a transformation does not diagonalize N elsewhere and it reappears in Eq. (C.46) at other events.

The geometrical meaning of the diagonal elements of N is more easily expressed in terms of the modified dyadic, $N - \tfrac{1}{2}(\text{tr } N)I$. The ath diagonal element

of this dyadic gives the rate of "twist" around the $_a\lambda^u$ direction applied to the triads in propagating them in the $_a\lambda^u$ direction itself. In a word, then, one might refer to these as the "torsions" of the spacelike congruence net. ·

5. Conditions on the Auxiliary Congruences. The ∇ Operator.

An orthonormal tetrad field aligned along a "given" congruence, generating three orthogonal but otherwise arbitrary auxiliary congruences, constitutes a complex geometrical structure. We wish, in this section, to remark about specializations of this auxiliary structure, some of which may be imposed in general, others only when the preferred congruence has special properties. While this discussion is not at all complete, it should at least show that the necessary equations for investigating such points are at hand in the dyadic notation. We first briefly discuss some specializations which are always available, then summarize several special cases which may occur, and finally, introduce the useful vector differential operator, ∇, suggested by one such geometrical subcase.

The pertinent equations are, in fact, Eqs. (C.22), (C.24), and (C.25); for when an aligned but otherwise arbitrary tetrad field is initially introduced upon a given timelike congruence, the general orthogonal transformation O_{ab} in these equations can often be selected to give a second, in some way special or canonical, tetrad field having the same alignment. The dyadic notation then allows the further generation (with constant O_{ab}) of a family of tetrad fields 3-space rotated from this second one, as was expounded previously.

The first example of this, encountered in the previous section, is the prescription of Fermi–Walker propagated axes *everywhere*, the condition ω = 0. *That this may be done in general* is clear from inspection of Eq. (C.22), when we regard the $\bar{\omega}_a$ as arbitrarily given initial fields, set $\omega_a = 0$, and solve for the three independent components of $\dot{O}_{ab}$ everywhere. A choice of O_{ab} on one spacelike 3-surface then suffices to determine a solution. We thus demonstrate by direct construction a transformation leading to a new tetrad field with the desired property. Subsequent 3-space rotations (with O_{ab} constant everywhere) clearly will preserve this property.

A second example is the imposition of body-fixed axes *everywhere*, Ω − ω = 0, the justification of which follows in exactly similar fashion.

Another important case is the imposition of the set of conditions N = 0, L = 0, ω = 0 *on a single*

world line of the congruence. That this may be done follows again by construction of the required transformation. Given first $\bar{N}_{ab}$, $\bar{L}_a$, and $\bar{\omega}_a$, the 12 equations in (C.22), (C.24), and (C.25) can now be solved for the twelve partial derivatives of the three scalar fields in O_{ab} *on the line*. With a choice of O_{ab} at one point on the line it may by quadrature be suitably determined along and near the line to achieve any desired values of N, L, and ω.

An essential point is that while this last can always be done along a line or at a point, it cannot be done on manifolds of higher dimensions unless further integrability conditions are satisfied. Such conditions, however, introduce relations among the other 12 components of Γ_{rst} (viz., a, Ω, S, referring to the timelike congruence) and so require the timelike congruence to have special properties. A typical situation occurs when one attempts simultaneously to impose Fermi–Walker propagation *everywhere* while also taking N and L to vanish on a line: the result is a constraint on the timelike congruence along that line.

We now proceed to summarize some similar cases in which partial degrees of integrability, or holonomy, are imposed on the congruence structure throughout space–time. The conditions take the form of the global vanishing of certain components of the object of anholonomity. The various conditions are not derived *ab initio* in the following; they are to be found for general spaces in Ref. 3. We are primarily interested here in specializing them to the case of a (3 + 1)-dimensional metric space with orthonormal tetrad vectors and then transscribing them into dyadic notation.

We consider first the geometrical situation in which one given pair of the four congruences is 2-forming. That is to say, the two congruences mesh together so as to form a (two-parameter) family of 2-surfaces embedded in the four-dimensional manifold. The condition for the s congruence and the t congruence to be 2-forming is

$$\Omega^{rst} = 0 \qquad (r \neq s, r \neq t). \qquad (C.47)$$

(We emphasize again that these conditions are written for the case of orthonormal tetrads only.) For a given pair (s, t) the inequalities allow only two values for the index r, and so two independent conditions result. There are six possible ways of pairing the congruences, and if we were to ask that *all* congruence pairs be 2-forming, we would require exactly one-half of the 24 independent components of the object of anholonomity to vanish everywhere. In dyadic terms from Eqs. (C.9)–(C.12) the 12 condi-

tions given by Eq. (C.47) for this completely 2-forming case become:

$$\Omega = 0, \quad \omega = 0,$$

$$S_{ab} = 0 \ (a \neq b), \quad N_{aa} = 0 \ (a \text{ not summed}), \quad (C.48)$$

so that, in addition to the vanishing of the two angular velocities, S must be diagonal and N off-diagonal. The constraints on Ω and S are of particular significance, since they restrict the physical congruences $_0\lambda^\mu$ for which this situation may exist.

We may consider, alternatively, the possibility that a given set of three congruences is 3-forming. This is here equivalent to the condition that the fourth congruence be 3-normal; that is, the unit vector generating this fourth congruence is everywhere proportional to the gradient of a scalar function, ψ, and so orthogonal to the family of 3-dimensional hypersurfaces, $\psi = $ constant, which essentially define a holonomic coordinate in the space. The condition for the r congruence to be 3-normal is

$$\Omega^{rst} = 0 \quad (s \neq r, \ t \neq r), \quad (C.49)$$

which is very similar to (C.47) but differs in the effect of the inequalities. Here, when r is given, s and t are allowed three values each, but the antisymmetry on s and t reduces the number of independent, nontrivial conditions to three. If, in this case, we ask that all four congruences be 3-normal, we again require the vanishing of 12 components of the object of anholonomity; clearly, in fact, the same 12 as for the case of complete 2-forming. The dyadic conditions for complete 3-normality, then, are already given by Eq. (C.48).

A large class of conditions, less restrictive than the complete cases covered by (C.48), could be considered. In accord with a dyadic approach however, which confers a special position exclusively on the timelike congruence, only those intermediate situations treating the three spacelike congruences impartially are of interest. There are four such subcases; the constraints for them follow immediately from Eq. (C.47) and (C.49) and they need only to be listed:

(1) All spacelike congruences are 2-forming with $_0\lambda^\mu$.

$$\Omega - \omega = 0, \quad S_{ab} = 0 \ (a \neq b). \quad (C.50)$$

(2) All pairs of spacelike congruences are 2-forming.

$$\Omega = 0, \quad N_{aa} = 0 \ (a \text{ not summed}). \quad (C.51)$$

(3) All spacelike congruences are 3-normal.

$$\Omega - \omega = 0, \quad S_{ab} = 0 \ (a \neq b),$$

$$N_{aa} = 0 \ (a \text{ not summed}). \quad (C.52)$$

(4) The timelike congruence is 3-normal.

$$\Omega = 0. \quad (C.53)$$

It is worth noting that the 12 components of Ω^{rst} which are not concerned in *any* of the constraint equations presented in Eqs. (C.47)–(C.53) are a, L, the diagonal elements of S, and the off-diagonal elements of N. As we have brought out in previous discussions, these are precisely the 12 components of the first curvature vectors of the four congruences. The entire vanishing of the object of anholonomity is secured, then, by the requirements that all four congruences be 3-normal *and geodesic*. As we remarked in Sec. B, this would imply the vanishing of the Riemann tensor and the introduction of holonomic Minkowski coordinates.

In Case 4, Eq. (C.53), the separation of space and time is accomplished globally—space–time is a sandwich of spacelike 3-manifolds, each normal to the (everywhere nonrotating) timelike congruence. The Riemannian structure of space–time allows invariant measurements in any one of these 3-manifolds; it is, consequently, a Riemannian 3-manifold with an induced intrinsic metric and a second fundamental form (just S) describing its immersion in the 4-space—the mathematics of this emerge naturally in Sec. D2. N and L now express exactly the nine components of the anholonomic affinity generated by an arbitrary triad field in a Riemannian 3-space. Even in the general case, this interpretation of N and L has much heuristic value, and completes our geometric discussion of these arrays.

If we pursue this last interpretation by introducing a vector operator ∇ to denote triad-strangled three-dimensional covariant differentiation as in Eq. (B.11), e.g.,

$$\nabla_c M_{ab} = D_c M_{ab} + \Gamma_{c \cdot a}^{\cdot d} M_{db} + \Gamma_{c \cdot b}^{\cdot d} M_{ad}, \quad (C.54)$$

we greatly simplify the notation in the dyadic differential equations to be presented in Sec. D. We denote ∇ the *three-dimensional covariant differentiation* operator, although of course this interpretation is only immediately accessible geometrically in Case (4) (as differentiation in immersed subspaces). Without inquiring further here into the geometries of quotient subspaces, we merely regard the ∇ operator in the general case as a useful notation. From the defining Eq. (C.54) we may calculate and tabulate the following useful formulas, where V is an arbitrary vector field, and M an arbitrary *symmetric* dyadic field:

$$\nabla V = DV - [N - \tfrac{1}{2}(\text{tr } N)I - L \times I] \times V, \quad (C.55)$$

$$\nabla \cdot \mathbf{V} = \mathbf{D} \cdot \mathbf{V} - 2\mathbf{L} \cdot \mathbf{V}, \qquad (\text{C.56})$$

$$\nabla \times \mathbf{V} = \mathbf{D} \times \mathbf{V} - \mathbf{N} \cdot \mathbf{V} - \mathbf{L} \times \mathbf{V}, \qquad (\text{C.57})$$

$$\nabla \times \mathbf{M} = \mathbf{D} \times \mathbf{M} - \mathbf{M} \cdot \mathbf{N} - 2\mathbf{N} \cdot \mathbf{M} - \mathbf{L} \times \mathbf{M}$$
$$+ \mathbf{L} \cdot \mathbf{M} \times \mathbf{I} + \tfrac{1}{2}(\text{tr } \mathbf{N})\mathbf{M} + (\text{tr } \mathbf{M})\mathbf{N}$$
$$+ (\mathbf{N} : \mathbf{M})\mathbf{I} - \tfrac{1}{2}(\text{tr } \mathbf{N})(\text{tr } \mathbf{M})\mathbf{I}, \qquad (\text{C.58})$$

$$\nabla \times \mathbf{M} - \mathbf{M} \times \nabla = \mathbf{D} \times \mathbf{M} - \mathbf{M} \times \mathbf{D} - 3\mathbf{M} \cdot \mathbf{N}$$
$$- 3\mathbf{N} \cdot \mathbf{M} - \mathbf{L} \times \mathbf{M} + \mathbf{M} \times \mathbf{L} + (\text{tr } \mathbf{N})\mathbf{M}$$
$$+ 2(\text{tr } \mathbf{M})\mathbf{N} + 2(\mathbf{N} : \mathbf{M})\mathbf{I} - (\text{tr } \mathbf{N})(\text{tr } \mathbf{M})\mathbf{I}, \qquad (\text{C.59})$$

$$\nabla \cdot \mathbf{M} = \mathbf{D} \cdot \mathbf{M} - 3\mathbf{L} \cdot \mathbf{M} - \mathbf{N} \dot{\times} \mathbf{M} + (\text{tr } \mathbf{M})\mathbf{L}. \qquad (\text{C.60})$$

D. THE DYADIC PARTIAL DIFFERENTIAL EQUATIONS AND INTERPRETATION

1. The Dyadic Components of the Riemann Tensor

In this section we first introduce and discuss two alternate splittings of strangled components of the Riemann or curvature tensor into dyadic arrays.

Accordingly as they contain two, one, or no zeros, the strangled components of the symmetrized Riemann tensor in Eq. (B.8) may be gathered into four arrays with the property of covariance under 3-space rotation:

$$P_{ab} = \tfrac{1}{2}\epsilon_{acf}\epsilon_{bdg}S^{cdfg}, \qquad (\text{D.1})$$

$$Q_{ab} = 3S_{a00b}, \qquad (\text{D.2})$$

$$B_{ab} + \epsilon_{acb}t^c = \epsilon_{..b}^{dc}S_{ad0c}. \qquad (\text{D.3})$$

We thus describe the 20 components of the curvature field of general relativity by three symmetric dyadics P, Q, B (the last is traceless) and a vector t. In Sec. D2, when we write all the dyadic partial differential equations, we interpret P in terms of the intrinsic curvature of the spacelike 3-manifolds of a normal congruence. In Sec. D3 we derive several results allowing physical interpretation of the differential equations; in particular we there interpret Q as giving the tidal acceleration between neighboring test particles. An interpretation of B and t also appears in Sec. D3—they determine the differential (tidal) precession between neighboring (inertially oriented) test particles. It should be noted that, like N, the dyadic B has a pseudocharacter under 3-space inversion.

The alternate splitting up of dyadic components of the Riemann tensor is suggested by considering the canonical resolution of this tensor, in four dimensions, into three irreducible tensorial parts with the same algebraic symmetries.[11] We write this in

strangled form as

$$R_{rstu} = C_{rstu} + (\eta_{r[u}H_{t]s} - \eta_{s[u}H_{t]r})$$
$$+ \tfrac{1}{12}R(\eta_{r[u}\eta_{t]s} - \eta_{s[u}\eta_{t]r}), \qquad (\text{D.4})$$

where

$$H_{rs} = R_{rs} - \tfrac{1}{4}R\eta_{rs}, \qquad (\text{D.5})$$

the strangled Ricci tensor is

$$R_{rs} = R_{.rst}^{t}, \qquad (\text{D.6})$$

and its scalar contraction is the curvature scalar

$$R = R_{.r}^{r}. \qquad (\text{D.7})$$

C_{rstu} is the conformal curvature tensor (strangled) of Weyl; it is antidouble-dual; all its contractions are zero; it in general exists for Riemannian geometries in four or more dimensions, where its vanishing is the necessary and sufficient condition for the metric to be conformally flat. In four dimensions C_{rstu} has ten independent components; upon resolution into proper dyadic arrays, according as the Lorentz indices contain one or two zeros, we obtain two symmetric dyadics (traceless, so having five components each) A and again the B of Eq. (D.3):

$$A_{ab} = C_{0a0b} = -\tfrac{1}{4}\epsilon_{acd}\epsilon_{bfg}C^{cdfg}, \qquad (\text{D.8})$$

$$B_{ab} = \tfrac{1}{2}\epsilon_{..b}^{cd}C_{0acd}. \qquad (\text{D.9})$$

The dyadic A, expressed in terms of the previous set, is one-half the traceless sum of P and Q:

$$\mathbf{A} = \tfrac{1}{2}[\mathbf{P} + \mathbf{Q} - \tfrac{1}{3}(\text{tr } \mathbf{P} + \text{tr } \mathbf{Q})\mathbf{I}]. \qquad (\text{D.10})$$

To complete this alternate splitting, the ten components of the Ricci tensor may also be resolved into dyadic arrays. For physical reasons we prefer to introduce these from the strangled form of the Einstein tensor $R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu}$, which, in Einstein theory, is identified with the negative of the non-gravitational stress-momentum-energy tensor, $T_{\mu\nu}$. [We have already adopted a unit of length such that the velocity of light $c = 1$; now we adopt a unit of mass such that the Newtonian constant of gravitation γ is $(4\pi)^{-1}$.] In dyadic form we have then a symmetric stress dyadic T, a momentum-density vector t, and an energy-density scalar ρ:

$$T_{ab} = \tfrac{1}{2}R_{ab} - \tfrac{1}{4}R\eta_{ab}, \qquad t_a = \tfrac{1}{2}R_{a0}, \qquad (\text{D.11})$$
$$\rho = -\tfrac{1}{2}R_{00} - \tfrac{1}{4}R.$$

The vector t was introduced previously in Eq. (D.3). The local proper system of a fluid is *defined* by the condition that $_0\lambda^\mu$ be an eigenvector of $T_{\mu\nu}$:[12]

[11] J. Géhéniau and R. Debever, Bull. Acad. Roy. Belg. Cl. Sci. 42, 114, 252, 313, 608 (1956).

[12] J. L. Synge, Proc. London Math. Soc. 43, 376 (1937).

$$T_{\mu,0}\lambda^\nu = -\rho_0\lambda_\mu, \qquad (D.12)$$

or simply

$$t = 0. \qquad (D.13)$$

In this proper system, ρ is the proper energy or rest-mass density. The condition (D.13) is invariant under 3-space rotation. It is of especial importance in formulating many relativistic problems where the preferred congruence is of both kinematical and dynamical significance.

The dyadic T is, up to its trace, one-half the difference of P and Q,

$$\mathsf{T} = \tfrac{1}{2}[-\mathsf{P} + \mathsf{Q} - (\mathrm{tr}\,\mathsf{Q})\mathsf{I}], \qquad (D.14)$$

and ρ is minus one-half the trace of P. We note finally that the curvature scalar R of Eq. (D.7) is given in terms of each set by

$$\tfrac{1}{2}R = -\mathrm{tr}\,\mathsf{T} - \rho = \mathrm{tr}\,\mathsf{P} + \mathrm{tr}\,\mathsf{Q}. \qquad (D.15)$$

We have then two entirely equivalent sets of curvature dyadics—it is difficult to say which is to be preferred. In Einstein's theory the ten components of the Einstein tensor, T, t, and ρ, express the true (or non-self-excited) *sources* of the total gravitational curvature, and the ten components of conformal curvature, A and B, express the expected ten components of a spin-2 gravitational *field*. From this point of view the second splitting is the more fundamental. Nevertheless the essential nonlinearity of Einsteinian theory appears both in the Bianchi Identities of Sec. D2, in 16 equations of which all these source and field terms are inextricably mixed, and again in the operational physical equations of test particle motion which are given in Sec. D3. In both of these, the more natural splitting of the Riemann tensor appears to be that first given, into the dyadics P, Q, B, and t, Eqs. (D.1)–(D.3).

The various possible radiative characters of Einsteinian gravitational fields are expressed, in close analogy with those of Maxwell fields, in the algebraically special forms of $C^\mu_{\cdot\nu\sigma\tau}$. The algebraic hierarchy for this due to Petrov, Pirani, and Sachs[13] leads, as might be expected, to simple canonical forms for our A and B.

Summarizing this briefly, for a Type II field, the conform tensor has a singly degenerate principal null direction, which, strangled in any local proper frame, defines a unit 3-vector of propagation, say

$\hat{\mathbf{v}}$; take this to be normal to a plane defined by otherwise arbitrary but orthogonal unit vectors $\hat{\mathbf{w}}$ and $\hat{\mathbf{u}}$; and then it may be shown that the field dyadics must be of the form

$$\mathsf{A} = (\tilde{a} - a)\hat{\mathbf{u}}\hat{\mathbf{u}} + (\tilde{a} + a)\hat{\mathbf{w}}\hat{\mathbf{w}} - 2\tilde{a}\hat{\mathbf{v}}\hat{\mathbf{v}}$$
$$+ c(\hat{\mathbf{u}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{u}}) + b(\hat{\mathbf{v}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{v}}), \qquad (D.16)$$

$$\mathsf{B} = (\tilde{c} + c)\hat{\mathbf{u}}\hat{\mathbf{u}} + (\tilde{c} - c)\hat{\mathbf{w}}\hat{\mathbf{w}} - 2\tilde{c}\hat{\mathbf{v}}\hat{\mathbf{v}}$$
$$+ a(\hat{\mathbf{u}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{u}}) + b(\hat{\mathbf{u}}\hat{\mathbf{v}} + \hat{\mathbf{v}}\hat{\mathbf{u}}). \qquad (D.17)$$

Here a, $\tilde{a}$, c, $\tilde{c}$, and b are arbitrary scalars under 3-rotations. $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$, $\hat{\mathbf{w}}$ are taken to form a right-handed orthonormal triad.

For a Type III algebraically special field the conform tensor has a doubly degenerate principal null direction—again denoting this by a unit $\hat{\mathbf{v}}$ we find that

$$\mathsf{A} = a(\hat{\mathbf{w}}\hat{\mathbf{w}} - \hat{\mathbf{u}}\hat{\mathbf{u}}) + c(\hat{\mathbf{u}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{u}})$$
$$+ b(\hat{\mathbf{v}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{v}}), \qquad (D.18)$$

$$\mathsf{B} = c(\hat{\mathbf{u}}\hat{\mathbf{u}} - \hat{\mathbf{w}}\hat{\mathbf{w}}) + a(\hat{\mathbf{u}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{u}})$$
$$+ b(\hat{\mathbf{u}}\hat{\mathbf{v}} + \hat{\mathbf{v}}\hat{\mathbf{u}}), \qquad (D.19)$$

which results from Eqs. (D.16), (D.17) on setting $\tilde{a} = \tilde{c} = 0$.

For a type-N algebraically special field the conform tensor has but one principal null direction, triply degenerate, and the canonical forms simplify further ($b = a = 0$) to

$$\mathsf{A} = c(\hat{\mathbf{u}}\hat{\mathbf{w}} + \hat{\mathbf{w}}\hat{\mathbf{u}}), \qquad (D.20)$$

$$\mathsf{B} = c(\hat{\mathbf{u}}\hat{\mathbf{u}} - \hat{\mathbf{w}}\hat{\mathbf{w}}). \qquad (D.21)$$

The quadrupole character of this extreme far zone radiative gravitational field is nicely shown by these last forms, in conjunction with the test particle equations to be given in Sec. D3. Roy and Rhadakrishna[14] have obtained equivalent forms in a recent paper, together with elegant results for gravitational and electromagnetic–gravitational shock fronts. They characterize the type N field, Eqs. (D.20)–(D.21), by saying that the 3-space quadrics associated with A and B are equal hyperbolic cylinders, coaxial (the $\hat{\mathbf{v}}$ direction!), with their other principal directions inclined at 45°. The scalar c characterizes the gravitational field strength seen by an observer whose world line is $_0\lambda^\mu$; by itself, a type N conform tensor has no nontrivial invariants. All of which is nicely analogous to the case of a null electromagnetic field.

[13] A. Z. Petrov, Sci. Trans. Kazan State University **114**, 55 (1954) [Translation by M. Karweit: Astron. Information, Trans. No. 29, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California (1963)]; F. A. E. Pirani, Phys. Rev. **105**, 1089 (1957); R. K. Sachs, Z. Phys. **157**, 462 (1960).

[14] S. R. Roy and L. Radhakrishna, Proc. Roy. Soc. (London) **A275**, 245 (1963).

2. The Dyadic Partial Differential Equations

We now write the four sets of differential relations which must hold between our dyadic fields in full generality, the application and analysis of which are the essence of this dyadic formalism for general relativistic physics. These are, respectively, (a) the Differential Identities—16 equations (one scalar three vector, one dyadic) arising from Eq. (B.7) metric and curvature independent; (b) the Curvature Equations—20 equations (one vector, three dyadic, the first traceless) introducing the Riemann tensor components, from Eq. (B.8); (c) the Bianchi Identities—20 equations (three vector, two dyadic, the first traceless) relating the derivatives of the Riemann components, from the integrability conditions Eq. (B.13); and (d) the Commutation Formulas for anholonomic space and time differentiation, special cases of Eq. (B.12).

(a) Differential Identities

$$\nabla \cdot \mathbf{\Omega} = \mathbf{a} \cdot \mathbf{\Omega}, \qquad (D.22)$$

$$\tfrac{1}{2}\nabla \times \mathbf{a} - (\dot{\mathbf{\Omega}} + \boldsymbol{\omega} \times \mathbf{\Omega}) = -\mathbf{S} \cdot \mathbf{\Omega} + (\operatorname{tr} \mathbf{S})\mathbf{\Omega}, \qquad (D.23)$$

$$\nabla \cdot \mathbf{N} + \nabla \times \mathbf{L} = -2\mathbf{L} \cdot \mathbf{N} + (\operatorname{tr} \mathbf{N})\mathbf{L}$$
$$- 2\mathbf{S} \cdot \mathbf{\Omega} + 2\boldsymbol{\omega} \times \mathbf{\Omega} \qquad (D.24)$$

$$2\dot{\mathbf{L}} = (\nabla + \mathbf{a}) \cdot [\mathbf{S}^{\star T} - (\operatorname{tr} \mathbf{S})\mathbf{I}] - \mathbf{S}^{\star} \dot{\times} \mathbf{N}^{\star}, \qquad (D.25)$$

$$\dot{\mathbf{N}} - \tfrac{1}{2}(\operatorname{tr} \dot{\mathbf{N}})\mathbf{I} = (\nabla + \mathbf{a}) \cdot (\mathbf{\Omega} - \boldsymbol{\omega})\mathbf{I}$$
$$+ \tfrac{1}{2}\mathbf{S}^{\star T} \times (\nabla + \mathbf{a}) - \tfrac{1}{2}(\nabla + \mathbf{a}) \times \mathbf{S}^{\star}$$
$$- \tfrac{1}{2}\mathbf{S}^{\star} \cdot \mathbf{N}^{\star} - \tfrac{1}{2}\mathbf{N}^{\star T} \cdot \mathbf{S}^{\star T}. \qquad (D.26)$$

To shorten Eqs. (D.25)–(D.26) we have used the notation $\mathbf{S}^{\star} \equiv \mathbf{S} - (\mathbf{\Omega} - \boldsymbol{\omega}) \times \mathbf{I}$ and $\mathbf{N}^{\star} \equiv \mathbf{N} - \tfrac{1}{2}(\operatorname{tr} \mathbf{N})\mathbf{I} - \mathbf{L} \times \mathbf{I}$. The superscript T denotes a transposed dyadic. The trace of Eq. (D.26) may be written in addition:

$$\operatorname{tr} \dot{\mathbf{N}} + 2\nabla \cdot (\mathbf{\Omega} - \boldsymbol{\omega}) = 2\mathbf{N} : \mathbf{S} - (\operatorname{tr} \mathbf{N})(\operatorname{tr} \mathbf{S})$$
$$- 2\mathbf{a} \cdot (\mathbf{\Omega} - \boldsymbol{\omega}) - 4\mathbf{L} \cdot (\mathbf{\Omega} - \boldsymbol{\omega}). \qquad (D.27)$$

The first two of these equations are remarkably simple, curvature-independent, general identities satisfied by the proper kinematic observables of any timelike congruence. The third, Eq. (D.24), expresses integrability conditions on the spatial parts, L and N, of the anholonomic affinity. The remaining three relate the time derivatives of L and N to the properties of the preferred congruence.

(b) Curvature Equations

$$\nabla \cdot \mathbf{S} - \nabla(\operatorname{tr} \mathbf{S}) + \nabla \times \mathbf{\Omega} = 2\mathbf{\Omega} \times \mathbf{a} - 2\mathbf{t}, \qquad (D.28)$$

$$\tfrac{1}{2}(\nabla \times \mathbf{S} - \mathbf{S} \times \nabla) - \tfrac{1}{2}(\nabla\mathbf{\Omega} + \mathbf{\Omega}\nabla)$$
$$= \mathbf{a}\mathbf{\Omega} + \mathbf{\Omega}\mathbf{a} - \mathbf{a} \cdot \mathbf{\Omega}\mathbf{I} - \mathbf{B}, \qquad (D.29)$$

$$\tfrac{1}{2}(\nabla \times \mathbf{N} - \mathbf{N} \times \nabla) - \tfrac{1}{2}(\nabla\mathbf{L} + \mathbf{L}\nabla)$$
$$= -\mathbf{N} \cdot \mathbf{N} + \tfrac{1}{2}(\operatorname{tr} \mathbf{N})\mathbf{N} - \mathbf{L}\mathbf{L} - [\tfrac{1}{8}(\operatorname{tr} \mathbf{N})^2$$
$$- \tfrac{1}{4}\mathbf{N} : \mathbf{N} - \tfrac{1}{2}\mathbf{L} \cdot \mathbf{L}]\mathbf{I} + \mathbf{E} - \tfrac{1}{2}(\operatorname{tr} \mathbf{E})\mathbf{I}, \qquad (D.30)$$

$$\dot{\mathbf{S}} + \boldsymbol{\omega} \times \mathbf{S} - \mathbf{S} \times \boldsymbol{\omega} - \tfrac{1}{2}(\nabla\mathbf{a} + \mathbf{a}\nabla)$$
$$= -\mathbf{S} \cdot \mathbf{S} + \mathbf{a}\mathbf{a} - \mathbf{\Omega}\mathbf{\Omega} + (\mathbf{\Omega} \cdot \mathbf{\Omega})\mathbf{I} - \mathbf{Q}. \qquad (D.31)$$

The traces of Eqs. (D.30) and (D.31) may be written in addition:

$$2\nabla \cdot \mathbf{L} = -\tfrac{1}{4}(\operatorname{tr} \mathbf{N})^2 + \tfrac{1}{2}\mathbf{N} : \mathbf{N} - \mathbf{L} \cdot \mathbf{L} + \operatorname{tr} \mathbf{E}, \qquad (D.32)$$

$$\nabla \cdot \mathbf{a} - \operatorname{tr} \dot{\mathbf{S}} = \mathbf{S} : \mathbf{S} - \mathbf{a} \cdot \mathbf{a} - 2\mathbf{\Omega} \cdot \mathbf{\Omega} + \operatorname{tr} \mathbf{Q}. \qquad (D.33)$$

Equation (D.30) may be referred to as the generalized equation of Gauss (c.f. Ref. 3, p. 278 and Ref. 4, p. 146). It contains only the spatial parts of the anholonomic affinity, L and N, and the dyadic E, defined as

$$\mathbf{E} \equiv -(\mathbf{P} + \tfrac{1}{2}\mathbf{S} \overset{\times}{\times} \mathbf{S} + \mathbf{\Omega}\mathbf{\Omega} + \boldsymbol{\omega}\mathbf{\Omega} + \mathbf{\Omega}\boldsymbol{\omega}). \qquad (D.34)$$

In our case (4), when $\mathbf{\Omega} = 0$, the preferred congruence is 3-space normal, and Eq. (D.30) then comprises the six curvature equations for an imbedded Riemannian 3-space. The dyadic E reduces to

$$\mathbf{E} = -\mathbf{P} - \tfrac{1}{2}\mathbf{S} \overset{\times}{\times} \mathbf{S} \qquad (\mathbf{\Omega} = 0), \qquad (D.35)$$

and is precisely the strangled Einstein 3-tensor for this imbedded space. The form explicitly reveals the dependence of the metric properties of the subspace on the four-dimensional curvature components P (which we have accordingly dubbed the induced curvature dyadic), and on the second fundamental form S, the rate-of-strain of the timelike congruence. Upon taking the covariant divergence of Eq. (D.35), the dyadic equations may be used to show further that

$$\nabla \cdot \mathbf{E} = 0 \qquad (\mathbf{\Omega} = 0), \qquad (D.36)$$

a vector equation expressing the three independent Bianchi Identities for a Riemannian 3-space. Finally, the scalar curvature of the subspace, $-2 \operatorname{tr} \mathbf{E}$, is related to the spatial anholonomic affinity by Eq. (D.32).

Equations (D.28 and D.29) may together be referred to as the generalized equations of Codazzi (cf. Ref. 3, p. 278 and Ref. 4, p. 146) inasmuch as, again when $\mathbf{\Omega} = 0$, they are the usual eight partial differential equations for the second fundamental form of the imbedded 3-space. A special case of Eq. (D.28) in tensor form has been used by Rayner[15] in discussing Born-type rigid motions ($\mathbf{S} = 0$) in general relativity, (c.f. Ref. 2).

---

[15] C. B. Rayner, Compt. Rend. 248, 929 (1959).

Equation (D.31) is essentially a kinematic relation for the preferred congruence; we return to its physical interpretation in Sec. D3. Its trace, Eq. (D.33), reduces for incoherent matter ($T = 0$, $t = 0$, $a = 0$) to an equation whose tensor equivalent is found in Raychaudhuri's work.[16]

The quantities L and N do not appear explicitly in eighteen of the thirty-six equations, (D.22) to (D.33), althought they still play an implicit role in the "covariant" derivative, ∇. It is often convenient to collect this particular set of equations in two nonsymmetric dyadic equations as follows:

$$\nabla a - (\dot{S} + \omega \times S - S \times \omega) + (\dot{\Omega} + \omega \times \Omega) \times I$$
$$= S \cdot S - \Omega \times S - S \times \Omega - aa$$
$$+ \Omega\Omega - (\Omega \cdot \Omega)I + Q, \qquad (D.37)$$

and

$$\nabla \Omega + S \times \nabla$$
$$= -2a\Omega + (a \cdot \Omega)I + B + t \times I. \qquad (D.38)$$

(c) Bianchi Equations

These follow from Eq. (B.13), but more directly can be obtained in dyadic form by differentiation of Eqs. (D.22)–(D.33), using the commutation formulas to be given in the following subsection.

$$\nabla \cdot Q - \nabla(\text{tr } Q) - 2(\dot{t} + \omega \times t)$$
$$= -S \dot{\times} B - 3\Omega \cdot B - \Omega \times t + 3S \cdot t$$
$$+ (\text{tr } S)t + a \cdot [P - Q - (\text{tr } P - \text{tr } Q)I], \qquad (D.39)$$

$$\nabla \cdot B - \nabla \times t$$
$$= S \dot{\times} P + 2\Omega \cdot Q + \Omega \cdot P - (\text{tr } P)\Omega, \qquad (D.40)$$

$$\nabla \cdot P = -S \dot{\times} B - 3\Omega \cdot B$$
$$- 3\Omega \times t + S \cdot t - (\text{tr } S)t, \qquad (D.41)$$

$$\nabla \times Q - Q \times \nabla - 2(\dot{B} + \omega \times B - B \times \omega)$$
$$= (P + Q) \times a - a \times (P + Q) - t \times S + S \times t$$
$$+ 3\Omega t + 3t\Omega - 2\Omega \cdot tI - \Omega \times B + B \times \Omega$$
$$- 3S \cdot B - 3B \cdot S + 4(\text{tr } S)B + 2S : BI, \qquad (D.42)$$

$$-\nabla \times B + B \times \nabla - \nabla t - t\nabla + 2\nabla \cdot tI$$
$$- 2(\dot{P} + \omega \times P - P \times \omega) = 2a \times B - 2B \times a$$
$$+ 2ta + 2at - 4a \cdot tI - \Omega \times P + P \times \Omega$$
$$- P \cdot S - S \cdot P + 2(\text{tr } S)P - 2S \dot{\times} Q. \qquad (D.43)$$

The trace of Eq. (D.43) is of independent interest:

$$2\nabla \cdot t - \text{tr } \dot{P} = -S : P + S : Q + (\text{tr } S)(\text{tr } P)$$
$$- (\text{tr } S)(\text{tr } Q) - 4a \cdot t. \qquad (D.44)$$

[16] A. Raychaudhuri, Phys. Rev. 98, 1123 (1955).

The scalar Eq. (D.44) may be joined with a vector equation which is the difference of (D.39) and (D.41), to give four familiar equations for the stress dyadic T, momentum density vector t and energy density ρ:

$$\nabla \cdot t + [\dot{\rho} + (\text{tr } S)\rho] = T : S - 2a \cdot t, \qquad (D.45)$$

$$\nabla \cdot T - [\dot{t} + \omega \times t + (\text{tr } S)t]$$
$$= S \cdot t + \Omega \times t - T \cdot a + \rho a. \qquad (D.46)$$

These are the "contracted Bianchi Identities" in dyadic form, commonly interpreted as conservation laws for energy and momentum.

(d) Commutation Formulas

A large variety of these may readily be inferred from Eq. (B.12). As was remarked, it is an inconvenience that neither the D nor ∇ operator commutes with itself, or with time differentiation. We will give here only three which are of frequent occurrence in manipulating the intrinsic derivative operator D; ϕ and V are arbitrary scalar and vector fields, respectively.

$$(D\phi)^{\cdot} - D(\dot{\phi}) = a\dot{\phi} - S \cdot D\phi$$
$$+ (\Omega - \omega) \times D\phi, \qquad (D.47)$$

$$D \times D\phi = 2\Omega\dot{\phi} + N \cdot D\phi + L \times D\phi, \qquad (D.48)$$

$$D \cdot (D \times V) = 2\Omega \cdot \dot{V} + N : DV + L \cdot D \times V. \qquad (D.49)$$

It is convenient however to give a quite complete tabulation of such formulas for the 3-space covariant operator ∇; here M is an arbitrary symmetric dyadic. For the time–space commutation relations we have:

$$(\nabla\phi)^{\cdot} - \nabla(\dot{\phi}) = a\dot{\phi} - S^{\star} \cdot \nabla\phi, \qquad (D.50)$$

$$(\nabla V)^{\cdot} - \nabla(\ ) = a\dot{V} - S^{\star} \cdot \nabla V$$
$$- [S^{\star T} \times (\nabla + a) + (\nabla + a) \cdot (\Omega - \omega)I] \times V, \qquad (D.51)$$

$$(\nabla \times M)^{\cdot} - \nabla \times (\dot{M}) = a \times \dot{M} - S^{\star} \dot{\times} \nabla M$$
$$+ [(\nabla + a) \times S^{\star}] \cdot M + M \cdot [(\nabla + a) \times S^{\star}]$$
$$+ [(\nabla + a) \times S^{\star} - (\nabla + a) \cdot (\Omega - \omega)I]$$
$$\cdot [M - (\text{tr } M)I] - [(\nabla + a) \times S^{\star}] : MI. \qquad (D.52)$$

The analogous commutators for $(\nabla \cdot V)^{\cdot}$, $(\nabla \times V)^{\cdot}$, and $(\nabla \cdot M)^{\cdot}$ follow directly from Eqs. (D.51) and (D.52) by contraction and antisymmetrization, and so need not be exhibited. We have for convenience again introduced the nonsymmetric dyadic $S^{\star}$ and its transpose $S^{\star T}$:

$$S^{\star} = S - (\Omega - \omega) \times I, \quad S^{\star T} = S + (\Omega - \omega) \times I. \qquad (D.53)$$

The commutation relations for spacelike direc-

tions are:

$$\nabla \times (\nabla \phi) = 2\Omega \cdot \{I\phi\}, \qquad (D.54)$$

$$\nabla \cdot (\nabla \times V) = 2\Omega \cdot \{\dot{V} + S^* \cdot V\}, \qquad (D.55)$$

$$\nabla \times (\nabla V)$$
$$= -E \times V + 2\Omega \cdot \{I\dot{V} - \tfrac{1}{2}S^* \times (I \times V)\}, \qquad (D.56)$$

$$\nabla \cdot (\nabla \times M) = -E \dot{\times} M$$
$$+ 2\Omega \cdot \{\dot{M} + \tfrac{3}{2}S^* \cdot [M - \tfrac{1}{3}(\operatorname{tr} M)I]\}, \qquad (D.57)$$

$$\nabla \cdot [\nabla \times (I \times V)]$$
$$= -E \dot{\times} (I \times V) + 2\Omega \cdot \{I \times \dot{V} + \tfrac{1}{2}S^* \cdot (I \times V)\}. \qquad (D.58)$$

These general relations appear quite complicated. Again, however, when $\Omega = 0$ and the timelike congruence is 3-space normal, we discover simple, perspicuous equations. Equations (D.54) and (D.55) become the familiar vector identities; the rest reduce to dyadic forms of the Ricci identities in a Riemannian 3-space, with the Einstein dyadic E acting for the curvature tensor.

3. Physical Interpretations

Let us consider further the relative separation r of two closely adjacent particles of the $_0\lambda^\mu$ congruence, Eq. (C.41). This is a local Cartesian vector equation, correct to first order in r; S, Ω, and ω are to be evaluated on one line of the congruence. Taking N and L to vanish on the line was tacitly necessary for interpretation of Eq. (C.40), for this condition implies that the spatial triad system is taken locally Cartesian and flat, and we in fact required this in order to write Eqs. (C.41)–(C.43), where the displacement r is a vector. We may thus say that Eq. (C.41) is not just pointwise valid, but rather is valid to first order in a flat metric 3-space carried along with the local observer. The observer is accelerating, and since we do not specialize ω along the world line, his reference triad is arbitrarily rotating.

Differentiating Eq. (C.41) with respect to time, and substituting $\dot{S}$ from Eq. (D.31) and $\dot{\Omega}$ from Eq. (D.23), we can eliminate all such quantities relating to the whole congruence in favor of the local kinematic observables of one particle-observer (or of one line of the congruence with its reference tetrad), viz., a and $-\omega$. These are respectively the vectorial reading of a linear accelerometer and the vector angular velocity of a (gyroscopically stabilized, or untorqued) "stable-platform."

We find as a result an equation for the observed spatial variation of a:

$$a_1 \equiv a + r \cdot \nabla a = a(1 - a \cdot r) + \ddot{r} + 2\omega \times \dot{r}$$
$$+ \omega \times (\omega \times r) + \dot{\omega} \times r + Q \cdot r. \qquad (D.59)$$

This is a quasi-Newtonian equation for a_1, the accelerometer reading at the adjacent point r, in terms of the accelerometer reading a at the origin of spatial coordinates and the relative acceleration $\ddot{r}$. It is entirely written in local, proper or "operational" terms, and is immediately useful for the analysis of experiments. The usual centrifugal, Coriolis, and angular acceleration terms will be recognized. A special relativistic clock rate correction factor $(1 - a \cdot r/c^2)$—where $c^2 = 1$ in our units—is but another manifestation of the "red shift" predicted by special relativity for accelerating frames and recently verified in local terrestrial experiments using the Mössbauer effect (compare Ref. 5, p. 411).

The term $Q \cdot r$ is the general relativistic term expressing the tidal effect of the curvature tensor on the relative acceleration. When Q is written in terms of our second set of dyadics this term becomes

$$Q \cdot r = [A + T + \tfrac{1}{3}(\rho - 2\operatorname{tr} T)I] \cdot r. \qquad (D.60)$$

In this form the contributions of the "source" and "field" parts of the Riemann tensor are separately revealed: for source-free regions one has just $A \cdot r$. If the test particles are free ($a = a_1 = 0$), Equation (D.59) reduces to the equation of geodesic deviation of Synge.[5] If on the other hand they are parts of a stress system obeying Hooke's law and the absolute accelerations a, a_1 are related to the stresses, one obtains the dynamical equations of Weber.[17] The dyadic partial differential equations, such as those for ∇a and $\nabla \Omega$, Eqs. (D.37) and (D.38), provide a generally valid instrument, expressed in an operational language, for the treatment of similar problems on the motion of macroscopic, continuous "test" bodies.

A similar equation may be found for the stable-platform angular velocity $-\omega_1$, at r, in terms of that at the origin, $-\omega$. From Eqs. (D.28), (D.29), (D.25), and (D.26) and again (C.41), and setting $N = 0$ and $L = 0$, we obtain

$$-\omega_1 \equiv -\omega + r \cdot \nabla(-\omega) = (-\omega)(1 - a \cdot r)$$
$$+ a \times (\dot{r} + \omega \times r) - B \cdot r + t \times r. \qquad (D.61)$$

Here all terms leading to a difference of $-\omega_1$ and $-\omega$ are nonclassical, of special or general relativistic origin. We again find a clock rate correction factor. The second special relativistic term is the differential Thomas precession. These two terms combined can be derived from the usual Thomas precession formula, in the differential limit, if care is taken to express all precession rates in terms of local proper

[17] J. Weber, *General Relativity and Gravitational Waves* (Interscience Publishers, Inc., New York, 1961), Chap. 8.

times. In the last two terms we again note separate contributions from the field and source parts of the Riemann tensor: a "spin" term $-\mathbf{B} \cdot \mathbf{r}$, arising from the conformal tensor, and an "orbital" term $\mathbf{t} \times \mathbf{r}$, from the Einstein tensor. For geodesic observers, only these general relativistic terms will remain; they may be denoted the differential Fokker precession.[18]

Equations (D.59) and (D.61) show how in principle the fourteen Riemann components $\mathbf{Q}$, $\mathbf{B}$, and $\mathbf{t}$ may be experimentally determined from local differential kinematical measurements near, and on, one arbitrarily given timelike world line. As was remarked previously, the remaining six components, in the induced curvature dyadic $\mathbf{P}$, are in principle determinable from local spatial surveying in a triad system, Eq. (D.30); this means that their geometric effects will be second order in the spatial displacement components r_a. An experimental approach to the measurement of $\mathbf{P}$ would no doubt instead involve kinematical experiments on $\mathbf{Q}$, $\mathbf{B}$, and $\mathbf{t}$ as above, but made by two or more point-observers in rapid relative motion. These complications will not arise in source-free regions, however: for expressing $\mathbf{P}$ by

$$\mathbf{P} = \mathbf{A} - \mathbf{T} + \tfrac{1}{3}(\text{tr } \mathbf{T} - 2\rho)\mathbf{I} \qquad (D.62)$$

and recalling Eq. (D.60), we clearly have in this case $\mathbf{P} = \mathbf{Q} = \mathbf{A}$.

As a final illustration we obtain an equation for the quasi-Newtonian "gravitational field" of a nonrotating $(\mathbf{\Omega} = 0)$ static distribution of matter with proper energy density ρ and stress dyadic $\mathbf{T}$. The matter is represented by a congruence $_0\lambda^\mu$ defined by the condition $\mathbf{t} = 0$, and everywhere nonrotating $(\omega = 0)$ auxiliary triads are introduced. A static distribution is defined operationally by the condition that in this tetrad system the local time derivative of every kinematic observable must vanish. We, of course, already have $\dot{\mathbf{\Omega}} = \dot{\omega} = 0$, but specifically impose the further conditions $\dot{\mathbf{a}} = 0$ and $\mathbf{S} = 0$, the latter being required to ensure that all relative displacements $\mathbf{r}$ are time independent.

When all these conditions $(\mathbf{\Omega} = \omega = \mathbf{t} = \dot{\mathbf{a}} = \mathbf{S} = 0)$ are invoked, Eqs. (D.25) and (D.26) show that $\dot{\mathbf{L}} = \dot{\mathbf{N}} = 0$, and the other dyadic equations then directly yield the same result for the local time derivative of every remaining quantity. For instance, the scalar Bianchi identity Eq. (D.45) has the immediate consequence, $\dot{\rho} = 0$.

We now imagine a population of proper Newtonian

observers, each of whom prefers to ascribe his kinematic observations not to his own absolute acceleration $\mathbf{a}$, but rather to a "gravitational field of force" with intensity $\mathbf{F} = -\mathbf{a}$. The "gravitational field equation" is then just Eq. (D.33) which, under the imposed conditions, may be written

$$\nabla \cdot \mathbf{F} = -4\pi\gamma\left(\rho_M - \frac{1}{c^2}\text{ tr } \mathbf{T}\right) + \frac{1}{c^2}\mathbf{F} \cdot \mathbf{F}, \qquad (D.63)$$

where we have put $-\mathbf{F}$ for $\mathbf{a}$; substituted for tr $\mathbf{Q}$ its equivalent, $\rho - \text{tr } \mathbf{T}$; restored dimensional factors; and defined a proper mass density, $\rho_M \equiv \rho/c^2$.

When $\mathbf{\Omega} = 0$ it follows from Eq. (D.23) that $\nabla \times \mathbf{a} = 0$, and this, together with $\dot{\mathbf{a}} = 0$, is sufficient to permit expressing $\mathbf{F}$ as the gradient of a time-independent scalar:

$$\mathbf{F} = -\nabla\phi, \qquad \dot{\phi} = 0. \qquad (D.64)$$

Equation (D.63) will then take the form

$$\nabla^2\phi = 4\pi\gamma\left(\rho_M - \frac{1}{c^2}\text{ tr } \mathbf{T}\right) - \frac{1}{c^2}(\nabla\phi)^2. \qquad (D.65)$$

For the prescribed conditions this is an exact equation reducing to Poisson's equation in the nonrelativistic approximation. If we also rewrite Eq. (D.37) in these terms and for these conditions, we find the following expressions for the tidal acceleration dyadic $\mathbf{Q}$:

$$\mathbf{Q} = -\nabla\mathbf{F} + \frac{1}{c^2}\mathbf{F}\mathbf{F}$$

$$= \nabla\nabla\phi + \frac{1}{c^2}(\nabla\phi)(\nabla\phi). \qquad (D.66)$$

Note added in proof: In a private communication, Dr. F. A. E. Pirani has very kindly called our attention to the "method of projection" of Carlo Cattaneo.[19] We were completely unaware of this work, whose relation to the present formulation should be noted. Our operator ∇, denoted by us the operator of "spatial covariant differentiation," is precisely the covariant operator of "transverse differentiation" of Cattaneo, strangled. Those of our equations such as (D.37) and (D.38) not explicitly involving $\mathbf{N}$ and $\mathbf{L}$ can of course be immediately "unstrangled" by multiplication with $_a\lambda^\mu$, $_b\lambda^\nu$, etc., to give covariant equations not depending on a choice of auxiliary congruences; such equations are thus derivable by the method of projection. On the other hand, our equations (D.24), (D.25), (D.26), and (D.30) explicitly contain $\mathbf{N}$ and $\mathbf{L}$, and seem to be much less accessible in covariant language, while vital for the completeness of the total set.

[18] A. D. Fokker, Proc. Roy. Acad. (Amsterdam) **23**, 729 (1920).

[19] See, for example, C. Cattaneo, Compt. Rend. **248**, 197 (1959); I. Cattaneo-Gasparini, Compt. Rend. **252**, 3722 (1961).

Scattering Wavefunctions for a Dirac Particle in a Central Potential*

D. M. Fradkin, G. L. Hammer, and T. A. Weber

Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa

(Received 3 June, 1964)

The scattering wavefunction for a Dirac particle in a central potential is written in terms of a matrix acting on a plane-wave spinor whose momentum direction $\hat{p}$ and polarization direction are the assigned directions for the asymptotic incident plane wave. The matrix involves four functions, independent of polarization direction, which multiply the matrices 1, $\alpha \cdot (\hat{p} - \hat{r})$, $\alpha \cdot (\hat{p} + \hat{r})$, and $\hat{p} \cdot \hat{r} \cdot \sigma$. Differential relations for these four functions are directly obtained and asymptotic relations are given. In particular, the function multiplying $\alpha \cdot (\hat{p} + \hat{r})$ is asymptotically zero, so the potential scattering formulation is identical with that given previously for the Coulomb potential. The scattering wavefunction is a solution of the general differential relations subject to appropriate boundary conditions. For the Coulomb potential, these differential relations simplify, and an iterative solution is developed based on a Green's function technique with the Sommerfeld–Maue approximation as the zero-order solution.

INTRODUCTION

THE scattering wavefunction for a Dirac particle in a central potential is that eigenfunction of the Hamiltonian which has the asymptotic behavior of a plane wave plus an outgoing spherical wave. The usual method of constructing the scattering wavefunction is to find simultaneous eigenfunctions of the Hamiltonian, the Dirac operator $K = \beta(\sigma \cdot L + 1)$ whose eigenvalues characterize both the total angular momentum and the parity, and the z component of the angular momentum. The scattering wavefunction is then expanded in terms of an infinite series of these angular momentum eigenfunctions. The expansion coefficients are chosen to provide the proper asymptotic behavior. This infinite series expansion, though exact, has the disadvantage of being unwieldy for purposes of calculation and does not lend itself easily to approximation procedures. Consequently, it is desirable to express the scattering wavefunction in a different form.

For the special case of a Coulomb potential, the infinite series expansion has been reorganized into the form of a matrix acting on a plane-wave spinor of arbitrary polarization direction.[1,2] The momentum direction $\hat{p}$ and the assigned polarization direction of the plane-wave spinor are the corresponding directions associated with the asymptotic incident plane wave. The matrix can be written in terms of three functions, independent of polarization direction, multiplying the matrices 1, $\alpha \cdot (\hat{p} - \hat{r})$ and $\hat{p}_\wedge \hat{r} \cdot \sigma$. These three functions are given in

terms of an infinite series which may be summed to zero order in the interaction strength to give the Sommerfeld–Maue approximation. For large r, a correction to higher order in the interaction strength may also be given in terms of a finite number of functions.[3] In addition to the case of the Coulomb potential, the scattering solution of the Biedenharn symmetric Dirac–Coulomb Hamiltonian[4,5] has also been reorganized[6] into the form of a matrix acting on a plane-wave spinor, and the matrix has exactly the same form as the one for the Dirac–Coulomb case.

In the present paper, it is shown that the scattering solution for a general central potential may be reorganized into the form of a matrix acting on a plane wave spinor of arbitrary polarization direction. For this general case, the matrix is written in terms of four functions, independent of polarization direction, multiplying the matrices 1, $\alpha \cdot (\hat{p} - \hat{r})$, $\alpha \cdot (\hat{p} + \hat{r})$, and $\hat{p}_\wedge \hat{r} \cdot \sigma$. Once this fact is established, it is shown that the assumption of this form leads to differential relations among the four functions. The exact solution is then a matter of solving these partial differential equations subject to the appropriate boundary conditions. The advantage of casting the problem into this form is that approximation procedures may be developed directly from the differential equations. Also, certain asymptotic relations among the four functions can be directly determined by invoking the boundary conditions.

* Contribution No. 1519. Work was performed in the Ames Laboratory of the U. S. Atomic Energy Commission.
[1] A. Deloff, Nucl. Phys. 13, 136 (1959).
[2] W. R. Johnson and R. T. Deck, J. Math. Phys. 3, 319 (1962).

[3] D. M. Fradkin, T. A. Weber, and C. L. Hammer, Ann. Phys. (N. Y.) 27, 338 (1964). In the following, this paper will be referred to as FWH.
[4] L. C. Biedenharn, Bull. Am. Phys. Soc. 7, 314 (1962).
[5] L. C. Biedenharn and N. V. V. J. Swamy, Phys. Rev. 133, B1353 (1964).
[6] D. M. Fradkin, Phys. Rev. 135, B1085 (1964).

Restricting the discussion of the general equations to the case of the Coulomb potential, it is shown that the function multiplying the matrix $\boldsymbol{\alpha} \cdot (\hat{p} + \hat{r})$ is identically zero. Also, the whole problem of obtaining the scattering solution reduces to finding the solution of a single second-order partial differential equation. A solution of this equation to zero order in the square of the interaction strength yields the Sommerfeld–Maue approximation. The Green's function based on this zero-order solution is then developed. In this way, the problem of solving the differential equation subject to appropriate boundary conditions is transformed into a problem of solving an integral equation. This provides an iteration procedure for expanding the exact solution in powers of the square of the interaction strength for all values of the dynamical Born parameter. For large r, the approximation obtained from the first iteration is equivalent to that previously obtained in FWH.

THE GENERAL FORM OF THE WAVEFUNCTION

The Hamiltonian for a Dirac particle in a central potential is given by

$$H = -i\boldsymbol{\alpha} \cdot \nabla + \beta + \lambda V(r). \quad (1.1)$$

Here, units are used for which $\hbar = m = c = 1$, and the notation in FWH is followed. In particular, λ is the interaction strength parameter and $V(r)$ is the central potential. In this section, it will be proved that the scattering solution ψ which satisfies

$$H\psi(E, \mathbf{r}) = E\psi(E, \mathbf{r}) \quad (1.2)$$

may be written in the form

$$\psi = [G + i\lambda M\boldsymbol{\alpha} \cdot (\hat{p} - \hat{r})$$
$$+ i\lambda N\boldsymbol{\alpha} \cdot (\hat{p} + \hat{r}) + iL\hat{p}_{\wedge}\hat{r} \cdot \boldsymbol{\sigma}]U(\hat{p}), \quad (1.3)$$

where G, M, N, and L are functions independent of the polarization direction, and $U(\hat{p})$ is a plane-wave spinor whose momentum direction $\hat{p}$ and polarization direction are the directions associated with the asymptotic incident plane wave.

It is known[7] that the scattering solution for a Dirac particle in a central potential is given in terms of an angular momentum eigenfunction expansion by

$$\psi(E, \mathbf{r}) = 4\pi[\pi/(2Ep)]^{\frac{1}{2}}$$
$$\times \sum_{m,k,\mu} i^{l(k)} c_m e^{i\Delta_k} C(l(k), \tfrac{1}{2}, j; \mu - m, m)$$
$$\text{times } Y_{l(k)}^{\mu-m*}(\hat{p})\psi_k^\mu(\mathbf{r}, E). \quad (1.4)$$

[7] See, for example, M. E. Rose, *Relativistic Electron Theory* (John Wiley & Sons, Inc., New York, 1961), p. 207.

Here $l(k) = |k| + \tfrac{1}{2}(s_k - 1)$, $j = |k| - \tfrac{1}{2}$, $k = \pm 1$, $\pm 2, \cdots$, $s_k = \pm 1$ for $k \gtrless 0$, μ is a half integer, and the summation extends over all k, $m = \pm \tfrac{1}{2}$, and μ such that $|\mu - m| \leq \ell(k)$. Also, $\hat{p}$ is the direction of the asymptotic momentum, $p = [E^2 - 1]^{\frac{1}{2}}$, C is the Clebsch–Gordon coefficient, Δ_k is the difference between the phase shift for the potential $V(r)$ and zero potential, c_m are arbitrary constants limited only by the condition $\sum c_m^* c_m = 1$, and Y is the usual spherical harmonic. The angular momentum eigenfunction $\psi_k^\mu(\mathbf{r}, E)$ satisfies

$$(H - E)\psi_k^\mu(\mathbf{r}, E) = (K + k)\psi_k^\mu(\mathbf{r}, E) = 0, \quad (1.5)$$

where K is the Dirac operator $\beta(\boldsymbol{\sigma} \cdot \mathbf{L} + 1)$. It has the form

$$\psi_k^\mu(\mathbf{r}, E) = \begin{bmatrix} g_k(E, r)\chi_k^\mu(\hat{r}) \\ if_k(E, r)\chi_{-k}^\mu(\hat{r}) \end{bmatrix}. \quad (1.6)$$

Here,

$$\chi_k^\mu(\hat{r}) = \sum_{\tau = \pm \frac{1}{2}} C(l(k), \tfrac{1}{2}, j; \mu - \tau, \tau) Y_{l(k)}^{\mu-\tau*}(\hat{r})\chi^\tau, \quad (1.7)$$

$g_k(E, r)$, $f_k(E, r)$ are solutions of the appropriate radial equations for the given central potential, and χ^τ are the two component spin-up, spin-down functions.

In analogy to the treatment of Johnson and Deck,[2] a coordinate system is chosen so that the polar z axis is oriented along the direction of the asymptotic momentum vector $\hat{p}$. This choice conveniently gives

$$Y_{l(k)}^{\mu-m*}(\hat{p}) = \delta_{\mu,m}[(2l(k) + 1)/(4\pi)]^{\frac{1}{2}}. \quad (1.8)$$

Evaluating the Clebsch–Gordon coefficients and performing the sums over m and μ, one finds that the scattering solution has the form

$$\psi(E, r) = \begin{bmatrix} (b_1 + Ob_2)v \\ (b_3 + Ob_4)w \end{bmatrix}. \quad (1.9)$$

Here,

$$O = \begin{bmatrix} 0 & -e^{-i\phi} \\ e^{i\phi} & 0 \end{bmatrix} = i(\sin\theta)^{-1}\boldsymbol{\sigma} \cdot \hat{r}_{\wedge}\hat{p}, \quad (1.10)$$

$$b_1 = \sum \rho_k g_k |k| P_{l(k)}(\cos\theta),$$

$$b_2 = \sum \rho_k s_k g_k P_{l(k)}^1(\cos\theta),$$

$$b_3 = -i[(E + 1)/(E - 1)]^{\frac{1}{2}} \sum \rho_k f_k |k| P_{l(-k)}(\cos\theta),$$

$$b_4 = i[(E + 1)/(E - 1)]^{\frac{1}{2}} \sum \rho_k s_k f_k P_{l(-k)}^1(\cos\theta),$$

$$\rho_k = [\pi/(2Ep)]^{\frac{1}{2}} i^{l(k)} e^{i\Delta_k},$$

$$P_{l(k)}^1(\cos\theta) = (\sin\theta)(d/d\cos\theta) P_{l(k)}(\cos\theta),$$

$$v = \begin{pmatrix} c_{\frac{1}{2}} \\ c_{-\frac{1}{2}} \end{pmatrix},$$

$$w = \frac{\mathbf{\sigma \cdot p}}{E + 1} v = \left[\frac{E - 1}{E + 1} \right]^{\frac{1}{2}} \begin{pmatrix} c_{\frac{1}{2}} \\ -c_{-\frac{1}{2}} \end{pmatrix}.$$

The argument of the Legendre polynomial P is $\cos \theta = \hat{r} \cdot \hat{p}$.

The two component column matrices v and w are the upper and lower components of the plane wave spinor of arbitrary polarization direction (corresponding to the arbitrary nature of c_m). Thus,

$$U(\hat{p}) = [(E + 1)/(2E)]^{\frac{1}{2}} \begin{pmatrix} v \\ w \end{pmatrix}, \qquad (1.11)$$

where

$$(\mathbf{\alpha \cdot p} + \beta - E)U(\hat{p}) = 0. \qquad (1.12)$$

Consequently, the general expression for the scattering solution ψ given by Eq. (1.9) can be written in the form

$$\psi = E^{\frac{1}{2}}[2(E + 1)]^{-\frac{1}{2}}\{[b_1 + ib_2(\sin \theta)^{-1}\mathbf{\sigma \cdot \hat{r}_\wedge \hat{p}}](1 + \beta)$$

$$+ [b_3 + ib_4(\sin \theta)^{-1}\mathbf{\sigma \cdot \hat{r}_\wedge \hat{p}}](1 - \beta)\} U(\hat{p}). \quad (1.13)$$

This establishes the fact that for a central potential, the scattering solution has the form of an operator acting on a plane-wave spinor of arbitrary polarization. This operator is composed of the four Dirac matrices 1, β, $\mathbf{\sigma \cdot \hat{r}_\wedge \hat{p}}$, and $\beta \mathbf{\sigma \cdot \hat{r}_\wedge \hat{p}}$ multiplied by functions which are *independent* of polarization direction. Alternatively, one can replace β by $E - \mathbf{\alpha \cdot p}$ when acting on $U(\hat{p})$, and consider then the operator in terms of the Dirac matrices 1, $\mathbf{\sigma \cdot \hat{p}_\wedge \hat{r}}$, $\mathbf{\alpha \cdot (\hat{p} + \hat{r})}$, and $\mathbf{\alpha \cdot (\hat{p} - \hat{r})}$ with multiplying functions that are also polarization independent.

DIFFERENTIAL RELATIONS

Consider a scattering solution of the form given in Eq. (1.3), namely

$$\psi(E, \mathbf{r}) = \mathfrak{D}(E, \mathbf{r}, \hat{p})U(\hat{p}), \qquad (2.1)$$

where

$$\mathfrak{D}(E, \mathbf{r}, \hat{p}) = [G + i\lambda M\mathbf{\alpha \cdot (\hat{p} - \hat{r})} + i\lambda N\mathbf{\alpha \cdot (\hat{p} + \hat{r})}$$

$$+ iL\hat{p}_\wedge \hat{r} \cdot \mathbf{\sigma}]. \qquad (2.2)$$

The differential relations among G, M, N, and L (functions that are independent of polarization direction) can be obtained directly by the following technique.

Since the scattering solution is an eigenfunction of the Hamiltonian, given in Eq. (1.1), it must satisfy the relation

$$(H - E)\mathfrak{D}(E, \mathbf{r}, \hat{p})U(\hat{p}) = 0. \qquad (2.3)$$

As a consequence, it follows that

$$\text{Tr} [\gamma_A(H - E)\mathfrak{D}(E, \mathbf{r}, \hat{p})P_+(E, \hat{p})] = 0, \qquad (2.4)$$

where γ_A is any one of the sixteen independent Dirac matrices, and $P_+(E, \hat{p})$ is the free-particle positive energy projection operator given by

$$P_+(E, \hat{p}) = (2E)^{-1}(\mathbf{\alpha \cdot p} + \beta + E). \qquad (2.5)$$

In other words, the coefficients of the sixteen Dirac matrices appearing in the expression

$$(H - E)\mathfrak{D}(E, \mathbf{r}, \hat{p})P_+(E, \hat{p})$$

are all equal to zero. This process yields sixteen equations, of which only eight are independent since P_+ connects the coefficients of β with those for $\mathbf{\alpha \cdot p} + E$. The vector matrices, e.g., $\mathbf{\alpha}$, may be conveniently resolved in the independent directions $\hat{r}$, $\hat{p}$, and $\hat{r}_\wedge \hat{p}$. Also, it is convenient to use a spherical coordinate system

$$x = r \sin \theta \cos \phi,$$

$$y = r \sin \theta \sin \phi,$$

$$z = r \cos \theta,$$

where, as before, the polar direction is $\hat{p}$ so that

$$\cos \theta = \hat{r} \cdot \hat{p} \equiv a. \qquad (2.6)$$

It is found that the eight independent equations are:

$$\partial G/\partial \phi = \partial M/\partial \phi = \partial N/\partial \phi = \partial L/\partial \phi = 0, \qquad (2.7)$$

$$[-(1 - a)(\partial/\partial r) + (1 - a^2)r^{-1}(\partial/\partial a) + ip(1 - a)$$

$$- (2/r)]M + [(1 + a)(\partial/\partial r) + (1 - a^2)r^{-1}(\partial/\partial a)$$

$$+ ip(1 + a) + (2/r)]N + VG = 0, \qquad (2.8)$$

$$[(\partial/\partial r) + (1 - a)r^{-1}(\partial/\partial a) - ip]M + [(\partial/\partial r)$$

$$- (1 + a)r^{-1}(\partial/\partial a) + ip]N - VL = 0, \qquad (2.9)$$

$$[-(\partial/\partial r) - (1 - a)r^{-1}(\partial/\partial a) + ip]G$$

$$+ [(1 - a)(\partial/\partial r) - (1 - a^2)r^{-1}(\partial/\partial a) - ip(1 - a)$$

$$+ (1 + a)r^{-1}]L + [2\lambda(\lambda V - 2E)]N = 0, \qquad (2.10)$$

$$[(1 + a)(\partial/\partial r) + (1 - a^2)r^{-1}(\partial/\partial a)$$

$$+ ip(1 + a) + (1 - a)r^{-1}]L$$

$$+ [(\partial/\partial r) - (1 + a)r^{-1}(\partial/\partial a) + ip]G$$

$$+ [2\lambda(\lambda V - 2E)]M = 0. \qquad (2.11)$$

Substituting G and L from Eqs. (2.8) and (2.9) into Eqs. (2.10) and (2.11), one obtains the coupled second-order partial differential equations con-

taining only M and N:

$$-f(\partial/\partial a)(1 - a)M + [r^2(\partial^2/\partial r^2) + (3 - f)r(\partial/\partial r)$$

$$+ (\partial/\partial a)(1 - a)(\partial/\partial a)(1 + a) + (1 - f)(1 + ipr)$$

$$+ (pr)^2 + (\lambda rV)^2 - 2\nu pr(rV)]N = 0, \qquad (2.12)$$

$$f(\partial/\partial a)(1 + a)N + [r^2(\partial^2/\partial r^2) + (3 - f)r(\partial/\partial r)$$

$$+ (\partial/\partial a)(1 + a)(\partial/\partial a)(1 - a) + (1 - f)(1 - ipr)$$

$$+ (pr)^2 + (\lambda rV)^2 - 2\nu pr(rV)]M = 0, \qquad (2.13)$$

where

$$f = [r(d/dr) \ln rV],$$

$$\nu = \lambda E/p, \text{ the Born parameter.} \qquad (2.14)$$

In these and subsequent equations, all the differential operators act on everything to their right in a particular term. The sole exception to this convention is the function f which is a function of r only and not a differential operator.

It is apparent that by operating on Eq. (2.12) by $(\partial/\partial a)(1 + a)$ and using Eq. (2.13), one can obtain a fourth-order partial differential equation for M alone. Similarly, by operating on Eq. (2.13) by $(\partial/\partial a)(1 - a)$, a fourth-order partial differential equation for N can be obtained.

The first-order partial differential relations simplify somewhat in terms of the parabolic coordinate system associated with the nonrelativistic problem, namely,

$$\xi_1 = ipr(1 + a), \qquad \xi_2 = ipr(1 - a). \qquad (2.15)$$

In this coordinate system, Eqs. (2.8)–(2.11) can be written

$$[-\xi_2(\partial/\partial\xi_2) + \tfrac{1}{2}\xi_2 - 1]M$$

$$+ [\xi_1(\partial/\partial\xi_1) + \tfrac{1}{2}\xi_1 + 1]N + \tfrac{1}{2}(rV)G = 0, \qquad (2.16)$$

$$[\partial/\partial\xi_1) - \tfrac{1}{2}]M + [\partial/\partial\xi_2) + \tfrac{1}{2}]N$$

$$- (rV)(\xi_1 + \xi_2)^{-1}L = 0, \qquad (2.17)$$

$$[-(\partial/\partial\xi_1) + \tfrac{1}{2}]G + 2[\xi_2(\partial/\partial\xi_2) - \tfrac{1}{2}\xi_2 + 1](\xi_1 + \xi_2)^{-1}L$$

$$+ 2[\lambda^2(rV)(\xi_1 + \xi_2)^{-1} + i\nu]N = 0, \qquad (2.18)$$

$$[(\partial/\partial\xi_2) + \tfrac{1}{2}]G + 2[\xi_1(\partial/\partial\xi_1) + \tfrac{1}{2}\xi_1 + 1](\xi_1 + \xi_2)^{-1}L$$

$$+ 2[\lambda^2(rV)(\xi_1 + \xi_2)^{-1} + i\nu]M = 0. \qquad (2.19)$$

Also, the coupled second-order equations, Eqs. (2.12) and (2.13), become:

$$-f[(\partial/\partial\xi_1) - (\partial/\partial\xi_2)]\xi_2(\xi_1 + \xi_2)^{-1}M$$

$$+ \{\xi_1(\partial^2/\partial\xi_1^2) + \xi_2(\partial^2/\partial\xi_2^2)$$

$$+ [2 - f\xi_1(\xi_1 + \xi_2)^{-1}](\partial/\partial\xi_1)$$

$$+ [1 - f\xi_2(\xi_1 + \xi_2)^{-1}](\partial/\partial\xi_2) + \tfrac{1}{2}(1 - f) + i\nu(rV)$$

$$- \tfrac{1}{4}(\xi_1 + \xi_2) - [(\lambda rV)^2 + f](\xi_1 + \xi_2)^{-1}\}N = 0, \qquad (2.20)$$

$$f[(\partial/\partial\xi_1) - (\partial/\partial\xi_2)]\xi_1(\xi_1 + \xi_2)^{-1}N$$

$$+ \{\xi_1(\partial^2/\partial\xi_1^2) + \xi_2(\partial^2/\partial\xi_2^2)$$

$$+ [1 - f\xi_1(\xi_1 + \xi_2)^{-1}](\partial/\partial\xi_1)$$

$$+ [2 - f\xi_2(\xi_1 + \xi_2)^{-1}](\partial/\partial\xi_2) - \tfrac{1}{2}(1 - f) + i\nu(rV)$$

$$- \tfrac{1}{4}(\xi_1 + \xi_2) + [(\lambda rV)^2 - f](\xi_1 + \xi_2)^{-1}\}M = 0. \qquad (2.21)$$

ASYMPTOTIC RELATIONS

The scattering solution has the asymptotic behavior of a plane wave plus an outgoing spherical wave. Thus,

$$\lim_{r\to\infty} \psi(E, \mathbf{r}) = e^{i(\mathbf{p}\cdot\mathbf{r} + \delta_p)}U(\hat{p})$$

$$+ r^{-1}e^{i(pr + \delta_s)}\mathfrak{D}_s(E, \mathbf{r}, \hat{p})U(\hat{p}), \qquad (3.1)$$

where δ_p and δ_s are the plane-wave and spherical-wave phase factors, respectively, and the asymptotic matrix operator $\mathfrak{D}_s(E, \mathbf{r}, \hat{p})$ has the form

$$\mathfrak{D}_s(E, \mathbf{r}, \hat{p}) = [G_s + i\lambda M_s\boldsymbol{\alpha}\cdot(\hat{p} - \hat{r})$$

$$+ i\lambda N_s\boldsymbol{\alpha}\cdot(\hat{p} + \hat{r}) + iL_s\hat{p}_\wedge\hat{r}\cdot\boldsymbol{\sigma}]. \qquad (3.2)$$

The functions G_s, M_s, N_s, and L_s are obtained from the asymptotic form of G, M, N, and L, respectively.

When operating on the plane wave spinor $U(\hat{p})$ of definite polarization, the matrices associated with M, N, and L cannot produce the same plane-wave spinor. This follows from the fact that the only matrix operators, independent of polarization direction, that have this property are the identity and $(\boldsymbol{\alpha}\cdot\mathbf{p} + \beta)$. Consequently, since it is assumed that the asymptotic plane wave is characterized by $U(\hat{p})$ with definite assigned polarization, only the function G can contribute to the asymptotic plane wave.

Asymptotically, the outgoing spherical wave itself must be proportional to a plane-wave spinor propagating in the $\hat{r}$ direction. Thus it follows that

$$(p\boldsymbol{\alpha}\cdot\hat{r} + \beta - E)\mathfrak{D}_s(E, \mathbf{r}, \hat{p})U(\hat{p}) = 0. \qquad (3.3)$$

This equation implies certain relations among the asymptotic functions G_s, M_s, N_s, and L_s. By arguments similar to those given in the preceding section, the coefficients of the sixteen Dirac matrices appearing in the expression

$$(p\boldsymbol{\alpha}\cdot\hat{r} + \beta - E)\mathfrak{D}_s(E, \mathbf{r}, \hat{p})P_+(E, \hat{p})$$

are all equal to zero. Of the resulting sixteen equations, only two are independent. These asymptotic relations are

$$N_s = 0, \tag{3.4}$$

$$G_s + 2i\nu M_s + (1 + a)L_s = 0,$$

where again $\nu = \lambda E/p$, $a = \cos\theta = \hat{r}\cdot\hat{p}$.

Since N_s is zero, the asymptotic matrix operator is

$$\mathfrak{D}_s(E, \mathbf{r}, \hat{p}) = [G_s + i\lambda M_s \boldsymbol{\alpha}\cdot(\hat{p} - \hat{r}) + iL_s\hat{p}_\wedge\hat{r}\cdot\boldsymbol{\sigma}]. \tag{3.5}$$

This is the same form as given in FWH, so the potential scattering formulation for a general central potential—cross sections for single, double, and triple scattering, asymmetry functions, and change of polarization direction—may be taken over completely from that reference.

Specifically, the relation of N_s and N is

$$\lim_{r\to\infty} [re^{-i(pr+\delta_s)}N] = N_s = 0. \tag{3.6}$$

It is of interest to inquire under what conditions N itself is identically zero (for all r). If one assumes that N is zero, then from Eq. (2.12) it is found that

$$f(\partial/\partial a)(1 - a)M = 0, \tag{3.7}$$

where f is $[r(d/dr)\ln rV]$. Consequently, if rV is not constant, i.e., for potentials other than the Coulomb one, the assumption that N equals zero implies that $M = (1 - a)^{-1}h(r)$, where $h(r)$ is some function of r only. Substituting this form in Eqs. (2.13) and (2.8), one obtains the relations:

$$[r^2(d^2/dr^2) + (3 - f)r(d/dr) + (1 - ipr)(1 - f)$$

$$+ (pr)^2 + (\lambda rV)^2 - 2\nu pr(rV)]h(r) = 0, \tag{3.8}$$

$$G = (V)^{-1}[(d/dr) + r^{-1} - ip]h(r). \tag{3.9}$$

But G cannot be a function of r alone, as implied by Eq. (3.9), since its asymptotic form must provide a plane-wave contribution $\exp i(\mathbf{p}\cdot\mathbf{r} + \delta_p)$ which has angular dependence. Thus, for any central potential other than the Coulomb one, N cannot be identically zero.

REDUCTION TO THE COULOMB CASE

The discussion will now be restricted to the case of a Coulomb potential. Thus, $V = -1/r$ and $f = r[(d/dr)\ln rV] = 0$. The sign of the potential V has been chosen negative so that positive interaction strength λ corresponds to attractive scattering. The function N may now be chosen to be identically zero, and Eq. (2.12) is satisfied without imposing any restrictions on the angular dependence

of M. The remaining independent relations among G, M, and L [Eqs. (2.8), (2.9), and (2.13)] become

$$G = e^{ipr}[-(1 - a)r(\partial/\partial r)$$

$$+ (1 - a^2)(\partial/\partial a) - 2]e^{-ipr}M, \tag{4.1}$$

$$L = -e^{ipr}[r(\partial/\partial r) + (1 - a)(\partial/\partial a)]e^{-ipr}M, \tag{4.2}$$

$$\{\Theta(r, a) + (\lambda/r)^2\}[r(1 - a)]^{\frac{1}{2}}M = 0, \tag{4.3}$$

where $\Theta(r, a)$ is the self-adjoint differential operator

$$\Theta(r, a) = \{r^{-2}[(\partial/\partial r)r^2(\partial/\partial r) + (\partial/\partial a)(1 - a^2)(\partial/\partial a)$$

$$- \tfrac{1}{2}(1 - a)^{-1} - 2ipr(\tfrac{1}{2} + i\nu)] + p^2\}. \tag{4.4}$$

Thus, the problem of finding the scattering solution reduces to solving a single second-order partial-differential equation for M only, subject to the appropriate boundary conditions.

With the understanding that the operand is independent of the azimuthal angle ϕ, the operator $\Theta(r, a)$ can be written

$$\Theta(r, a) = \{\nabla^2 - [2r^2(1 - a)]^{-1}$$

$$- 2ip(\tfrac{1}{2} + i\nu)r^{-1} + p^2\}, \tag{4.5}$$

where ∇^2 is the Laplacian operator. Thus Eq. (4.3) is a Schrödinger equation, but the "potential" is such that it is separable only in spherical coordinates.[8]

In terms of the parabolic coordinates $\xi_1 = ipr(1 + a)$, $\xi_2 = ipr(1 - a)$, the differential relations for G, M, and L are:

$$G = [\xi_2 - 2 - 2\xi_2(\partial/\partial\xi_2)]M, \tag{4.6}$$

$$L = (\xi_1 + \xi_2)[\tfrac{1}{2} - \partial/\partial\xi_1)]M, \tag{4.7}$$

$$[\xi_1(\partial^2/\partial\xi_1^2) + \xi_2(\partial^2/\partial\xi_2^2) + (\partial/\partial\xi_1) + 2(\partial/\partial\xi_2)$$

$$- \tfrac{1}{4}(\xi_1 + \xi_2) - (\tfrac{1}{2} + i\nu) + \lambda^2(\xi_1 + \xi_2)^{-1}]M = 0. \tag{4.8}$$

It is seen that the function M is not separable in parabolic coordinates because of the term in λ^2. However, as a zero-order approximation, the term in λ^2 can be neglected. The resulting solution, characterized by M_0, that satisfies the scattering boundary conditions is the single eigenfunction $M_0(\xi_1, \xi_2)$

$$= -\tfrac{1}{2}\Gamma(1 - i\nu)e^{\nu\pi/2}e^{(\xi_1 - \xi_2)/2} {}_1F_1(1 + i\nu, 2, \xi_2). \tag{4.9}$$

Using Eqs. (4.6) and (4.7), one obtains the auxiliary results

$$G_0(\xi_1, \xi_2) = \Gamma(1 - i\nu)e^{\nu\pi/2}e^{(\xi_1 - \xi_2)/2} {}_1F_1(i\nu, 1, \xi_2), \tag{4.10}$$

$$L_0(\xi_1, \xi_2) = 0. \tag{4.11}$$

---

[8] L. P. Eisenhart, Phys. Rev. **74**, 87 (1948).

This zero-order approximation is the well known Sommerfeld–Maue approximation to the Dirac–Coulomb problem. In the preceding equations, $_1F_1$ is the regular confluent hypergeometric function and the normalization for M_0 has been chosen so the asymptotic plane wave derivable from G_0 corresponds to unit incident flux.

THE COULOMB GREEN'S FUNCTION

The partial differential equation [Eq. (4.3)] for $[r(1-a)]^{\frac{1}{2}}M(\mathbf{r})$ together with the boundary condition that asymptotically $M(\mathbf{r})$ behaves like a plane wave plus an outgoing spherical wave, can be replaced by the equivalent integral equation

$$[r(1-a)]^{\frac{1}{2}}M(\mathbf{r}) = [r(1-a)]^{\frac{1}{2}}M_0(\mathbf{r})$$
$$- \lambda^2 \int G(\mathbf{r},\mathbf{r}')(r')^{-2}[r'(1-a')]^{\frac{1}{2}}M(\mathbf{r}')\,d\mathbf{r}'. \quad (5.1)$$

Here, $M_0(\mathbf{r})$ is that solution (Eq. 4.9) of the homogeneous equation $\Theta(r,a)[r(1-a)]^{\frac{1}{2}}M_0(\mathbf{r}) = 0$ which asymptotically behaves like a plane wave plus an outgoing spherical wave. The Green's function $G(\mathbf{r},\mathbf{r}')$, which like $M(\mathbf{r})$ is independent of the azimuthal angle ϕ, must satisfy the partial differential equation

$$\Theta(r,a)G(\mathbf{r},\mathbf{r}') = (2\pi)^{-1}r^{-2}\delta(r-r')\delta(a-a'). \quad (5.2)$$

It must also satisfy the boundary conditions of regularity at $r = 0$ and asymptotic behavior of an outgoing spherical wave at $r \to \infty$.

In this formulation, the differential operator $\Theta(r,a)$ rather than $\Theta(r,a)[r(1-a)]^{\frac{1}{2}}$ is considered since the former is self-adjoint while the latter is not. In the construction of a Green's function, eigenfunctions of the relevant operator and also of the adjoint operator are required. Thus, the simplification of considering a self-adjoint operator is that only one set of eigenfunctions has to be used.

Since the inhomogeneous term of the integral equation involves the factor $(\lambda/r')^2$, it is most convenient to develope the Green's function in terms of spherical coordinates. For this purpose, one considers the solutions of the differential equation

$$\Theta(r,a)Q(r,a) = 0, \quad (5.3)$$

which are separable in spherical coordinates. These eigenfunctions have the form

$$Q_k(r,a) = r^{-1}\mathcal{R}_k(r)\mathcal{A}_k(a), \quad (5.4)$$

where the angular function $\mathcal{A}_k(a)$ and the radial function $\mathcal{R}_k(r)$ satisfy the equations

$$[(\partial/\partial a)(1-a^2)(\partial/\partial a) - \tfrac{1}{2}(1-a)^{-1}$$
$$- \tfrac{1}{4} + k^2]\mathcal{A}_k(a) = 0, \quad (5.5)$$

$$[(\partial^2/\partial r^2) + p^2 + 2ip(-\tfrac{1}{2} - i\nu)r^{-1}$$
$$+ (\tfrac{1}{4} - k^2)r^{-2}]\mathcal{R}_k(r) = 0. \quad (5.6)$$

The regular solution of the separated angular equation is

$$\mathcal{A}_k(a) = [(1-a)/(2k)]^{\frac{1}{2}}(d/da)[P_k(a) + P_{k-1}(a)],$$
$$k = 1, 2, \cdots, \quad (5.7)$$

where P_k is the Legendre polynomial of order k. These solutions possess the orthogonality and completeness relations:

$$\int_{-1}^{1} \mathcal{A}_k(a)\mathcal{A}_{k'}(a)\,da = \delta_{k,k'}, \quad (5.8)$$

$$\sum_{k=1}^{\infty} \mathcal{A}_k(a)\mathcal{A}_k(a') = \delta(a-a'), \quad -1 \le a, a' \le 1. \quad (5.9)$$

The differential equation for the radial function $\mathcal{R}_k(r)$ is just the Whittaker equation in which the independent variable is $2ipr$. The solution that is regular at the origin is the Whittaker function[9]

$$\mathcal{R}_k(r) = M_{-\frac{1}{2}-i\nu,k}(2pre^{i\pi/2})$$
$$= (2pre^{i\pi/2})^{\frac{1}{2}+k}e^{-ipr}$$
$$\times {}_1F_1(k+1+i\nu, 2k+1, 2pre^{i\pi/2}). \quad (5.10)$$

The solution that asymptotically behaves like an outgoing spherical wave is the other Whittaker function

$$\mathcal{R}_k(r) = W_{\frac{1}{2}+i\nu,k}(2pre^{-i\pi/2}),$$
$$= [\Gamma(k-i\nu)]^{-1}(2pre^{-i\pi/2})^{\frac{1}{2}-k}e^{ipr}$$
$$\times \int_0^\infty e^{-t}t^{k-i\nu-1}(2pre^{-i\pi/2} + t)^{k+i\nu}\,dt. \quad (5.11)$$

This has the asymptotic form

$$\lim_{r\to\infty} W_{\frac{1}{2}+i\nu,k}(2pre^{-i\pi/2}) = (2pre^{-i\pi/2})^{\frac{1}{2}+i\nu}e^{ipr}. \quad (5.12)$$

The Wronskian of these two solutions of the radial equation is

$$[M_{-\frac{1}{2}-i\nu,k}(x)](d/dx)[W_{\frac{1}{2}+i\nu,k}(xe^{-i\pi})]$$
$$- [W_{\frac{1}{2}+i\nu,k}(xe^{-i\pi})](d/dx)[M_{-\frac{1}{2}-i\nu,k}(x)]$$
$$= e^{i\pi(\frac{1}{2}-k)}\Gamma(2k+1)/\Gamma(k-i\nu), \quad (5.13)$$

where $x = 2pre^{i\pi/2}$.

The Green's function $G(\mathbf{r},\mathbf{r}')$ is now constructed by weighting the terms of the angular delta function with radial functions satisfying the boundary

---

[9] The notation for the Whittaker functions that is adopted here follows L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge University Press, Cambridge, England, 1960). The properties of these functions quoted here are developed in that reference.

conditions and having a discontinuity in slope at $r = r'$.[10] Thus, the Green's function is given by

$$G(\mathbf{r}, \mathbf{r}') = \sum_{k=1}^{\infty} c_k (rr')^{-1} \mathcal{C}_k(a) \mathcal{C}_k(a')$$

$$\times \begin{cases} M_{-\frac{1}{2}-i\nu,k}(2pre^{i\pi/2}) W_{\frac{1}{2}+i\nu,k}(2pr'e^{-i\pi/2}) \\ \qquad\qquad \text{for } r < r' \\ W_{\frac{1}{2}+i\nu,k}(2pre^{-i\pi/2}) M_{-\frac{1}{2}-i\nu,k}(2pr'e^{i\pi/2}) \\ \qquad\qquad \text{for } r > r'. \end{cases} \quad (5.14)$$

The constant c_k is determined by the condition

$$\int \Theta(r, a) G(\mathbf{r}, \mathbf{r}') \, d\mathbf{r} = 1, \quad (5.15)$$

which follows from Eq. (5.2). By integrating over the infinitesimal region $r' - \epsilon < r < r' + \epsilon$ for which the Green's function has a discontinuous slope, and using the Wronskian relation and the completeness of the $\mathcal{C}$'s, one finds that

$$c_k = \Gamma(k - i\nu) e^{-i\pi(k+1)} [4\pi p \Gamma(2k + 1)]^{-1}. \quad (5.16)$$

Now that the Green's function has been determined, the function $M_0(\mathbf{r}')$ given in Eq. (4.9) can be substituted for $M(\mathbf{r}')$ in the integral of Eq. (5.1) in order to obtain the first order of an iterative expansion in λ^2. For this purpose, it is convenient to express $M_0(\mathbf{r}')$ in terms of spherical coordinates instead of in terms of parabolic coordinates. By referring to FWH, it is found that $M_0(\mathbf{r})$ is given by the expansion

$$[r(1 - a)]^{\frac{1}{2}} M_0(\mathbf{r}) = e^{i\pi/4} e^{\nu\pi/2} (2p)^{-1} \sum_{k=1}^{\infty} (k/p)^{\frac{1}{2}}$$

times $[\Gamma(k - i\nu)/\Gamma(2k + 1)] r^{-1} M_{-\frac{1}{2}-i\nu,k}(2pre^{i\pi/2}) \mathcal{C}_k(a)$.
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.17)$$

---

[10] This technique is discussed by P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 825 ff.

Consequently,

$$\int G(\mathbf{r}, \mathbf{r}')(r')^{-2}[r'(1 - a')]^{\frac{1}{2}} M_0(\mathbf{r}') \, d\mathbf{r}'$$

$$= -\tfrac{1}{2}(p)^{-\frac{3}{2}} e^{i\pi/4} e^{\nu\pi/2} \sum_{k=1}^{\infty} (k)^{\frac{1}{2}}(-1)^k$$

$$\times [\Gamma(k - i\nu)/\Gamma(2k + 1)]^2 r^{-1} \mathcal{C}_k(a)$$

$$\times [W_{\frac{1}{2}+i\nu,k}(2pre^{-i\pi/2}) I_1(k, \nu; 2pr)$$

$$+ M_{-\frac{1}{2}-i\nu,k}(2pre^{i\pi/2}) I_2(k, \nu; 2pr)], \quad (5.18)$$

where

$$I_1(k, \nu; 2pr) = \lim_{\epsilon \to 0^+} \int_0^{2pr} e^{-\epsilon y} [y^{-1} M_{-\frac{1}{2}-i\nu,k}(ye^{i\pi/2})]^2 \, dy,$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.19)$$

$$I_2(k, \nu; 2pr) = \lim_{\epsilon \to 0^+} \int_{2pr}^{\infty} e^{-\epsilon y} y^{-2} W_{\frac{1}{2}+i\nu,k}$$

$$\times (ye^{-i\pi/2}) M_{-\frac{1}{2}-i\nu,k}(ye^{i\pi/2}) \, dy. \quad (5.20)$$

For applications to potential scattering, the asymptotic value for large r is of interest. In this case, $I_2(k, \nu; \infty) = 0$ and the definite integral defining $I_1(k, \nu; \infty)$ can be evaluated[11] in terms of known functions. This result is

$$I_1(k, \nu; \infty) = (-1)^k e^{-i\pi/2} e^{-\nu\pi} \Gamma(2k) \Gamma(2k + 1)$$

$$\times [\Gamma(k + 1 + i\nu) \Gamma(k - i\nu)]^{-1}$$

$$\times [i\pi + \psi_1(k + 1 + i\nu) - \psi_1(k - i\nu)], \quad (5.21)$$

where ψ_1 is the derivative of the logarithm of the gamma function. The λ^2 contribution to $M(\mathbf{r})$ for asymptotic r obtained in this fashion by use of the Green's function is in complete agreement with the Taylor's expansion of the exact solution developed in FWH.

---

[11] D. M. Fradkin, Ph.D. thesis, Iowa State University, Ames, Iowa, 1963 (unpublished).

Complex Lorentz Group with a Real Metric: Group Structure*

A. O. BARUT

Physics Department, University of Colorado, and National Bureau of Standards, Boulder, Colorado
(Received 3 March 1964; final manuscript received 21 May 1964)

In the attempts to connect the Lorentz group and the internal symmetry group of fundamental particles, a 16-parameter connected, noncompact group of rank 4 is studied in detail. The subgroup structure, Lie algebra and its complex extension (which is A_3 in Cartan's notation), little groups, the inhomogeneous groups, and the group invariants are discussed.

I. INTRODUCTION

IN previous notes[1,2] we have discussed the ways in which the quantum numbers of the real inhomogeneous Lorentz group (mass, spin) appear to be coupled to the internal quantum numbers, and the desirability of considering larger groups which contain the real Lorentz group as well as other internal quantum numbers. In this connection we study in this paper the mathematical properties of the complex Lorentz group with a real metric. We are interested mainly in the real form of the group. The complex extension of this group, as pointed out in the Appendix, is SU_4. However, as is known, the complex extension of a Lie group does not determine its various real forms, and, to our knowledge, the real form of the complex Lorentz group has not been discussed in the literature. Furthermore, the detailed relations obtained here, will be used when this group is considered as a possible exact symmetry group of elementary particle interactions.[2]

II. THE GROUP

The group under study is the complex Lorentz group with a real metric: the set of complex transformations Λ in a four-dimensional space satisfying the condition[3]

$$\Lambda^\dagger G \Lambda = G, \tag{II.1}$$

where G is the metric matrix

$$G = \begin{bmatrix} +1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix}$$

The transformations act on a space of complex 4-vectors z^μ with an invariant norm

$$|z|^2 \equiv z^\mu z_\mu = |z^0|^2 - |z^1|^2 - |z^2| - |z^3|^2 \tag{II.2}$$

which is always real. In contrast to this, the complex Lorentz group L used in the analytic continuation of mass shell amplitudes[4] has the invariant metric

$$z^2 = z^{0^2} - z^{1^2} - z^{2^2} - z^{3^2} \tag{II.3}$$

and satisfies the equation

$$L^T G L = G. \tag{II.4}$$

The group (II.1) is intended to connect the space-time and internal quantum numbers of elementary particles. Whereas the group (II.4) is isomorphic to a direct product of two 2-by-2 unimodular groups, the structure of the group (II.1) is much more complex.[5]

First we discuss the tensor calculus within the group and the definition of dotted and undotted, upper and lower indices.

Together with Λ we must also consider the transformations

$$\Lambda, \ \Lambda^{T^{-1}}, \ \Lambda^*, \ \Lambda^{\dagger^{-1}}.$$

All these representations satisfy the same group property $\Lambda^\dagger G \Lambda = G$.

We denote the indices of Λ as follows:

$$z_\mu' = \Lambda_\mu{}^\nu z_\nu.$$

It follows from (II.1) that Λ and $\Lambda^{\dagger^{-1}}$, and Λ^* and $\Lambda^{T^{-1}}$ are equivalent:

$$\Lambda = G\Lambda^{\dagger^{-1}}G, \tag{II.5}$$

$$\Lambda^* = G\Lambda^{T^{-1}}G. \tag{II.6}$$

[1] A. O. Barut, Nuovo Cimento 32, 234 (1964).
[2] A. O. Barut, Phys. Rev. 135, B839 (1964); in *Proceedings of the Coral Gables Conference on Symmetry* (W. H. Freeman and Company, San Francisco, 1964).
[3] The superscripts †, *, and T stand for Hermitian conjugate, complex conjugate, and transpose of a matrix, respectively.
[4] See, for example, R. Jost in *Theoretical Physics in the 20th Century*, edited by M. Fierz and V. Weisskopf (Interscience Publishers, Inc., New York, 1960).
[5] For a discussion of this 12-parameter complex Lorentz group (II.4) (plus 8 parameters of translations) from the point of view of internal quantum numbers, see A. O. Barut, in *Symposium on Lorentz Group* (University of Colorado Press, Boulder, Colorado, 1964).

But Λ and $\Lambda*$ are not. There is no fixed matrix C such that

$$\Lambda = C\Lambda*C;$$

in mathematical language, there is no inner automorphism of this form. We denote the tensors transforming under $\Lambda*$ by dotted indices

$$z'_{\dot\mu} = \Lambda_{\dot\mu}{}^{\dot\nu}z_{\dot\nu}(\Lambda_{\dot\mu}{}^{\dot\nu} \equiv \Lambda^*{}_\mu{}^\nu).$$

The invariant form can now be written as

$$z^{\dot\mu}z_{\dot\mu} = g_{\dot\mu\nu}z^{\dot\mu}z^\nu \qquad (\text{II.7})$$

or

$$g_{\dot\mu\nu}z'^{\dot\mu}z'^\nu = g_{\dot\mu\nu}\Lambda^{\dot\mu}{}_{\dot\rho}\Lambda^\nu{}_\sigma z^{\dot\rho}z^\sigma,$$

where

$$g_{\mu\dot\nu}\Lambda^{\dot\mu}{}_{\dot\rho}\Lambda'{}^\nu{}_\sigma = g_{\dot\rho\sigma} \mid \Lambda^{\dot\mu}{}_{\dot\rho}\Lambda_{\dot\mu\sigma} = g_{\dot\rho\sigma}$$

which is equivalent to (II.1). The elements of G have always mixed indices:

$$\Lambda_\mu{}^\nu = g_{\mu\dot\nu}\Lambda^{\dagger-1}{}^\nu{}_\sigma g^{\sigma\dot\nu}, \qquad (\text{II.5}')$$

with

$$\Lambda_{\dot\mu}{}^{\dot\nu} = g_{\dot\mu\nu}\Lambda^{T-1}{}^\nu{}_\sigma g^{\sigma\dot\nu}, \qquad (\text{II.6}')$$

$$g_{\dot\mu\nu}g^{\sigma\dot\nu} = \delta_{\dot\mu}{}^{\dot\nu}; \quad g_{\mu\dot\sigma}g^{\dot\sigma\nu} = \delta_\mu{}^\nu.$$

We have then the indices as follows:

$$\Lambda \quad : \Lambda_\mu{}^\nu \text{ transforming } z_\mu,$$
$$\Lambda^* \quad : \Lambda_{\dot\mu}{}^{\dot\nu} \quad \prime\prime \qquad z_{\dot\nu},$$
$$\Lambda^{\dagger-1} : \Lambda^\mu{}_\nu \quad \prime\prime \qquad z^\mu,$$
$$\Lambda^{\dagger-1} : \Lambda^{\dot\mu}{}_{\dot\nu} \quad \prime\prime \qquad z^{\dot\mu}.$$

III. GROUP PROPERTIES

A. Connectedness

The Λ group is connected, that is, all group elements can be reached by continuously varying the parameters of the group. From the defining equation (II.1) we get

$$\det \Lambda = e^{i\phi} \qquad (\text{III.1})$$

which is continuous. In particular it connects the four pieces of the real Lorentz group: I, P, T and PT. For example, the following element of Λ

$$\begin{bmatrix} e^{i\phi_1} & 0 & 0 & 0 \\ 0 & e^{i(\phi_2-\phi_1)} & 0 & 0 \\ 0 & 0 & \cos\phi_3 & \sin\phi_3 \\ 0 & 0 & -\sin\phi_3 & \cos\phi_3 \end{bmatrix}, \qquad (\text{III.2})$$

where ϕ_1, ϕ_2 and ϕ_3 are real parameters, connects continuously I, P, T, and PT.[6]

Together with Λ, $\Lambda e^{i\phi}$ belongs also to the group. We can therefore write

$$\Lambda = \Lambda_+ \otimes e^{i\phi}, \qquad (\text{III.3})$$
$$\det \Lambda_+ = 1.$$

In other words, Λ has a one-dimensional (Abelian) invariant subgroup. We show that Λ_+ is *simple*. The group is, however, not simply connected, because the path $e^{i\phi}$ cannot be shrunk to zero.

B. Infinitesimal Generators and Subgroups

If we write

$$\Lambda = e^\omega, \qquad (\text{III.4})$$

we obtain for infinitesimal ω

$$\omega^\dagger G = G\omega \qquad (\text{III.5})$$

or

$$\omega^*_{\sigma\mu}g_{\sigma\nu} = -g_{\mu\sigma}\omega_{\sigma\nu}$$

or

$$\omega_{00}{}^* = -\omega_{00},$$
$$\omega_{k0} = \omega_{0k}{}^*, \qquad (\text{III.6})$$
$$\omega_{ik}{}^* = -\omega_{ki}.$$

The infinitesimal generators are either Hermitian or skew Hermitian. There are 16 parameters in ω. We can choose the 16 infinitesimal generators as follows:

$$1. \ \omega_{ik}^* = -\omega_{ki}$$

Real matrices must be antisymmetric; pure imaginary ones symmetric. We take

$$R_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ R_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \ R_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

and

$$U_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ U_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \end{pmatrix}, \ U_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix}$$

and

$$C_1 = \begin{pmatrix} 0 & i & & \\ & & 0 & \\ & & & 0 \end{pmatrix}, \ C_2 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & i & 0 \end{pmatrix}, \ C_3 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & i \end{pmatrix}.$$

$$2. \ \omega_{ok} = \omega_{ko}^*$$

Real matrices are now symmetric and pure

_____

[6] The complex Lorentz group (II.4) connects only the pieces I and PT together and P and T together (see Ref. 4).

Table I. The commutation relations.

	R_1	R_2	R_3	L_1	L_2	L_3	M_1	M_2	M_3	U_1	U_2	U_3	C_1	C_2	C_3	C
R_1	0	R_3	$-R_2$	0	L_3	$-L_2$	0	M_3	$-M_2$	A	U_3	$-U_2$	$-U_1$	U_1	0	0
R_2		0	R_1	$-L_3$	0	L_1	$-M_3$	0	M_1	$-U_3$	B	U_1	0	$-U_2$	U_2	0
R_3			0	L_2	$-L_1$	0	M_2	$-M_1$	0	$-U_2$	U_1	D	$-U_3$	0	U_3	0
L_1				0	$-R_3$	R_2	a	U_3	$-U_2$	0	$-M_3$	M_2	0	0	M_1	$-M_1$
L_2					0	$-R_1$	U_3	b	$-U_1$	$-M_3$	0	M_1	M_2	0	0	$-M_2$
L_3						0	$-U_2$	$-U_1$	c	$-M_2$	$-M_1$	0	0	M_3	0	$-M_3$
M_1							0	$-R_3$	R_2	0	L_3	$-L_2$	0	0	$-L_1$	L_1
M_2								0	$-R_1$	L_3	0	$-L_1$	$-L_2$	0	0	L_2
M_3									0	L_2	L_1	0	0	$-L_3$	0	L_3
U_1										0	$-R_3$	$-R_2$	R_1	$-R_1$	0	0
U_2											0	$-R_1$	R_1	0	R_2	$-R_2$
U_3												0	R_3	0	$-R_3$	0
C_1													0	0	0	0
C_2														0	0	0
C_3															0	0
C																0

$$a = 2(C_3 - C), \ b = 2(C_1 - C), \ c = 2(C_2 - C)$$
$$A = 2(C_1 - C_2), \ B = 2(C_2 - C_3), \ D = 2(C_1 - C_3)$$

imaginary ones are antisymmetric. We choose

$$L_1 = \begin{pmatrix} 0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&0 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 0&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}, \quad L_3 = \begin{pmatrix} 0&0&-1&0\\0&0&0&0\\-1&0&0&0\\0&0&0&0 \end{pmatrix}$$

and

$$M_1 = \begin{pmatrix} 0&0&0&i\\0&0&0&0\\0&0&0&0\\-i&0&0&0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0&i&0&0\\-i&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 0&0&-i&0\\0&0&0&0\\i&0&0&0\\0&0&0&0 \end{pmatrix}$$

3. $\omega_{00} = \omega_{00}^{*}$

We take

$$C = \begin{pmatrix} i&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}.$$

There are four mutually commuting diagonal generators C, C_1, C_2, C_3. The group is of rank 4. The generator C commutes with all other generators; it belongs to the one-dimensional invariant Abelian subgroup. The remaining 15 generators form also a Lie algebra under the operation of commutation.

The three generators R_1, R_2, R_3 represent a rotation group, R_i and L_i ($i = 1, 2, 3$) span the real Lorentz group, and the six generators R_i, M_i span a particular subgroup of complex Lorentz transformations which is isomorphic to the real Lorentz group. None of these subgroups is an invariant subgroup.

Consider the subgroup of Λ of the form

$$\begin{pmatrix} a & 0 \\ 0 & B \end{pmatrix}.$$

We have then

$$a = e^{i\Psi} \quad \text{and} \quad B^{+}B = I;$$

Thus B is the 3×3 unitary group U_3. In Λ_+ we can choose $a = 1$, then $\det B = 1$. Hence, the subgroup

$$\begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix}$$

is the enlarged (to four dimensions) SU_3 group. The infinitesimal generators R_i, U_i and the two traceless combinations of C_i span SU_3. The commutation relations are shown in Table I and the subgroup structure schematically in Fig. 1.

The complex Lie algebra of Λ_+ and that of SU_4 are identical (see Appendix I). The complex Lie algebra can be brought easily into the Cartan's form,[7] and the roots can be determined by forming the combinations $R_j \pm iU_j$ and $L_j \pm iM_j$.[2,8]

From structure constants it can be shown that Λ_+ is semisimple.[8]
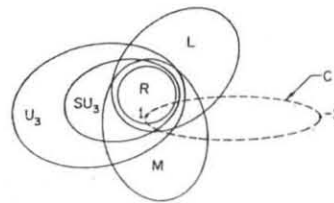
Fig. 1. Intersections of the subgroups.

[7] See G. Racah, "Group Theory and Spectroscopy," Institute for Advanced Study Lecture Notes (1951) (reprinted CERN 61-3); W. Pauli, "Continuous Groups in Quantum Mechanics," Lecture Notes, CERN-31; A. Salam, in *Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963).

[8] See a much enlarged version of this paper—University of Colorado preprint (unpublished)—in which other forms of the generators and the invariants are given.

IV. INHOMOGENEOUS GROUP AND INVARIANTS

We now consider inhomogeneous infinitesimal transformations of the form

$$z'_\mu = z_\mu + (\alpha_\mu{}^\nu + i\beta_\mu{}^\nu)z_\nu + a_\mu + ib_\mu, \qquad (IV.1)$$

where $\alpha_{\mu\nu}$ is real antisymmetric and $\beta_{\mu\nu}$, real symmetric:

$$\alpha_{\mu\nu} = -\alpha_{\nu\mu}, \qquad \beta_{\mu\nu} = \beta_{\nu\mu}. \qquad (IV.2)$$

The group property implies the following relations for the composition of the parameters of two infinitesimal transformations:

$$\begin{aligned}
\alpha_{\mu\nu} &= \alpha'_{\mu\nu} + \alpha''_{\mu\nu} + \alpha'_{\mu\sigma}\alpha''^\sigma{}_\nu - \beta'_{\mu\sigma}\beta''^\sigma{}_\nu, \\
\beta_{\mu\nu} &= \beta'_{\mu\nu} + \beta''_{\mu\nu} + \beta'_{\mu\sigma}\alpha''^\sigma{}_\nu + \alpha'_{\mu\sigma}\beta''^\sigma{}_\nu,
\end{aligned} \qquad (IV.3)$$

and

$$\begin{aligned}
a_{\mu\nu} &= a'_\mu + a''_\mu + \alpha'_{\mu\sigma}a''^\sigma - \beta'_{\mu\sigma}b''^\sigma, \\
b_\mu &= b'_\mu + b''_\mu + \beta'_{\mu\sigma}a''^\sigma + \alpha'_{\mu\sigma}b''^\sigma.
\end{aligned} \qquad (IV.3')$$

We write the representations of these infinitesimal transformations in the form

$$U = 1 - \tfrac{1}{2}\alpha^{\mu\nu}M_{\mu\nu} + \tfrac{1}{2}i\beta^{\mu\nu}N_{\mu\nu} + ia^\mu k_\mu + ib^\mu h_\mu,$$

where $M_{\mu\nu}$ is an antisymmetric set of real matrices and $N_{\mu\nu}$ is a symmetric set of pure imaginary matrices. The connection of these generators to those discussed in previous sections is the following:

$$M_{\mu\nu} = \begin{bmatrix} 0 & L_2 & -L_3 & L_1 \\ & 0 & R_1 & R_3 \\ & & 0 & R_2 \\ & & & 0 \end{bmatrix} = -M_{\nu\mu}, \qquad (IV.4)$$

$$N_{\mu\nu} = \begin{bmatrix} X_{00} & M_2 & -M_3 & M_1 \\ & X_{11} & U_1 & U_3 \\ & & X_{22} & U_2 \\ & & & X_{33} \end{bmatrix} = N_{\nu\mu}. \qquad (IV.5)$$

From the group property of the representations, $U = U'U''$, and the Eqs. (V.2) and (V.3), we obtain the following commutation relations:

$$\begin{aligned}
[M_{\mu\nu}, M_{\sigma\rho}] = &-g_{\nu\sigma}M_{\mu\rho} - g_{\mu\rho}M_{\nu\sigma} \\
&+ g_{\mu\sigma}M_{\nu\rho} + g_{\nu\rho}M_{\mu\sigma},
\end{aligned}$$

$$\begin{aligned}
[N_{\mu\nu}, N_{\sigma\rho}] = &\; g_{\nu\sigma}M_{\mu\rho} + g_{\mu\rho}M_{\nu\sigma} \\
&+ g_{\mu\sigma}M_{\nu\rho} + g_{\nu\rho}M_{\mu\sigma}, \qquad (IV.6)
\end{aligned}$$

$$\begin{aligned}
[M_{\mu\nu}, N_{\sigma\rho}] = &-g_{\nu\sigma}N_{\mu\rho} + g_{\mu\rho}N_{\nu\sigma} \\
&+ g_{\mu\sigma}N_{\nu\rho} - g_{\nu\rho}N_{\mu\sigma},
\end{aligned}$$

which agree with the previous form of the commutators of the homogeneous group, and

$$\begin{aligned}
[M_{\mu\nu}, k_\sigma] &= g_{\mu\sigma}k_\nu - g_{\sigma\nu}k_\mu, \\
[M_{\mu\nu}, h_\sigma] &= g_{\mu\sigma}h_\nu - g_{\sigma\nu}h_\mu, \\
[N_{\mu\nu}, k_\sigma] &= -i(g_{\mu\sigma}h_\nu + g_{\sigma\nu}h_\mu), \\
[N_{\mu\nu}, h_\sigma] &= i(g_{\mu\sigma}k_\nu + g_{\sigma\nu}k_\mu),
\end{aligned} \qquad (IV.7)$$

and

$$[k_\mu, k_\nu] = [h_\mu, h_\nu] = [k_\mu, h_\nu] = 0. \qquad (IV.8)$$

In terms of these new generators the invariant of the homogeneous group is given by

$$F^2 = \tfrac{1}{2}(M_{\mu\nu}M^{\nu\mu} - N_{\mu\nu}N^{\nu\mu}). \qquad (IV.9)$$

Now we can evaluate the invariants of the inhomogeneous group. From the first two equations we get

$$[M_{\mu\nu}, k^2] = [M_{\mu\nu}, h^2] = 0,$$

but k^2 and h^2 do not commute with $N_{\mu\nu}$. From the last two equations we obtain

$$[N_{\mu\nu}, k^2 + h^2] = 0. \qquad (IV.10)$$

Hence $k^2 + h^2$ is the first invariant of the inhomogeous group as might be expected from the invariant norm discussed in Sec. II.

The other two invariants of the inhomogeneous group are given by[9]

$$\begin{aligned}
C_1 = &\,(k^2 + h^2)N_\mu{}^\mu - N_{\mu\nu}(k^\mu k^\nu + h^\mu h^\nu) \\
&+ M_{\mu\nu}(k^\mu h^\nu - h^\nu k^\mu) \qquad (IV.11)
\end{aligned}$$

and

$$\begin{aligned}
C_2 = &\,(k^2 + h^2)\{(M_{\nu\alpha}k^\alpha + N_{\nu\alpha}h^\alpha)(M^{\nu\beta}k_\beta + N^{\nu\beta}h_\beta) \\
&+ (M_{\nu\alpha}h^\alpha - N_{\nu\alpha}k^\alpha)(M^{\nu\beta}h_\beta - N^{\nu\beta}k_\beta)\} \\
&- \tfrac{1}{2}\{M_{\mu\nu}(k^\mu h^\nu - k^\nu h^\mu) - N_{\mu\nu}(k^\mu k^\nu + h^\mu h^\nu)\}^2 \\
&- \tfrac{1}{2}(k^2 + h^2)^2(M_{\mu\nu}M^{\mu\nu} + N_{\mu\nu}N^{\mu\nu}). \qquad (IV.11')
\end{aligned}$$

V. LITTLE GROUPS

The concept of "little group" arises in the representation theory of the inhomogeneous group.[10] The little group of a "momentum" vector $p = k + ih$ is defined as the subgroup of the transformations Λ_p which leave p invariant:

$$\Lambda_p p = p. \qquad (V.1)$$

If we write, infinitesimally, $\Lambda_p = 1 + \omega_p$, we have

$$\omega_p p = 0. \qquad (V.2)$$

---

[9] This has been found independently by Dr. Y. Murai (private communication).
[10] E. P. Wigner, Ann. Math. 40, 149 (1939).

TABLE II. Commutation relations for the little
group for $M_0^2 = 0$.[a]

	a	b	c	d	e	R_1	U_1	f
a	0	0	0	0	d	b	$-c$	$-d$
b		0	0	0	c	$-a$	$-d$	c
c			0	0	$-b$	$-d$	a	$-b$
d				0	$-a$	c	b	a
e					0	0	0	0
R_1						0	f	$-U_1$
U_1							0	R_1
A								0

[a] The generator $M - C + C_3$ commutes with all the above eight and forms a one-dimensional invariant subgroup. The other four mutually commuting generators are $a = L_2 + R_1$, $b = L_2 - R_3$, $c = M_2 + U_3$, $d = M_3 - U_2$. Furthermore, $e = C_1 + C_2$, $f = (C_1 - C_2)/2$. The generators R_1, U_1, f form a rotation group.

It is then easy to identify the generators of the little group. Because the norm $p^2 = k^2 + h^2 = M_0^2$ is invariant, the little groups may be classified according to the values of this invariant:

(1) $M_0^2 > 0$: p can be brought to the form $(z^0, 0, 0, 0)$, and, consequently, by (V.2) the little group is the group U_3 consisting of the infinitesimal generators $\mathbf{R}$, $\mathbf{U}$, and $\mathbf{C}$.

(2) $M_0^2 < 0$: p can be brought to the form $(0, 0, 0, z^0)$, and the little group consists of the generators R_1, U_1, L_2, L_3, M_2, M_3, C_1, C_2, and C.

It is to be noted that the complex extension of Cartan's form for these two little groups is the same.

(3) $M_0^2 = 0$: the standard form of p is now $a(1, 0, 0, 1)$. The little group is now again a nine-parameter group, but is now of rank 5 instead of 3 as in the two previous cases. It consists of the infinitesimal generators: R_1, U_1, C_1, C_2 (i.e., the group U_2), $L_2 - R_3$, $R_2 + L_3$, $M_2 + U_3$, $M_3 - U_2$ and $M_1 - C + C_3$.

The commutation relations of this group is shown in Table II.

We also introduce the little "groups" with respect to the real part of p, the actual linear momentum k, i.e., transformations which leave k invariant:

$$\Lambda_k(k + ih) = (k + ih'). \qquad \text{(V.3)}$$

These transformations have the usual little groups of the real inhomogeneous group, i.e., part of the set $M_{\mu\nu}$ in (IV.4) and the whole of $N_{\mu\nu}$ in (IV.5). For $k^2 > 0$, we have the generators $\mathbf{R}$ and $N_{\mu\nu}$; for $k^2 = 0$, the set L_2, L_3, R_1, and $N_{\mu\nu}$. The transformations (V.3) do not form a Lie group unless in the limit $\mathbf{L} = 0$ in the case of $k^2 > 0$, and in the limit $L_1 = R_2 = R_3 = 0$ in the case $k^2 = 0$. In these limits the little "groups" are of rank 4.

APPENDIX I: $\Lambda_\mp$ GROUP AND SU_4

The infinitesimal generators ω of the unitary group in four dimensions, $U^\dagger U = I$, satisfy the relation $\omega^\dagger = -\omega$, or $\omega_{\mu\nu}^* = -\omega_{\mu\nu}$. The 16 generators can be chosen, therefore, in the notation of Sec. III. B, to be

$$\mathbf{R}, \mathbf{U}, \mathbf{C}, -i\mathbf{M} = \mathbf{M}', \quad i\mathbf{L} = \mathbf{L}', \quad \text{and} \quad C.$$

Because of the factors i in $\mathbf{M}$ and $\mathbf{L}$, the commutation relations of these generators, and hence the real Lie algebra, are quite different. However, as far as the complex Lie algebra, e.g., Cartan's form, is concerned, we can form exactly the same linear combinations with complex coefficients replacing $\mathbf{M} = i\mathbf{M}'$ and $\mathbf{L} = -i\mathbf{L}'$ everywhere. Thus the complex extensions of the Lie algebras of the two groups Λ and U_4 (or Λ_+ and SU_4) coincide. This is an example of the known fact that to a given complex Lie algebra there correspond, in general, more than one real Lie algebra.[11]

[11] L. Pontrayagin, *Topological Groups* (Princeton University Press, Princeton, New Jersey, 1958), p. 265.

Algebra of Dirac Bilinears

R. PENNEY

Scientific Laboratory, Ford Motor Company, Dearborn, Michigan
(Received 9 June, 1964)

All of the possible quadratic relations among the Dirac bilinear covariants which one may construct using both the usual bispinor and its charge-conjugate are given. Many interesting purely algebraic results are found for a general Dirac field.

I. INTRODUCTION

THE tensor bilinears which one may construct from the Dirac γ matrices and the Dirac bispinor ψ are well known.[1] Such bilinears are the ingredients for constructing interaction Lagrangians, for example. The bilinears also occur in any analysis of the hydrodynamics of the electron field,[2] or, indeed, in any physical theory of the spinor particles. Yet it is not generally recognized that the bilinear quantities are not algebraically independent.

Some of the relations, quadrilinear in ψ, which occur among the bilinears have been found by Pauli,[3] and by Kofink.[4] These relations have been used in the Vigier[2] theory of elementary particles. A very limited use of the Pauli–Kofink relations has also occurred in an attempted[5] geometric theory of neutrinos.

It is the author's belief that any relations existing among the Dirac bilinears are of importance for physical understanding of the Dirac field, and for this reason, a completely exhaustive study of such relations has been made. In the present paper, we present all possible quadrilinear combinations of the usual Dirac bilinears, and of those bilinears one may construct with the use of the charge-conjugate field. We will not attempt any physical interpretation which may follow from such relations, but merely point out that any proposed neutrino theory of light or geometric theory of neutrinos must use the given relations to some extent.

We will use Minkowski coordinates as is usual, and our notation will be that used by Roman[1] in his book.

II. USUAL BILINEARS

The Dirac equation may be written as

$$\gamma_\mu \partial_\mu \psi = -\kappa \psi$$

with

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\delta_{\mu\nu}.$$

The adjoint spinor $\bar{\psi}$ is defined by

$$\bar{\psi} \equiv \psi^+ \gamma_4$$

as usual. The tensor bilinears one may construct are

$$S = \bar{\psi}\psi, \qquad\qquad P = i\bar{\psi}\gamma_5\psi,$$

$$V_\mu = i\bar{\psi}\gamma_\mu\psi, \qquad\qquad A_\mu = i\bar{\psi}\gamma_5\gamma_\mu\psi,$$

$$T_{\mu\nu} = \frac{i}{2}\bar{\psi}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)\psi, \quad *T_{\mu\nu} = \frac{i}{2}\bar{\psi}\gamma_5(\gamma_\mu\gamma_\mu - \gamma_\nu\gamma_\mu)\psi,$$

where γ_5 is $\gamma_1\gamma_2\gamma_3\gamma_4$ and $*T_{\mu\nu}$ is the dual tensor which may also be expressed by

$$*T_{\mu\nu} = -\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}T_{\alpha\beta}.$$

As Pauli[3] and Kofink[4] showed, there exist certain algebraic relations among these bilinears. For example, we have that

$$V_\mu A_\mu = 0,$$

$$V_\mu V_\mu = -A_\mu A_\mu.$$

III. CHARGE-CONJUGATE BILINEARS

As is well known, the charge-conjugate bispinor, defined by

$$\psi^c \equiv C\bar{\psi}^T, \qquad -\gamma_\mu^T = C^{-1}\gamma_\mu C,$$

obeys the Dirac equation. Using this new bispinor, we may construct new bilinears. For example we may construct

$$S^{(1)} \equiv \bar{\psi}^c\psi,$$

which would be a new scalar. Or we could construct

$$V_\mu^{(2)} \equiv i\bar{\psi}\gamma_\mu\gamma^c$$

a new vector.

In general, we denote a bilinear constructed with $\bar{\psi}^c$ and ψ by the superscript (1); a bilinear using $\bar{\psi}$, ψ^c by the superscript (2). A bilinear which

[1] P. Roman, *Theory of Elementary Particles* (Interscience Publishers, Inc., New York, 1961), 2nd ed., p. 112.
[2] L. de Broglie, *Vigier Theory of Elementary Particles* (Elsevier Publishing Company, New York, 1963).
[3] W. Pauli, Ann. Inst. Henri Poincaré **6**, 109 (1936).
[4] W. Kofink, Ann. Physik **30**, 91 (1937); **37**, 421 (1940).
[5] A. Inomata, Bull. Am. Phys. Soc. **9**, 86 (1964).

TABLE I. A list of all quadratic relations among the Dirac bilinears.

	V_ν	A_ν	$T_{\rho\nu}$	$*T_{\rho\nu}$	$V_\nu^{(1)}$	$V_\nu^{(2)}$	$T_{\rho\nu}^{(1)}$	$T_{\rho\nu}^{(2)}$	$*T_{\rho\nu}^{(1)}$	$*T_{\rho\nu}^{(2)}$
V_ν	$-(S^2+P^2)$	0	PA_ρ	iSA_ρ	0	0	$iSV_\rho^{(1)}$	$-iSV_\rho^{(2)}$	$PV_\rho^{(1)}$	$-PV_\rho^{(2)}$
A_ν	0	S^2+P^2	PV_ρ	iSV_ρ	0	0	$PV_\rho^{(1)}$	$PV_\rho^{(2)}$	$-iSV_\rho^{(1)}$	$iSV_\rho^{(2)}$
$T_{\mu\nu}$	PA_μ	PV_μ	$V_\mu V_\rho - A_\mu A_\rho + S^2\delta_{\mu\rho}$	$-iSP\delta_{\mu\rho}$	$-iSV_\mu^{(1)}$	$iSV_\mu^{(2)}$	b	b	d	d
$*T_{\mu\nu}$	iSA_μ	iSV_μ	$-iSP\delta_{\mu\rho}$	a	$-PV_\mu^{(1)}$	$PV_\mu^{(2)}$	c	c	e	e
$V_\nu^{(1)}$	0	0	$-iSV_\rho^{(1)}$	$-PV_\rho^{(1)}$	0	$2(S^2+P^2)$	0	$-2iSV_\rho$ $-2PA_\rho$	0	$-2iSA_\rho$ $-2PV_\rho$
$V_\nu^{(2)}$	0	0	$iSV_\rho^{(2)}$	$PV_\rho^{(2)}$	$2(S^2+P^2)$	0	$2iSV_\rho$ $-2PA_\rho$	0	$-2iSA_\rho$ $+2PV_\rho$	0
$T_{\mu\nu}^{(1)}$	$iSV_\mu^{(1)}$	$PV_\mu^{(1)}$	b	c	0	$2iSV_\mu$ $-2PA_\mu$	$V_\mu^{(1)} V_\rho^{(1)}$	f	0	h
$T_{\mu\nu}^{(2)}$	$-iSV_\mu^{(2)}$	$PV_\mu^{(2)}$	b	c	$-2iSV_\mu$ $-2PA_\mu$	0	f	$V_\mu^{(2)} V_\rho^{(2)}$	i	0
$*T_{\mu\nu}^{(1)}$	$PV_\mu^{(1)}$	$-iSV_\mu^{(1)}$	d	e	0	$-2iSA_\mu$ $+2PV_\mu$	0	i	a	g
$*T_{\mu\nu}^{(2)}$	$-PV_\mu^{(2)}$	$iSV_\mu^{(2)}$	d	e	$-2PV_\mu$ $-2iSA_\mu$	0	h	0	g	a

a. $T_{\mu\nu}T_{\rho\nu} + *T_{\mu\nu}*T_{\rho\nu} = \frac{1}{2}(T_{\alpha\beta}T_{\alpha\beta})\delta_{\mu\rho}$.
b. $T_{\mu\nu}T_{\rho\nu}^{(a)} + T_{\mu\nu}^{(a)}T_{\rho\nu} = V_\mu V_\rho^{(a)} + V_\rho V_\mu^{(a)}$.
c. $*T_{\mu\nu}T_{\rho\nu}^{(a)} + T_{\mu\nu}^{(a)}*T_{\rho\nu} = A_\mu V_\rho^{(a)} + A_\rho V_\mu^{(a)}$.
d. $T_{\mu\nu}*T_{\rho\nu}^{(a)} + *T_{\mu\nu}^{(a)}T_{\rho\nu} = -A_\mu V_\rho^{(a)} - A_\rho V_\mu^{(a)}$.
e. $*T_{\mu\nu}*T_{\rho\nu}^{(a)} + *T_{\mu\nu}^{(a)}*T_{\rho\nu} = -V_\mu V_\rho^{(a)} - V_\rho V_\mu^{(a)}$.
f. $T_{\mu\nu}^{(1)}T_{\rho\nu}^{(2)} + T_{\mu\nu}^{(2)}T_{\rho\nu}^{(1)} = 2V_\mu V_\rho + 2A_\mu A_\rho + 2(P^2 - S^2)\delta_{\mu\rho}$.
g. $*T_{\mu\nu}^{(1)}*T_{\rho\nu}^{(2)} + *T_{\rho\nu}^{(1)}*T_{\mu\nu}^{(2)} = -2V_\mu V_\rho - 2A_\mu A_\rho + 2(P^2 - S^2)\delta_{\mu\rho}$.
h. $T_{\mu\nu}^{(1)}*T_{\rho\nu}^{(2)} + T_{\rho\nu}^{(1)}*T_{\mu\nu}^{(2)} = 2A_\mu V_\rho + 2A_\rho V_\mu + 4iSP\delta_{\mu\rho}$.
i. $T_{\mu\nu}^{(2)}*T_{\rho\nu}^{(1)} + T_{\rho\nu}^{(2)}*T_{\mu\nu}^{(1)} = -2A_\mu V_\rho - 2A_\rho V_\mu + 4iSP\delta_{\mu\rho}$.
a = 1, 2

uses $\bar\psi^c$, ψ^c will be denoted by a superscript (C). By simple calculation, we easily see that

$$S^{(1)} = S^{(2)} = P^{(1)} = P^{(2)} = A_\mu^{(1)} = A_\mu^{(2)} = 0,$$

and that the bilinears using both $\bar\psi^c$ and ψ^c are the same, except for a sign, as the original bilinears. We thus have only 6 new bilinears, viz.,

$$V_\mu^{(1)}, \quad V_\mu^{(2)},$$
$$T_{\mu\nu}^{(1)}, \quad T_{\mu\nu}^{(2)},$$
$$*T_{\mu\nu}^{(1)}, \quad *T_{\mu\nu}^{(2)}.$$

In all, therefore, we have 12 bilinear quantities. Two of these are scalars, and we thus have 10 bilinear vectors or antisymmetric tensors to combine. These are therefore at most 55 quadrilinear expressions obtainable by tensor contraction.

IV. QUADRATIC RELATIONS

We have calculated all of the possible combinations of the bilinears and give our results in Table I. Insofar as was possible, the relations have been subjected to internal consistency checks. Thus, any errors in the table could only consist of sign errors, which have been eliminated as nearly as possible.

In constructing the charge-conjugate quantities, we have made a choice for the C matrix (which is not unique) which must be noted. If we use the spinorial representation, the defining equation for C gives the result

$$C = a\gamma_2\gamma_4, \quad a^*a = 1,$$

and we have taken the value $a = +1$.

We also must point out that neither $V_\mu^{(1)}$ nor $V_\mu^{(2)}$ has the reality conditions of a velocity vector in Minkowski space, but the combinations

$$V_\mu^{(1)} + V_\mu^{(2)}, \quad i(V_\mu^{(1)} - V_\mu^{(2)})$$

do have. That is, the spatial components are Hermitian, with the time component anti-Hermitian. Similar remarks apply to the tensors $T_{\mu\nu}^{(1)}$, $T_{\mu\nu}^{(2)}$.

Of particular interest on the table are those entities which vanish. We see that $V_\mu^{(1)}$, $V_\mu^{(2)}$ are null vectors which are orthogonal to both V_μ and A_μ. Similarly, both $T_{\mu\nu}^{(1)}$ and $T_{\mu\nu}^{(2)}$ are null tensors.

Some of the physical consequences of these purely algebraic relations will be reported in a later paper.

Two Classes of New Conservation Laws for the Electromagnetic Field and for Other Massless Fields

Thomas A. Morgan*

Department of Physics, Syracuse University, Syracuse, New York
(Received 9 June 1964)

The existence of a class of conserved tensors $T_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m}$ and the existence of a class of conserved tensor densities $V_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m}$ is exhibited for the electromagnetic field. They are differential generalizations of the energy momentum tensor in the sense that they are bilinear in E and B, contain $n + m$ derivatives and are symmetric, trace free and divergenceless on their $\mu\nu$ pair of indices. Corresponding conserved quantities for the two-component neutrino field, linear gravitational field and indeed for all massless free fields are also exhibited.

INTRODUCTION

RECENTLY, Lipkin has established a new conservation law for the electromagnetic field.[1] He showed the existence of a third-rank tensor density $Z_{\mu\nu\alpha}$ which is a bilinear function of **E** and **B** containing one derivative and which is conserved in the sense that $Z_{\mu\nu}{}^{\alpha}{}_{,\alpha} = 0$, where the comma denotes differentiation and thus for bounded fields

$$\frac{d}{dt} \int Z_{\mu\nu0} \, dx^3 = 0. \tag{1}$$

In this article, a class of conserved tensors and a class of conserved tensor densities is exhibited. Dr. Lipkin's expression is simply related to one of the conserved tensor densities.

TENSOR CONSERVATION LAWS

The Maxwell energy momentum tensor $T_{\mu\nu}$ can be written

$$T_{\mu\nu} = -\tfrac{1}{2}(F_{\mu\alpha}F^{\alpha}{}_{\nu} + F_{\mu\alpha}{}^{*}F^{\alpha}{}_{\nu}{}^{*}), \tag{2}$$

where $F_{\mu\nu}{}^{*} = \tfrac{1}{2}F^{\alpha\beta}\epsilon_{\mu\nu\alpha\beta}$ is the dual of the electromagnetic field strengths $F_{\mu\nu}$. The identity[2]

$$A_{\mu\sigma}B^{\sigma}{}_{\nu} - A_{\nu\sigma}{}^{*}B^{\sigma}{}_{\mu}{}^{*} = \tfrac{1}{2}g_{\mu\nu}(A_{\rho\sigma}B^{\sigma\rho}) \tag{3}$$

implies that $T_{\mu\nu}$ is symmetric in $\mu\nu$. If the energy momentum tensor is generalized to

$$T_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m} = -\tfrac{1}{2}(F_{\mu\sigma,\alpha_1\cdots\alpha_n}F^{\sigma}{}_{\nu,\beta_1\cdots\beta_m}$$
$$+ F_{\mu\sigma,\alpha_1\cdots\alpha_n}{}^{*}F^{\sigma}{}_{\nu,\beta_1\cdots\beta_m}{}^{*}), \tag{4}$$

then the same identity implies that its generalization is also symmetric in $\mu\nu$. The identity

$$A_{\rho\sigma}B^{\rho\sigma} = -A_{\rho\sigma}{}^{*}B^{\rho\sigma}{}^{*} \tag{5}$$

implies that both $T_{\mu\nu}$ and its generalization are trace free on the $\mu\nu$ pair of indices. The proof that $T_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m}$ is divergence free on $\mu\nu$, is a trivial extension of the proof that $T_{\mu\nu}$ is divergence free and is obtained by using Maxwell's equations in the form

$$F^{\mu\nu}{}_{,\nu} = 0; \qquad F^{\mu\nu}{}^{*}{}_{,\nu} = 0 \tag{6}$$

to show that

$$T^{\mu}{}_{\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m,\mu} = -\tfrac{1}{2}(F^{\mu}{}_{\sigma,\alpha_1\cdots\alpha_n}F^{\sigma}{}_{\nu,\mu\beta_1\cdots\beta_m}$$
$$+ F^{\mu}{}_{\sigma,\alpha_1\cdots\alpha_n}{}^{*}F^{\sigma}{}_{\nu,\mu\beta_1\cdots\beta_m}{}^{*}), \tag{7}$$

and then by using Maxwell's equations in the equivalent form

$$F_{\mu\nu,\sigma} + F_{\nu\sigma,\mu} + F_{\sigma\mu,\nu} = 0$$
$$= F_{\mu\nu,\sigma}{}^{*} + F_{\nu\sigma,\mu}{}^{*} + F_{\sigma\mu,\nu}{}^{*} \tag{8}$$

to show that

$$T^{\mu}{}_{\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m,\mu} = +\tfrac{1}{4}(F_{\mu\sigma,\alpha_1\cdots\alpha_m}F^{\mu\sigma}{}_{,\nu\beta_1\cdots\beta_m}$$
$$+ F_{\mu\sigma,\alpha_1\cdots\alpha_n}{}^{*}F^{\mu\sigma}{}_{,\nu\beta_1\cdots\beta_m}{}^{*}); \tag{9}$$

but this expression is zero by the identity Eq. (5). The indices $\alpha_1 \ldots \alpha_n; \beta_1 \ldots \beta_m$ are essentially spectators and this allows one to generalize the energy momentum tensor in a simple and perhaps trivial manner. In general these new conserved quantities cannot be expressed as derivatives of the energy momentum tensor and they are thus in this sense independent of it. When a current J^{μ} is present then

$$T^{\mu}{}_{\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m,\mu} = \tfrac{1}{2}(J_{\sigma,\alpha_1\cdots\alpha_n}F^{\sigma}{}_{\nu,\beta_1\cdots\beta_m}$$
$$+ J_{\sigma,\beta_1\cdots\beta_m}F^{\sigma}{}_{\nu,\alpha_1\cdots\alpha_n}). \tag{10}$$

MASSLESS FREE FIELDS

In the Pauli–Fierz formulation of massless free fields, a field of spin s is represented by the completely

* Present Address: University of Nebraska, Lincoln, Nebraska.
[1] D. M. Lipkin, J. Math. Phys. 5, 698 (1964).
[2] Where $A^{\mu\nu}$, $B^{\rho\sigma}$ are antisymmetric tensors, and $g_{\mu\nu}$ is the flat-space metric.

symmetric spinor $\phi_{A_1 A_2 \cdots A_{2s}}$ with $2s$ indices, and the field equations are

$$\partial^{A\dot{B}}\phi_{AA_2 \cdots A_n} = 0; \qquad \partial^{A\dot{B}}\phi_{\dot{B}\dot{B}_2 \cdots \dot{B}_n} = 0, \qquad (11)$$

where $\partial^{A\dot{B}}$ is the derivative operator in spinor form. The current of the two-component neutrino field is $\phi_A \phi_{\dot{B}} + \text{H.c.}$ where ϕ_A is the field strength and H.c. denotes the Hermitian conjugate. It is obvious that Eq. (11) implies that the current is conserved. The expression for the current of the two-component neutrino may be generalized to give a conservation law for any Pauli–Fierz field. Consider

$$\phi_{A_1 \cdots A_{2s}} \phi_{\dot{B}_1 \cdots \dot{B}_{2s}} + \text{H.c.} \qquad (12)$$

It is a completely symmetric and a completely trace-free tensor of rank $2s$ which is divergenceless on all indices. In particular the expression in Eq. (12) is for the electromagnetic field just the Maxwell energy momentum tensor.

The expression for the current for the two-component neutrino can also be generalized to give derivative conservation laws. Consider

$$\phi_{A,\alpha_1 \cdots \alpha_n} \phi_{\dot{B},\beta_1 \cdots \beta_m} + \text{H.c.} \qquad (13)$$

It is a tensor of rank of $n + m + 1$ which is divergenceless on the tensor index corresponding to the pair of spin indices A, $\dot{B}$. Both generalizations can be performed simultaneously to yield the derivative conservation law exhibited in Eq. (14) for a field of spin s.

$$\phi_{A_1 \cdots A_{2s}, \alpha_1 \cdots \alpha_m}\phi_{\dot{B}_1 \cdots \dot{B}_{2s}, \beta_1 \cdots \beta_m} + \text{H.c.} \qquad (14)$$

For the case of the electromagnetic field, the expression in Eq. (14) and the expression in Eq. (4) are identical.

There is a second class of conservation laws for the two-component neutrino. The energy momentum tensor of the two-component neutrino

$$i(\phi_A \partial_{C\dot{D}}\phi_{\dot{B}} - \text{H.c.}) \qquad (15)$$

is also conserved. The expression in Eq. (15) is again simply generalized to the arbitrary Pauli–Fierz field of spin s

$$i(\phi_{A_1 \cdots A_{2s}}\partial_{C\dot{D}}\phi_{\dot{B}_1 \cdots \dot{B}_{2s}} - \text{H.c.}). \qquad (16)$$

It can also be generalized by adding spectator derivatives to yield

$$i(\phi_{A,\alpha_1 \cdots \alpha_n}\partial_{C\dot{D}}\phi_{B,\beta_1 \cdots \beta_m} - \text{H.c.}). \qquad (17)$$

Combining the generalizations made in Eq. (16) and Eq. (17), one obtains the conserved tensor densities

$$i(\phi_{A_1 \cdots A_{2s},\alpha_1 \cdots \alpha_n}\partial_{C\dot{D}}\phi_{\dot{B}_1 \cdots \dot{B}_{2s},\beta_1 \cdots \beta_m} - \text{H.c.}). \qquad (18)$$

CONSERVED TENSOR DENSITIES

The tensor equivalent of Eq. (18) is for the electromagnetic field

$$V_{\mu\nu\alpha_1 \cdots \alpha_n\beta_1 \cdots \beta_m} = F_{\mu\sigma,\alpha_1 \cdots \alpha_n}F^{\sigma}{}_{\nu,\beta_1 \cdots \beta_n}^{*}$$
$$- F_{\mu\sigma,\alpha_1 \cdots \alpha_n}^{*}F^{\sigma}{}_{\nu,\beta_1 \cdots \beta_n}. \qquad (19)$$

This tensor density can be seen to be symmetric in μ and ν by means of the identity in Eq. (3). It is more convenient for this purpose to write the identity in Eq. (3) in the form

$$A_{\mu\sigma}B^{\sigma}{}_{\nu}^{*} + A_{\nu\sigma}^{*}B^{\sigma}{}_{\mu} = \tfrac{1}{2}g_{\mu\nu}(A_{\rho\sigma}^{*}B^{\sigma\rho}). \qquad (20)$$

$V_{\mu\nu\alpha_1 \cdots \alpha_n\beta_1 \cdots \beta_m}$ is traceless in the $\mu\nu$ pair of indices because of the identity

$$A_{\rho\sigma}^{*}B^{\rho\sigma} = A_{\rho\sigma}B^{\rho\sigma*}. \qquad (21)$$

The proof that $V_{\mu\nu\alpha_1 \cdots \alpha_n\beta_1 \cdots \beta_m}$ is divergenceless in $\mu\nu$ and hence represents a conserved quantity has exactly the same steps as were performed in Eqs. (6) through (9) except now the identity in Eq. (21) is used in place of identity in Eq. (5).

A particularly interesting conserved tensor density is $V_{\mu\nu\alpha}$ which can be shown to be trace free and divergenceless on all three indices. $V_{\mu\nu\alpha}$ is essentially the conserved tensor density discovered by Dr. Lipkin, except that it can be easily shown to be divergenceless on all indices, a fact not mentioned by him. It is interesting that all the conserved tensor expressions and all of the conserved tensor density expressions are, like the Maxwell energy momentum tensor, invariant under duality rotations.

Note added in proof: The existence of an infinite number of conserved quantities for a free field follows from the fact that the number of quanta associated with each mode of a free field is constant in time.