JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 5, NUMBER 12 DECEMBER 1964

Theorem Relating the Eigenvalue Density for Random Matrices to the Zeros of the Classical Polynomials

BURT V. BRONK
*Palmer Physical Laboratory, Princeton University, Princeton, New Jersey**
(Received 12 May 1964)

A theorem due to Stielties shows that the problem of locating the zeros of the classical polynomials is equivalent to finding the electrostatic equilibrium positions for a set of interacting point charges. Defining a density function for these same charges we find that the density of zeros for the nth Hermite polynomial is the same as the density of eigenvalues for an ensemble of n-dimensional Hermitian matrices. Similarly the location of the zeros of the nth Laguerre polynomial determines the density of eigenvalues for an ensemble of n-dimensional positive matrices, and the zeros of the $\frac{1}{2}n$th Tchebichef polynomial determine the density for the real part of the eigenvalues for an ensemble of n-dimensional unitary matrices.

A MATRIX ensemble is defined by giving a class of matrices to be considered, and a measure on the sample space of those matrices. The eigenvalue distributions arising from the Gaussian ensemble[1] for Hermitian matrices, and the exponential ensemble[2] for positive matrices relate to the present paper. Rather than repeat their definitions, however, we define here two somewhat different ensembles which we might call spherical. The study of these led directly to the present result. Over the class of all Hermitian matrices, we define the probability of finding a matrix with elements near the elements of a given matrix H, as

$$\begin{aligned} D(H) &= C_1\, dH \quad \text{for} \quad \mathrm{Tr}\,(H^2) \leq \tfrac{1}{2}n^2, \\ &= 0 \qquad \text{for} \quad \mathrm{Tr}\,(H^2) > \tfrac{1}{2}n^2. \end{aligned} \tag{1}$$

The ijth element of H is given by

$$h_{ij} \equiv r_{ij} + is_{ij}, \qquad r_{ij} = r_{ji}, \quad s_{ij} = -s_{ji}, \tag{2}$$

and

$$dH = dr_{11}\, dr_{12} \cdots dr_{nn}\, ds_{12} \cdots ds_{(n-1)n}. \tag{3}$$

Call this ensemble I. Although (1) leads to a semicircle law, one should notice that we do not have independence of elements which was required in those ensembles Wigner studied.[1c]

Now we define a different ensemble over the class of all positive Hermitian matrices, any member of which can be written[2]

$$P = (A^{\dagger}A). \tag{4}$$

A is a complex matrix. In this case our sample space is the set of all n-dimensional complex matrices. The ijth element of one of these complex matrices is given by two real numbers, a_{ij} and b_{ij}, and is written $a_{ij} + ib_{ij}$. If we take as our volume element, the product of differentials for all $2n^2$ real and imaginary components,

$$dA = da_{11}\, da_{12}\, da_{21} \cdots da_{nn}\, db_{11}\, db_{12} \cdots db_{nn}, \tag{5}$$

then we take as the probability of finding an n-

* Present address: Physics Department, State University of New York at Stony Brook, Stony Brook, New York.

[1] (a) E. P. Wigner, "Distribution Laws for Roots of a Random Hermitian Matrix" (unpublished). The pertinent results and definitions are available in (b) N. Rosenzweig, "Brandeis Summer Institute 1962, Statistical Physics" (W. A. Benjamin, Inc., New York, 1963). (c) E. P. Wigner, Ann. Math. **67**, 325 (1958).

[2] B. V. Bronk, "Exponential Ensemble for Random Matrices," J. Math. Phys. (to be published).

dimensional positive matrix which can be formed as in (4) from a complex matrix A' with elements in the neighborhood of elements of A, the following:

$$\begin{aligned} D(P)\,dA &\equiv C_2[\det(A^\dagger A)]^{\alpha+1}\,dA \\ &\quad \text{for } \operatorname{Tr}(A^\dagger A) \le n(n+\alpha-1), \\ &= 0 \quad \text{for } \operatorname{Tr}(A^\dagger A) > n(n+\alpha-1). \end{aligned} \tag{6}$$

Call this ensemble II. C_1 and C_2 are normalization constants for the two ensembles.

Using the results of Refs. 1 and 2, we can transform to the representation given by eigenvalues and eigenvectors, and integrating out the latter, we obtain the joint eigenvalue distribution functions.

Ensemble I:

$$\begin{aligned} &P(\lambda_1, \cdots \lambda_n)\,d\lambda_1 \cdots d\lambda_n \\ &\quad = C_1' \prod_{i<j} (\lambda_i - \lambda_j)^2\,d\lambda_1 \cdots d\lambda_n \text{ for } \sum_i \lambda_i^2 \le \tfrac{1}{2}n^2 \\ &\quad = 0 \qquad\qquad \text{for } \sum_i \lambda_i^2 > \tfrac{1}{2}n^2\,. \end{aligned} \tag{7}$$

Ensemble II:

$$\begin{aligned} &P(\kappa_1, \cdots \kappa_n)\,d\kappa_1 \cdots d\kappa_n \\ &\quad = C_2' \prod_i \kappa_i^{\alpha+1} \prod_{i<j} (\kappa_i - \kappa_j)^2\,d\kappa_1 \cdots d\kappa_n \\ &\qquad \text{for } \sum \kappa_i \le n(n+\alpha-1) \\ &\quad = 0 \quad \text{for } \sum \kappa_i > n(n+\alpha-1). \end{aligned} \tag{8}$$

The κ are the eigenvalues of P, and thus are positive.

We also will refer to the unitary ensemble of Dyson[3] which we will call ensemble III. The measure of this ensemble is given by the invariant group measure of the unitary group. The joint eigenvalue distribution function for this ensemble is given by

$$\begin{aligned} &P(e^{i\theta_1}, \cdots, e^{i\theta_n})\,d\theta_1 \cdots d\theta_n \\ &= C_3' \prod_{i<j} |e^{i\theta_i} - e^{i\theta_j}|^2\,d\theta_1 \cdots d\theta_n, \text{ for } 0 < \theta_i \le 2\pi. \end{aligned} \tag{9}$$

We can now state our result.

Theorem. If we define a density function $\sigma(x)$, which gives the approximate number of zeros per unit interval in the argument x, of the nth orthogonal polynomial, we find as n becomes large,

(1) The density of zeros of the nth Hermite polynomial,[4a] $H_n(x)$, is identical with the single eigenvalue density for ensemble I, for n-dimensional Hermitian matrices, and is given by

$$\begin{aligned} \sigma_{n\mathrm{I}}(x) &= (1/\pi)(2n - x^2)^{\frac{1}{2}}, \qquad x \le (2n)^{\frac{1}{2}} \\ &= 0, \qquad\qquad x > (2n)^{\frac{1}{2}}. \end{aligned} \tag{10}$$

(2) The density of zeros of the associated Laguerre polynomial,[4b] $L_n^\alpha(x)$, is identical with the single eigenvalue density for ensemble II, for positive Hermitian matrices and is given by

$$\begin{aligned} \sigma_{n\mathrm{II}}(x) &= (1/4\pi x)[-x^2 + (4n + 2\alpha - 2)x - (\alpha-1)^2]^{\frac{1}{2}} \\ &\qquad \text{for } \alpha^2/4n < x < 4n + 2\alpha \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{11}$$

The constant α is real, and we consider those α for which $1 \ll \alpha \ll n$.

(3) The density of zeros of the Tchebichef polynomials[4c] $T_{n/2}(x)$ [which is the Jacobi polynomial $P_{n/2}^{(-\frac{1}{2},-\frac{1}{2})}(x)$], is identical with the density of the real part of the eigenvalues for ensemble III, and is given by

$$\begin{aligned} \sigma_{n\mathrm{III}}(x) &= (n/\pi(1 - x^2)^{\frac{1}{2}}), \qquad |x| \le 1 \\ &= 0, \qquad\qquad |x| > 1. \end{aligned} \tag{12}$$

For the first two parts of the above theorem, we need a theorem due to Stielties, the proof of which is given by Szegö.[5] Only a few changes of constants are needed to bring it to the form stated here.

Theorem. Consider a system of n unit masses located at the variable points $\{x_1, x_2, \cdots, x_n\}$ in the interval $[-\infty, \infty]$, such that their moment of inertia satisfies

$$(\sum x_i^2/n) \le \tfrac{1}{2}(n-1), \tag{13}$$

then the unique maximum of the function

$$V(x_1, \cdots, x_n) = \prod_{i<j} (x_i - x_j)^2 \tag{14}$$

is attained if the $\{x_i\}$ are the zeros of the Hermite polynomial $H_n(x)$. For the Laguerre case, the x_i are allowed to vary on the interval $[0, \infty]$, and the average values of the positions are restricted by

$$(\sum x_i/n) \le n(n+\alpha). \tag{15}$$

In the second case, (14) is maximized if the x_i are the zeros of the Laguerre polynomial $L_n^\alpha(x)$.

It is clear that $-\log(V)$ may be considered to be the energy of a set of n point charges interacting among themselves with a logarithmic repulsive po-

[3] F. J. Dyson, J. Math. Phys. **3**, 140 (1962). Notice the results of the present paper apply to all cases of Dyson's classification of ensembles by transformation properties. The formulas here are stated for the case $\beta = 2$.

[4] *Bateman Manuscript Project*, edited by A. Erdléyi (McGraw-Hill Book Company, Inc., New York, 1953). See (a) Sec. 10.13; (b) Sec. 10.12; (c) Sec. 10.11.

[5] G. Szegö, Am. Math. Soc. Colloq. Publ., **23**, 139 (1959).

tential, and the theorem gives the locations for electrostatic equilibrium. At equilibrium, the equality[6] sign holds for (13) and (15), since if it were "less than," then we would be free to extend the outermost charges, increasing (14).

Suppose we now consider the set of charges as a charged fluid, for which a density $\sigma(x)$ may be defined. Then we assume as in classical statistical mechanics that the charges distribute themselves in such a way as to maximize the logarithm of the joint distribution function. This is equivalent to the usual statement that the overwhelming majority of states for a classical system are very near to the most probable state. Considering first the Hermite case, we must find the function $\sigma_I(x)$ which maximizes

$$I = \int_{-A}^{A} dx \int_{-A}^{A} dy\ \sigma_I(x)\sigma_I(y)\ln|x - y| \tag{16}$$

subject to

$$\int_{-A}^{A} \sigma_I(x)\,dx = n \tag{17}$$

and

$$\int_{-A}^{A} x^2\sigma_I(x)\,dx = \frac{n^2}{2} \tag{18}$$

and $\sigma_I(x) \geq 0$ everywhere.

Equation (18) is the relation equivalent to (13) for a charged fluid. The integrand in (16) gives minus the energy of repulsion for the charges near y, due to the charges near x, assuming that the density near y isn't affected by the density near x. That is, correlations are neglected.

Varying σ, and using Lagrange multipliers for the two constraining equations, we obtain

$$C_1 - 2C_2y^2 + 2\int dx\ \sigma(x)\ln|x - y| = 0, \tag{19}$$

which is the same integral equation solved by Wigner,[7] except that in this case Eqs. (17) and (18) give two simultaneous equations determining A and C_2. These are easily solved to give the constants in (10).

We can write out a set of three equations determining the level density for ensemble II.

$$I = \int_0^A dx \int_0^A dy\ \sigma_{II}(x)\sigma_{II}(y)\ln|x - y|, \tag{20}$$

$$\int_0^A \sigma_{II}(x)\,dx = n, \tag{21}$$

$$\int_0^A x\sigma_{II}(x)\,dx = n(n + \alpha - 1). \tag{22}$$

This leads to an integral equation which is the same as that which determines the most probable distribution for the exponential ensemble. We therefore have for σ_{II}, the same density function obtained by different means in Ref. 2. We only need to check that the function σ_{II}, which we give in (11), satisfies (21) and (22). These are integrals of the form

$$J = \oint x^\alpha(-Ax^2 + 2Bx - C)^{\beta/2}\,dx. \tag{23}$$

The general evaluation of these is given by Born,[8] and is written,

$$J = 2\pi i\{-\text{res}_0\,[x^\alpha(-Ax^2 + 2Bx - C)^{\beta/2}] + \text{res}_0\,[y^\alpha(-A + 2By - Cy^2)^{\beta/2}]\}, \tag{24}$$

where $y = 1/x$. Using (11) for σ_{II}, and evaluating the integrals by means of (24), we see that (21) and (22) are satisfied.

It is worth noting a previous result[9] which states that only one eigenvalue is contained in the region around and beyond $x = (2n)^{\frac{1}{2}}$, for the Gaussian ensemble. This indicates that the procedure of equating the density of the charged fluid with the average position of the point charges is valid even rather near this point, since the largest zero of $H_n(x)$ is located there.

Turning to ensemble III, the product in (9) is a maximum if the n charges are evenly spaced at intervals $(2\pi/n)$ apart around the unit circle.[3] Now the $\frac{1}{2}n$th Tchebichef polynomial is defined

$$T_{n/2}(x) = \cos[\tfrac{1}{2}n \arccos(x)]. \tag{25}$$

If x is $\cos(\theta)$, we see that $T_{n/2}(\cos(\theta))$ has zeros at

$$\theta = \pm\pi/n;\ \pm 3\pi/n;\ \pm 5\pi/n \cdots \tag{26}$$

so that the $\frac{1}{2}n$ zeros of $T_{n/2}$ are the (double) projections on some correctly chosen real axis, of n points evenly spaced around the unit circle.[10] For large n, the distance between two consecutive zeros is given by the derivative of the cosine, and the reciprocal of this gives the well-known density function (12).

[6] From this point on the proof also applies to Rosenzweig's "Fixed Strength" ensemble, see Ref. 1(b), p. 110.

[7] E. P. Wigner, *Proceedings of the Canadian Mathematical Congress*, (1954), p. 174 (unpublished).

[8] M. Born, *Mechanics of the Atom*, translated by Fisher and Hartree, (G. Bell and Sons, London, 1960), Appendix II.

[9] B. Bronk, J. Math. Phys. 5, 215 (1964).

[10] We have actually proved for ensemble III a special case of a much more general theorem, which can be stated roughly, that the zeros of a set of polynomials orthogonalized with respect to an *arbitrary* weight function on $[-1, 1]$, are cosines of angles distributed uniformly around the circle. See Ref. 5, Theorem 12.7.2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

If Microns Were Fermis*

H. C. CORBEN
Quantum Physics Laboratory, Physical Research Division
TRW/Space Technology Laboratories, Inc., Redondo Beach, California
(Received 21 May 1964)

Evidence is rapidly accumulating that elementary particle states are simply the rotational levels of a symmetrical top, as derived from relativistic quantum mechanics. Some of this evidence is presented for the case of zero-strangeness bosons. An analogy is drawn between this point of view and the nonrelativistic theory of the rotational levels of a rigid molecule.

IF the masses of elementary particles were a billion times as large as they are known to be, atomic and molecular dimensions would be decreased by the same ratio, and "infrared" quanta emitted by molecules would have energies measured in the hundreds of MeV.[1] Research with accelerators in the BeV range would now be revealing that molecules are not the immutable entities that they had been believed to be. A number of different "particles" (we call them molecular rotational levels) would have been observed, and we would have measured their various scattering cross-sections, selection rules, life-times, parities, angular momenta J, and energies E. Someone would have plotted the values of $(J + \frac{1}{2})^2$ against E, obtained several series of collinear points, and found that the straight lines so defined were parallel. By analytical continuation into the complex angular momentum plane the common slope of these Regge trajectories would have been related to the average radius R of the bound states of the various particles that had been discovered, through the formula[2]

$$\hbar^2 d(\alpha + \tfrac{1}{2})^2/dp^2 = R^2, \qquad (1)$$

where $J = \text{Re}\ (\alpha)$, $p^2 = 2mE$.

The common radius for all of the particles which were observed to transform into each other might have caused us to suspect that we were dealing with a single molecule of mass m, radius R, and moment of inertia $\sim mR^2$. If this were the case it would lead to an understanding of the conservation of "molecule number" in all observed reactions. It would also add support to the point of view that, since the various particles were excited states of the one fundamental entity, they could be regarded as forming a "democracy" in which no one of them appeared as more fundamental than any of the others. Some scientists, however, would have taken a more phenomenological approach, and developed an algebra based on the lowest observed levels in order to predict the existence of higher states. In particular, since experimentalists had observed only one $J = 0$ level, two $J = 1$ levels, and three $J = 2$ levels for a particular molecule, group theoretical arguments would have predicted that there would be four levels with $J = 3$, and would have yielded a formula, later verified experimentally, for the energies of these states.

As an example, Fig. 1 shows the Regge trajectories and the positions of the Regge poles for real integral values of J for the lowest rotational levels of the molecule NH_3. The common slope of the trajectories indicates a radius of order 1 Å (in real units). It is very satisfying that a radius of just this value is in agreement with data from high-energy (ultraviolet) scattering experiments.

Fortunately, it never occurred to scientists forty or fifty years ago that the molecular rotational levels could be anything but what they are, and they promptly wrote down the appropriate equations for a rotator in the Bohr–Sommerfeld theory and later in the "new wave mechanics," and found that the Schrödinger equation of a free top with moments of inertia A, A, C leads to the rotational levels[3]

$$E = \frac{\hbar^2}{2}\left[\frac{J(J+1)}{A} + \left(\frac{1}{C} - \frac{1}{A}\right)K^2\right]$$

$$J = 0, 1, 2, \cdots, \qquad K = 0, 1, 2, \cdots, J,$$

so that

$$(J + \tfrac{1}{2})^2 = 2EA/\hbar^2 + \tfrac{1}{4} - (\epsilon - 1)K^2 \qquad (2)$$

* Research supported by the STL Company Independent Research Program and in part by the Office of Naval Research.

[1] Of course, if all matter were made of these hypothetical particles, all but a few of us would be less than 20 Å tall, and the change of scale would pass unnoticed.

[2] T. Regge, Nuovo Cimento **14**, 951 (1959); **18**, 947 (1960).

[3] D. M. Dennison, Phys. Rev. **28**, 318 (1926).

where $\epsilon = A/C$. Thus

$$\hbar^2 d(J + \tfrac{1}{2})^2/dp^2 = A/m \qquad (3)$$

which is essentially the same as Eq. (1). For the various values of K it is clear that Eq. (2) defines a series of Regge trajectories for the symmetrical top—even in nonrelativistic wave mechanics such a top does not have a unique value for its spin or its energy. Furthermore, it is not necessary to postulate any interactions in order to define such a set of curves—a free rotating molecule has its own built-in Regge trajectories, the shape and slope of which depend on the radii of gyration of the molecule. Finally, once Eq. (2) is derived, it is not necessary to draw these curves anyhow!

We have stressed the fact that no potentials or bound states are required to derive Regge trajectories for a free symmetrical top in nonrelativistic wave mechanics, and the same conclusion is valid in relativistic wave mechanics. It is surprising that the relativistic generalization of the Schrödinger equation for a symmetrical top has been overlooked until recently.[4] All that we had in relativistic quantum mechanics was the generalization of Pauli spin theory, but whether added to the Schrödinger equation or deduced from the Dirac equation this will never lead to Regge trajectories, or states of different spin values—it is the special case $A \to \infty$, above, in which the rest energy is independent of the angular momentum.

The slopes of Regge trajectories imply a finite "bound-state" radius. A finite radius implies a finite moment of inertia, and this in its turn implies the existence of rotational levels, whether for a molecule or an elementary particle. Independently of this, the *observed* finite size of elementary particles warns us to expect rotational levels. We should be surprised if they were not there.

Generalized field theory is nothing but the relativistic quantum mechanics of the symmetrical top, in the same way that ordinary field theory (e.g., the Dirac theory of the electron) is the relativistic quantum theory of a spinning top in the limit $A \to \infty$. Neither relativity theory nor quantum mechanics demands this restraint, which in relativistic quantum theory has focused our attention on wave equations which are irreducible representations of the Lorentz group, each of which describes a particle with a unique mass and a unique spin. As pointed out by Chew and Frautschi,[5]

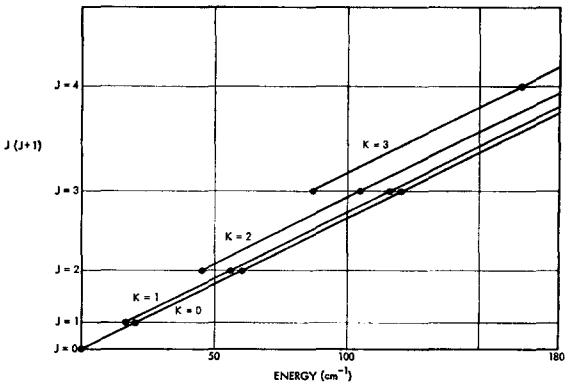


FIG. 1. Regge trajectories for lowest rotational levels of NH_3. The common slope corresponds to a moment of inertia $A = 2.816 \times 10^{-40}$ gm cm^2.

field theories of this type necessarily single out certain particles as fundamental, leaving others to be computed as complex aggregates, an artificial and very unsatisfactory distinction which has caused many physicists to abandon hope that relativistic field theories can lead to a meaningful description of elementary particle states.

Generalized field theory does not suffer from this defect. The Schrödinger equation for a symmetrical top is replaced in this theory by linear relativistic wave equations which are a generalization of the wave equations of standard field theory. They describe the rotational states of a particle in relativistic quantum theory and lead to hierarchies of states of different values of angular momentum and rest energy, just as in the nonrelativistic case.

The free-particle wave equation which we consider in this paper is the generalized Kemmer equation for bosons

$$[i\beta_\mu p_\mu + mc - m_0 c\beta_{\mu\nu}\beta'_{\mu\nu}]\psi = 0, \qquad (4)$$

where β_μ, β'_μ are commuting sets of Kemmer operators and $\beta_{\mu\nu} = (\beta_\mu, \beta_\nu)$, $\beta'_{\mu\nu} = (\beta'_\mu, \beta'_\nu)$. The relation between this equation and the relativistic classical mechanics of a symmetrical top has been discussed elsewhere.[6] It corresponds to the case $A = \hbar^2/2m_0c^2$.

The conserved charge-current density is assumed to be

$$j_\mu = \tfrac{1}{2}ie_0 c\bar{\psi}(1 + \eta'_5)\beta_\mu\psi, \qquad (5)$$

where

$$\bar{\psi} = \psi^*\eta_4\eta'_4,$$

$$\eta_\mu = 2\beta_\mu^2 - 1, \qquad \eta'_\mu = 2\beta'^2_\mu - 1, \qquad \eta'_5 = \eta'_1\eta'_2\eta'_3\eta'_4.$$

[4] H. C. Corben, Proc. Natl. Acad. Sci. U.S. **48**, 1559 (1962).

[5] G. F. Chew, and S. C. Frautschi, Phys. Rev. Letters **7**, 394 (1961).

[6] H. C. Corben, Phys. Rev. **131**, 2219 (1963).

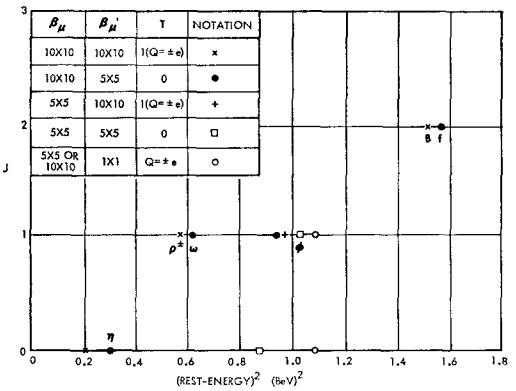


FIG. 2. Chew–Frautschi plot of the solutions of the equation $[i\beta_\mu p_\mu + mc - m_0 c\beta_{\mu\nu}\beta'_{\mu\nu}]\psi = 0$ for $m = 1041$ MeV, $m_0 = 106$ MeV.

The 4-vector

$$s_\mu = \bar{\psi}\beta_\mu\psi \tag{6}$$

is also conserved by Eq. (4). In the rest system, Eq. (4) becomes

$$[\beta_4 E - \epsilon - (\mathbf{\Sigma}\cdot\mathbf{\Sigma}' + \mathbf{\lambda}\cdot\mathbf{\lambda}')]\psi = 0, \tag{7}$$

where

$$W = 2m_0c^2E, \qquad \epsilon = 2m/m_0,$$

$$(\beta_{23}, \beta_{31}, \beta_{12}) = i\mathbf{\Sigma}, \qquad (\beta_{14}, \beta_{24}, \beta_{34}) = i\mathbf{\lambda}.$$

The solutions are characterized by the eigenvalues ± 1 of the commuting operators $\eta_4\eta'_4$ and η'_5. Thus, charged states correspond to $\eta'_5 = +1$ and neutral states to $\eta'_5 = -1$. We adopt the normalization

$$\int \psi^*\beta_4\psi \, dV = 1$$

so that the charge $Q = -(i/c)\int j_4 \, dV$ is given by $Q = 0(\eta'_5 = -1, \eta_4\eta'_4 = \pm 1)$, $Q = e_0(\eta'_5 = +1, \eta_4\eta'_4 = +1)$, and $Q = -e_0(\eta'_5 = +1, \eta_4\eta'_4 = -1)$. Thus each state is characterized by its charge, its spin J and z component $J_z = \hbar(\mathbf{\Sigma} + \mathbf{\Sigma}')_z$ and by its rest energy as given by the eigenvalue equation (7). However, a neutral state may appear in different representations with opposite signs of $\eta_4\eta'_4$ but with the same rest energy.

We now find that neutral states with $\eta_4\eta'_4 = +1$ are sometimes accompanied in the same representation by states of the same spin, the same mass [to terms of order $(m_0/m)^2$], but with charge $\pm e_0$, whereas neutral states with $\eta_4\eta'_4 = -1$ always appear alone. To verify that these cases correspond to $T = 1$, $T = 0$, respectively, we write

$$Q = T_3 = \tfrac{1}{2}\eta_4\eta'_4(1 + \eta'_5)$$

and find that without inventing a special space to accommodate them, we can define operators T_1, T_2 such that

$$\mathbf{T}\times\mathbf{T} = i\mathbf{T}.$$

For each triplet of solutions $\psi_+\psi_0\psi_-$ which occur as indicated above, it is then found that, if we neglect terms of order $(m_0/m)^2$, the components of $\mathbf{T}$ in fact transform these states into each other in precisely the manner required of an isospin operator with $T = 1$. Isospin is therefore a dynamical consequence of Eq. (7); it is not necessary to postulate its existence in some "isospin space." On the other hand, it is seen as an approximation, leading to the usual isospin only when mass differences between the components are neglected.

The rest-energy eigenvalues of Eq. (4) are exhibited in Fig. 2 for the choice $mc^2 = 1041$ MeV, $m_0c^2 = 106$ MeV of the only parameters in the theory.[7] The experimental values of the identified states are not shown, because on this plot they are indistinguishable from the theoretical values.[8] The structure of the states is independent of the choice of these parameters, and as the values of m, m_0 are varied the points in Fig. 2 move horizontally. In the limit $m \to 0$, all points move to infinity except the f state and the states marked with open circles, which move along $J = 1$ and $J = 0$ to zero. The above choice of m and m_0 fits the masses for the f and η states, but *every other number on the left half of Table I is computed from Eqs. (5) and (7).* A search for the unobserved states around 1 BeV and 450 MeV is strongly recommended.[9] In Table II the same data are exhibited in a form which shows how the states appear in the various representations of β_μ and β'_μ. There is no doubt here that the f, ω, and η appear in the same representation, and that the ϕ appears separately, a question which is unresolved in the unitary symmetry model.

From the nonrelativistic example discussed at the beginning of this note it is clear that the solutions for the relativistic case may be discussed in terms of Regge poles, but this is valuable only when a dynamical theory is lacking. In the present case, the single equation (4) is seen to simultaneously describe states of different angular momenta and

[7] H. C. Corben, Phys. Rev. Letters **10**, 555 (1963). The parameters m, m_0 in this note were ill-chosen.

[8] Except for the spin of the B state, here predicted as 2 and the $\eta - 2\pi$ here predicted as unity.

[9] Reported states with $S = 0$, $T = 0$ or 1 are $\chi_1(T = 1, 1050$ MeV$)\kappa_{22}(T = 0, J$ even, 1040 MeV$)$ $\phi_3(885$ MeV$)$ $\phi_2(T = 0, 520$ MeV$)$ $\phi_1(T = 0, 395$ MeV$)$ $\zeta(T = 1$ 564 MeV$)$ [M. Roos, Rev. Mod. Phys. **35**, 318 (1963)]. See also M. Roos, Phys. Letters **8**, 1 (1964); and G. Goldhaber *et al.*, Phys. Rev. Letters **12**, 336 (1964); G. R. Kalbfleisch *et al.*, *ibid.*, p. 527; M. Goldberg *et al.*, *ibid.*, p. 546.

TABLE I. Normal modes of a free symmetrical top according to relativistic quantum mechanics, for the case of the generalized Kemmer equation for bosons. Theory for spin 3 state obtained from Eq. (4) by replacing $\beta_{\mu\nu}'$ by $\beta_{\mu\nu}' + \beta_{\mu\nu}''$. Recently reported unnamed states marked ?.[a]

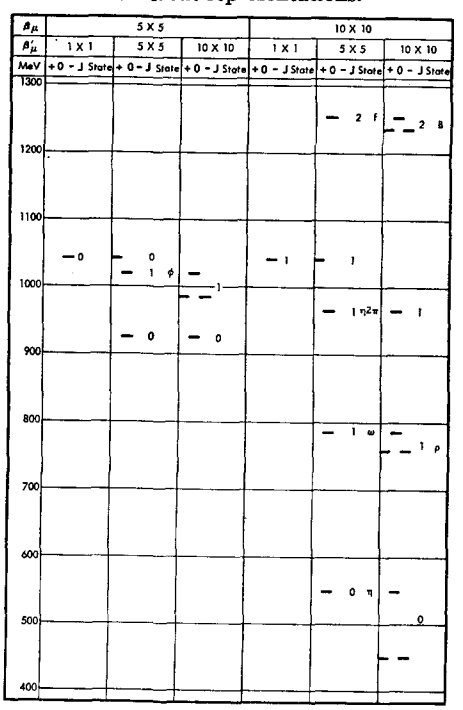
	Theory				Experiment			
β_μ	Charge	Spin	Isospin	Rest-energy (MeV)	Particle	Charge	Spin	Isospin	Rest-energy (MeV)
	0	2	0	Input	f	0	2	0	1253 ± 20
	0	0	0	Input	η	0	0	0	548.5 ± 0.6
	0	1	0	785	ω	0	1	0	783 ± 2
	$\pm e$ 0	2	1	1235 1253	B	$\pm e$ 0	?	1	1215, 1220
10×10	$\pm e$ 0	1	1	758 785	ρ	$\pm e$ 0	1	1	757 ± 5 754 ± 5, 770
	$\pm e$ 0	0	1	448 548	?	0	?	0	520 ± 20
	0	1	0	965	$\eta - 2\pi$	0	?	0 or 1	960
	$\pm e$	1	0	1041	W?	$\pm e$	1	0	?
	0	1	0	1019	ϕ	0	1	0	1019.5 ± 0.3
5×5	$\pm e$ 0	1	1	985 1019	?	$-e$	?	≥ 1	1000 ± 10
	0	0	0	925	?	0	?	0	922 ± 30
	$\pm e$	0	0	1041					
10×10		3		1460 ± 30	?	0		≤ 1	1410

[a] M. Ross, Phys. Letters 8, 1 (1964).

rest energies, to provide a dynamical basis for isospin, and, if the other states predicted by the theory are found, to yield a remarkably accurate description of the observed levels of bosons with $S = 0$. In any case, the simplicity of Eq. (4) allows us to use it as a basis on which to construct a fully quantized version of generalized meson field theory in which a number of states of different spin, mass, and charge are described by the one equation, their masses being expressed in terms of only two parameters.

In the case of molecules, the spectra are complicated by vibrational and electronic transitions which are related to internal properties of the molecule, and mesons and baryons have a right to be considered "elementary" only if they are ultimately found not to possess such levels. But the rotational modes must be there anyhow, and their existence puts at least part of the elementary particle spectrum on a simple footing. We could not expect to compute the properties of internal levels, if they exist, without knowing more about the structure of the particles, which would also give us the reason for the values of the parameters m, m_0 which fit the observed data so well. On the other hand, as far as the rotational states are concerned, we should not be any more surprised at being able to compute them in terms of these parameters than we are at being able to compute the rotational states of a symmetrical molecule in terms of the parameters A and C which describe its moments of inertia. Further, the analogy between elementary particle processes and the Raman

TABLE II. Charge, spin, and rest-energy states of Eq. (4) for different representations.



effect has already been pointed out.[10] If the point of view advanced in this note is correct, we should be investigating the relativistic generalization of the Raman spectrum of a symmetrical top.[11]

Rotational levels defined by equations similar to Eq. (4) have been applied to bosons of odd S and to fermions of even and odd S, and there seems no reason why similar but more complicated equations could not be applied to atomic nuclei throughout the periodic table. This would be consistent with the philosophy of "bootstrap" calculations; vice versa, rotational levels of atomic nuclei are known to exist, and if similar techniques are applicable to states with baryon number zero and to states with baryon number of more than two hundred, it follows that rotational levels must also exist for $B = 0$, and it is just a matter of computing them relativistically. This is what has been done here.

Note added in proof: This equation also leads to a fine structure at 782, 789, 806 MeV and states at 1220, 1653 MeV.

[10] V. Weisskopf, Phys. Today **16**, 26 (1963).
[11] G. Placzek and E. Teller, Z. Physik **81**, 209 (1933).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Electromagnetic Radiation in the Presence of Moving Simple Media

K. S. H. Lee and C. H. Papas
California Institute of Technology, Pasadena, California
(Received 9 June 1964)

The radiation pattern of an arbitrary source immersed in a moving simple medium is calculated by deducing the differential equation for the potential 4-vector in the rest frame of the source and then solving the equation in terms of a Green's function. As an illustrative example, the case where the source is an oscillating dipole is worked out in detail.

1. INTRODUCTION

By reason of its relevance to astrophysical and spatial studies, the problem of calculating the influence of a relativistic wind on the radiation properties of an arbitrary electromagnetic source is of considerable importance. In the present paper we describe a method of solving this problem for the special case where the wind is a simple medium (i.e., a homogeneous, isotropic, nondispersive dielectric) moving at uniform velocity with respect to the source.

To formulate the problem we introduce the rest frame K of the source and the rest frame K' of the medium, and note that the relative motion between the source and the wind can be specified by stating that K' moves at a uniform velocity $\mathbf{v}$ with respect to K or that K moves at a uniform velocity $-\mathbf{v}$ with respect to K'. Clearly, K and K' are then inertial frames of reference and hence, in accord with Minkowski's postulate, the macroscopic Maxwell equations are covariant under the Lorentz transformation that connects the space–time coordinates of K with those of K'. Since the covariance of Maxwell's equations permits the combining of the vector and scalar potentials into a 4-vector, namely, the potential 4-vector, and since the field can be obtained by differentiating the potential 4-vector, knowing the potential 4-vector is tantamount to knowing the field. Thus the problem of calculating the effects of a relativistic wind on the radiation from a source amounts to one of finding the potential 4-vector in the rest frame K of the source.

One may try to calculate the potential 4-vector in K by first finding the potential 4-vector in K' and then transforming the result from K' to K by the proper Lorentz transformation. However, such an approach leads to difficulties, because the integral representations for the retarded potentials in K' contain the time implicitly in a complicated form. A simpler method consists of deriving the differential equation for the potential 4-vector in K and then solving it by the Green's function technique. In this method the differential equation in K is found by transforming the already known differential equation for the potential 4-vector in K'.

Accordingly, we first find the differential equation for the potential 4-vector in the rest frame K of the source, and then obtain the appropriate solution

effect has already been pointed out.[10] If the point of view advanced in this note is correct, we should be investigating the relativistic generalization of the Raman spectrum of a symmetrical top.[11]

Rotational levels defined by equations similar to Eq. (4) have been applied to bosons of odd S and to fermions of even and odd S, and there seems no reason why similar but more complicated equations could not be applied to atomic nuclei throughout the periodic table. This would be consistent with the philosophy of "bootstrap" calculations; vice versa, rotational levels of atomic nuclei are known to exist, and if similar techniques are applicable to states with baryon number zero and to states with baryon number of more than two hundred, it follows that rotational levels must also exist for $B = 0$, and it is just a matter of computing them relativistically. This is what has been done here.

Note added in proof: This equation also leads to a fine structure at 782, 789, 806 MeV and states at 1220, 1653 MeV.

[10] V. Weisskopf, Phys. Today **16**, 26 (1963).
[11] G. Placzek and E. Teller, Z. Physik **81**, 209 (1933).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Electromagnetic Radiation in the Presence of Moving Simple Media

K. S. H. Lee and C. H. Papas
California Institute of Technology, Pasadena, California
(Received 9 June 1964)

The radiation pattern of an arbitrary source immersed in a moving simple medium is calculated by deducing the differential equation for the potential 4-vector in the rest frame of the source and then solving the equation in terms of a Green's function. As an illustrative example, the case where the source is an oscillating dipole is worked out in detail.

1. INTRODUCTION

By reason of its relevance to astrophysical and spatial studies, the problem of calculating the influence of a relativistic wind on the radiation properties of an arbitrary electromagnetic source is of considerable importance. In the present paper we describe a method of solving this problem for the special case where the wind is a simple medium (i.e., a homogeneous, isotropic, nondispersive dielectric) moving at uniform velocity with respect to the source.

To formulate the problem we introduce the rest frame K of the source and the rest frame K' of the medium, and note that the relative motion between the source and the wind can be specified by stating that K' moves at a uniform velocity $\mathbf{v}$ with respect to K or that K moves at a uniform velocity $-\mathbf{v}$ with respect to K'. Clearly, K and K' are then inertial frames of reference and hence, in accord with Minkowski's postulate, the macroscopic Maxwell equations are covariant under the Lorentz transformation that connects the space–time coordinates of K with those of K'. Since the covariance of Maxwell's equations permits the combining of the vector and scalar potentials into a 4-vector, namely, the potential 4-vector, and since the field can be obtained by differentiating the potential 4-vector, knowing the potential 4-vector is tantamount to knowing the field. Thus the problem of calculating the effects of a relativistic wind on the radiation from a source amounts to one of finding the potential 4-vector in the rest frame K of the source.

One may try to calculate the potential 4-vector in K by first finding the potential 4-vector in K' and then transforming the result from K' to K by the proper Lorentz transformation. However, such an approach leads to difficulties, because the integral representations for the retarded potentials in K' contain the time implicitly in a complicated form. A simpler method consists of deriving the differential equation for the potential 4-vector in K and then solving it by the Green's function technique. In this method the differential equation in K is found by transforming the already known differential equation for the potential 4-vector in K'.

Accordingly, we first find the differential equation for the potential 4-vector in the rest frame K of the source, and then obtain the appropriate solution

of the equation in terms of a Green's function. For the purpose of illustrating the method, the effect of a relativistic wind on the radiation pattern of an oscillating dipole is worked out.

2. THE DIFFERENTIAL EQUATION FOR THE POTENTIAL 4-VECTOR

In the rest frame K' of the wind the vector potential $\mathbf{A}'(\mathbf{r}', t')$ and the scalar potential $\phi'(\mathbf{r}', t')$, due to the source current density $\mathbf{J}'(\mathbf{r}', t')$ and the source charge density $\rho'(\mathbf{r}', t')$, obey the inhomogeneous wave equations

$$\left(\nabla'^2 - \frac{n'^2}{c^2}\frac{\partial^2}{\partial t'^2}\right)\mathbf{A}' = -\mu'\mathbf{J}', \qquad \left(\nabla'^2 - \frac{n'^2}{c^2}\frac{\partial^2}{\partial t'^2}\right)\phi' = -\frac{1}{\epsilon'}\rho', \tag{2.1}$$

where μ' and ϵ' are the permeability and the dielectric constant of the simple medium and $n'[= c(\mu'\epsilon')^{\frac{1}{2}}]$ is its index of refraction. (In this section, primed quantities are referred to K' and unprimed ones to K.) It clearly follows from these equations that the potential 4-vector $A'_\alpha = [\mathbf{A}', (i/c)\phi']$ satisfies the equation

$$[1 + \kappa c^2\delta_{4\alpha}]L'A'_\alpha = -\mu' J'_\alpha \quad (\alpha = 1, 2, 3, 4), \tag{2.2}$$

where $\delta_{4\alpha}$ is the Kronecker delta, κ is a shorthand for $(n'^2 - 1)/c^2$, and L' is the operator $\nabla'^2 - (n'^2/c^2)\partial^2/\partial t'^2$. The source term on the right side is the current density 4-vector $J'_\alpha = (\mathbf{J}', ic\rho')$.

To transform Eq. (2.2) to the reference frame K we note that since A'_α and J'_α are 4-vectors they are related to their counterparts A_β, J_β in K by

$$A'_\alpha = a_{\alpha\beta}A_\beta, \qquad J'_\alpha = a_{\alpha\beta}J_\beta, \tag{2.3}$$

where $a_{\alpha\beta}$ are the coefficients of the proper Lorentz transformation that carries K into K'. Furthermore, by writing L' as

$$L' = \nabla'^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t'^2} - \kappa\frac{\partial^2}{\partial t'^2}, \tag{2.4}$$

and then noting that the first two terms constitute an invariant and that the third term transforms according to the rule

$$\partial/\partial t' = \gamma(\partial/\partial t) + \gamma\mathbf{v}\cdot\nabla, \tag{2.5}$$

where $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$, we see that L' transforms into

$$L = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \kappa\gamma^2\left(\mathbf{v}\cdot\nabla + \frac{\partial}{\partial t}\right)^2. \tag{2.6}$$

In view of Eqs. (2.3) and (2.6) it is clear that Eq. (2.2) becomes

$$[1 + \kappa c^2\delta_{4\alpha}]a_{\alpha\beta}LA_\beta = -\mu' a_{\alpha\beta}J_\beta. \tag{2.7}$$

Applying the orthogonality relation $a_{\alpha\nu}a_{\alpha\beta} = \delta_{\nu\beta}$ to this equation, we then obtain

$$LA_\nu = -\mu' J_\nu + (\kappa c^2/n'^2)\mu' a_{4\nu}a_{4\beta}J_\beta, \tag{2.8}$$

which in terms of the velocity 4-vector $U_\nu = (\gamma\mathbf{v}, i\gamma c)$ yields

$$LA_\nu = -\mu' J_\nu - \frac{\kappa}{n'^2}\mu' J_\alpha U_\alpha U_\nu. \tag{2.9}$$

In three dimensions Eq. (2.9) takes the form

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \kappa\gamma^2\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)^2\right]\mathbf{A} = -\mu'\mathbf{J} - \frac{\mu'\kappa}{n'^2}\gamma^2\mathbf{v}\mathbf{v}\cdot\mathbf{J} + \frac{\mu'\kappa c^2}{n'^2}\gamma^2\rho\mathbf{v},$$

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \kappa\gamma^2\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)^2\right]\phi = -c^2\mu'\left(1 - \frac{\kappa c^2}{n'^2}\gamma^2\right)\rho - \frac{\mu'\kappa c^2}{n'^2}\gamma^2\mathbf{v}\cdot\mathbf{J}. \tag{2.10}$$

Thus we see that the inhomogeneous wave equations (2.1) for $\mathbf{A}'$, ϕ' in K' transform into Eqs. (2.10) for $\mathbf{A}$, ϕ in K.[1]

3. GREEN'S FUNCTION REPRESENTATION OF RADIATION FIELD

To solve Eqs. (2.10) we represent $\mathbf{A}$, ϕ, ρ, and $\mathbf{J}$ by the Fourier integrals

$$\mathbf{A}(\mathbf{r}, t) = \int_{-\infty}^{\infty}\mathbf{A}_\omega(\mathbf{r})e^{-i\omega t}\,d\omega, \qquad \phi(\mathbf{r}, t) = \int_{-\infty}^{\infty}\phi_\omega(\mathbf{r})e^{-i\omega t}\,d\omega,$$
$$\rho(\mathbf{r}, t) = \int_{-\infty}^{\infty}\rho_\omega(\mathbf{r})e^{-i\omega t}\,d\omega, \qquad \mathbf{J}(\mathbf{r}, t) = \int_{-\infty}^{\infty}\mathbf{J}_\omega(\mathbf{r})e^{-i\omega t}\,d\omega, \tag{2.11}$$

and note that the continuity equation $\nabla\cdot\mathbf{J}(\mathbf{r}, t) = -(\partial/\partial t)\rho(\mathbf{r}, t)$ yields

$$\nabla\cdot\mathbf{J}_\omega(\mathbf{r}) = i\omega\rho_\omega(\mathbf{r}). \tag{2.12}$$

Substituting representations (2.11) into Eqs. (2.10) and using Eq. (2.12) we see that $\mathbf{A}_\omega(\mathbf{r})$ and $\phi_\omega(\mathbf{r})$

[1] Actually, the same result can be achieved by using the tensor form of Maxwell's equations as the point of departure. See J. M. Jauch and K. M. Watson, Phys. Rev. 74, 950, 1485 (1948).

satisfy

$$[\nabla^2 + \omega^2/c^2 - \kappa\gamma^2(\mathbf{v}\cdot\nabla - i\omega)^2]\phi_\omega(\mathbf{r}) = \mathbf{M}\cdot\mathbf{J}_\omega(\mathbf{r}),$$
$$[\nabla^2 + \omega^2/c^2 - \kappa\gamma^2(\mathbf{v}\cdot\nabla - i\omega)^2]\mathbf{A}_\omega(\mathbf{r}) = \mathbf{N}\cdot\mathbf{J}_\omega(\mathbf{r}), \quad (2.13)$$

where the vector $\mathbf{M}$ and the dyadic $\mathbf{N}$ are given by

$$\mathbf{M} = -\frac{c^2\mu'}{i\omega}\left(1 - \frac{\kappa c^2}{n'^2}\gamma^2\right)\nabla - \frac{\mu'\kappa c^2}{n'^2}\gamma^2\mathbf{v}, \quad (2.14)$$

$$\mathbf{N} = -\mu'\mathbf{U} - \frac{\mu'\kappa}{n'^2}\gamma^2\mathbf{v}\mathbf{v} + \frac{\mu'\kappa c^2}{i\omega n'^2}\gamma^2\mathbf{v}\nabla, \quad (2.15)$$

and $\mathbf{U}$ is the unit dyadic.

From the linearity of these equations it is clear that their solutions are

$$\phi_\omega(\mathbf{r}) = -\mathbf{M}\cdot\int \mathbf{J}_\omega(\mathbf{r}')G_\omega(\mathbf{r},\mathbf{r}')\,dV',$$
$$\mathbf{A}_\omega(\mathbf{r}) = -\mathbf{N}\cdot\int \mathbf{J}_\omega(\mathbf{r}')G_\omega(\mathbf{r},\mathbf{r}')\,dV', \quad (2.16)$$

provided the function $G_\omega(\mathbf{r},\mathbf{r}')$ is a solution of

$$[\nabla^2 + \omega^2/c^2 - \kappa\gamma^2(\mathbf{v}\cdot\nabla - i\omega)^2]G_\omega(\mathbf{r},\mathbf{r}') = -\delta(\mathbf{r}-\mathbf{r}'). \quad (2.17)$$

The solution of (2.17) is the Green's function

$$G_\omega(\mathbf{r},\mathbf{r}') = \frac{1}{8\pi^3}\iiint_{-\infty}^{\infty} \frac{e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}}{k^2 - \omega^2/c^2 - \kappa\gamma^2(\mathbf{v}\cdot\mathbf{k} - \omega)^2}\,d^3k. \quad (2.18)$$

From a knowledge of the Green's function $G_\omega(\mathbf{r},\mathbf{r}')$, the potentials $\phi_\omega(\mathbf{r})$ and $A_\omega(\mathbf{r})$, due to an arbitrary current density $\mathbf{J}_\omega(\mathbf{r})$, can be found from relations (2.16). In turn, the electric vector $\mathbf{E}_\omega(\mathbf{r})$ and the magnetic vector $\mathbf{B}_\omega(\mathbf{r})$ can be found from a knowledge of $\phi_\omega(\mathbf{r})$ and $\mathbf{A}_\omega(\mathbf{r})$ by using the relations

$$\mathbf{E}_\omega = -\nabla\phi_\omega + i\omega\mathbf{A}_\omega, \quad \mathbf{B}_\omega = \nabla\times\mathbf{A}_\omega. \quad (2.19)$$

Thus the calculation of the electromagnetic field $\mathbf{E}_\omega$, $\mathbf{B}_\omega$ radiated by the current $\mathbf{J}_\omega(\mathbf{r})$ hinges on the evaluation of the Green's function (2.18).

4. EVALUATION OF THE GREEN'S FUNCTION

Without loss of generality we choose the z axis of the reference frame K to be parallel to the velocity $\mathbf{v}$ of the moving medium. With this choice the integral representation for the Green's function becomes

$$G_\omega(\mathbf{r},\mathbf{r}') = \frac{1}{8\pi^3}\iiint_{-\infty}^{\infty} \frac{e^{ik_x(x-x')}e^{ik_y(y-y')}e^{ik_z(z-z')}}{k^2 - k_0^2 - \kappa\gamma^2(vk_z - \omega)^2}\,dk_x\,dk_y\,dk_z, \quad (2.20)$$

where $k_0^2 = \omega^2/c^2$. In preparation for the k_z integration we write the denominator of the integrand in the factored form

$$k^2 - k_0^2 - \kappa\gamma^2(vk_z - \omega)^2 = \gamma^2(1 - n'^2\beta^2)(k_z - \tau_+)(k_z - \tau_-), \quad (2.21)$$

where

$$\tau_\pm = \frac{-k_0\beta\gamma^2(n'^2 - 1) \pm [n'^2k_0^2 - \gamma^2(1 - n'^2\beta^2)(k_x^2 + k_y^2)]^{\frac{1}{2}}}{\gamma^2(1 - n'^2\beta^2)}. \quad (2.22)$$

Here $\beta = v/c$ and hence $n'\beta = v/v'_{\rm ph}$ where $v'_{\rm ph} = (\mu'\epsilon')^{-\frac{1}{2}}$ denotes the phase velocity of the radiation in K'. In the complex k_z plane the integrand has poles at $k_z = \tau_+$ and at $k_z = \tau_-$.

From expression (2.22) we see that when $n'\beta < 1$ the pole τ_+ may lie either on the real axis or in the upper k_z-plane, and the pole τ_- may lie either on the real axis or in the lower k_z-plane. When $n'\beta > 1$ both poles τ_+ and τ_- always lie on the real axis. The question of how the contour should be indented to circumvent poles on the real axis can be settled by assuming the medium to be slightly lossy. Thus it can be shown that when $n'\beta < 1$, the path of integration must always lie below τ_+ and above τ_-. However, when $n'\beta > 1$ the path of integration must always lie below both τ_+ and τ_-. In any case the contour is closed in the upper half-plane for $z - z' > 0$ and in the lower half-plane for $z - z' < 0$.

Accordingly, when $n'\beta < 1$ and $z - z' > 0$, expression (2.20) becomes

$$G_\omega(\mathbf{r},\mathbf{r}') = \frac{1}{8\pi^3}\frac{2\pi i}{\gamma^2(1 - n'^2\beta^2)} \times \iint_{-\infty}^{\infty} \frac{e^{i\tau_+(z-z')}e^{ik_x(x-x')}e^{ik_y(y-y')}}{\tau_+ - \tau_-}\,dk_x\,dk_y. \quad (2.23)$$

Introducing the polar coordinates (s, χ) by the relations $k_x = s\cos\chi$, $k_y = s\sin\chi$ and integrating with respect to χ from $\chi = 0$ to $\chi = 2\pi$, we get

$$G_\omega(\mathbf{r},\mathbf{r}') = \frac{a}{4\pi}e^{-ibk(z-z')}\int_0^\infty \frac{J_0\{s[(x-x')^2 + (y-y')^2]^{\frac{1}{2}}\}\exp[ia(k^2 - s^2)^{\frac{1}{2}}(z-z')]}{(s^2 - k^2)^{\frac{1}{2}}}\,s\,ds, \quad (2.24)$$

where

$$a = [\gamma(1 - n'^2\beta^2)^{\frac{1}{2}}]^{-1}, \qquad b = a\beta\gamma^2[(n'^2 - 1)/n'], \qquad k = an'k_0. \tag{2.25}$$

Then by evaluating the integral we have

$$G_\omega(\mathbf{r}, \mathbf{r}') = \frac{a}{4\pi} e^{-ibk(z-z')} \times \frac{\exp\{ik[(x-x')^2 + (y-y')^2 + a^2(z-z')^2]^{\frac{1}{2}}\}}{[(x-x')^2 + (y-y')^2 + a^2(z-z')^2]^{\frac{1}{2}}}. \tag{2.26}$$

It turns out that this result is valid for $z - z' < 0$ also. Hence, when $n'\beta < 1$ we see that in general

$$G_\omega(\mathbf{r}, \mathbf{r}') = \frac{a}{4\pi} \exp\left[-i\left(\frac{bk}{v}\right)\mathbf{v}\cdot(\mathbf{r} - \mathbf{r}')\right] \times \frac{\exp\left\{ik\left[(\mathbf{r}-\mathbf{r}')^2 + \frac{a^2-1}{v^2}[\mathbf{v}\cdot(\mathbf{r}-\mathbf{r}')]^2\right]^{\frac{1}{2}}\right\}}{\left[(\mathbf{r}-\mathbf{r}')^2 + \frac{a^2-1}{v^2}[\mathbf{v}\cdot(\mathbf{r}-\mathbf{r}')]^2\right]^{\frac{1}{2}}}. \tag{2.27}$$

When $n'\beta > 1$, expression (2.20) yields

$$G_\omega(\mathbf{r}, \mathbf{r}') = \frac{a}{4\pi^2} e^{ibk(z-z')} \times \int_0^\infty \int_0^{2\pi} \frac{\sin[a(z-z')(s^2+k^2)^{\frac{1}{2}}]}{(s^2+k^2)^{\frac{1}{2}}} \times \exp[is(x-x')\cos\chi + is(y-y')\sin\chi] s\, d\chi\, ds \tag{2.28}$$

for $z - z' > 0$, and

$$G_\omega(\mathbf{r}, \mathbf{r}') = 0 \tag{2.29}$$

for $z - z' < 0$. Here $a = [\gamma(n'^2\beta^2 - 1)^{\frac{1}{2}}]$. Carrying out the χ integration we get

$$G_\omega(\mathbf{r}, \mathbf{r}') = \frac{a}{2\pi} e^{ibk(z-z')} \int_0^\infty \frac{J_0\{s[(x-x')^2 + (y-y')^2]^{\frac{1}{2}}\} \sin[a(z-z')(s^2+k^2)^{\frac{1}{2}}]}{(s^2+k^2)^{\frac{1}{2}}} s\, ds \tag{2.30}$$

which in turn yields

$$G_\omega(\mathbf{r}, \mathbf{r}') = \frac{a}{2\pi} e^{ibk(z-z')} \times \frac{\cos\{k[-(x-x')^2 - (y-y')^2 + a^2(z-z')^2]^{\frac{1}{2}}\}}{[-(x-x')^2 - (y-y')^2 + a^2(z-z')^2]^{\frac{1}{2}}} \tag{2.31}$$

for

$$z - z' > [(x-x')^2 + (y-y')^2]^{\frac{1}{2}}a^{-1},$$

and

$$G_\omega(\mathbf{r}, \mathbf{r}') = 0 \tag{2.32}$$

for $z - z' < [(x-x')^2 + (y-y')^2]^{\frac{1}{2}}a^{-1}$. This shows that when $n'\beta > 1$ the Green's function is different from zero only within a conical region in the direction of the wind (Fig. 1). The half-angle θ_0 of the cone is given by

$$\tan\theta_0 = a = [\gamma(n'^2\beta^2 - 1)^{\frac{1}{2}}]^{-1} \tag{2.33}$$

5. RADIATION FIELD OF AN OSCILLATING DIPOLE

Let us now consider the far-zone field of an oscillating dipole of frequency ω for the case where $n'\beta < 1$. Let the dipole be located at the origin of a spherical coordinate system (r, θ, ϕ). If $\mathbf{p}$ is its dipole moment, then $\mathbf{J}_\omega(\mathbf{r}') = -i\omega\mathbf{p}\delta(\mathbf{r}')$. Accordingly, with the aid of expression (2.27) we see that

$$\int \mathbf{J}_\omega(\mathbf{r}')G_\omega(\mathbf{r}, \mathbf{r}')\, dV' = -i\omega\mathbf{p}ae^{-ibkr\cos\theta} \times \frac{\exp[ikr(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}]}{4\pi r(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}}, \tag{2.34}$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{r}$. Substituting expression (2.34) into Eqs. (2.16) one can obtain the scalar and vector potentials of the dipole, and then from these potentials can deduce by use of the first of Eqs. (2.19) the electric field of the dipole.

Assuming that $\mathbf{p}$ is perpendicular to $\mathbf{v}$, i.e., $\mathbf{p} = \mathbf{e}_x p$ and $\mathbf{v} = \mathbf{e}_z v$, we find that the far-zone electric field is given by

$$E_\theta = \omega^2\mu' p \frac{a^3}{\sin^2\theta + a^2\cos^2\theta}\cos\theta\cos\phi e^{-ibkr\cos\theta} \times \frac{\exp[ikr(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}]}{4\pi r(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}},$$

$$E_\phi = -\omega^2\mu' pa\sin\phi e^{-ibkr\cos\theta} \times \frac{\exp[ikr(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}]}{4\pi r(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}},$$

$$E_r = 0, \tag{2.35}$$

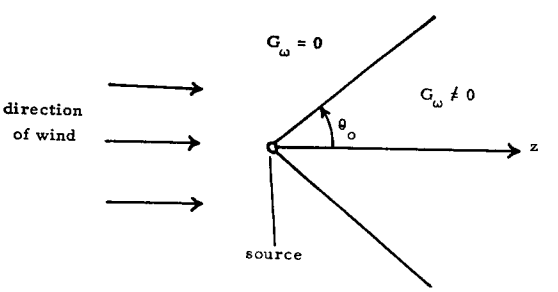


FIG. 1. Conical region where Green's function is not identically zero.

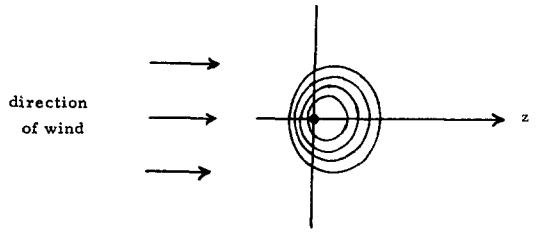


FIG. 2. Surfaces of constant phase dragged by wind.

where, as before

$$a = \frac{1}{\gamma(1 - n'^2\beta^2)^{\frac{1}{2}}}, \quad b = a\beta\gamma^2\frac{n'^2 - 1}{n'}, \quad k = an'\frac{\omega}{c}. \tag{2.36}$$

In the case where $\mathbf{p}$ is parallel to $\mathbf{v}$, i.e., $\mathbf{p} = \mathbf{e}_z p$ and $\mathbf{v} = \mathbf{e}_z v$, we find that

$$E_\theta = -\omega^2\mu' p \frac{a^3}{(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}} \sin\theta e^{-ibkr\cos\theta} \times \frac{\exp[ikr(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}]}{4\pi r(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}}, \tag{2.37}$$

$$E_\phi = E_r = 0.$$

In examining the above expressions for the far-zone field components we see that the phase of each is given by

$$\psi = -bkz + k(x^2 + y^2 + a^2z^2)^{\frac{1}{2}} - \omega t. \tag{2.38}$$

The surfaces of constant phase, i.e., surfaces defined by ψ = const, are oblate spheroids. If the medium were not moving, the surfaces would be spheres. That is, the motion of the medium contracts these spheres in the direction of motion and thus deforms them into oblate spheroids. The axis of rotational symmetry of each spheroid passes through the source and is parallel to the direction of the wind. As time increases, a spheroidal surface of constant phase expands and is dragged by the wind (Fig. 2).

For the complementary case where $n'\beta > 1$ the Green's function that enters the calculation is given by Eq. (2.31), and the far-zone radiation field can be found in much the same way as in the case where $n'\beta < 1$.

ACKNOWLEDGMENT

The authors are grateful to the U. S. Air Force Office of Scientific Research for supporting this research.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

High-Field Magnetoresistance of Inhomogeneous Semiconductors and Plasmas. II. Two-Dimensional Inhomogeneity Distributions

J. A. MORRISON AND H. L. FRISCH
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 11 February 1964; final manuscript received 24 July 1964)

We extend in this paper the treatment of the high field magnetoresistance of a previously described classical model of a semiconductor (plasma) containing a two-dimensional distribution of inhomogeneities. The basic assumptions on the classical model are that the scale of the inhomogeneities is large compared to the mean thermal wavelength of an electron and the Landau level spacing is large compared to kT. The magnetic field $\mathcal{H}$ is taken parallel to the z coordinate and the inhomogeneity distribution is characterized by a sufficiently smooth potential $\varphi(x, z)$. The 4-moment equations are solved asymptotically for large $\mathcal{H}$, and an equivalent asymptotic solution is obtained, subject to certain mathematical assumptions, for the transport equation. The magnetoresistance is shown, in general, not to saturate, but to increase, as $\mathcal{H}^2$, with increasing $\mathcal{H}$.

1. INTRODUCTION

THIS is the second in a series of papers in which we examine the possibility that the high-field nonsaturation of the magnetoresistance of a semiconductor (or a model plasma) has a classical origin in the presence in the sample of a more or less random distribution of spatial inhomogeneities. The first paper,[1] hereafter referred to as I, dealt only with the special case of a stratified medium. Here we extend the theory developed in I to two-dimensional inhomogeneity distributions which vary both along the x and z coordinates, the latter being

[1] H. L. Frisch and J. A. Morrison, Ann. Phys. 26, 181 (1964). Equations referring to this paper are prefixed with a I.

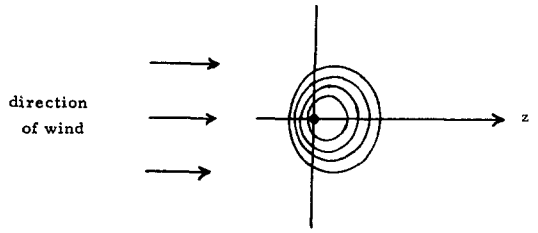


FIG. 2. Surfaces of constant phase dragged by wind.

where, as before

$$a = \frac{1}{\gamma(1 - n'^2\beta^2)^{\frac{1}{2}}}, \quad b = a\beta\gamma^2\frac{n'^2 - 1}{n'}, \quad k = an'\frac{\omega}{c}. \tag{2.36}$$

In the case where $\mathbf{p}$ is parallel to $\mathbf{v}$, i.e., $\mathbf{p} = \mathbf{e}_z p$ and $\mathbf{v} = \mathbf{e}_z v$, we find that

$$E_\theta = -\omega^2\mu' p \frac{a^3}{(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}} \sin\theta e^{-ibkr\cos\theta} \times \frac{\exp[ikr(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}]}{4\pi r(\sin^2\theta + a^2\cos^2\theta)^{\frac{1}{2}}}, \tag{2.37}$$

$$E_\phi = E_r = 0.$$

In examining the above expressions for the far-zone field components we see that the phase of each is given by

$$\psi = -bkz + k(x^2 + y^2 + a^2z^2)^{\frac{1}{2}} - \omega t. \tag{2.38}$$

The surfaces of constant phase, i.e., surfaces defined by ψ = const, are oblate spheroids. If the medium were not moving, the surfaces would be spheres. That is, the motion of the medium contracts these spheres in the direction of motion and thus deforms them into oblate spheroids. The axis of rotational symmetry of each spheroid passes through the source and is parallel to the direction of the wind. As time increases, a spheroidal surface of constant phase expands and is dragged by the wind (Fig. 2).

For the complementary case where $n'\beta > 1$ the Green's function that enters the calculation is given by Eq. (2.31), and the far-zone radiation field can be found in much the same way as in the case where $n'\beta < 1$.

ACKNOWLEDGMENT

The authors are grateful to the U. S. Air Force Office of Scientific Research for supporting this research.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

High-Field Magnetoresistance of Inhomogeneous Semiconductors and Plasmas. II. Two-Dimensional Inhomogeneity Distributions

J. A. MORRISON AND H. L. FRISCH
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 11 February 1964; final manuscript received 24 July 1964)

We extend in this paper the treatment of the high field magnetoresistance of a previously described classical model of a semiconductor (plasma) containing a two-dimensional distribution of inhomogeneities. The basic assumptions on the classical model are that the scale of the inhomogeneities is large compared to the mean thermal wavelength of an electron and the Landau level spacing is large compared to kT. The magnetic field $\mathcal{H}$ is taken parallel to the z coordinate and the inhomogeneity distribution is characterized by a sufficiently smooth potential $\varphi(x, z)$. The 4-moment equations are solved asymptotically for large $\mathcal{H}$, and an equivalent asymptotic solution is obtained, subject to certain mathematical assumptions, for the transport equation. The magnetoresistance is shown, in general, not to saturate, but to increase, as $\mathcal{H}^2$, with increasing $\mathcal{H}$.

1. INTRODUCTION

THIS is the second in a series of papers in which we examine the possibility that the high-field nonsaturation of the magnetoresistance of a semiconductor (or a model plasma) has a classical origin in the presence in the sample of a more or less random distribution of spatial inhomogeneities. The first paper,[1] hereafter referred to as I, dealt only with the special case of a stratified medium. Here we extend the theory developed in I to two-dimensional inhomogeneity distributions which vary both along the x and z coordinates, the latter being

[1] H. L. Frisch and J. A. Morrison, Ann. Phys. 26, 181 (1964). Equations referring to this paper are prefixed with a I.

parallel to the external magnetic field $\mathcal{H}$. The experimentally more-usually found three-dimensional inhomogeneity distributions require separate treatment. The notation and definitions employed are those of I and only the bare definitions are repeated here.

Both the theory and the principal results found by us are sufficiently similar to those obtained in I so that the discussion given in I of the underlying physics of our classical model,[2] and the nonsaturating high-field magnetoresistance which it exhibits, suffices for the purposes of this paper. Section 2 reviews tersely the two- and three-dimensional versions of our basic transport equation and certain associated equations. The principal result, found in Sec. 3, from an asymptotic (large $\mathcal{H}$) solution of the 4-moment equations (which is essentially related to the asymptotic solution of the transport equation as in I) is that the trace of the resistivity tensor in the presence of $\mathcal{H}$ increases, in general, in proportion to the square of the dimensionless product of the cyclotron frequency and the constant relaxation time, characterizing electron–phonon collisions, provided the impurity potential[1] is sufficiently smooth. An exceptional case arises, when the magnetoresistance does saturate, if the initial, unperturbed by the external electric field, local electron density is such that its z coordinate average is independent of x.

The asymptotic (large $\mathcal{H}$) solution of the 4-moment equations may be obtained in a similar, but more involved, manner when the inhomogeneity distribution is characterized by a sufficiently smooth potential $\varphi(x, z + \lambda y)$. This analysis will be given in a separate paper.

2. THE TRANSPORT AND MOMENT EQUATIONS

We first consider the three-dimensional modified Boltzmann equation satisfied by the stationary electron distribution $f(\mathbf{v}, \mathbf{x})$. As in I (2.5) this is

$$\mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{x}} + \mathbf{a}\cdot\frac{\partial f}{\partial \mathbf{v}} + \frac{q}{m}\mathbf{E}\cdot\frac{\partial f}{\partial \mathbf{v}} + \frac{q}{mc}(\mathbf{v}\times\mathbf{H})\cdot\frac{\partial f}{\partial \mathbf{v}} = \frac{1}{\tau}[n(\mathbf{x})f_M - f], \tag{2.1}$$

where

$$\mathbf{a}(\mathbf{x}) = -\frac{q}{m}\frac{\partial \psi}{\partial \mathbf{x}}; \quad \nabla^2(\psi - V) = \frac{q}{D}[\bar{n}_0 - n(\mathbf{x})], \tag{2.2}$$

and

$$f_M = (2\pi)^{-\frac{3}{2}}e^{-\mathbf{v}^2/2}; \qquad n(\mathbf{x}) = \int f(\mathbf{v}, \mathbf{x})\, d\mathbf{v}. \tag{2.3}$$

[2] The high-field magnetoresistance exhibited by a fully quantum mechanical version of the basic model introduced in I is currently under investigation.

We retain only terms up to the first order in the components of the applied small constant electric field E and set

$$f = (f_0 + f_1); \qquad n = (n_0 + n_1); \qquad \psi = (\psi_0 + \psi_1); \qquad \mathbf{a} = \mathbf{a}^{(0)} + \mathbf{a}^{(1)}, \tag{2.4}$$

where the subscript (or superscript) 0 corresponds to the value of the entity in the absence of the electric field. In (2.2) the quantity $\bar{n}_0$ is the spatial average of $n_0(\mathbf{x})$, i.e.,

$$\bar{n}_0 = \langle n_0(\mathbf{x})\rangle_x, \tag{2.5}$$

and $V(x)$ is the impurity potential. Throughout, averages with respect to spatial variables will be indicated by corresponding subscripts to the averaging bracket.

Thus, from (2.1),

$$\mathbf{v}\cdot\frac{\partial f_0}{\partial \mathbf{x}} + \mathbf{a}^{(0)}\cdot\frac{\partial f_0}{\partial \mathbf{v}} + \frac{q}{mc}(\mathbf{v}\times\mathbf{H})\cdot\frac{\partial f_0}{\partial \mathbf{v}} = \tau^{-1}[n_0(\mathbf{x})f_M - f_0], \tag{2.6}$$

the appropriate solution of which is

$$f_0 = n_0(\mathbf{x})f_M, \tag{2.7}$$

where

$$\frac{\partial n_0}{\partial \mathbf{x}} = n_0(\mathbf{x})\mathbf{a}^{(0)} = -\frac{q}{m}n_0(\mathbf{x})\frac{\partial \psi_0}{\partial \mathbf{x}}, \tag{2.8}$$

from (2.2). Hence,

$$n_0(\mathbf{x}) = \frac{\bar{n}_0 \exp[-q\psi_0(\mathbf{x})/m]}{\langle \exp[-q\psi_0(\mathbf{x})/m]\rangle_x}. \tag{2.9}$$

Also, from (2.2),

$$\nabla^2 V = \left\{\nabla^2\psi_0 - \frac{q}{D}\bar{n}_0\left[1 - \frac{\exp(-q\psi_0/m)}{\langle \exp(-q\psi_0/m)\rangle_x}\right]\right\}. \tag{2.10}$$

We assume that $\psi_0(\mathbf{x})$, rather than $V(\mathbf{x})$, is prescribed and regard (2.10) as an equation for $V(\mathbf{x})$. Of course, $\psi_0(\mathbf{x})$ cannot be entirely arbitrary if an appropriately bounded solution of (2.10) is to exist. However, we are not restricted to strictly periodic potentials $\psi_0(\mathbf{x})$, but can consider a wider class, analogous to that discussed in I, which we refer to as the random case.

In component form, we have $\mathbf{x} = (x, y, z)$ and $\mathbf{v} = (u, v, w)$. We will take the z axis to be in the direction of the magnetic field, and set

$$\mathbf{H} = (0, 0, \mathcal{H}); \qquad \Omega = q\mathcal{H}/mc. \tag{2.11}$$

Then the electric field perturbed part of the electron

distribution, f_1, satisfies

$$\left(u\frac{\partial f_1}{\partial x}+v\frac{\partial f_1}{\partial y}+w\frac{\partial f_1}{\partial z}\right)+\left[a_x^{(0)}\frac{\partial f_1}{\partial u}+a_y^{(0)}\frac{\partial f_1}{\partial v}+a_z^{(0)}\frac{\partial f_1}{\partial w}\right]+\Omega\left(v\frac{\partial f_1}{\partial u}-u\frac{\partial f_1}{\partial v}\right)+\frac{f_1}{\tau}=n_1\frac{f_M}{\tau}-\left[a_x^{(1)}\frac{\partial f_0}{\partial u}+a_y^{(1)}\frac{\partial f_0}{\partial v}+a_z^{(1)}\frac{\partial f_0}{\partial w}\right]-\frac{q}{m}\left(E_x\frac{\partial f_0}{\partial u}+E_y\frac{\partial f_0}{\partial v}+E_z\frac{\partial f_0}{\partial w}\right),\tag{2.12}$$

where, from (2.2)–(2.4),

$$\mathbf{a}^{(1)}=-\frac{q}{m}\frac{\partial\psi_1}{\partial\mathbf{x}};\qquad \nabla^2\psi_1=-\frac{q}{D}n_1;\qquad n_1=\int f_1(\mathbf{v},\mathbf{x})\,d\mathbf{v}.\tag{2.13}$$

As in I, we introduce a reduced distribution function. Thus, let

$$f_1=f_0\left(\alpha u+\beta v+\gamma w-\frac{q}{m}\psi_1+G\right),\tag{2.14}$$

where

$$\alpha=\frac{q\tau(E_x+\omega E_y)}{m(1+\omega^2)},\qquad \beta=\frac{q\tau(E_y-\omega E_x)}{m(1+\omega^2)},\qquad \gamma=\frac{q\tau}{m}E_z,\qquad \omega=\Omega\tau.\tag{2.15}$$

Then,

$$\left(\frac{n_1}{n_0}+\frac{q}{m}\psi_1\right)=\frac{1}{(2\pi)^{\frac{3}{2}}}\iiint_{-\infty}^{\infty}e^{-\frac{1}{2}(u^2+v^2+w^2)}G\,du\,dv\,dw\equiv Q,\tag{2.16}$$

and

$$\left(u\frac{\partial G}{\partial x}+v\frac{\partial G}{\partial y}+w\frac{\partial G}{\partial z}\right)+\left[a_x^{(0)}\frac{\partial G}{\partial u}+a_y^{(0)}\frac{\partial G}{\partial v}+a_z^{(0)}\frac{\partial G}{\partial w}\right]+\Omega\left(v\frac{\partial G}{\partial u}-u\frac{\partial G}{\partial v}\right)+\frac{(G-Q)}{\tau}+[\alpha a_x^{(0)}+\beta a_y^{(0)}+\gamma a_z^{(0)}]=0.\tag{2.17}$$

A quantity of particular interest is the local particle current $\mathbf{J}(\mathbf{x})$, which is given by the first-order velocity moment

$$J_i(\mathbf{x})=\int \mathrm{v}_i f(\mathbf{v},\mathbf{x})\,d\mathbf{v}=[J_i^{(0)}+J_i^{(1)}],\tag{2.18}$$

corresponding to (2.4). From (2.7) it follows that $J_i^{(0)}\equiv 0$, and, from (2.14), that

$$J_x=n_0(\mathbf{x})\left[\alpha+\frac{1}{(2\pi)^{\frac{3}{2}}}\iiint_{-\infty}^{\infty}ue^{-\frac{1}{2}(u^2+v^2+w^2)}G\,du\,dv\,dw\right],$$

$$J_y=n_0(\mathbf{x})\left[\beta+\frac{1}{(2\pi)^{\frac{3}{2}}}\iiint_{-\infty}^{\infty}ve^{-\frac{1}{2}(u^2+v^2+w^2)}G\,du\,dv\,dw\right],$$

$$J_z=n_0(\mathbf{x})\left[\gamma+\frac{1}{(2\pi)^{\frac{3}{2}}}\iiint_{-\infty}^{\infty}we^{-\frac{1}{2}(u^2+v^2+w^2)}G\,du\,dv\,dw\right].\tag{2.19}$$

The second-order velocity moment is

$$P_{ij}(\mathbf{x})=\int v_iv_jf(\mathbf{v},\mathbf{x})\,d\mathbf{v}=[P_{ij}^{(0)}+P_{ij}^{(1)}].\tag{2.20}$$

Equations for the velocity moments are obtained by taking moments of the transport equation (2.1), and the first few moment equations are given in I (2.11)–(2.14). From (2.7) it follows that

$$P_{ij}^{(0)}=n_0(x)\delta_{ij}.\tag{2.21}$$

The electric field perturbed moments $J_i^{(1)}=J_i$ and $P_{ij}^{(1)}$ satisfy, as in I (2.29a) and I (2.29b),

$$\partial J_i/\partial x_i=0,\tag{2.22}$$

and

$$\frac{\partial P_{ij}^{(1)}}{\partial x_i}-[a_j^{(1)}n_0+a_j^{(0)}n_1]-\frac{q}{m}E_jn_0-\frac{q}{mc}\epsilon_{jrs}J_rH_s+\frac{J_j}{\tau}=0,\tag{2.23}$$

with ϵ_{jrs} the three-index totally antisymmetric unit tensor.

We will be concerned in this paper with the 4-moment approximation wherein (2.23) is truncated by setting

$$P_{ij}^{(1)}=n_1(\mathbf{x})\delta_{ij}.\tag{2.24}$$

If the quantity Q defined in (2.16) is then introduced, (2.23) becomes, using (2.8) and (2.13),

$$n_0\frac{\partial Q}{\partial x_j}=\frac{q}{m}n_0E_j+\frac{q}{mc}\epsilon_{jrs}J_rH_s-\frac{J_j}{\tau}.\tag{2.25}$$

We now take $\mathbf{H}$ as in (2.11) and introduce normalized

quantities by

$$x = \tau\xi, \quad y = \tau\eta, \quad z = \tau\zeta, \quad \frac{q}{m}\psi_0 = \varphi, \quad Q = \frac{q\tau}{m}S. \tag{2.26}$$

Then, from (2.25),

$$J_x = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_x + \omega E_y) - \left(\frac{\partial S}{\partial \xi} + \omega\frac{\partial S}{\partial \eta}\right)\right],$$

$$J_y = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_y - \omega E_x) - \left(\frac{\partial S}{\partial \eta} - \omega\frac{\partial S}{\partial \xi}\right)\right],$$

$$J_z = \frac{q\tau}{m}n_0\left(E_z - \frac{\partial S}{\partial \zeta}\right), \tag{2.27}$$

where ω is given by (2.15). From (2.9),

$$n_0 = \bar{n}_0 e^{-\varphi}/\langle e^{-\varphi}\rangle_{\xi,\eta,\zeta}. \tag{2.28}$$

Finally, from (2.22) and (2.27),

$$\left(\frac{\partial^2 S}{\partial \xi^2} + \frac{\partial^2 S}{\partial \eta^2}\right) + (1+\omega^2)\frac{\partial^2 S}{\partial \zeta^2}$$
$$= \frac{\partial \varphi}{\partial \xi}\left[\left(\frac{\partial S}{\partial \xi} + \omega\frac{\partial S}{\partial \eta}\right) - (E_x + \omega E_y)\right]$$
$$+ \frac{\partial \varphi}{\partial \eta}\left[\left(\frac{\partial S}{\partial \eta} - \omega\frac{\partial S}{\partial \xi}\right) - (E_y - \omega E_x)\right]$$
$$+ (1+\omega^2)\frac{\partial \varphi}{\partial \zeta}\left(\frac{\partial S}{\partial \zeta} - E_z\right). \tag{2.29}$$

3. THE HIGH-FIELD MAGNETORESISTANCE

In the remainder of the paper we will consider the two-dimensional case where $\varphi = \varphi(\xi, \zeta)$, i.e., φ is independent of η. Moreover, we are concerned with the high-field magnetoresistance, corresponding to $\omega \gg 1$. The asymptotic solution of (2.29), with S and φ independent of η, is obtained in Appendix 1, for sufficiently smooth φ, in the form

$$S = [(E_x + \omega E_y)M + E_z N], \tag{3.1}$$

where M and N are expressed in reciprocal powers of

$$\Theta = (1+\omega^2). \tag{3.2}$$

From these results we can derive the asymptotic form of the spatial averages of the components of the local current. Note that from (2.27), in the present case,

$$J_x = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_x + \omega E_y) - \frac{\partial S}{\partial \xi}\right],$$

$$J_y = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_y - \omega E_x) + \omega\frac{\partial S}{\partial \xi}\right]. \tag{3.3}$$

Hence, with φ independent of η,

$$\langle J_y\rangle_{\xi,\zeta} = \left[\frac{q\tau\bar{n}_0}{m}E_y - \omega\langle J_x\rangle_{\xi,\zeta}\right]. \tag{3.4}$$

Also, from (2.27), (2.28), and (3.1)–(3.3),

$$J_x = \frac{q\tau\bar{n}_0 e^{-\varphi}}{m\Theta\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left[(E_x + \omega E_y)\left(1 - \frac{\partial M}{\partial \xi}\right) - E_z\frac{\partial N}{\partial \xi}\right],$$

$$J_z = \frac{q\tau\bar{n}_0 e^{-\varphi}}{m\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left[E_z\left(1 - \frac{\partial N}{\partial \zeta}\right) - (E_x + \omega E_y)\frac{\partial M}{\partial \zeta}\right]. \tag{3.5}$$

Let us express the spatial averages of J_x and J_z in the form

$$\langle J_x\rangle_{\xi,\zeta} = \frac{q\tau\bar{n}_0}{m\Theta}[a(E_x + \omega E_y) + bE_z],$$

$$\langle J_z\rangle_{\xi,\zeta} = \frac{q\tau\bar{n}_0}{m}\left[\frac{c}{\Theta}(E_x + \omega E_y) + dE_z\right]. \tag{3.6}$$

Inverting the relationships in (3.4) and (3.6),

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \varrho\cdot\begin{bmatrix} \langle J_x\rangle_{\xi,\zeta} \\ \langle J_y\rangle_{\xi,\zeta} \\ \langle J_z\rangle_{\xi,\zeta} \end{bmatrix}, \tag{3.7}$$

where

$$\varrho(\omega) = \frac{m}{q\tau\bar{n}_0}\begin{bmatrix} \left(\frac{d}{\Delta} - \omega^2\right) & -\omega & -\frac{b}{\Theta\Delta} \\ \omega & 1 & 0 \\ -\frac{c}{\Theta\Delta} & 0 & \frac{a}{\Theta\Delta} \end{bmatrix}, \tag{3.8}$$

with

$$\Delta = (ad/\Theta - bc/\Theta^2). \tag{3.9}$$

The magnetoresistance ratio in direction i is

$$[\varrho(\omega) - \varrho(0)]_{ii}/[\varrho(0)]_{ii}. \tag{3.10}$$

We do not consider here the calculation of $\varrho(0)$, but do find the asymptotic behavior of $\varrho(\omega)$ for $\omega \gg 1$. Since, as will be seen, the quantity $[\varrho(\omega)]_{zz}$ is (generally) asymptotically proportional to ω^2, we are then able to conclude that the magnetoresistance does not saturate, but rather increases as the square of the strength of the magnetic field. Note that, from (3.8), the magnetoresistance ratio in the y direction vanishes identically.

Now, from (3.5) and (3.6),

$$a = \frac{1}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left\langle e^{-\varphi}\left(1 - \frac{\partial M}{\partial \xi}\right)\right\rangle_{\xi,\zeta},$$

$$b = \frac{-1}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left\langle e^{-\varphi}\frac{\partial N}{\partial \xi}\right\rangle_{\xi,\zeta},$$

$$c = \frac{-\Theta}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left\langle e^{-\varphi}\frac{\partial M}{\partial \zeta}\right\rangle_{\xi,\zeta}, \tag{3.11}$$

$$d = \frac{1}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left\langle e^{-\varphi}\left(1 - \frac{\partial N}{\partial \zeta}\right)\right\rangle_{\xi,\zeta}.$$

Turning to the results of Appendix 1, from (A1.6), (A1.9), (A1.10), (A1.14), and (A1.16),

$$b = O(1), \qquad d = [[\langle e^{-\varphi}\rangle_{\xi,\zeta}]^{-1}\langle[\langle e^{\varphi}\rangle_{\zeta}]^{-1}\rangle_{\xi} + O(1/\Theta)], \tag{3.12}$$

and the leading term in d does not vanish. Also, from (A1.5), (A1.22), (A1.25), and (A1.26),

$$c = O(1), \qquad a = \{[\langle e^{-\varphi}\rangle_{\xi,\zeta}\langle[\langle e^{-\varphi}\rangle_{\zeta}]^{-1}\rangle_{\xi}]^{-1} + O(1/\Theta)\}, \tag{3.13}$$

and the leading term in a does not vanish. Thus, from (3.9),

$$\Delta = [ad/\Theta + O(1/\Theta^2)], \qquad a/\Theta\Delta = O(1), \qquad d/\Delta = [\Theta/a + O(1)]. \tag{3.14}$$

Hence, from (3.2), (3.8), (3.13), and (3.14),

$$\frac{q\tau\tilde{n}_0}{m}[\varrho(\omega)]_{xx} = \{\omega^2[\langle\langle e^{-\varphi}\rangle_{\zeta}\rangle_{\xi}\langle[\langle e^{-\varphi}\rangle_{\zeta}]^{-1}\rangle_{\xi} - 1] + O(1)\}, \tag{3.15}$$

and $[\varrho(\omega)]_{zz} = O(1)$.

The result of (3.15) reduces to that obtained in I when φ is independent of ζ, and in fact $e^{-\varphi}$ in that case is merely replaced by $\langle e^{-\varphi}\rangle_{\zeta}$. The coefficient of ω^2 in (3.15) does not vanish in general, although it does if $\langle e^{-\varphi}\rangle_{\zeta}$ is independent of ξ. It is interesting to note that such is the case, in particular, if $\varphi = \varphi(\zeta + \lambda\xi)$, i.e., for a stratified medium in which the magnetic field is not parallel to the sheets of inhomogeneities comprising the stratification, and then the magnetoresistance saturates. Incidentally, an exact solution to (2.29) of the form $S = S(\zeta + \lambda\xi)$ is easily obtained when $\varphi = \varphi(\zeta + \lambda\xi)$. In general, though, the magnetoresistance ratio in the x direction is, from (3.10) and (3.15), asymptotically proportional to

$$\omega^2[\langle e^{-\varphi}\rangle_{\xi,\zeta}\langle[\langle e^{-\varphi}\rangle_{\zeta}]^{-1}\rangle_{\xi} - 1] = \omega^2[\langle\tilde{n}_0/\langle n_0(x,z)\rangle_z\rangle_x - 1], \tag{3.16}$$

from (2.26) and (2.28), φ being independent of y.

There are strong indications that the same asymptotic magnetoresistance effect is derivable from the transport equation as is given by (3.16) from the 4-moment approximation. Thus, in Appendices 2 and 3 we consider the asymptotic solution of the equation for the reduced distribution function, with $\varphi = \varphi(\xi, \zeta)$, i.e., φ independent of η. At certain stages of the asymptotic development (see Appendix 3) we make certain assumptions about the forms of the solutions of two partial integro-differential equations, and leave open the question of the consistency of these assumptions with the determination of subsequent terms in the asymptotic expansion. We remark that (3.4), which was derived in the 4-moment approximation, is actually an exact consequence of the linearized transport equation, when φ is independent of η, and is derivable from (2.23), as shown in Appendix 2. We retain the definitions of (3.6), so that the relations (3.7)–(3.9) still hold. Then, from (A2.20) and (A3.31),

$$b = O(1), \qquad c = O(1), \qquad d = O(1), \tag{3.17}$$

and

$$a = \{[\langle e^{-\varphi}\rangle_{\xi,\zeta}\langle[\langle e^{-\varphi}\rangle_{\zeta}]^{-1}\rangle_{\xi}]^{-1} + O(1/\omega^2)\}. \tag{3.18}$$

From (3.2), (3.8), and (3.9) it follows that (3.14) and (3.15) again hold, thus giving the same asymptotic magnetoresistance effect, subject, of course, to the validity of the above mentioned assumptions.

ACKNOWLEDGMENTS

We are indebted to Dr. H. Grad and Dr. C. Herring for helpful discussions.

APPENDIX 1. ASYMPTOTIC SOLUTION OF THE 4-MOMENT PROBLEM

We are here concerned with the asymptotic solution, for $\omega \gg 1$, of the equation

$$\frac{\partial^2 S}{\partial\xi^2} + (1+\omega^2)\frac{\partial^2 S}{\partial\zeta^2} = \frac{\partial\varphi}{\partial\xi}\left[\frac{\partial S}{\partial\xi} - (E_x + \omega E_y)\right] + (1+\omega^2)\frac{\partial\varphi}{\partial\zeta}\left(\frac{\partial S}{\partial\zeta} - E_z\right). \tag{A1.1}$$

Let

$$S = [(E_x + \omega E_y)M + E_z N], \quad \Theta = (1+\omega^2). \tag{A1.2}$$

Then

$$\frac{\partial^2 M}{\partial\xi^2} + \Theta\frac{\partial^2 M}{\partial\zeta^2} = \frac{\partial\varphi}{\partial\xi}\left(\frac{\partial M}{\partial\xi} - 1\right) + \Theta\frac{\partial\varphi}{\partial\zeta}\frac{\partial M}{\partial\zeta}, \tag{A1.3}$$

$$\frac{\partial^2 N}{\partial\xi^2} + \Theta\frac{\partial^2 N}{\partial\zeta^2} = \frac{\partial\varphi}{\partial\xi}\frac{\partial N}{\partial\xi} + \Theta\frac{\partial\varphi}{\partial\zeta}\left(\frac{\partial N}{\partial\zeta} - 1\right). \tag{A1.4}$$

We will seek asymptotic expansions of the form

$$M = \left(M_0 + \frac{M_1}{\Theta} + \frac{M_2}{\Theta^2} + \cdots\right), \tag{A1.5}$$

$$N = \left(N_0 + \frac{N_1}{\Theta} + \frac{N_2}{\Theta^2} + \cdots\right). \tag{A1.6}$$

We first consider (A1.4). Then, from (A1.6),

$$\frac{\partial^2 N_0}{\partial\zeta^2} = \frac{\partial\varphi}{\partial\zeta}\left(\frac{\partial N_0}{\partial\zeta} - 1\right); \quad \frac{\partial}{\partial\zeta}\left[e^{-\varphi}\left(\frac{\partial N_0}{\partial\zeta} - 1\right)\right] = 0. \tag{A1.7}$$

It follows that

$$e^{-\varphi}(\partial N_0/\partial\zeta - 1) = V_0(\xi); \qquad \partial N_0/\partial\zeta = [1 + V_0(\xi)e^{\varphi}]. \tag{A1.8}$$

Averaging over ζ we find

$$V_0(\xi)\langle e^{\varphi}\rangle_\zeta = -1; \quad \partial N_0/\partial\zeta = [1 - e^{\varphi}/\langle e^{\varphi}\rangle_\zeta]. \tag{A1.9}$$

Hence

$$N_0 = [\rho_0(\xi, \zeta) + W_0(\xi)]; \qquad \rho_0 = \int_0^\zeta [1 - e^{\varphi}/\langle e^{\varphi}\rangle_\zeta]\, d\zeta. \tag{A1.10}$$

Note that in the random case, conditions are imposed on φ in order that ρ_0 should remain bounded as $\zeta \to \pm\infty$. Similar conditions arise at each subsequent stage of the analysis. Having determined N_0 to within a function of ξ, we proceed by iteration.

Thus, for $n \geq 0$, from (A1.4) and (A1.6),

$$\left(\frac{\partial^2 N_{n+1}}{\partial\zeta^2} - \frac{\partial\varphi}{\partial\zeta}\frac{\partial N_{n+1}}{\partial\zeta}\right) = \left(\frac{\partial\varphi}{\partial\xi}\frac{\partial N_n}{\partial\xi} - \frac{\partial^2 N_n}{\partial\xi^2}\right). \tag{A1.11}$$

This may be written in the form

$$\frac{\partial}{\partial\zeta}\left(e^{-\varphi}\frac{\partial N_{n+1}}{\partial\zeta}\right) = -\frac{\partial}{\partial\xi}\left(e^{-\varphi}\frac{\partial N_n}{\partial\xi}\right). \tag{A1.12}$$

Suppose that we have

$$N_n = [\rho_n(\xi, \zeta) + W_n(\xi)], \tag{A1.13}$$

where $W_n(\xi)$ is as yet undetermined. Now, from (A1.12),

$$\frac{\partial}{\partial\xi}\left\langle e^{-\varphi}\frac{\partial N_n}{\partial\xi}\right\rangle_\zeta = 0; \quad \left\langle e^{-\varphi}\frac{\partial N_n}{\partial\xi}\right\rangle_\zeta = C_n, \tag{A1.14}$$

where C_n is a constant. Hence, from (A1.13),

$$\langle e^{-\varphi}\rangle_\zeta \frac{dW_n}{d\xi} + \left\langle e^{-\varphi}\frac{\partial\rho_n}{\partial\xi}\right\rangle_\zeta = C_n; \qquad \frac{dW_n}{d\xi} = [\langle e^{-\varphi}\rangle_\zeta]^{-1}\left[C_n - \left\langle e^{-\varphi}\frac{\partial\rho_n}{\partial\xi}\right\rangle_\zeta\right]. \tag{A1.15}$$

Averaging over ξ we find

$$C_n\langle[\langle e^{-\varphi}\rangle_\zeta]^{-1}\rangle_\xi = \left\langle\left\{[\langle e^{-\varphi}\rangle_\zeta]^{-1}\left\langle e^{-\varphi}\frac{\partial\rho_n}{\partial\xi}\right\rangle_\zeta\right\}\right\rangle_\xi. \tag{A1.16}$$

Thus $W_n(\xi)$ is determined to within an irrelevant constant.

From (A1.12),

$$\frac{\partial N_{n+1}}{\partial\zeta} = e^{\varphi}\left[V_{n+1}(\xi) - \int_0^\zeta \frac{\partial}{\partial\xi}\left(e^{-\varphi}\frac{\partial N_n}{\partial\xi}\right)d\zeta\right]. \tag{A1.17}$$

Averaging over ζ we find

$$\langle e^{\varphi}\rangle_\zeta V_{n+1}(\xi) = \left\langle e^{\varphi}\int_0^\zeta \frac{\partial}{\partial\xi}\left(e^{-\varphi}\frac{\partial N_n}{\partial\xi}\right)d\zeta\right\rangle_\zeta. \tag{A1.18}$$

Then integrating (A1.17) we obtain

$$N_{n+1} = [\rho_{n+1}(\xi, \zeta) + W_{n+1}(\xi)], \tag{A1.19}$$

where $W_{n+1}(\xi)$ is as yet unknown, and $\rho_{n+1}(\xi, \zeta)$ is explicitly determined in terms of $\rho_n(\xi, \zeta)$, from (A1.13) and (A1.15)–(A1.18).

We now turn to (A1.3). Then, from (A1.5),

$$\frac{\partial^2 M_0}{\partial\zeta^2} = \frac{\partial\varphi}{\partial\zeta}\frac{\partial M_0}{\partial\zeta}; \qquad \frac{\partial}{\partial\zeta}\left(e^{-\varphi}\frac{\partial M_0}{\partial\zeta}\right) = 0. \tag{A1.20}$$

It follows that

$$\partial M_0/\partial\zeta = T_0(\xi)e^{\varphi}; \qquad T_0(\xi)\langle e^{\varphi}\rangle_\zeta = 0. \tag{A1.21}$$

Hence,

$$\partial M_0/\partial\zeta = 0; \qquad M_0 = U_0(\xi). \tag{A1.22}$$

From (A1.3) and (A1.5), the next term in the asymptotic expansion satisfies

$$\left(\frac{\partial^2 M_1}{\partial\zeta^2} - \frac{\partial\varphi}{\partial\zeta}\frac{\partial M_1}{\partial\zeta}\right) = \left[\frac{\partial\varphi}{\partial\xi}\left(\frac{\partial M_0}{\partial\xi} - 1\right) - \frac{\partial^2 M_0}{\partial\xi^2}\right], \tag{A1.23}$$

or, using (A1.22),

$$\frac{\partial}{\partial\zeta}\left(e^{-\varphi}\frac{\partial M_1}{\partial\zeta}\right) = -\frac{\partial}{\partial\xi}\left[e^{-\varphi}\left(\frac{dU_0}{d\xi} - 1\right)\right]. \tag{A1.24}$$

Thus, averaging over ζ,

$$\frac{\partial}{\partial\xi}\left[\left(\frac{dU_0}{d\xi} - 1\right)\langle e^{-\varphi}\rangle_\zeta\right] = 0; \quad \left(\frac{dU_0}{d\xi} - 1\right)\langle e^{-\varphi}\rangle_\zeta = D_0, \tag{A1.25}$$

where D_0 is a constant. Hence,

$$\frac{dU_0}{d\xi} = \left[1 + \frac{D_0}{\langle e^{-\varphi}\rangle_\zeta}\right]; \quad D_0\langle[\langle e^{-\varphi}\rangle_\zeta]^{-1}\rangle_\xi = -1, \tag{A1.26}$$

averaging over ξ.

If we let

$$\mu(\xi, \zeta) = \int_0^\zeta \frac{\partial}{\partial\xi}\left[e^{-\varphi}\left(\frac{dU_0}{d\xi} - 1\right)\right]d\zeta, \tag{A1.27}$$

then (A1.24) gives

$$\partial M_1/\partial\zeta = e^{\varphi}[T_1(\xi) - \mu(\xi, \zeta)]. \tag{A1.28}$$

Averaging over ζ we find

$$T_1(\xi)\langle e^{\varphi}\rangle_\zeta = \langle\mu(\xi, \zeta)e^{\varphi}\rangle_\zeta. \tag{A1.29}$$

Hence, from (A1.28),

$$M_1 = [\nu_1(\xi, \zeta) + U_1(\xi)], \tag{A1.30}$$

where $\nu_1(\xi, \zeta)$ is given explicitly in terms of φ, from (A1.26)–(A1.29), and $U_1(\xi)$ is as yet undetermined. We now proceed by iteration. Thus, for $n \geq 1$, from (A1.3) and (A1.5),

$$\left(\frac{\partial^2 M_{n+1}}{\partial\zeta^2} - \frac{\partial\varphi}{\partial\zeta}\frac{\partial M_{n+1}}{\partial\zeta}\right) = \left(\frac{\partial\varphi}{\partial\xi}\frac{\partial M_n}{\partial\xi} - \frac{\partial^2 M_n}{\partial\xi^2}\right). \tag{A1.31}$$

We suppose that

$$M_n = [\nu_n(\xi, \zeta) + U_n(\xi)], \tag{A1.32}$$

where $U_n(\xi)$ is as yet undetermined. The iteration proceeds just as in the case for N_{n+1}.

APPENDIX 2. ASYMPTOTIC SOLUTION OF THE TRANSPORT EQUATION

We are here concerned with the asymptotic solution, for $\omega \gg 1$, of the equation for the reduced distribution function G, when $\varphi = \varphi(\xi, \zeta)$, i.e., φ is independent of η. From (2.8), (2.15), (2.17), and (2.26),

$$\left(u\frac{\partial G}{\partial \xi} + w\frac{\partial G}{\partial \zeta}\right) - \left(\frac{\partial \varphi}{\partial \xi}\frac{\partial G}{\partial u} + \frac{\partial \varphi}{\partial \zeta}\frac{\partial G}{\partial w}\right) + (G - Q) + \omega\left(v\frac{\partial G}{\partial u} - u\frac{\partial G}{\partial v}\right) - \left(\alpha\frac{\partial \varphi}{\partial \xi} + \gamma\frac{\partial \varphi}{\partial \zeta}\right) = 0, \tag{A2.1}$$

where Q is given by (2.16). Under the transformation

$$u = r\sin\theta, \qquad v = -r\cos\theta, \tag{A2.2}$$

this becomes

$$\omega\frac{\partial G}{\partial \theta} = (G - Q) + w\frac{\partial G}{\partial \zeta} - \frac{\partial \varphi}{\partial \zeta}\left(\gamma + \frac{\partial G}{\partial w}\right) + r\sin\theta\frac{\partial G}{\partial \xi} - \frac{\partial \varphi}{\partial \xi}\left(\alpha + \frac{\cos\theta}{r}\frac{\partial G}{\partial \theta} + \sin\theta\frac{\partial G}{\partial r}\right). \tag{A2.3}$$

Now let

$$G = [\alpha\omega^2 g^{(\alpha)} + \gamma g^{(\gamma)}]. \tag{A2.4}$$

Then,

$$\omega\frac{\partial g^{(\alpha)}}{\partial \theta} = (g^{(\alpha)} - q^{(\alpha)}) + w\frac{\partial g^{(\alpha)}}{\partial \zeta} - \frac{\partial \varphi}{\partial \zeta}\frac{\partial g^{(\alpha)}}{\partial w} + r\sin\theta\frac{\partial g^{(\alpha)}}{\partial \xi} - \frac{\partial \varphi}{\partial \xi}\left(\frac{1}{\omega^2} + \frac{\cos\theta}{r}\frac{\partial g^{(\alpha)}}{\partial \theta} + \sin\theta\frac{\partial g^{(\alpha)}}{\partial r}\right), \tag{A2.5}$$

$$\omega\frac{\partial g^{(\gamma)}}{\partial \theta} = (g^{(\gamma)} - q^{(\gamma)}) + w\frac{\partial g^{(\gamma)}}{\partial \zeta} - \frac{\partial \varphi}{\partial \zeta}\left(1 + \frac{\partial g^{(\gamma)}}{\partial w}\right) + r\sin\theta\frac{\partial g^{(\gamma)}}{\partial \xi} - \frac{\partial \varphi}{\partial \xi}\left(\frac{\cos\theta}{r}\frac{\partial g^{(\gamma)}}{\partial \theta} + \sin\theta\frac{\partial g^{(\gamma)}}{\partial r}\right), \tag{A2.6}$$

where, from (2.16) and (A2.2)–(A2.4),

$$q^{(\alpha,\gamma)} = \frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{2\pi}\int_0^{\infty}\int_{-\infty}^{\infty} re^{-\frac{1}{2}(r^2+w^2)}g^{(\alpha,\gamma)}\,dw\,dr\,d\theta. \tag{A2.7}$$

From (2.15), (2.19), (2.28), (A2.2), and (A2.4),

$$J_x = \frac{\bar{n}_0 e^{-\varphi}}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}[\alpha(1 + \omega^2\sigma^{(\alpha)}) + \gamma\sigma^{(\gamma)}],$$

$$J_v = \frac{\bar{n}_0 e^{-\varphi}}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}\left[\frac{q\tau}{m}E_v + \alpha\omega(\omega\rho^{(\alpha)} - 1) + \gamma\rho^{(\gamma)}\right],$$

$$J_z = \frac{\bar{n}_0 e^{-\varphi}}{\langle e^{-\varphi}\rangle_{\xi,\zeta}}[\alpha\omega^2\nu^{(\alpha)} + \gamma(1 + \nu^{(\gamma)})], \tag{A2.8}$$

where

$$\begin{Bmatrix}\sigma^{(\alpha,\gamma)}\\ \rho^{(\alpha,\gamma)}\\ \nu^{(\alpha,\gamma)}\end{Bmatrix} = \frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{2\pi}\int_0^{\infty}\int_{-\infty}^{\infty} r\begin{Bmatrix} r\sin\theta\\ -r\cos\theta\\ w\end{Bmatrix} \times e^{-\frac{1}{2}(r^2+w^2)}g^{(\alpha,\gamma)}\,dw\,dr\,d\theta. \tag{A2.9}$$

From (2.8), (2.11), (2.15), (2.22), (2.23), and (2.26),

$$\partial J_x/\partial\xi + \partial J_z/\partial\zeta = 0, \tag{A2.10}$$

$$\partial P_{xv}/\partial\xi + \partial P_{vz}/\partial\zeta = \left[\frac{q\tau}{m}E_v n_0 - (\omega J_x + J_v)\right]. \tag{A2.11}$$

Averaging (A2.11) over ξ and ζ leads to (3.4). Hence, from (A2.8) and (A2.10),

$$\frac{\partial}{\partial\xi}\left[e^{-\varphi}\left(\sigma^{(\alpha)} + \frac{1}{\omega^2}\right)\right] + \frac{\partial}{\partial\zeta}(e^{-\varphi}\nu^{(\alpha)}) = 0, \tag{A2.12}$$

$$\frac{\partial}{\partial\xi}(e^{-\varphi}\sigma^{(\gamma)}) + \frac{\partial}{\partial\zeta}[e^{-\varphi}(1 + \nu^{(\gamma)})] = 0, \tag{A2.13}$$

and, using (3.4),

$$\omega\langle e^{-\varphi}\sigma^{(\alpha,\gamma)}\rangle_{\xi,\zeta} + \langle e^{-\varphi}\rho^{(\alpha,\gamma)}\rangle_{\xi,\zeta} = 0. \tag{A2.14}$$

We now seek asymptotic solutions to (A2.5) and (A2.6) of the form

$$g^{(\alpha,\gamma)} = g_0^{(\alpha,\gamma)} + \frac{1}{\omega}g_1^{(\alpha,\gamma)} + \frac{1}{\omega^2}g_2^{(\alpha,\gamma)} + \cdots, \tag{A2.15}$$

with similar expansions for $q^{(\alpha,\gamma)}$, $\sigma^{(\alpha,\gamma)}$, $\rho^{(\alpha,\gamma)}$, and $\nu^{(\alpha,\gamma)}$, and first consider (A2.6). Thus,

$$\partial g_0^{(\gamma)}/\partial\theta = 0; \qquad g_0^{(\gamma)} = e_0^{(\gamma)}(\xi, \zeta; r, w). \tag{A2.16}$$

From (A2.9) it follows that

$$\sigma_0^{(\gamma)} = 0 = \rho_0^{(\gamma)}, \tag{A2.17}$$

and then from (A2.13) that

$$(1 + \nu_0^{(\gamma)}) = A_0^{(\gamma)}(\xi)e^{\varphi}. \tag{A2.18}$$

Returning to (A2.6),

$$\frac{\partial g_1^{(\gamma)}}{\partial\theta} = (e_0^{(\gamma)} - q_0^{(\gamma)}) + w\frac{\partial e_0^{(\gamma)}}{\partial\zeta} - \frac{\partial\varphi}{\partial\zeta}\left(1 + \frac{\partial e_0^{(\gamma)}}{\partial w}\right) + \sin\theta\left(r\frac{\partial e_0^{(\gamma)}}{\partial\xi} - \frac{\partial\varphi}{\partial\xi}\frac{\partial e_0^{(\gamma)}}{\partial r}\right). \tag{A2.19}$$

From single-valuedness, $\int_0^{2\pi} \partial g_1^{(\gamma)}/\partial\theta\, d\theta = 0$. Hence,

$$(e_0^{(\gamma)} - q_0^{(\gamma)}) + w\frac{\partial e_0^{(\gamma)}}{\partial\zeta} - \frac{\partial\varphi}{\partial\zeta}\left(1 + \frac{\partial e_0^{(\gamma)}}{\partial w}\right) = 0, \tag{A2.20}$$

and

$$g_1^{(\gamma)} = \left[e_1^{(\gamma)}(\xi, \zeta; r, w) + \cos\theta\left(\frac{\partial\varphi}{\partial\xi}\frac{\partial e_0^{(\gamma)}}{\partial r} - r\frac{\partial e_0^{(\gamma)}}{\partial\xi}\right)\right]. \tag{A2.21}$$

From (A2.9) it follows that

$$\sigma_1^{(\gamma)} = 0, \tag{A2.22}$$

which is consistent with (A2.14) in view of (A2.17). Note that (A2.18) is a consequence of (A2.20).

Next, from (A2.6) and (A2.21), one obtains an expression for $\partial g_2^{(\gamma)}/\partial\theta$ which is rather lengthy. We consequently omit the details, but remark that some simplification is obtained by making use of the partial derivatives with respect to r and ξ of (A2.20). From the single-valuedness condition $\int_0^{2\pi} \partial g_2^{(\gamma)}/\partial\theta\, d\theta = 0$ is obtained the equation

$$(e_1^{(\gamma)} - q_1^{(\gamma)}) + w\frac{\partial e_1^{(\gamma)}}{\partial\zeta} - \frac{\partial\varphi}{\partial\zeta}\frac{\partial e_1^{(\gamma)}}{\partial w} = 0. \tag{A2.23}$$

Also, the coefficient of $\sin\theta$ in $g_2^{(\gamma)}$ is found to be

$$\left\{\frac{\partial^2\varphi}{\partial\xi\partial\zeta}\left[w\frac{\partial e_0^{(\gamma)}}{\partial r} - r\left(1 + \frac{\partial e_0^{(\gamma)}}{\partial w}\right)\right] - r\frac{\partial q_0^{(\gamma)}}{\partial\xi}\right\}. \tag{A2.24}$$

From (A2.9) and (A2.16) it then follows, after some integrations by parts, that

$$\sigma_2^{(\gamma)} = -\left[\frac{\partial q_0^{(\gamma)}}{\partial\xi} + \frac{\partial^2\varphi}{\partial\xi\partial\zeta}(1 + \nu_0^{(\gamma)})\right]. \tag{A2.25}$$

Hence, from (A2.13) and (A2.18),

$$\frac{\partial}{\partial\zeta}(e^{-\varphi}\nu_2^{(\gamma)}) = \frac{\partial}{\partial\xi}\left[e^{-\varphi}\frac{\partial q_0^{(\gamma)}}{\partial\xi} + A_0^{(\gamma)}(\xi)\frac{\partial^2\varphi}{\partial\xi\partial\zeta}\right]. \tag{A2.26}$$

Averaging (A2.26) over ζ it follows that

$$\frac{\partial}{\partial\xi}\left\langle e^{-\varphi}\frac{\partial q_0^{(\gamma)}}{\partial\xi}\right\rangle_\zeta = 0; \quad \left\langle e^{-\varphi}\frac{\partial q_0^{(\gamma)}}{\partial\xi}\right\rangle_\zeta = B_0^{(\gamma)}, \tag{A2.27}$$

where $B_0^{(\gamma)}$ is a constant.

Thus, (A2.20) is a partial integrodifferential equation for $e_0^{(\gamma)}$, wherein $q_0^{(\gamma)}$ is expressed in terms of $e_0^{(\gamma)}$ by means of (A2.7) and (A2.16), and in addition $q_0^{(\gamma)}$ satisfies the condition of (A2.27). For our purposes it is not necessary to solve for $e_0^{(\gamma)}$, nor to carry the asymptotic expansion any further. We are concerned with the spatial averages of the components of the local current, and retain the definition of (3.6). Hence, from (2.15), (3.2), (3.6), and (A2.8)

$$b = (1 + \omega^2)\langle e^{-\varphi}\sigma^{(\gamma)}\rangle_{\xi,\zeta}/\langle e^{-\varphi}\rangle_{\xi,\zeta}; \qquad d = [1 + \langle e^{-\varphi}\nu^{(\gamma)}\rangle_{\xi,\zeta}/\langle e^{-\varphi}\rangle_{\xi,\zeta}]. \tag{A2.28}$$

Then, from (A2.17), (A2.18), and (A2.22),

$$b = O(1), \qquad d = [\langle A_0^{(\gamma)}(\xi)\rangle_\xi/\langle e^{-\varphi}\rangle_{\xi,\zeta} + O(1/\omega)]. \tag{A2.29}$$

The determination of $A_0^{(\gamma)}(\xi)$ involves $e_0^{(\gamma)}$. However, we do not need to know it explicitly, but do assume that $\langle A_0^{(\gamma)}(\xi)\rangle_\xi \neq 0$. We consider the asymptotic solution of (A2.5) in the next Appendix.

APPENDIX 3. ASYMPTOTIC SOLUTION OF THE TRANSPORT EQUATION

Turning to the asymptotic solution of (A2.5) we have, from (A2.15),

$$\partial g_0^{(\alpha)}/\partial\theta = 0; \qquad g_0^{(\alpha)} = e_0^{(\alpha)}(\xi, \zeta; r, w). \tag{A3.1}$$

From (A2.9) it follows

$$\sigma_0^{(\alpha)} = 0 = \rho_0^{(\alpha)}, \tag{A3.2}$$

and then from (A2.12) that

$$\nu_0^{(\alpha)} = A_0^{(\alpha)}(\xi)e^{\varphi}. \tag{A3.3}$$

Returning to (A2.5),

$$\frac{\partial g_1^{(\alpha)}}{\partial\theta} = (e_0^{(\alpha)} - q_0^{(\alpha)}) + w\frac{\partial e_0^{(\alpha)}}{\partial\zeta} - \frac{\partial\varphi}{\partial\zeta}\frac{\partial e_0^{(\alpha)}}{\partial w} + \sin\theta\left(r\frac{\partial e_0^{(\alpha)}}{\partial\xi} - \frac{\partial\varphi}{\partial\xi}\frac{\partial e_0^{(\alpha)}}{\partial r}\right). \tag{A3.4}$$

From single valuedness we obtain

$$(e_0^{(\alpha)} - q_0^{(\alpha)}) + w\frac{\partial e_0^{(\alpha)}}{\partial\zeta} - \frac{\partial\varphi}{\partial\zeta}\frac{\partial e_0^{(\alpha)}}{\partial w} = 0. \tag{A3.5}$$

Then,

$$g_1^{(\alpha)} = \left[e_1^{(\alpha)}(\xi, \zeta; r, w) + \cos\theta\left(\frac{\partial\varphi}{\partial\xi}\frac{\partial e_0^{(\alpha)}}{\partial r} - r\frac{\partial e_0^{(\alpha)}}{\partial\xi}\right)\right]. \tag{A3.6}$$

From (A2.9) it follows that

$$\sigma_1^{(\alpha)} = 0, \tag{A3.7}$$

which is consistent with (A2.14) in view of (A3.2). Then, from (A2.12) and (A3.7),

$$\nu_1^{(\alpha)} = A_1^{(\alpha)}(\xi)e^{\varphi}. \tag{A3.8}$$

Next, from (A2.5) and (A3.6), one obtains an expression for $\partial g_2^{(\alpha)}/\partial\theta$, but as before we omit the details. From the single-valuedness condition

$$\int_0^{2\pi} \frac{\partial g_2^{(\alpha)}}{\partial\theta}\, d\theta = 0$$

is obtained the equation

$$(e_1^{(\alpha)} - q_1^{(\alpha)}) + w\frac{\partial e_1^{(\alpha)}}{\partial \zeta} - \frac{\partial \varphi}{\partial \zeta}\frac{\partial e_1^{(\alpha)}}{\partial w} = 0. \tag{A3.9}$$

Also, the coefficient of $\sin\theta$ in $g_2^{(\alpha)}$ is found to be

$$\left[\frac{\partial^2 \varphi}{\partial \xi \partial \zeta}\left(w\frac{\partial e_0^{(\alpha)}}{\partial r} - r\frac{\partial e_0^{(\alpha)}}{\partial w}\right) - r\frac{\partial q_0^{(\alpha)}}{\partial \xi}\right]. \tag{A3.10}$$

Hence, from (A2.9) and (A3.1), after some integrations by parts,

$$\sigma_2^{(\alpha)} = -\left(\frac{\partial q_0^{(\alpha)}}{\partial \xi} + \frac{\partial^2 \varphi}{\partial \xi \partial \zeta}\nu_0^{(\alpha)}\right), \tag{A3.11}$$

and from (A2.12) and (A3.3),

$$\frac{\partial}{\partial \zeta}(e^{-\varphi}\nu_2^{(\alpha)}) = \frac{\partial}{\partial \xi}\left[e^{-\varphi}\left(\frac{\partial q_0^{(\alpha)}}{\partial \xi} - 1\right) + A_0^{(\alpha)}(\xi)\frac{\partial^2 \varphi}{\partial \xi \partial \zeta}\right]. \tag{A3.12}$$

Averaging (A3.12) over ζ it follows that

$$\begin{aligned} &\frac{\partial}{\partial \xi}\left\langle e^{-\varphi}\left(\frac{\partial q_0^{(\alpha)}}{\partial \xi} - 1\right)\right\rangle_\zeta = 0; \\ &\left\langle e^{-\varphi}\left(\frac{\partial q_0^{(\alpha)}}{\partial \xi} - 1\right)\right\rangle_\zeta = B_0^{(\alpha)}, \end{aligned} \tag{A3.13}$$

where $B_0^{(\alpha)}$ is a constant.

It is at this stage that we make an assumption regarding the solution to (A3.5), namely that $e_0^{(\alpha)}$ is a function of ξ alone,

$$e_0^{(\alpha)} = m_0^{(\alpha)}(\xi). \tag{A3.14}$$

Then, from (A2.7) and (A3.1),

$$q_0^{(\alpha)} = m_0^{(\alpha)}(\xi), \tag{A3.15}$$

so that (A3.14) is certainly a solution of (A3.5). Now, from (A3.13) we obtain

$$\begin{aligned} &(dm_0^{(\alpha)}/d\xi - 1)\langle e^{-\varphi}\rangle_\zeta = B_0^{(\alpha)}; \\ &dm_0^{(\alpha)}/d\xi = [1 + B_0^{(\alpha)}/\langle e^{-\varphi}\rangle_\zeta]. \end{aligned} \tag{A3.16}$$

Averaging over ξ,

$$B_0^{(\alpha)}\langle[\langle e^{-\varphi}\rangle_\zeta]^{-1}\rangle_\xi = -1. \tag{A3.17}$$

From (A2.9), (A3.1), (A3.11), (A3.14), and (A3.15),

$$\nu_0^{(\alpha)} = 0; \qquad \sigma_2^{(\alpha)} = -dm_0^{(\alpha)}/d\xi. \tag{A3.18}$$

We now proceed with the asymptotic expansion, under the assumption of (A3.14). In this case,

$$g_2^{(\alpha)} = \left[e_2^{(\alpha)}(\xi, \zeta; r, w) + \cos\theta\left(\frac{\partial \varphi}{\partial \xi}\frac{\partial e_1^{(\alpha)}}{\partial r} - r\frac{\partial e_1^{(\alpha)}}{\partial \xi}\right) - r\sin\theta\frac{dm_0^{(\alpha)}}{d\xi} + \frac{r^2}{4}\cos 2\theta\frac{d^2 m_0^{(\alpha)}}{d\xi^2}\right]. \tag{A3.19}$$

Then (A2.5) leads to an equation for $\partial g_3^{(\alpha)}/\partial\theta$. As before we omit the details, but remark that some simplification is obtained by making use of the partial derivatives with respect to r and ξ of (A3.9). From the single-valuedness condition

$$\int_0^{2\pi} \partial g_3^{(\alpha)}/\partial\theta\, d\theta = 0,$$

is obtained the equation

$$\begin{aligned} &(e_2^{(\alpha)} - q_2^{(\alpha)}) + w\frac{\partial e_2^{(\alpha)}}{\partial \zeta} - \frac{\partial \varphi}{\partial \zeta}\frac{\partial e_2^{(\alpha)}}{\partial w} \\ &\quad = \left[\frac{r^2}{2}\frac{d^2 m_0^{(\alpha)}}{d\xi^2} - \frac{\partial \varphi}{\partial \xi}\left(\frac{dm_0^{(\alpha)}}{d\xi} - 1\right)\right]. \end{aligned} \tag{A3.20}$$

Also, the coefficient of $\sin\theta$ in $g_3^{(\alpha)}$ is found to be

$$\left[\frac{\partial^2 \varphi}{\partial \xi \partial \zeta}\left(w\frac{\partial e_1^{(\alpha)}}{\partial r} - r\frac{\partial e_1^{(\alpha)}}{\partial w}\right) - r\frac{\partial q_1^{(\alpha)}}{\partial \xi}\right]. \tag{A3.21}$$

Hence, from (A2.9) and (A3.6), after some integrations by parts,

$$\sigma_3^{(\alpha)} = -\left(\frac{\partial q_1^{(\alpha)}}{\partial \xi} + \frac{\partial^2 \varphi}{\partial \xi \partial \zeta}\nu_1^{(\alpha)}\right), \tag{A3.22}$$

and from (A2.12) and (A3.8),

$$\frac{\partial}{\partial \zeta}(e^{-\varphi}\nu_3^{(\alpha)}) = \frac{\partial}{\partial \xi}\left[e^{-\varphi}\frac{\partial q_1^{(\alpha)}}{\partial \xi} + A_1^{(\alpha)}(\xi)\frac{\partial^2 \varphi}{\partial \xi \partial \zeta}\right]. \tag{A3.23}$$

Averaging over ζ it follows that

$$\frac{\partial}{\partial \xi}\left\langle e^{-\varphi}\frac{\partial q_1^{(\alpha)}}{\partial \xi}\right\rangle_\zeta = 0; \quad \left\langle e^{-\varphi}\frac{\partial q_1^{(\alpha)}}{\partial \xi}\right\rangle_\zeta = B_1^{(\alpha)}, \tag{A3.24}$$

where $B_1^{(\alpha)}$ is a constant.

We now make an assumption regarding the solution to (A3.9), namely that $e_1^{(\alpha)}$ is a function of ξ alone,

$$e_1^{(\alpha)} = m_1^{(\alpha)}(\xi). \tag{A3.25}$$

Then, from (A2.7) and (A3.6),

$$q_1^{(\alpha)} = m_1^{(\alpha)}(\xi), \tag{A3.26}$$

so that (A3.25) is certainly a solution of (A3.9). Hence, from (A3.24),

$$\frac{dm_1^{(\alpha)}}{d\xi}\langle e^{-\varphi}\rangle_\zeta = B_1^{(\alpha)}; \quad \frac{dm_1^{(\alpha)}}{d\xi} = \frac{B_1^{(\alpha)}}{\langle e^{-\varphi}\rangle_\zeta}. \tag{A3.27}$$

Averaging over ξ,

$$B_1^{(\alpha)} = 0; \qquad e_1^{(\alpha)} = 0 = q_1^{(\alpha)}, \tag{A3.28}$$

from (A3.25)–(A3.27), omitting an irrelevant constant. Thus, from (A2.9), (A3.6), (A3.22), and (A3.28),

$$\nu_1^{(\alpha)} = 0, \qquad \sigma_3^{(\alpha)} = 0. \tag{A3.29}$$

At this time we leave open the question of the consistency of the assumptions in (A3.14) and (A3.25) with the determination of the subsequent terms in the asymptotic expansion.

Turning to the spatial averages of the components of the local current, we find from (2.15), (3.2), (3.6), and (A2.8) that

$$a = [1 + \omega^2\langle e^{-\varphi}\sigma^{(\alpha)}\rangle_{\xi,\zeta}/\langle e^{-\varphi}\rangle_{\xi,\zeta}], \qquad c = \omega^2\langle e^{-\varphi}\nu^{(\alpha)}\rangle_{\xi,\zeta}/\langle e^{-\varphi}\rangle_{\xi,\zeta}. \tag{A3.30}$$

Thus, from (A3.2), (A3.7), (A3.16)–(A3.18), and (A3.29),

$$c = O(1), \quad a = \{[\langle e^{-\varphi}\rangle_{\xi,\zeta}\langle[\langle e^{-\varphi}\rangle_\xi]^{-1}\rangle_\xi]^{-1} + O(1/\omega^2)\}. \tag{A3.31}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

High-Field Magnetoresistance of Inhomogeneous Semiconductors and Plasmas. III. Two-Dimensional Inhomogeneity Distributions

J. A. Morrison and H. L. Frisch
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 24 March, 1964)

In previous papers we have studied the high magnetic field magnetoresistance of a classical model semiconductor (or a plasma) produced by the presence in the sample of more or less random spatial inhomogeneities on a scale small compared to the size of the sample. We extend our computations of the leading terms of the components of the resistivity tensor to the two-dimensional case in which the external magnetic field $\mathfrak{H}$ is inclined (but not perpendicular) to the plane of variation of the spatial inhomogeneity distribution. The field $\mathfrak{H}$ is applied along the z axis, and the inhomogeneity distribution is characterized by a sufficiently smooth potential $\varphi(x, z + \lambda y)$. While the magnetoresistance ratios saturate in the y and z directions, one finds that this ratio is proportional to $\mathfrak{H}^2$ (as $\mathfrak{H} \to \infty$) in the x direction.

1. INTRODUCTION

THIS is the third in a series of papers in which we examine the possibility that the high-field nonsaturation of the magnetoresistance of a semiconductor (or a model plasma) has a classical origin in the presence in the sample of a more or less random distribution of spatial inhomogeneities. The first paper,[1] hereafter referred to as I, dealt only with the special case of a stratified medium. In the second paper,[2] hereafter referred to as II, the theory was extended to two-dimensional inhomogeneity distributions which vary along both the x and z coordinates, the latter being parallel to the external magnetic field $\mathfrak{H}$. We here extend the analysis of the 4-moment equations to the two-dimensional case in which the magnetic field is inclined to the plane of variation of the inhomogeneity distribution, but not perpendicular to it. The underlying physics of our classical model is given in I. The notation and definitions employed are those of II, and only the necessary definitions are repeated here.

Thus, introduce normalized quantities

$$x = \tau\xi, \quad y = \tau\eta, \quad z = \tau\zeta; \quad \omega = q\tau\mathfrak{H}/mc, \tag{1.1}$$

where τ is a constant relaxation time (in dimensions in which kT/m is unity), and x, y, z are position coordinates. Then, from II (2.28), in the full three-dimensional case,

$$n_0 = \bar{n}_0 e^{-\varphi}/\langle e^{-\varphi}\rangle_{\xi,\eta,\zeta} \tag{1.2}$$

is the unperturbed (by the electric field $\mathbf{E}$) local number density of electrons, where $\varphi = q\psi_0/m$, with ψ_0 the unperturbed effective impurity potential. Throughout, averages with respect to spatial variables will be indicated by corresponding subscripts to the averaging bracket, as in (1.2). Also, from II (2.27) and II (2.29), the components of the local current, in the 4-moment approximation, are given by

$$J_x = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_x + \omega E_y) - \left(\frac{\partial S}{\partial \xi} + \omega\frac{\partial S}{\partial \eta}\right)\right],$$

[1] H. L. Frisch and J. A. Morrison, Ann. Phys. (N. Y.), 26, 181 (1964).
[2] J. A. Morrison and H. L. Frisch, J. Math. Phys. 5, 1672 (1964) (preceding paper).

$$\nu_1^{(\alpha)} = 0, \qquad \sigma_3^{(\alpha)} = 0. \tag{A3.29}$$

At this time we leave open the question of the consistency of the assumptions in (A3.14) and (A3.25) with the determination of the subsequent terms in the asymptotic expansion.

Turning to the spatial averages of the components of the local current, we find from (2.15), (3.2), (3.6), and (A2.8) that

$$a = [1 + \omega^2\langle e^{-\varphi}\sigma^{(\alpha)}\rangle_{\xi,\zeta}/\langle e^{-\varphi}\rangle_{\xi,\zeta}], \qquad c = \omega^2\langle e^{-\varphi}\nu^{(\alpha)}\rangle_{\xi,\zeta}/\langle e^{-\varphi}\rangle_{\xi,\zeta}. \tag{A3.30}$$

Thus, from (A3.2), (A3.7), (A3.16)–(A3.18), and (A3.29),

$$c = O(1), \quad a = \{[\langle e^{-\varphi}\rangle_{\xi,\zeta}\langle[\langle e^{-\varphi}\rangle_\xi]^{-1}\rangle_\xi]^{-1} + O(1/\omega^2)\}. \tag{A3.31}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

High-Field Magnetoresistance of Inhomogeneous Semiconductors and Plasmas. III. Two-Dimensional Inhomogeneity Distributions

J. A. Morrison and H. L. Frisch
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 24 March, 1964)

In previous papers we have studied the high magnetic field magnetoresistance of a classical model semiconductor (or a plasma) produced by the presence in the sample of more or less random spatial inhomogeneities on a scale small compared to the size of the sample. We extend our computations of the leading terms of the components of the resistivity tensor to the two-dimensional case in which the external magnetic field $\mathfrak{H}$ is inclined (but not perpendicular) to the plane of variation of the spatial inhomogeneity distribution. The field $\mathfrak{H}$ is applied along the z axis, and the inhomogeneity distribution is characterized by a sufficiently smooth potential $\varphi(x, z + \lambda y)$. While the magnetoresistance ratios saturate in the y and z directions, one finds that this ratio is proportional to $\mathfrak{H}^2$ (as $\mathfrak{H} \to \infty$) in the x direction.

1. INTRODUCTION

THIS is the third in a series of papers in which we examine the possibility that the high-field nonsaturation of the magnetoresistance of a semiconductor (or a model plasma) has a classical origin in the presence in the sample of a more or less random distribution of spatial inhomogeneities. The first paper,[1] hereafter referred to as I, dealt only with the special case of a stratified medium. In the second paper,[2] hereafter referred to as II, the theory was extended to two-dimensional inhomogeneity distributions which vary along both the x and z coordinates, the latter being parallel to the external magnetic field $\mathfrak{H}$. We here extend the analysis of the 4-moment equations to the two-dimensional case in which the magnetic field is inclined to the plane of variation of the inhomogeneity distribution, but not perpendicular to it. The underlying physics of our classical model is given in I. The notation and definitions employed are those of II, and only the necessary definitions are repeated here.

Thus, introduce normalized quantities

$$x = \tau\xi, \quad y = \tau\eta, \quad z = \tau\zeta; \quad \omega = q\tau\mathfrak{H}/mc, \tag{1.1}$$

where τ is a constant relaxation time (in dimensions in which kT/m is unity), and x, y, z are position coordinates. Then, from II (2.28), in the full three-dimensional case,

$$n_0 = \bar{n}_0 e^{-\varphi}/\langle e^{-\varphi}\rangle_{\xi,\eta,\zeta} \tag{1.2}$$

is the unperturbed (by the electric field $\mathbf{E}$) local number density of electrons, where $\varphi = q\psi_0/m$, with ψ_0 the unperturbed effective impurity potential. Throughout, averages with respect to spatial variables will be indicated by corresponding subscripts to the averaging bracket, as in (1.2). Also, from II (2.27) and II (2.29), the components of the local current, in the 4-moment approximation, are given by

$$J_x = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_x + \omega E_y) - \left(\frac{\partial S}{\partial \xi} + \omega\frac{\partial S}{\partial \eta}\right)\right],$$

[1] H. L. Frisch and J. A. Morrison, Ann. Phys. (N. Y.), 26, 181 (1964).
[2] J. A. Morrison and H. L. Frisch, J. Math. Phys. 5, 1672 (1964) (preceding paper).

$$J_y = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_y - \omega E_x) - \left(\frac{\partial S}{\partial \eta} - \omega\frac{\partial S}{\partial \xi}\right)\right],$$
$$J_z = \frac{q\tau n_0}{m}\left(E_z - \frac{\partial S}{\partial \zeta}\right), \tag{1.3}$$

where S is the bounded solution (indeterminate to within a constant) of the equation

$$\left(\frac{\partial^2 S}{\partial \xi^2} + \frac{\partial^2 S}{\partial \eta^2}\right) + (1+\omega^2)\frac{\partial^2 S}{\partial \zeta^2}$$
$$= \frac{\partial \varphi}{\partial \xi}\left[\left(\frac{\partial S}{\partial \xi} + \omega\frac{\partial S}{\partial \eta}\right) - (E_x + \omega E_y)\right]$$
$$+ \frac{\partial \varphi}{\partial \eta}\left[\left(\frac{\partial S}{\partial \eta} - \omega\frac{\partial S}{\partial \xi}\right) - (E_y - \omega E_x)\right]$$
$$+ (1+\omega^2)\frac{\partial \varphi}{\partial \zeta}\left(\frac{\partial S}{\partial \zeta} - E_z\right). \tag{1.4}$$

This last equation merely expresses the condition that div $\mathbf{J} = 0$.

In II, we treated the two-dimensional case $\varphi = \varphi(\xi, \zeta)$. If the magnetic field, which is in the ζ direction, is inclined at an angle $\tan^{-1}\lambda$ to the plane of variation of two-dimensional inhomogeneity distribution, then, by suitable choice of axes, we have $\varphi = \varphi(\xi, \zeta + \lambda\eta)$. Since λ is to be finite, the case $\varphi = \varphi(\xi, \eta)$ is an exceptional one, and is not included in the analysis. The remainder of the paper considers the case $\varphi = \varphi(\xi, \zeta + \lambda\eta)$. The principal result, found from an asymptotic, large ω (or $\mathfrak{H}$), solution of the 4-moment equations, is that the magnetoresistance ratio in the ξ (or x) direction increases, in general, as the square of ω, provided the impurity potential is sufficiently smooth. The result obtained reduces to those found in I and II, in the special cases considered therein.

2. THE HIGH-FIELD MAGNETORESISTANCE

We here consider the two-dimensional case where $\varphi = \varphi(\xi, \zeta + \lambda\eta)$, with λ a (finite) constant. Moreover, we are concerned with the high-field magnetoresistance and will obtain the asymptotic (as $\omega \to \infty$) solution of (1.4), for sufficiently smooth φ. We let

$$\sigma = (\zeta + \lambda\eta), \tag{2.1}$$

so that

$$\varphi = \varphi(\xi, \sigma); \qquad S = S(\xi, \sigma). \tag{2.2}$$

Then, from (1.3) and (1.4),

$$J_x = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_x + \omega E_y) - \left(\frac{\partial S}{\partial \xi} + \lambda\omega\frac{\partial S}{\partial \sigma}\right)\right],$$
$$J_y = \frac{q\tau n_0}{m(1+\omega^2)}\left[(E_y - \omega E_x) - \left(\lambda\frac{\partial S}{\partial \sigma} - \omega\frac{\partial S}{\partial \xi}\right)\right],$$
$$J_z = \frac{q\tau n_0}{m}\left(E_z - \frac{\partial S}{\partial \sigma}\right), \tag{2.3}$$

where

$$\frac{\partial^2 S}{\partial \xi^2} + (1+\lambda^2+\omega^2)\frac{\partial^2 S}{\partial \sigma^2}$$
$$= \frac{\partial \varphi}{\partial \xi}\left[\left(\frac{\partial S}{\partial \xi} + \lambda\omega\frac{\partial S}{\partial \sigma}\right) - (E_x + \omega E_y)\right]$$
$$+ \frac{\partial \varphi}{\partial \sigma}\left[(1+\lambda^2+\omega^2)\frac{\partial S}{\partial \sigma} - \lambda\omega\frac{\partial S}{\partial \xi}\right.$$
$$\left. - \lambda(E_y - \omega E_x) - (1+\omega^2)E_z\right]. \tag{2.4}$$

This last equation may be conveniently written as

$$P[S] \equiv \frac{\partial}{\partial \xi}\left[e^{-\varphi}\left(\frac{\partial S}{\partial \xi} + \lambda\omega\frac{\partial S}{\partial \sigma}\right)\right]$$
$$+ \frac{\partial}{\partial \sigma}\left\{e^{-\varphi}\left[(1+\lambda^2+\omega^2)\frac{\partial S}{\partial \sigma} - \lambda\omega\frac{\partial S}{\partial \xi}\right]\right\}$$
$$= \frac{\partial}{\partial \xi}\left[e^{-\varphi}(E_x + \omega E_y)\right]$$
$$+ \frac{\partial}{\partial \sigma}\left\{e^{-\varphi}[\lambda(E_y - \omega E_x) + (1+\omega^2)E_z]\right\}. \tag{2.5}$$

We express S in the form

$$S = [E_x L + \omega(E_y M + E_z N)], \tag{2.6}$$

and remark that N here differs by a factor ω from the N defined in II (3.1). From (2.5) and (2.6),

$$P[L] = \left[\frac{\partial}{\partial \xi}(e^{-\varphi}) - \lambda\omega\frac{\partial}{\partial \sigma}(e^{-\varphi})\right], \tag{2.7}$$

$$P[M] = \left[\frac{\partial}{\partial \xi}(e^{-\varphi}) + \frac{\lambda}{\omega}\frac{\partial}{\partial \sigma}(e^{-\varphi})\right], \tag{2.8}$$

$$P[N] = \left(\omega + \frac{1}{\omega}\right)\frac{\partial}{\partial \sigma}(e^{-\varphi}). \tag{2.9}$$

Consequently,

$$M = (L + \lambda N). \tag{2.10}$$

In Appendix 1 we obtain asymptotic expansions for L and N, in reciprocal powers of ω.

Now introduce

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} f = \left\langle e^{-\varphi}\frac{\partial L}{\partial \xi}\right\rangle_{\xi,\sigma}; \quad \langle e^{-\varphi}\rangle_{\xi,\sigma} r = \omega\left\langle e^{-\varphi}\frac{\partial L}{\partial \sigma}\right\rangle_{\xi,\sigma}, \tag{2.11}$$

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} g = \left\langle e^{-\varphi}\frac{\partial M}{\partial \xi}\right\rangle_{\xi,\sigma}; \langle e^{-\varphi}\rangle_{\xi,\sigma} s = \omega\left\langle e^{-\varphi}\frac{\partial M}{\partial \sigma}\right\rangle_{\xi,\sigma}, \tag{2.12}$$

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} h = \left\langle e^{-\varphi}\frac{\partial N}{\partial \xi}\right\rangle_{\xi,\sigma}; \quad \langle e^{-\varphi}\rangle_{\xi,\sigma} t = \omega\left\langle e^{-\varphi}\frac{\partial N}{\partial \sigma}\right\rangle_{\xi,\sigma}. \tag{2.13}$$

From (2.10) it follows that

$$g = (f + \lambda h); \qquad s = (r + \lambda t). \tag{2.14}$$

From (1.2) and (2.2),

$$n_0 = \bar{n}_0 e^{-\varphi}/\langle e^{-\varphi}\rangle_{\xi,\sigma}. \tag{2.15}$$

Hence, from Eqs. (2.3), (2.6), (2.11)–(2.13), and (2.15), we have

$$\begin{bmatrix} \langle J_x\rangle_{\xi,\sigma} \\ \langle J_y\rangle_{\xi,\sigma} \\ \langle J_z\rangle_{\xi,\sigma} \end{bmatrix} = \frac{q\tau\bar{n}_0}{m} \begin{bmatrix} \frac{(1-f-\lambda r)}{(1+\omega^2)} & \frac{\omega(1-g-\lambda s)}{(1+\omega^2)} & -\frac{\omega(h+\lambda t)}{(1+\omega^2)} \\ \frac{(\omega^2 f-\omega^2-\lambda r)}{\omega(1+\omega^2)} & \frac{(\omega^2 g+1-\lambda s)}{(1+\omega^2)} & \frac{(\omega^2 h-\lambda t)}{(1+\omega^2)} \\ -\frac{r}{\omega} & -s & (1-t) \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}. \tag{2.16}$$

Equation (2.16) may be inverted to give

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \varrho(\omega) \cdot \begin{bmatrix} \langle J_x\rangle_{\xi,\sigma} \\ \langle J_y\rangle_{\xi,\sigma} \\ \langle J_z\rangle_{\xi,\sigma} \end{bmatrix}. \tag{2.17}$$

We write

$$q\tau\bar{n}_0\varrho/m = \mathbf{II}/\Gamma; \qquad \Gamma = (1+\omega^2)\Delta, \tag{2.18}$$

where Δ is the determinant of the square matrix in (2.16). The magnetoresistance ratio in direction i is

$$[\varrho(\omega) - \varrho(0)]_{ii}/[\varrho(0)]_{ii}. \tag{2.19}$$

As in I and II, we do not consider the calculation of $\varrho(0)$, but do investigate the asymptotic behavior of $\varrho(\omega)$, for $\omega \to \infty$. This is done in Appendix 2, where it is shown that ρ_{xy}, ρ_{xz}, ρ_{yz}, and ρ_{zz} are at most $O(\omega)$. Also, ρ_{yy}, ρ_{yz}, ρ_{zy}, and ρ_{zz} are at most $O(1)$. Thus, in particular, the magnetoresistance ratios in the y and z directions saturate as $\omega \to \infty$.

On the other hand, from (A2.24), (A2.34), and (A2.35),

$$\frac{q\tau\bar{n}_0}{m}(1+\lambda^2)\langle AC/K\rangle[\varrho(\omega)]_{xx} \sim \omega^2\{\langle AC/K\rangle[\langle A\rangle\langle 1/A\rangle - 1] + \lambda^2\langle A\rangle[\langle AC/K\rangle\langle C/AK\rangle - \langle C/K\rangle^2]\}, \tag{2.20}$$

where

$$A(\xi) = \langle e^{-\varphi}\rangle_\sigma; \qquad C(\xi) = [\langle e^{\varphi}\rangle_\sigma]^{-1}, \tag{2.21}$$

and

$$K(\xi) = [(1+\lambda^2)A - \lambda^2 C], \tag{2.22}$$

it being understood that the averages in (2.20) are over ξ. The result of (2.20) reduces to those obtained previously in I and II. Thus, if φ is independent of σ, then $C = A = K$ and

$$q\tau\bar{n}_0\,[\varrho(\omega)]_{xx}/m \sim \omega^2[\langle e^{-\varphi}\rangle_\xi\langle e^{\varphi}\rangle_\xi - 1], \tag{2.23}$$

as obtained in I. If $\lambda = 0$, then

$$q\tau\bar{n}_0\,[\varrho(\omega)]_{xx}/m \sim \omega^2[\langle A\rangle\langle 1/A\rangle - 1], \tag{2.24}$$

as obtained in II. The coefficient of ω^2 in (2.20) does not vanish in general, although it does if A is independent of ξ, that is, a constant. Such is the case if φ is independent of ξ, and then, as shown in II, the magnetoresistance saturates. In general, though, the magnetoresistance ratio in the x direction is asymptotically proportional to ω^2, and the result of (2.20) may be expressed in terms of n_0 by means of (2.15).

APPENDIX 1. ASYMPTOTIC SOLUTION OF THE 4-MOMENT EQUATIONS

We first consider the asymptotic (as $\omega \to \infty$) solution of (2.7), where P is the operator defined in (2.5). Thus,

$$\frac{\partial}{\partial\xi}\left[e^{-\varphi}\left(\frac{\partial L}{\partial\xi} + \lambda\omega\frac{\partial L}{\partial\sigma}\right)\right] + \frac{\partial}{\partial\sigma}\left\{e^{-\varphi}\left[(1+\lambda^2+\omega^2)\frac{\partial L}{\partial\sigma} - \lambda\omega\frac{\partial L}{\partial\xi}\right]\right\} = \left[\frac{\partial}{\partial\xi}(e^{-\varphi}) - \lambda\omega\frac{\partial}{\partial\sigma}(e^{-\varphi})\right]. \tag{A1.1}$$

We seek an asymptotic solution of the form

$$L = \left(L_0 + \frac{L_1}{\omega} + \frac{L_2}{\omega^2} + \cdots\right). \tag{A1.2}$$

Hence,

$$\frac{\partial}{\partial\sigma}\left(e^{-\varphi}\frac{\partial L_0}{\partial\sigma}\right) = 0; \qquad \frac{\partial L_0}{\partial\sigma} = T_0(\xi)e^{\varphi}. \tag{A1.3}$$

Averaging over σ,

$$T_0(\xi)\langle e^{\varphi}\rangle_\sigma = 0, \tag{A1.4}$$

which implies that

$$\partial L_0/\partial\sigma = 0; \qquad L_0 = R_0(\xi). \tag{A1.5}$$

The next equation arising from (A1.1) and (A1.2) is, using (A1.5),

$$\frac{\partial}{\partial\sigma}\left[e^{-\varphi}\left(\frac{\partial L_1}{\partial\sigma} - \lambda\frac{dR_0}{d\xi} + \lambda\right)\right] = 0. \tag{A1.6}$$

Hence,

$$\frac{\partial L_1}{\partial\sigma} = \left[\lambda\left(\frac{dR_0}{d\xi} - 1\right) + T_1(\xi)e^{\varphi}\right]. \tag{A1.7}$$

Averaging over σ determines $T_1(\xi)$, and then

$$\frac{\partial L_1}{\partial\sigma} = \lambda\left(\frac{dR_0}{d\xi} - 1\right)\left(1 - \frac{e^{\varphi}}{\langle e^{\varphi}\rangle_\sigma}\right). \tag{A1.8}$$

Proceeding to the next term in the expansion of (A1.2) we will see that $dR_0/d\xi$ can be determined, and so on.

Thus, from (A1.1), using (A1.5), we have

$$\frac{\partial}{\partial\xi}\left[e^{-\varphi}\left(\frac{dR_0}{d\xi} + \lambda\frac{\partial L_1}{\partial\sigma} - 1\right)\right] + \frac{\partial}{\partial\sigma}\left[e^{-\varphi}\left(\frac{\partial L_2}{\partial\sigma} - \lambda\frac{\partial L_1}{\partial\xi}\right)\right] = 0. \tag{A1.9}$$

Averaging over σ we obtain

$$\frac{\partial}{\partial\xi}\left\langle e^{-\varphi}\left(\frac{dR_0}{d\xi} + \lambda\frac{\partial L_1}{\partial\sigma} - 1\right)\right\rangle_\sigma = 0; \tag{A1.10}$$

$$\left(\frac{dR_0}{d\xi} - 1\right)\langle e^{-\varphi}\rangle_\sigma + \lambda\left\langle e^{-\varphi}\frac{\partial L_1}{\partial\sigma}\right\rangle_\sigma = B_0,$$

where B_0 is a constant. Using (A1.8) it follows that

$$\left(\frac{dR_0}{d\xi} - 1\right)\left((1 + \lambda^2)\langle e^{-\varphi}\rangle_\sigma - \frac{\lambda^2}{\langle e^{\varphi}\rangle_\sigma}\right) = B_0. \tag{A1.11}$$

Hence,

$$dR_0/d\xi = \{1 + B_0[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}\}, \tag{A1.12}$$

where

$$\chi(\xi) = \{1 - [\langle e^{-\varphi}\rangle_\sigma\langle e^{\varphi}\rangle_\sigma]^{-1}\}. \tag{A1.13}$$

Averaging (A1.12) over ξ determines B_0, and then

$$\frac{dR_0}{d\xi} = \left\{1 - \frac{[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}}{\langle[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}\rangle_\xi}\right\}. \tag{A1.14}$$

Thus $\partial L_1/\partial\sigma$ is determined, and L_0 is determined to within an irrelevant constant. From (A1.8) and (A1.14),

$$L_1 = [\delta_1(\xi, \sigma) + R_1(\xi)], \tag{A1.15}$$

where $\delta_1(\xi, \sigma)$ is explicitly known. Note that in the random (nonperiodic) case, as mentioned in I and II, certain conditions must be imposed on φ. Here the requirement is that $\int_0^\sigma [1 - e^{\varphi}/\langle e^{\varphi}\rangle_\sigma]\, d\sigma$ should remain bounded as $\sigma \to \pm\infty$, and analogous conditions arise at each stage of the analysis. There is also a restriction imposed on φ in order that $R_0(\xi)$ should remain bounded as $\xi \to \pm\infty$. Now, from (A1.9) and (A1.13)–(A1.15),

$$\frac{\partial L_2}{\partial\sigma} = \left[\lambda\frac{dR_1}{d\xi} + \epsilon_2(\xi, \sigma) + T_2(\xi)e^{\varphi}\right], \tag{A1.16}$$

where $\epsilon_2(\xi, \sigma)$ is known explicitly. Averaging over σ determines $T_2(\xi)$, and then

$$\frac{\partial L_2}{\partial\sigma} = \left\{\lambda\frac{dR_1}{d\xi}[1 - e^{\varphi}/\langle e^{\varphi}\rangle_\sigma] + \kappa_2(\xi, \sigma)\right\}, \tag{A1.17}$$

where $\kappa_2(\xi, \sigma)$ is known. We can now proceed by induction.

Thus, for $n \geq 2$, from (A1.1),

$$\frac{\partial}{\partial\xi}\left[e^{-\varphi}\left(\frac{\partial L_{n-1}}{\partial\xi} + \lambda\frac{\partial L_n}{\partial\sigma}\right)\right] + \frac{\partial}{\partial\sigma}\left\{e^{-\varphi}\left[\frac{\partial L_{n+1}}{\partial\sigma} - \lambda\frac{\partial L_n}{\partial\xi} + (1 + \lambda^2)\frac{\partial L_{n-1}}{\partial\sigma}\right]\right\} = 0. \tag{A1.18}$$

We suppose that

$$L_{n-1} = [\delta_{n-1}(\xi, \sigma) + R_{n-1}(\xi)], \tag{A1.19}$$

and

$$\frac{\partial L_n}{\partial\sigma} = \left\{\lambda\frac{dR_{n-1}}{d\xi}\left(1 - \frac{e^{\varphi}}{\langle e^{\varphi}\rangle_\sigma}\right) + \kappa_n(\xi, \sigma)\right\}, \tag{A1.20}$$

where $\delta_{n-1}(\xi, \sigma)$ and $\kappa_n(\xi, \sigma)$ are known, and $R_{n-1}(\xi)$ is as yet undetermined. Now, averaging (A1.18) over σ,

$$\frac{\partial}{\partial\xi}\left\langle e^{-\varphi}\left(\frac{\partial L_{n-1}}{\partial\xi} + \lambda\frac{\partial L_n}{\partial\sigma}\right)\right\rangle_\sigma = 0; \qquad \left\langle e^{-\varphi}\left(\frac{\partial L_{n-1}}{\partial\xi} + \lambda\frac{\partial L_n}{\partial\sigma}\right)\right\rangle_\sigma = B_{n-1}, \tag{A1.21}$$

where B_{n-1} is a constant. From (A1.19)–(A1.21) it follows that

$$\frac{dR_{n-1}}{d\xi}\left((1 + \lambda^2)\langle e^{-\varphi}\rangle_\sigma - \frac{\lambda^2}{\langle e^{\varphi}\rangle_\sigma}\right) + \left\langle e^{-\varphi}\left(\frac{\partial\delta_{n-1}}{\partial\xi} + \lambda\kappa_n\right)\right\rangle_\sigma = B_{n-1}. \tag{A1.22}$$

Hence, from (A1.13),

$$\frac{dR_{n-1}}{d\xi} = \left[B_{n-1} - \left\langle e^{-\varphi}\left(\frac{\partial\delta_{n-1}}{\partial\xi} + \lambda\kappa_n\right)\right\rangle_\sigma\right] \times [\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}. \tag{A1.23}$$

Averaging over ξ determines B_{n-1}, and thus $dR_{n-1}/d\xi$. So $\partial L_n/\partial\sigma$ is now determined, and L_{n-1} is deter-

mined to within an irrelevant constant. Hence,

$$L_n = [\delta_n(\xi, \sigma) + R_n(\xi)], \tag{A1.24}$$

where $\delta_n(\xi, \sigma)$ is known and $R_n(\xi)$ is as yet undetermined. Then, from (A1.18),

$$\frac{\partial L_{n+1}}{\partial \sigma} = \left[\lambda \frac{dR_n}{d\xi} + \epsilon_{n+1}(\xi, \sigma) + T_{n+1}(\xi)e^{\varphi}\right], \tag{A1.25}$$

where $\epsilon_{n+1}(\xi, \sigma)$ is known. Averaging over σ determines $T_{n+1}(\xi)$, and then

$$\frac{\partial L_{n+1}}{\partial \sigma} = \left\{\lambda \frac{dR_n}{d\xi}[1 - e^{\varphi}/\langle e^{\varphi}\rangle_\sigma] + \kappa_{n+1}(\xi, \sigma)\right\}, \tag{A1.26}$$

where $\kappa_{n+1}(\xi, \sigma)$ is known. This completes the induction process.

We now consider the asymptotic solution of (2.9), that is

$$\frac{\partial}{\partial \xi}\left[e^{-\varphi}\left(\frac{\partial N}{\partial \xi} + \lambda\omega \frac{\partial N}{\partial \sigma}\right)\right] + \frac{\partial}{\partial \sigma}\left\{e^{-\varphi}\left[(1 + \lambda^2 + \omega^2)\frac{\partial N}{\partial \sigma} - \lambda\omega \frac{\partial N}{\partial \xi}\right]\right\} = \left(\omega + \frac{1}{\omega}\right)\frac{\partial}{\partial \sigma}(e^{-\varphi}). \tag{A1.27}$$

We seek an asymptotic solution of the form

$$N = \left(N_0 + \frac{N_1}{\omega} + \frac{N_2}{\omega^2} + \cdots\right). \tag{A1.28}$$

Hence, proceeding as before,

$$\frac{\partial}{\partial \sigma}\left(e^{-\varphi}\frac{\partial N_0}{\partial \sigma}\right) = 0, \tag{A1.29}$$

which implies

$$N_0 = W_0(\xi). \tag{A1.30}$$

Then,

$$\frac{\partial}{\partial \sigma}\left[e^{-\varphi}\left(\frac{\partial N_1}{\partial \sigma} - \lambda \frac{dW_0}{d\xi} - 1\right)\right] = 0, \tag{A1.31}$$

which leads to

$$\frac{\partial N_1}{\partial \sigma} = \left(1 + \lambda \frac{dW_0}{d\xi}\right)\left(1 - \frac{e^{\varphi}}{\langle e^{\varphi}\rangle_\sigma}\right). \tag{A1.32}$$

Next, using (A1.30),

$$\frac{\partial}{\partial \xi}\left[e^{-\varphi}\left(\frac{dW_0}{d\xi} + \lambda \frac{\partial N_1}{\partial \sigma}\right)\right] + \frac{\partial}{\partial \sigma}\left[e^{-\varphi}\left(\frac{\partial N_2}{\partial \sigma} - \lambda \frac{\partial N_1}{\partial \xi}\right)\right] = 0. \tag{A1.33}$$

Averaging over σ,

$$\frac{\partial}{\partial \xi}\left\langle e^{-\varphi}\left(\frac{dW_0}{d\xi} + \lambda \frac{\partial N_1}{\partial \sigma}\right)\right\rangle_\sigma = 0, \qquad \frac{dW_0}{d\xi}\langle e^{-\varphi}\rangle_\sigma + \lambda\left\langle e^{-\varphi}\frac{\partial N_1}{\partial \sigma}\right\rangle_\sigma = C_0, \tag{A1.34}$$

where C_0 is a constant. Hence, from (A1.32),

$$(dW_0/d\xi)\,[(1 + \lambda^2)\langle e^{-\varphi}\rangle_\sigma - \lambda^2/\langle e^{\varphi}\rangle_\sigma] + \lambda[\langle e^{-\varphi}\rangle_\sigma - 1/\langle e^{\varphi}\rangle_\sigma] = C_0. \tag{A1.35}$$

Thus, from (A1.13),

$$dW_0/d\xi = \{C_0[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1} - \lambda\chi(1 + \lambda^2\chi)^{-1}\}. \tag{A1.36}$$

Averaging over ξ determines C_0, and then

$$\frac{dW_0}{d\xi} = \lambda\left\{\langle\chi(1 + \lambda^2\chi)^{-1}\rangle_\xi \frac{[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}}{\langle[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}\rangle_\xi} - \chi(1 + \lambda^2\chi)^{-1}\right\}. \tag{A1.37}$$

The asymptotic expansion proceeds by iteration in a manner similar to that for L, but for our purposes it is not necessary to go any further.

APPENDIX 2. ASYMPTOTIC EXPANSION OF SOME SPATIAL AVERAGES

We are first concerned with the asymptotic expansions of the quantities defined in (2.11) and (2.13). From (A1.5) and (A1.30) it follows that f, h, r, and t are at most $O(1)$ as $\omega \to \infty$. Corresponding to (A1.2) and (A1.28) we have the asymptotic expansions

$$f = \left(f_0 + \frac{f_1}{\omega} + \frac{f_2}{\omega^2} + \cdots\right), \qquad h = \left(h_0 + \frac{h_1}{\omega} + \frac{h_2}{\omega^2} + \cdots\right), \tag{A2.1}$$

with similar expansions for r and t. From (A1.5) and (A1.8),

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} f_0 = \left\langle e^{-\varphi}\frac{dR_0}{d\xi}\right\rangle_{\xi,\sigma} = \left\langle \langle e^{-\varphi}\rangle_\sigma \frac{dR_0}{d\xi}\right\rangle_\xi, \tag{A2.2}$$

and

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} r_0 = \lambda\left\langle e^{-\varphi}\left(\frac{dR_0}{d\xi} - 1\right)[1 - e^{\varphi}/\langle e^{\varphi}\rangle_\sigma]\right\rangle_{\xi,\sigma} = \lambda\left\langle\{\langle e^{-\varphi}\rangle_\sigma - [\langle e^{\varphi}\rangle_\sigma]^{-1}\}\left(\frac{dR_0}{d\xi} - 1\right)\right\rangle_\xi. \tag{A2.3}$$

From (A1.13) and (A1.14) it then follows that

$$\langle e^{-\varphi}\rangle_{\xi,\sigma}(1 - f_0) = \frac{\langle(1 + \lambda^2\chi)^{-1}\rangle_\xi}{\langle[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}\rangle_\xi}, \tag{A2.4}$$

and

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} r_0 = \frac{-\lambda\langle\chi(1 + \lambda^2\chi)^{-1}\rangle_\xi}{\langle[\langle e^{-\varphi}\rangle_\sigma(1 + \lambda^2\chi)]^{-1}\rangle_\xi}. \tag{A2.5}$$

From (2.13), (A1.30), and (A1.32),

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} h_0 = \left\langle e^{-\varphi}\frac{dW_0}{d\xi}\right\rangle_{\xi,\sigma} = \left\langle \langle e^{-\varphi}\rangle_\sigma \frac{dW_0}{d\xi}\right\rangle_\xi, \tag{A2.6}$$

and

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} t_0 = \left\langle e^{-\varphi}\left(1 + \lambda \frac{dW_0}{d\xi}\right)\left(1 - \frac{e^{\varphi}}{\langle e^{\varphi}\rangle_{\sigma}}\right)\right\rangle_{\xi,\sigma}$$
$$= \left\langle \{\langle e^{-\varphi}\rangle_{\sigma} - [\langle e^{\varphi}\rangle_{\sigma}]^{-1}\}\left(1 + \lambda \frac{dW_0}{d\xi}\right)\right\rangle_{\xi}. \tag{A2.7}$$

Thus, from (A1.37),

$$\langle e^{-\varphi}\rangle_{\xi,\sigma} h_0 = \lambda\left\{\frac{\langle (1 + \lambda^2\chi)^{-1}\rangle_{\xi}\langle \chi(1 + \lambda^2\chi)^{-1}\rangle_{\xi}}{\langle [\langle e^{-\varphi}\rangle_{\sigma}(1 + \lambda^2\chi)]^{-1}\rangle_{\xi}} - \langle \langle e^{-\varphi}\rangle_{\sigma}\chi(1 + \lambda^2\chi)^{-1}\rangle_{\xi}\right\}, \tag{A2.8}$$

and

$$\{\langle e^{-\varphi}\rangle_{\xi,\sigma}(1 - t_0) - \langle [\langle e^{\varphi}\rangle_{\sigma}]^{-1}\rangle_{\xi}\}$$
$$= \lambda^2\left\{\langle \langle e^{-\varphi}\rangle_{\sigma}\chi^2(1 + \lambda^2\chi)^{-1}\rangle_{\xi} - \frac{[\langle \chi(1 + \lambda^2\chi)^{-1}\rangle_{\xi}]^2}{\langle [\langle e^{-\varphi}\rangle_{\sigma}(1 + \lambda^2\chi)]^{-1}\rangle_{\xi}}\right\}. \tag{A2.9}$$

We remark that, as is readily verified,

$$(h_0 + r_0 + \lambda t_0) = 0. \tag{A2.10}$$

Now, from (2.18), where Δ is the determinant of the square matrix in (2.16), we have

$$\Gamma = \{[(1 - t)(1 - f) - rh] - \lambda[s(1 - f) + rg]\}. \tag{A2.11}$$

From (2.14), eliminating g and s, it is found that

$$\Gamma = \{(1 + \lambda^2)[(1 - t)(1 - f) - rh] - \lambda[r + \lambda(1 - f)]\}. \tag{A2.12}$$

Expanding Γ in inverse powers of ω,

$$\Gamma = \left(\Gamma_0 + \frac{\Gamma_1}{\omega} + \frac{\Gamma_2}{\omega^2} + \cdots\right), \tag{A2.13}$$

it follows that

$$\Gamma_0 = \{(1 + \lambda^2)[(1 - t_0)(1 - f_0) - r_0 h_0] - \lambda[r_0 + \lambda(1 - f_0)]\}. \tag{A2.14}$$

At this stage we introduce the quantities

$$A(\xi) = \langle e^{-\varphi}\rangle_{\sigma}, \quad B(\xi) = (1 + \lambda^2\chi)^{-1}, \quad C(\xi) = [\langle e^{\varphi}\rangle_{\sigma}]^{-1}, \quad D(\xi) = [\langle e^{-\varphi}\rangle_{\sigma}(1 + \lambda^2\chi)]^{-1}. \tag{A2.15}$$

Note that, from Schwarz's inequality,

$$A/C \geq 1. \tag{A2.16}$$

Also we have

$$\lambda^2\chi B = (1 - B); \quad B = AD; \quad (1 - \chi)A = C, \tag{A2.17}$$

the last following from (A1.13). Further, from (A2.17),

$$(1 + \lambda^2)\chi AB = (A - BC), \quad (1 - \chi)B = CD. \tag{A2.18}$$

Next, from (A2.4), (A2.5), (A2.8), (A2.9), and (A2.15),

$$\langle A\rangle(1 - f_0) = \langle B\rangle/\langle D\rangle, \quad \langle A\rangle r_0 = -\lambda\langle \chi B\rangle/\langle D\rangle, \tag{A2.19}$$

and

$$\langle A\rangle h_0 = \lambda[\langle B\rangle\langle \chi B\rangle/\langle D\rangle - \langle \chi AB\rangle], \tag{A2.20}$$

$$\langle A\rangle(1 - t_0) = \{\langle C\rangle + \lambda^2[\langle \chi^2 AB\rangle - \langle \chi B\rangle^2/\langle D\rangle]\}, \tag{A2.21}$$

it being understood that the averages are over ξ. From (A2.14) and (A2.17)–(A2.21), after some reduction,

$$\langle A\rangle^2\langle D\rangle\Gamma_0 = (1 + \lambda^2)\{\langle B\rangle[\langle C\rangle + \lambda^2\langle \chi^2 AB\rangle] - \lambda^2\langle \chi B\rangle\langle \chi AB\rangle\} + \lambda^2\langle A\rangle[\langle \chi B\rangle - \langle B\rangle]$$
$$= \{\langle B\rangle\langle C\rangle + \lambda^2[\langle \chi B\rangle\langle BC\rangle - \langle B\rangle\langle \chi BC\rangle]\} = \langle BC\rangle. \tag{A2.22}$$

That is,

$$[\langle e^{-\varphi}\rangle_{\xi,\sigma}]^2\Gamma_0 = \frac{\langle [(1 + \lambda^2)\langle e^{\varphi}\rangle_{\sigma} - \lambda^2/\langle e^{-\varphi}\rangle_{\sigma}]^{-1}\rangle_{\xi}}{\langle [(1 + \lambda^2)\langle e^{-\varphi}\rangle_{\sigma} - \lambda^2/\langle e^{\varphi}\rangle_{\sigma}]^{-1}\rangle_{\xi}}. \tag{A2.23}$$

Note that $\Gamma_0 \neq 0$. Also, we know that f, g, h, r, s, and t are all $O(1)$, at most, as $\omega \to \infty$. Hence, from (2.16)–(2.18), ρ_{xy}, ρ_{xz}, ρ_{yx}, and ρ_{zz} are $O(\omega)$, at most. Also, ρ_{yy}, ρ_{yz}, ρ_{zy}, and ρ_{zz} are $O(1)$, at most. Finally,

$$\frac{q\tau\bar{n}_0}{m}[\varrho(\omega)]_{xx} \sim \frac{\omega^2\Sigma_0}{\Gamma_0}, \quad \Sigma_0 = [g_0(1 - t_0) + s_0 h_0]. \tag{A2.24}$$

It remains to determine Σ_0. Since, from (2.14),

$$g_0 = (f_0 + \lambda h_0), \quad s_0 = (r_0 + \lambda t_0), \tag{A2.25}$$

we have

$$\Sigma_0 = [f_0(1 - t_0) + h_0(\lambda + r_0)]. \tag{A2.26}$$

From (A2.17)–(2.21), after some reduction, it is found that

$$\langle A\rangle^2\langle D\rangle\Sigma_0 = \{\langle A\rangle\langle D\rangle[\langle C\rangle + \lambda^2\langle \chi^2 AB\rangle - \lambda^2\langle \chi AB\rangle] - \langle B\rangle[\langle C\rangle + \lambda^2\langle \chi^2 AB\rangle] - \lambda^2\langle A\rangle\langle \chi B\rangle^2 + \lambda^2[\langle A\rangle\langle B\rangle\langle \chi B\rangle + \langle \chi B\rangle\langle \chi AB\rangle]\}, \tag{A2.27}$$

and, hence,

$$\langle A\rangle^2\langle D\rangle\Sigma_0 = \{\langle A\rangle\langle D\rangle[\langle C\rangle - \lambda^2\langle \chi BC\rangle] + [\langle B\rangle\langle \chi AB\rangle - \langle \chi B\rangle\langle AB\rangle - \langle A\rangle\langle B\rangle\langle (1 - \chi)B\rangle]\}$$
$$= \{[\langle A\rangle\langle D\rangle - \langle AD\rangle]\langle BC\rangle + \langle CD\rangle[\langle AB\rangle - \langle A\rangle\langle B\rangle]\}. \tag{A2.28}$$

We proceed to express Σ_0 in a form which clearly

exhibits the fact that it is nonnegative. Thus, let

$$K = [(1 + \lambda^2)A - \lambda^2 C], \qquad \text{(A2.29)}$$

so that, from (A1.13) and (A2.15),

$$B = A/K; \qquad D = 1/K. \qquad \text{(A2.30)}$$

Note that, from (A2.16),

$$K \geq A \geq C > 0. \qquad \text{(A2.31)}$$

Now, from (A2.28), we have

$$\langle A\rangle^2\langle 1/K\rangle\Sigma_0 = \{\langle AC/K\rangle[\langle A\rangle\langle 1/K\rangle - \langle A/K\rangle] + \langle C/K\rangle[\langle A^2/K\rangle - \langle A\rangle\langle A/K\rangle]\}. \qquad \text{(A2.32)}$$

Hence, using (A2.29),

$$\{(1 + \lambda^2)\langle A\rangle^2\langle 1/K\rangle\Sigma_0 - \langle AC/K\rangle[\langle A\rangle\langle 1/A\rangle - 1]\} = \lambda^2\{\langle AC/K\rangle[\langle A\rangle\langle C/AK\rangle - \langle C/K\rangle] + \langle C/K\rangle[\langle AC/K\rangle - \langle A\rangle\langle C/K\rangle]\}. \qquad \text{(A2.33)}$$

Thus,

$$(1 + \lambda^2)\langle A\rangle^2\langle 1/K\rangle\Sigma_0 = \{\langle AC/K\rangle[\langle A\rangle\langle 1/A\rangle - 1] + \lambda^2\langle A\rangle[\langle AC/K\rangle\langle C/AK\rangle - \langle C/K\rangle^2]\}, \qquad \text{(A2.34)}$$

which is seen to be nonnegative by invoking Schwarz's inequality. From (A2.22) and (A2.30),

$$\langle A\rangle^2\langle 1/K\rangle\Gamma_0 = \langle AC/K\rangle. \qquad \text{(A2.35)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Numerical Method for the Exact Expansion of Generating Functions

C. ISENBERG
Wheatstone Laboratory, King's College, London, England
(Received 5 May 1964; final manuscript received 3 June 1964)

A new method is developed for the evaluation of coefficients in multinomial generating functions. This method requires solely numerical techniques to derive the coefficients in the expansion of the generating function, and thus avoids the organizational problems that occur when algebraic methods are used to expand these functions. With the use of a digital computer this method enables the number of coefficients to be extended considerably. As an example of its application some of the coefficients in a four-variable generating function are calculated.

INTRODUCTION

IN many branches of solid-state physics, particularly those involving the properties of crystal lattices, e.g., lattice dynamics,[1] cooperative phenomena,[2] polymer science,[3,4] one often encounters enumerative problems which can be formulated in terms of a generating function. Three typical examples of these generating functions are

$$G_l(x, y, z) = [(x + x^{-1})(y + y^{-1})(z + z^{-1})]^l, \qquad (1)$$

$$G_m(x, y, z, \beta) = \mathrm{Tr}\begin{bmatrix} xy & \beta x^{-2} & 5 \\ y^{-1}z & xz & y^{-1} \\ (\beta^2 x + 1) & z^{-1} & x^2\beta \end{bmatrix}^m, \qquad (2)$$

$$G_n(x, y) = \det\begin{bmatrix} (x^3y + y^{-1}) & (x^{-2} + y^{-4} + x^5y) \\ (y^{-4} + 1) & (x^2 + xy + y^2 + x + 3) \end{bmatrix}^n, \qquad (3)$$

where l, m, and n are integers. Tr and det refer to the trace and determinant, respectively.

All these examples are expressible as a finite multinomial expression and would require a computer program with a complex organizational system to evaluate the coefficients occurring in their expansions,[5,6] which are the solution to the enumerative problem, for all but the lowest values of l, m, and n.[6a] In the course of the algebraic evaluation of the generating functions, all the intermediate powers

[1] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *The Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, Inc., New York, 1963).
[2] C. Domb, Advan. Phys. **34**, 35 (1960).
[3] H. N. V. Temperley, Phys. Rev. **103**, 1 (1956).
[4] M. E. Fisher and M. F. Sykes, Phys. Rev. **114**, 45 (1959).
[5] W. S. Brown, Bell System Tech. J. **42**, 2081 (1963).
[6] W. S. Brown, J. P. Hyde, and B. A. Tagne, (to be published).
[6a] *Added in proof*: The author is grateful to Dr. H. N. V. Temperly for pointing out that the coefficients in the first example can be obtained in closed form, consequently this example is trivial.

exhibits the fact that it is nonnegative. Thus, let

$$K = [(1 + \lambda^2)A - \lambda^2 C], \qquad \text{(A2.29)}$$

so that, from (A1.13) and (A2.15),

$$B = A/K; \qquad D = 1/K. \qquad \text{(A2.30)}$$

Note that, from (A2.16),

$$K \geq A \geq C > 0. \qquad \text{(A2.31)}$$

Now, from (A2.28), we have

$$\langle A\rangle^2\langle 1/K\rangle\Sigma_0 = \{\langle AC/K\rangle[\langle A\rangle\langle 1/K\rangle - \langle A/K\rangle] + \langle C/K\rangle[\langle A^2/K\rangle - \langle A\rangle\langle A/K\rangle]\}. \qquad \text{(A2.32)}$$

Hence, using (A2.29),

$$\{(1 + \lambda^2)\langle A\rangle^2\langle 1/K\rangle\Sigma_0 - \langle AC/K\rangle[\langle A\rangle\langle 1/A\rangle - 1]\} = \lambda^2\{\langle AC/K\rangle[\langle A\rangle\langle C/AK\rangle - \langle C/K\rangle] + \langle C/K\rangle[\langle AC/K\rangle - \langle A\rangle\langle C/K\rangle]\}. \qquad \text{(A2.33)}$$

Thus,

$$(1 + \lambda^2)\langle A\rangle^2\langle 1/K\rangle\Sigma_0 = \{\langle AC/K\rangle[\langle A\rangle\langle 1/A\rangle - 1] + \lambda^2\langle A\rangle[\langle AC/K\rangle\langle C/AK\rangle - \langle C/K\rangle^2]\}, \qquad \text{(A2.34)}$$

which is seen to be nonnegative by invoking Schwarz's inequality. From (A2.22) and (A2.30),

$$\langle A\rangle^2\langle 1/K\rangle\Gamma_0 = \langle AC/K\rangle. \qquad \text{(A2.35)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Numerical Method for the Exact Expansion of Generating Functions

C. ISENBERG
Wheatstone Laboratory, King's College, London, England
(Received 5 May 1964; final manuscript received 3 June 1964)

A new method is developed for the evaluation of coefficients in multinomial generating functions. This method requires solely numerical techniques to derive the coefficients in the expansion of the generating function, and thus avoids the organizational problems that occur when algebraic methods are used to expand these functions. With the use of a digital computer this method enables the number of coefficients to be extended considerably. As an example of its application some of the coefficients in a four-variable generating function are calculated.

INTRODUCTION

IN many branches of solid-state physics, particularly those involving the properties of crystal lattices, e.g., lattice dynamics,[1] cooperative phenomena,[2] polymer science,[3,4] one often encounters enumerative problems which can be formulated in terms of a generating function. Three typical examples of these generating functions are

$$G_l(x, y, z) = [(x + x^{-1})(y + y^{-1})(z + z^{-1})]^l, \qquad (1)$$

$$G_m(x, y, z, \beta) = \mathrm{Tr}\begin{bmatrix} xy & \beta x^{-2} & 5 \\ y^{-1}z & xz & y^{-1} \\ (\beta^2 x + 1) & z^{-1} & x^2\beta \end{bmatrix}^m, \qquad (2)$$

$$G_n(x, y) = \det\begin{bmatrix} (x^3y + y^{-1}) & (x^{-2} + y^{-4} + x^5y) \\ (y^{-4} + 1) & (x^2 + xy + y^2 + x + 3) \end{bmatrix}^n, \qquad (3)$$

where l, m, and n are integers. Tr and det refer to the trace and determinant, respectively.

All these examples are expressible as a finite multinomial expression and would require a computer program with a complex organizational system to evaluate the coefficients occurring in their expansions,[5,6] which are the solution to the enumerative problem, for all but the lowest values of l, m, and n.[6a] In the course of the algebraic evaluation of the generating functions, all the intermediate powers

[1] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *The Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, Inc., New York, 1963).
[2] C. Domb, Advan. Phys. **34**, 35 (1960).
[3] H. N. V. Temperley, Phys. Rev. **103**, 1 (1956).
[4] M. E. Fisher and M. F. Sykes, Phys. Rev. **114**, 45 (1959).
[5] W. S. Brown, Bell System Tech. J. **42**, 2081 (1963).
[6] W. S. Brown, J. P. Hyde, and B. A. Tagne, (to be published).
[6a] *Added in proof*: The author is grateful to Dr. H. N. V. Temperly for pointing out that the coefficients in the first example can be obtained in closed form, consequently this example is trivial.

have to be evaluated (e.g., $n = 1, 2, \cdots, N$) even if only one coefficient in the final expansion is required; there is no way of bypassing all the intermediate coefficients.

The only common feature of these generating functions is that, for numerical values of the parameters x, y, z, etc., they can be readily evaluated by existing computational routines.

In this paper it is shown that by a series of numerical evaluations of the generating function its multinomial coefficients can be derived. This enables a systematic method to be applied to the most complicated generating functions for the evaluation of the coefficients in its expansion. The application of this method enables the work on existing generating functions to be extended considerably.

FINITE MULTINOMIAL GENERATING FUNCTIONS

Consider the finite multinomial generating function $G(x_1, x_2, \cdots x_s)$, in the s variables x_1 to x_s, which can be expressed by an operation $\mathbf{O}$ on the variables $x_1, \cdots x_s$. Then

$$G(x_1, \cdots x_s) = \mathbf{O}[x_1, \cdots x_s]. \tag{4}$$

Since this can be expressed as a finite multinomial, we may write

$$\sum_k \sum_{\alpha_k=t_k}^{t_k+N_k} A_{\alpha_1\alpha_2\cdots\alpha_s} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_s^{\alpha_s}, \tag{5}$$

where the integer powers α_k ($k = 1, \cdots, s$) range from t_k to $t_k + N_k$, and $A_{\alpha_1\cdots\alpha_s}$ are the numerical coefficients in the multinomial expansion of $G(x_1, \cdots x_s)$.

In order to obtain the coefficient of $x_1^{\alpha_1^0} x_2^{\alpha_2^0} \cdots x_s^{\alpha_s^0}$, multiply (5) by $x_1^{-\alpha_1^0} x_2^{-\alpha_2^0} \cdots x_s^{-\alpha_s^0}$, so that we now require the constant term, which is independent of $x_1, x_2, \cdots, x_s$, in this expansion. Now (5) becomes

$$\left(\prod_{k=1}^{s} x_k^{-\alpha_k^0}\right)\mathbf{O}[x_1, \cdots, x_s] = \sum_k \sum_{\alpha_k=t_k}^{t_k+N_k} A_{\alpha_1\cdots\alpha_s} \prod_{j=1}^{s} x_j^{\alpha_j-\alpha_j^0}. \tag{6}$$

Changing our variables to $\alpha_j' = \alpha_j - \alpha_j^0$ so that $t_k' = t_k - \alpha_k^0$, we obtain

$$= \sum_k \sum_{\alpha_k'=t_k'}^{t_k'+N_k} A_{\alpha_1\cdots\alpha_s} \prod_j x_j^{\alpha_j'}. \tag{7}$$

Now substituting

$$x_k = \exp\{2\pi i[r_{kj}/(N_k' + 1)]\}, \quad k = 1, \cdots, s, \tag{8}$$

so that

$$x_k^{\alpha_k} = \exp\{2\pi i[\alpha_k r_{kj}/(N_k' + 1)]\}. \tag{9}$$

This gives from (6) and (7)

$$\exp\left[-2\pi i \sum_k \frac{\alpha_k^0 r_{kj}}{N_k' + 1}\right]\mathbf{O}\left[\exp\left(2\pi i \frac{\alpha_1 r_{1j}}{N_1' + 1}\right), \cdots, \exp\left(2\pi i \frac{\alpha_s r_{sj}}{N_s' + 1}\right)\right] = \sum_k \sum_{\alpha_k=t_k'}^{t_k'+N_k} A_{\alpha_1\cdots\alpha_s} \exp\left[2\pi i \sum_k \frac{\alpha_k r_{kj}}{N_k' + 1}\right]. \tag{10}$$

Using the result,

$$\sum_{r_{kj}=0}^{N_k'} \exp\left[2\pi i \frac{\alpha_k' r_{kj}}{N_k' + 1}\right] = (N_k' + 1)\delta_{\alpha_k',0} \tag{11}$$

(where $r_{kj} = 0, 1, \cdots, N_k'$ for $j = 0, 1, \cdots, N_k'$) providing

$$|\alpha_k'| < (N_k' + 1) \tag{12}$$

for all values of α_k. Summing Eq. (10) over all r_{kj},

$$\sum_k \sum_{r_{kj}=0}^{N_k'} \exp\left[-2\pi i \sum_k \frac{\alpha_k^0 r_{kj}}{N_k' + 1}\right]\mathbf{O}\left[\exp\left(2\pi i \frac{\alpha_1 r_{1j}}{N_1' + 1}\right), \cdots, \exp\left(2\pi i \frac{\alpha_s r_{sj}}{N_s' + 1}\right)\right] = \sum_k \sum_{r_{kj}=0}^{N_k'} \sum_{\alpha_k'=t_k'}^{t_k'+N_k} A_{\alpha_1\cdots\alpha_s} \exp\left[2\pi i \sum_k \frac{\alpha_k' r_{kj}}{N_k' + 1}\right], \tag{13}$$

which reduces, from (11), to

$$= \sum_k \sum_{\alpha_k'=t_k'}^{t_k'+N_k} A_{\alpha_1\cdots\alpha_s} \prod_{j=1}^{s} (N_j' + 1)\delta_{\alpha_j',0} \tag{14}$$

$$= \left[\prod_{j=1}^{s} (N_j' + 1)\right] A_{\alpha_1^0\cdots\alpha_s^0}. \tag{15}$$

Hence

$$A_{\alpha_1^0\cdots\alpha_s^0} = \frac{\sum_k \sum_{r_{kj}=0}^{N_k'} \exp\left(-2\pi i \sum \frac{\alpha_k^0 r_{kj}}{N_k' + 1}\right)\mathbf{O}\left[\exp\left(2\pi i \frac{\alpha_1 r_{1j}}{N_1' + 1}\right), \cdots, \exp\left(2\pi i \frac{\alpha_s r_{sj}}{N_s' + 1}\right)\right]}{\left[\prod_{j=1}^{s} (N_j' + 1)\right]}. \tag{16}$$

Thus, by evaluating the numerator numerically over all values of r_{kj}, choosing N_k' so that condition (12), i.e., $|\alpha_k| < (N_k' + 1)$ for all α_k, is satisfied, the coefficient $A_{\alpha_1^0\cdots\alpha_s^0}$ is obtained. The optimum value of N_k', which will reduce the number of terms in the summations in Eq. (16), is given by the maxi-

mum value of the integer $|\alpha_k|$. If we choose N_k' so that (12) is satisfied for all coefficients $A_{\alpha_1 \cdots \alpha_s}$, we can evaluate the entire generating function, or by choosing appropriate values of N_k' restrict the size of the computation so that only those coefficients required are calculated. The amount of computation can be further reduced by considering a reduced generating function. For example by considering the generating function

$$G'(x_1, \cdots, x_s) = \left(\prod_{k=1}^{s} x_k^{-t_k - \{\frac{1}{2}N_k\}}\right) G(x_1, \cdots, x_s), \quad (17)$$

where $\{\frac{1}{2}N_k'\}$ has the values

$$\{\tfrac{1}{2}N_k\} = \tfrac{1}{2}N_k \quad \text{if} \quad N_k \quad \text{is even}, \quad (18)$$

$$= \tfrac{1}{2}N_k + \tfrac{1}{2} \text{ if } N_k \text{ is odd}, \quad (19)$$

the values of N_k', Eq. (12), can be decreased, thus reducing the number of elements in the summation from 0 to N_k' in Eq. (16).

In many physical problems the generating function has considerable symmetry in the variables.[7] This enables us to restrict the summation in Eq. (16) providing we weight each point in an irreducible volume of $x_1 \cdots x_s$ space suitably. For example, a point that occurs on a boundary face of the reduced volume will only have, in general, half the weight of a point inside this volume. The symmetry can also be used to restrict the calculation of the coefficients, $A_{\alpha_1 \cdots \alpha_s}$, as many coefficients may be equal, or multiples, of other coefficients.

Storage is often an important factor in programming this calculation for a digital computer if the generating function contains many variables. Thus it is frequently convenient to store all the elements of $\mathbf{O}(x_1, \cdots, x_s)$ in the double sum

$$\sum_k \sum_{r_{ki}=0}^{N_k'} \exp\{-2\pi i \sum \alpha_k^0 r_{ki}/(N_k' + 1)\} \mathbf{O}(x_1 \cdots x_s),$$

Eq. (16), or those that occur only in one irreducible symmetry element of the $x_1, \cdots, x_s$ space, over which the summation in (16) is taken, and obtain the

$$\sum_k \sum_{r_{ki}=0}^{N_k} \exp\{-2\pi i \sum \alpha_k^0 r_{ki}/(N_k' + 1)\} \mathbf{O}(x_1, \cdots, x_s)$$

by retrieving the elements of $\mathbf{O}(x_1, \cdots, x_s)$ from the store to perform this sum for each coefficient $A_{\alpha_1 \cdots \alpha_s}$.

The calculation of the coefficients arising in the expansion of generating functions will, in general, require the use of complex arithmetic, but in many problems x_i^r and x_i^{-r} (r being an integer) appear symmetrically or asymmetrically so that the trigometrical functions can be employed. This enables much of the calculation to be carried out with the use of real numbers.

THE EXTENSION TO INFINITE MULTINOMIAL GENERATING FUNCTIONS

The finite multinomial generating functions arise primarily in the statistical mechanics of lattice problems involving short, finite-range interatomic forces. When the range of the force extends to infinity the generating function is expressible as an infinite multinomial.[8] Consequently the method developed for infinite multinomial generating functions has to be modified before the coefficients $A_{\alpha_1 \cdots \alpha_s}$ can be evaluated.

The analysis of the last section requires only a small modification to include the case of infinite multinomial generating functions. If we replace x_k in Eqs. (8) and (9) by

$$x_k = C_k \exp(2\pi i \theta_k), \quad (20)$$

where the numerical value of the constants C_k will be chosen later. Equation (11) must be replaced by

$$\int_0^1 \exp(2\pi i \alpha_k \theta_k)\, d\theta_k = \delta\alpha_{k,0}, \quad (21)$$

there now being no restrictions on α_k [see Eq. (12)]. Equation (16) becomes for both finite and infinite multinomial expansions

$$A_{\alpha_1^0 \cdots \alpha_s^0} = (\pi C_k) \int_0^1 \cdots \int_0^1 \exp(-2\pi i \sum \alpha_k^0 \theta_k)$$

$$\times \mathbf{O}[C_1 e^{2\pi i \alpha_1 \theta_1}, \cdots C_s e^{2\pi i \alpha_s \theta_s}]\, d\theta_1 \cdots d\theta_s, \quad (22)$$

providing the integrand in (22) does not contain any poles within the region of integration[9] in $x_1 \cdots x_s$ space other than at the origin. For finite multinomial generating function this condition is satisfied for all values of C_k as the integrand contains no poles, other than at the origin of $x_1 \cdots x_s$ space. Thus it is convenient to put $C_k = 1$ ($k = 1 \cdots s$). In general, infinite multinomial generating functions will contain poles within the region integration in $x_1 \cdots x_s$ space. Thus the numerical values of C_k must be chosen to exclude all poles other than the one at the origin before the integration can be performed.

Multidimensional integrals of this type, with cubic

[7] See for example, L. Brillouin, *Wave Propagation in Periodic Structures*, (McGraw-Hill Book Company, Inc., London, 1946).

[8] See Refs. 1–4.

[9] E. G. Phillips, *Functions of a Complex Variable* (Oliver and Boyd, London, 1958).

boundary conditions, which cannot be evaluated in closed form have been studied by Miller,[10] Tyler,[11] and Irwin.[12] They have derived cubature formulas for these integrals to enable their numerical value to be rapidly computed.

A ONE-DIMENSIONAL GENERATING FUNCTION

In order to illustrate the general analysis of the previous sections, we consider a generating function containing only one parameter, x [$k = 1$ in Eq. (16)], given by

$$G_p(x) = \mathrm{Tr}\,[\mathbf{C}(x)]^p, \tag{23}$$

where

$$\mathbf{C}(x) = \mathbf{A} + x\mathbf{B}. \tag{24}$$

A and **B** are square matrices containing numerical elements and having no symmetry properties.

The coefficient of the (α_1^0)th power of x in the expansion of $G_p(x)$ is, from (16),

$$A_{\alpha_1^0} = (N_1' + 1)^{-1} \sum_{r_1 i=0}^{N_1'} \exp\left(-2\pi i \frac{\alpha_1^0 r_{1i}}{N_1' + 1}\right) \times \mathrm{Tr}\left[\mathbf{C}\left(\exp 2\pi i \frac{\alpha_1 r_{1i}}{N_1' + 1}\right)\right]^p. \tag{25}$$

Now we require a suitable value of N_1' which satisfies condition (12). The pth power of **C**, i.e., $\mathbf{A} + x\mathbf{B}$ contains matrix elements which are polynomials in x of degree p. Consequently $G_p(x)$ will be a polynomial of degree p in x. To evaluate $A_{\alpha_1^0}$ all the powers of x, α_1 ($\alpha_1 = 0, 1, \cdots, p$), in the expansion of $G_p(x)$ must, from (12) satisfy

$$|\alpha_1'| < N_1' + 1, \tag{26}$$

where

$$\alpha_1' = \alpha_1 - \alpha_1^0. \tag{27}$$

However, we require all the coefficients $A_{\alpha_1^0}$ ($\alpha_1^0 = 0, 1, \cdots, p$) so that condition (26) must apply for all values of α_1^0. The greatest value of $|\alpha_1'|$ is p with the result that (26) becomes for all α_1^0

$$p < N_1' + 1. \tag{28}$$

Hence the optimum choice of N_1', which reduces the number of terms in the summation (25) to a minimum, is $N_1' = p$. All the coefficients of $x^{\alpha_1^0}$ ($\alpha_1^0 = 0, 1, \cdots, p$) in $G_p(x)$ are consequently given by substituting $N_1' = p$ in (25) which reduces to

$$A_{\alpha_1^0} = (p + 1)^{-1} \sum_{r_1 i=0}^{p} \exp\left(-2\pi i \frac{\alpha_1^0 r_{1i}}{(p + 1)}\right) \times \mathrm{Tr}\left[\mathbf{C}\left(\exp 2\pi i \frac{\alpha_1 r_{1i}}{p + 1}\right)\right]. \tag{29}$$

[10] J. C. P. Miller, Math. Comp. **14**, 13, 130, 240 (1960).
[11] G. W. Tyler, J. Can. Math. Soc. **4**, 393 (1953). I am grateful to G. S. Joyce for this reference.
[12] J. O. Irwin, Tracts for Computers **10** (Cambridge 1923).

TABLE I. The coefficients of the polynomial, independent of x_1, x_2, x_3, in powers of x_4^s ($s = 0, 1, \ldots, 14$) in the expansion of $G_{14}(x_1, x_2, x_3, x_4)$.

s	A_{000s}			
0	3	05033	88756	64384
1	40	90267	18814	38528
2	259	03590	73046	99616
3	1026	52767	95700	32640
4	2843	09035	27336	71040
5	5821	81091	67063	60576
6	9094	84138	00806	48192
7	11027	91102	44089	71264
8	10459	70589	35290	52160
9	7760	44513	27372	56960
10	4467	75401	24402	46528
11	1956	87232	21870	01856
12	625	21282	41987	96288
13	132	40573	63169	58720
14	14	18632	88911	02720

The digital computation of this sum can now be performed with the most economic use of computing time by evaluating, and storing, the elements of Tr $[\mathbf{C}(x)]^p$ for all values of r_{1i} ($r_{1i} = 0, 1, \cdots, p$) and subsequently evaluating the coefficients $A_{\alpha_1^0}$ ($\alpha_1^0 = 0, 1, \cdots, p$) by performing the summation in Eq. (29) using the values of Tr $[\mathbf{C}(x)]^p$ already calculated.

Previously the evaluation of generating functions of the form (26) could only be performed for small values of n, the order of the matrix **C**. Equation (29) enables computations, on the fastest electronic computers, to be extended to matrices where $n \sim 50$ and $p \sim 50$.

A NUMERICAL EXAMPLE OF THE USE OF THE METHOD

An example which illustrates the use of the method for the numerical evaluation of coefficients in the expansion of a finite multinomial generating function is taken from the field of lattice dynamics. Here one requires the polynomials in a four-variable generating function, $G_n(x_1, x_2, x_3, x_4)$, which are independent of x_1, x_2, and x_3.

$G_n(x_1, x_2, x_3, x_4)$ is defined by[13]

$$G_n(x_1, x_2, x_3, x_4) = \tfrac{1}{3}\,\mathrm{Tr}\,(\mathbf{D}^n), \tag{30}$$

where the 3×3 matrix **D** has elements

$$d_{11} = 8 - (x_2 + x_2^{-1}) + (x_3 + x_3^{-1}) + (x_1x_2^{-1} + x_1^{-1}x_2) + (x_1x_3^{-1} + x_1^{-1}x_3) + x_4y \tag{31}$$

$$d_{23} = (x_2x_3^{-1} + x_2^{-1}x_3) - (x_1 + x_1^{-1}); \tag{32}$$

[13] T. H. K. Barron, Phil. Mag. **46**, 720 (1955); thesis, Oxford (1956) (unpublished).

the remaining elements can be obtained by cyclically permuting the suffixes, and y is given by

$$y = 12 - \sum_{\text{cyclic}} (x_1 + x_1^{-1}) + (x_1 x_2^{-1} + x_1^{-1} x_2), \quad (33)$$

where the summation is taken over all cyclic permutations of x_1, x_2, and x_3.

Here the **D** matrix has the symmetry of the face-centered cubic lattice in the x_1, x_2, and x_3 coordinates.[7] Consequently it is not necessary to evaluate all of the r_{ki} in Eq. (16). One can restrict the calculation to $\frac{1}{48}$th of the x_1, x_2, x_3 space, provided that all the points in this volume are given appropriate weights as discussed previously.

The first 14 polynomials in x_4 ($n = 1, 2, \cdots, 14$), independent of x_1, x_2, x_3 were computed to double-length accuracy on the Manchester University Atlas computer in 50 sec. This machine is inefficient for double-length arithmetic, taking about ten times as long to do a double-length operation as a single-length one.

The 14th polynomial is tabulated in Table I. Previous techniques have only enabled the first four polynomials to be obtained.[12,13] The limit to the number of polynomials now obtainable is dependent only on the machine computation time, which increases approximately as n^d, the index d is the dimensionality of the problem. For this example, $d = 4$.

ACKNOWLEDGMENTS

I would like to thank Professor C. Domb for his criticisms of the paper and his interest in the future applications of the method. Thanks are also due to the Department of Scientific and Industrial Research for the award of a maintenance grant, and to Manchester University for the use of their Atlas computer.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Generalization of the Variational Method of Kahan, Rideau, and Roussopoulos and Its Application to Neutron Transport Theory*

MORTON D. KOSTIN†
School of Engineering and Applied Science, Princeton University, Princeton, New Jersey
AND
HARVEY BROOKS
Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts
(Received 9 June 1964)

The variational method of Kahan, Rideau, and Roussopoulos (KRR) frequently used in neutron transport theory to estimate weighted averages is extended and generalized. In the KRR variational method a first variation in the trial functions produces a second variation in the estimate of the weighted average. Two generalized variational functionals which depend on trial operators instead of trial functions are given. A first variation in the trial operators produces an nth variation in the estimate of the weighted average when an nth order generalized variational functional is used. Both perturbation theory and the KRR variational method are derived as special cases of the generalized variational method. Several examples including calculations of transport equation spatial moments using diffusion equation solutions as trial operators are studied with good results.

1. INTRODUCTION

ONE of the first applications of the variational method to inhomogeneous neutron transport equations of the form

$$f(x) = \int k(x, x') f(x')\, dx' + s(x) \quad (1)$$

was made by Marshak[1] who defined the functional

$$j_M[f_t] = \frac{\int dx\, f_t(x)\left[f_t(x) - \int dx'\, k(x, x') f_t(x') \right]}{\left[\int dx\, f_t(x) s(x) \right]^2}, \quad (2)$$

where $f(x)$ is the neutron flux and here $k(x, x')$ is a positive symmetric kernel. By writing the trial func-

* Research supported in part by the Office of Naval Research.
† Sloan Postdoctoral Fellow.

[1] R. E. Marshak, Phys. Rev. 71, 688 (1947).

the remaining elements can be obtained by cyclically permuting the suffixes, and y is given by

$$y = 12 - \sum_{\text{cyclic}} (x_1 + x_1^{-1}) + (x_1 x_2^{-1} + x_1^{-1} x_2), \quad (33)$$

where the summation is taken over all cyclic permutations of x_1, x_2, and x_3.

Here the **D** matrix has the symmetry of the face-centered cubic lattice in the x_1, x_2, and x_3 coordinates.[7] Consequently it is not necessary to evaluate all of the r_{ki} in Eq. (16). One can restrict the calculation to $\frac{1}{48}$th of the x_1, x_2, x_3 space, provided that all the points in this volume are given appropriate weights as discussed previously.

The first 14 polynomials in x_4 ($n = 1, 2, \cdots, 14$), independent of x_1, x_2, x_3 were computed to double-length accuracy on the Manchester University Atlas computer in 50 sec. This machine is inefficient for double-length arithmetic, taking about ten times as long to do a double-length operation as a single-length one.

The 14th polynomial is tabulated in Table I. Previous techniques have only enabled the first four polynomials to be obtained.[12,13] The limit to the number of polynomials now obtainable is dependent only on the machine computation time, which increases approximately as n^d, the index d is the dimensionality of the problem. For this example, $d = 4$.

ACKNOWLEDGMENTS

I would like to thank Professor C. Domb for his criticisms of the paper and his interest in the future applications of the method. Thanks are also due to the Department of Scientific and Industrial Research for the award of a maintenance grant, and to Manchester University for the use of their Atlas computer.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Generalization of the Variational Method of Kahan, Rideau, and Roussopoulos and Its Application to Neutron Transport Theory*

MORTON D. KOSTIN†
School of Engineering and Applied Science, Princeton University, Princeton, New Jersey
AND
HARVEY BROOKS
Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts
(Received 9 June 1964)

The variational method of Kahan, Rideau, and Roussopoulos (KRR) frequently used in neutron transport theory to estimate weighted averages is extended and generalized. In the KRR variational method a first variation in the trial functions produces a second variation in the estimate of the weighted average. Two generalized variational functionals which depend on trial operators instead of trial functions are given. A first variation in the trial operators produces an nth variation in the estimate of the weighted average when an nth order generalized variational functional is used. Both perturbation theory and the KRR variational method are derived as special cases of the generalized variational method. Several examples including calculations of transport equation spatial moments using diffusion equation solutions as trial operators are studied with good results.

1. INTRODUCTION

ONE of the first applications of the variational method to inhomogeneous neutron transport equations of the form

$$f(x) = \int k(x, x') f(x')\, dx' + s(x) \quad (1)$$

was made by Marshak[1] who defined the functional

$$j_M[f_t] = \frac{\int dx\, f_t(x)\left[f_t(x) - \int dx'\, k(x, x') f_t(x') \right]}{\left[\int dx\, f_t(x) s(x) \right]^2}, \quad (2)$$

where $f(x)$ is the neutron flux and here $k(x, x')$ is a positive symmetric kernel. By writing the trial func-

* Research supported in part by the Office of Naval Research.
† Sloan Postdoctoral Fellow.

[1] R. E. Marshak, Phys. Rev. 71, 688 (1947).

tion as $f_t(x) = f(x) + \delta f(x)$, where $f(x)$ satisfies (1) and $\delta f(x)$ is the correction function, he showed that the functional (2) is equal to the weighted average

$$\left[\int dx\, f(x)s(x)\right]^{-1} \tag{3}$$

plus terms which are quadratic in the correction function.

Marshak then demonstrated that from the weighted average (3) an estimate of the asymptotic neutron density for the Milne problem could be obtained. Using a very simple trial function, he calculated the estimate $f_M(x) - x \to 0.7083$ as $x \to \infty$ which compares favorably with the exact value $f(x) - x \to 0.7104461$ as $x \to \infty$. With a more sophisticated trial function, LeCaine[2] computed the result $f_{LC}(x) - x \to 0.7104457$ as $x \to \infty$.

The variational functional discussed by Marshak yields an estimate of the weighted average of the neutron flux with the source term as a weighting function. Employing a modification of the Schwinger variational functional[3,4] used for the calculation of nuclear phase shifts, Francis, Stewart, Bohl, and Krieger[5] calculated arbitrary averages of the neutron distribution. They considered the variational expression

$$j_{SF}[g_t, f_t] = \frac{\int g_t(x)s(x)\,dx \int w(x)f_t(x)\,dx}{\int g_t(x)f_t(x)\,dx - \iint g_t(x)k(x, x')f_t(x')\,dx\,dx'} \tag{4}$$

and its associated supplementary problem

$$g(x) = \int k(x, x')g(x')\,dx' + w(x), \tag{5}$$

where $w(x)$ is the weighting function and here $k(x, x')$ is a symmetric kernel. If the trial functions $f_t(x)$ and $g_t(x)$ closely approximate the exact solutions of (1) and (5),

$$f_t(x) = f(x) + \delta f(x), \tag{6}$$

$$g_t(x) = g(x) + \delta g(x), \tag{7}$$

then the functional (4) is equal to the desired weighted average

$$\int w(x)f(x)\,dx \tag{8}$$

plus terms which depend on the product of the errors in the trial functions. This variational expression is stationary about (8) for small but arbitrary variations of the trial functions about the exact solutions $f(x)$ and $g(x)$. If the weighting function $w(x)$ is set equal to the source function $s(x)$ in (4) and (5), then the supplementary problem (5) is identical to (1) and the functional (4) reduces to the reciprocal of Marshak's functional (2):

$$j_{SF}[f_t, f_t] = \frac{\left[\int s(x)f_t(x)\,dx\right]^2}{\int dx\, f_t(x)\left[f_t(x) - \int dx'\, k(x, x')f_t(x')\right]}.$$

The variational functional

$$j_{KRR}[g_t, f_t] = \int g_t(x)^*s(x)\,dx + \int w(x)f_t(x)\,dx - \int g_t(x)^*Lf_t\,dx$$

$$= (g_t, s) + (w, f_t) - (g_t, Lf_t), \tag{9}$$

explicitly introduced by Kahan, Rideau, and Roussopoulos,[6-8] (KRR) was first applied to problems in transport theory by Devooght[9] and Selengut.[10] Corngold[11] has used a related functional to construct a variational principle for the resonance escape probability. Here, L represents a linear operator and the symbol (g_t, s) denotes the Hermitian scalar product $\int g_t(x)^*s(x)\,dx$. Since we are dealing with real quantities, the complex conjugate sign is not used.

It has been shown[8,10,12] that the functional (9) may be transformed into a form which is independent of the normalization of the trial functions by setting

$$f_t(x) = a\bar{f}_t(x), \tag{10}$$

$$g_t(x) = b\bar{g}_t(x), \tag{11}$$

where a and b are multiplicative trial parameters and $f_t(x)$ and $g_t(x)$ are the new trial functions.

$$j_{KRR}[b\bar{g}_t, a\bar{f}_t] = b(\bar{g}_t, s) + a(w, \bar{f}_t) - ba(\bar{g}_t, L\bar{f}_t). \tag{12}$$

Requiring that (12) be stationary with respect to variations in a and b yields

[2] J. LeCaine, Phys. Rev. **72**, 564 (1947).
[3] J. Schwinger, Phys. Rev. **72**, 742 (1947); Hectographed notes on nuclear physics, Harvard, 1947.
[4] J. M. Blatt and J. D. Jackson, Phys. Rev. **76**, 18 (1949).
[5] N. C. Francis, J. C. Stewart, L. S. Bohl, and T. J. Krieger, Progr. Nucl. Energy, Ser. I **3**, 360 (1959).
[6] T. Kahan and G. Rideau, Compt. Rend., **233**, 849 (1951).
[7] T. Kahan and G. Rideau, J. Phys. Radium **13**, 326 (1952).
[8] P. Roussopoulos, Compt. Rend. **236**, 1858 (1953).
[9] J. Devooght, Phys. Rev., **111** 665 (1958).
[10] D. S. Selengut, Hanford Laboratories Report HW-59126, 1959 (unpublished).
[11] N. Corngold, Proc. Phys. Soc. (London) **A70**, 793 (1957).
[12] G. Rowlands, J. Nucl. Energy **13**, 176 (1961).

$$a = (\bar{g}_t, s)/(\bar{g}_t, L\bar{f}_t), \tag{13}$$

$$b = (w, \bar{f}_t)/(\bar{g}_t, L\bar{f}_t). \tag{14}$$

The normalization-independent functional is obtained by combining (13) and (14) with (12).

$$\bar{j}_{KRR}[\bar{g}_t, \bar{f}_t] = (\bar{g}_t, s)(w, \bar{f}_t)/(\bar{g}_t, L\bar{f}_t). \tag{15}$$

If the operator L is self-adjoint, then (15) is equivalent to the functional (4) used by Francis *et al.*

2. THE KRR VARIATIONAL FUNCTIONAL

Let us investigate the KRR variational functional

$$j_{KRR}[g_t, f_t] = (g_t, s) + (w, f_t) - (g_t, Lf_t) \tag{16}$$

in more detail. We consider small variations of the trial functions $f_t(x)$ and $g_t(x)$ about the functions $u_1(x)$ and $u_2(x)$.

$$f_t(x) = u_1(x) + \delta f(x), \tag{17}$$

$$g_t(x) = u_2(x) + \delta g(x). \tag{18}$$

We may express $f_t(x)$ as a member of a one-parameter family with small parameter ϵ_1,

$$f_t(x) = u_1(x) + \epsilon_1 v_1(x), \tag{19}$$

where $\epsilon_1 v_1(x) = \delta f(x)$ is called a first variation of the function. Similarly, $g_t(x)$ may be expressed as a member of another one-parameter family,

$$g_t(x) = u_2(x) + \epsilon_2 v_2(x). \tag{20}$$

Substituting (19) and (20) into the functional (16), we obtain

$$\begin{aligned} j_{KRR}[g_t, f_t] &= (w, u_1) + (u_2, s) - (u_2, Lu_1) \\ &\quad + \epsilon_1[(w, v_1) - (L^+u_2, v_1)] \\ &\quad + \epsilon_2[(v_2, s) - (v_2, Lu_1)] - \epsilon_1\epsilon_2(v_2, Lv_1) \\ &= j + \delta j + \delta^2 j, \end{aligned} \tag{21}$$

where

$$\delta j = \epsilon_1[(w, v_1) - (L^+u_2, v_1)] + \epsilon_2[(v_2, s) - (v_2, Lu_1)], \tag{22}$$

which depends linearly on ϵ_1 and ϵ_2 is called a first variation of the functional, and

$$\delta^2 j = -\epsilon_1\epsilon_2(v_2, Lv_1), \tag{23}$$

which depends on the product of the small parameters ϵ_1 and ϵ_2 is called a second variation of the functional.

If $u_1(x)$ is a solution of

$$Lf(x) = s(x), \tag{24}$$

and $u_2(x)$ of

$$L^+g(x) = w(x), \tag{25}$$

where L^+ is the operator adjoint to L, i.e., $u_1(x) = f(x)$ and $u_2(x) = g(x)$, then the first variation of the functional vanishes leaving the weighted average (w, f) and the second variation of the functional

$$j_{KRR}[g_t, f_t] = (w, f) + \delta^2 j. \tag{26}$$

Thus, a first variation of the trial functions about the solutions of (24) and (25) produces a second variation of the functional (16) about the weighted average (w, f).

3. THE GENERALIZED VARIATIONAL FUNCTIONAL j_A

The equations of linear neutron transport theory can be represented by the operator equation

$$Lf(x) = s(x), \tag{27}$$

where the operator L relates the flux distribution $f(x)$ to the neutron source distribution $s(x)$. A second important operator is the Green's operator G, the inverse of L,

$$LG = GL = I, \tag{28}$$

which produces the flux distribution from the source distribution,

$$Gs(x) = f(x). \tag{29}$$

Two generalizations of the variational method are presented. Consider n trial Green's operators G_1, G_2, $\cdots$, G_n where each trial Green's operator G_i differs from the exact Green's operator G by the small operator variations δG_i.

$$G_i = G - \delta G_i \qquad i = 1, 2, \cdots, n \tag{30}$$

We seek a functional $j_{An}[G_1, \cdots, G_n]$ which differs from the exact weighted average $(w, Gs) = (w, f)$ by a term which depends on the product of the δG_i. The operator, now defined,

$$E_n = L\delta G_1 L\delta G_2 \cdots L\delta G_n = \prod_{i=1}^{n} L\delta G_i \tag{31}$$

involves such a product, and the desired functional is obtained:

$$\begin{aligned} j_{An}[G_1, \cdots, G_n] &= (w, G[I - E_n]s) \\ &= \left(w, G\left[I - \prod_{i=1}^{n} L\delta G_i\right]s\right) \\ &= \left(w, G\left[I - \prod_{i=1}^{n} (LG - LG_i)\right]s\right) \\ &= \left(w, G\left[I - \prod_{i=1}^{n} (I - LG_i)\right]s\right). \end{aligned} \tag{32}$$

When (32) is written out explicitly,

$$j_{A1} = (w, G_1 s), \tag{33}$$

$$j_{A2} = (w, [G_1 + G_2 - G_1 L G_2] s), \tag{34}$$

$$j_{A3} = (w, [G_1 + G_2 + G_3 - G_1 L G_2 - G_1 L G_3 - G_2 L G_3 + G_1 L G_2 L G_3] s), \tag{35}$$

we see that it does not depend on G. The derivation of (32) shows that first variations of the trial operators G_i about the exact Green's operator G result in a nth variation of the functional j_{An} about the weighted average (w, Gs),

$$j_{An} = (w, Gs) - (w, \delta G_1 L \delta G_2 \cdots L \delta G_n s). \tag{36}$$

The variational functional (32) has the property that the first, second, third, $\cdots$, $(n-1)$th variations of the functional all vanish,

$$\delta^m j_{An} = 0 \quad \text{for} \quad m = 1, 2, \cdots, n-1. \tag{37}$$

The KRR variational functional can be derived from j_{A2}:

$$j_{A2} = (w, [G_1 + G_2 - G_1 L G_2] s) = (G_1^+ w, s) + (w, G_2 s) - (G_1^+ w, L G_2 s). \tag{38}$$

Since G_1 and G_2 are independent trial operators, we can let them generate the trial functions

$$g_t = G_1^+ w, \qquad f_t = G_2 s. \tag{39}$$

Combining (39) with (38) we arrive at the KRR functional (16).

From (32) we see that the nth-order generalized variational functional j_{An} is reduced to one of $(n-1)$th order by setting one of the trial operators equal to the null operator O. For example, $j_{A3}[G_1, G_2, G_3]$ which is given in (35) becomes $j_{A2}[G_1, G_2]$ if we set $G_3 = O$.

The functional (32) involves the n trial operators $G_1, G_2, \cdots, G_n$. If we set them equal to a single trial operator G_t,

$$G_1 = G_2 = \cdots = G_n = G_t, \tag{40}$$

then (32) reduces to

$$j_{An}[G_t] = (w, G[I - (I - L G_t)^n] s). \tag{41}$$

This functional may be expressed in terms of binomial coefficients.

$$j_{A1} = (w, G_t s), \tag{42}$$

$$j_{A2} = 2(w, G_t s) - (w, G_t L G_t s), \tag{43}$$

$$j_{A3} = 3(w, G_t s) - 3(w, G_t L G_t s) + (w, G_t [L G_t]^2 s), \tag{44}$$

$$j_{An} = \sum_{m=1}^{n} \binom{n}{m} (w, G_t [-L G_t]^{m-1} s). \tag{45}$$

The variational method can be used to obtain an approximate value of the flux $f(x)$ itself by setting the weighting function equal to the delta function:

$$w(x') = \delta(x - x'). \tag{46}$$

From (34) and (39) we obtain a functional

$$j[h_t, f_t] = f_t(x) + \int h_t(x, x') s(x')\, dx' - \int h_t(x, x') L f_t(x')\, dx', \tag{47}$$

which depends on a trial function $f_t(x)$ and a trial Green's function $h_t(x, x')$ and is similar to one used by Devooght.

The relation between the perturbation method and the generalized variational method is easily established.

$$LG = GL = I, \qquad L = L_o + L_p \tag{48}$$

is the perturbed problem; L_p is the perturbing operator, and the unperturbed Green's operator G_o satisfies the equation

$$L_o G_o = G_o L_o = I, \tag{49}$$

where I is the identity operator. Let G_o, without any trial parameters, be the trial operator in (41). Then

$$\begin{aligned} j_{An}[G_o] &= (w, G[I - (I - L G_o)^n] s) \\ &= (w, G_o [L G_o]^{-1} [I - (-L_p G_o)^n] s) \\ &= (w, G_o [I + L_p G_o]^{-1} [I - (-L_p G_o)^n] s) \\ &= (w, G_o [I - (L_p G_o) + (L_p G_o)^2 - \cdots + (-L_p G_o)^{n-1}] s) \end{aligned} \tag{50}$$

becomes the familiar nth-order perturbation series. When the source term is also the weighting function

$$w(x) = s(x), \tag{51}$$

a lower bound to the desired weighted average $(w, f) = (s, Gs)$ can be obtained if L is a self-adjoint, positive definite operator. If we require the trial operator G_t to be self-adjoint, then $\delta G = \delta G^+$. From (51), for even n, we see that the variational functional

$$\begin{aligned} j_{An} &= (s, Gs) - (s, \delta G [L \delta G]^{1/(2n-1)} L [\delta G L]^{1/(2n-1)} \delta G s) \\ &= (s, Gs) - ([\delta G L]^{1/(2n-1)} \delta G s, L [\delta G L]^{1/(2n-1)} \delta G s) \end{aligned} \tag{52}$$

must be less than or equal to the exact weighted average (s, Gs) if L is positive definite.

4. EXAMPLE

The ideas developed in the previous sections are illustrated by applying them to the diffusion equation for a homogeneous isotropic infinite medium with a macroscopic absorption cross section Σ_a, a diffusion coefficient D, and an infinite plane source emitting 1 neutron/cm^2 sec located at $z = z'$:

$$(\Sigma_a - D\, d^2/dz^2)g(z, z') = \delta(z - z'). \quad (53)$$

The Green's function $g(z, z')$ gives the steady-state flux distribution at position z resulting from the source at z'. The weighted average $(w, f) = (w, Gs)$ can be expressed explicitly in terms of the Green's function,

$$(w, Gs) = \iint w(z)g(z, z')s(z')\, dz\, dz'. \quad (54)$$

If we let the source and weighting functions be delta functions

$$s(z) = w(z) = \delta(z - z_s), \quad (55)$$

then the desired weighted average becomes the value of the flux at the source position.

$$(s, Gs) = g(z_s, z_s). \quad (56)$$

For the diffusion equation (53), the Green function is known exactly.

$$g(z, z') = (1/2kD) \exp(-k\,|z - z'|), \quad (57)$$

where $k^2 = \Sigma_a/D$. Let $kD = \frac{1}{2}$ so that the exact weighted average (56) has the value $(s, Gs) = 1$.

Using a trial Green's operator G_t with the second- and fourth-order variational methods we obtain estimates of the weighted average (56) and compare them with the exact answer. For simplicity we consider the trial Green's function

$$g_t(z, z') = (a_t/2k_1D) \exp(-k_1\,|z - z'|), \quad (58)$$

where $k_1 = \Sigma_{a1}/D$. This Green's function corresponds to a Green's operator in the form $G_t = a_tH$ where a_t is a variational or trial parameter and H is a fixed Green's operator. Equation (58) has a definite interpretation; it is the Green's function for a medium with the same diffusion coefficient as the original medium but with a different absorption cross section and a plane source emitting a_t neutron·cm^2/sec. The dimensionless constant

$$\beta = (k^2 - k_1^2)/k_1^2 \quad (59)$$

expresses the difference between the original medium and the medium corresponding to (58).

We start with the second-order variational method:

$$j_{A2} = 2a_t(s, Hs) - a_t^2(s, HLHs). \quad (60)$$

The Ritz condition $dj_{A2}/da_t = 0$ determines the variational parameter and leads us to the expression

$$j_{A2} = (s, Hs)^2/(s, HLHs). \quad (61)$$

After a straightforward evaluation of the terms (s, Hs) and $(s, HLHs)$ for $\beta = 0.1$, we find that the variational answer $j_{A2} = 0.9988$ compares favorably with the exact result, $(s, Gs) = 1.0000$.

Since H is a self-adjoint operator, L is a self-adjoint positive definite operator, and the source and weighting terms are identical, (52) shows that all the even-order variational weighted averages will yield a lower bound for (s, Gs).

The fourth-order variational functional

$$j_{A4} = 4a_t(s, Hs) - 6a_t^2(s, HLHs) + 4a_t^3(s, H(LH)^2s) - a_t^4(s, H(LH)^3s), \quad (62)$$

a quartic function of variational parameter, takes its maximum value for $a_t = 0.953$. This gives a variational answer $j_{A4} = 0.999998$ which shows a considerable improvement over the second order variational method. Additional comparisons between j_{A2} and j_{A4} are presented in Table I.

TABLE I. Variational estimates for the weighted average (s, Gs) obtained from second- and fourth-order variational functionals.

β	Second-order variational functional j_{A2}	Fourth-order variational functional j_{A4}
0.1	0.9988	0.999998
0.3	0.991	0.99989
0.5	0.980	0.99939
0.7	0.966	0.9983
0.9	0.951	0.9964
Exact result $(s, Gs) = 1.000000$		

5. THE GENERALIZED VARIATIONAL FUNCTIONAL j_B

Another generalized variational function can be derived from

$$j_{A2}[\bar{G}_t] = (w, [2\bar{G}_t - \bar{G}_tL\bar{G}_t]s) \quad (63)$$

by using the n-parameter trial operator

$$\bar{G}_t = \sum_{i=1}^{n} a_iG_t[LG_t]^{i-1} \quad (64)$$

with the variational parameters a_i. Trial functions resembling (64) have been used by Biedenharn and

Blatt,[13] Kikuta,[14] Goldhammer and Feenberg[15] in eigenvalue problems. The trial operator $\tilde{G}_t(G_t; a_i)$ given in (64) is a function of a single operator G_t and the variational parameters a_i. A more general trial operator $\tilde{G}_t(G_1, G_2, \cdots, G_m; a_i)$ which is a function of several independent trial operators G_1, $G_2, \cdots, G_m$ and additional variational parameters could also be used in the procedure to follow. This would lead to a more general although more complicated result.

We require the trial operator G_t be such that all the w_i defined below in (66) be finite and that the matrix $\mathbf{A}_n$ given in (70) be nonsingular. Substituting (64) into (63) we obtain

$$j_{A2} = 2\sum_{i=1}^{n} a_i w_i - \sum_{i=1}^{n} a_i a_j w_{i+j}, \tag{65}$$

where

$$w_i = (w, G_t[LG_t]^{i-1}s). \tag{66}$$

On applying the Ritz conditions

$$\partial j_{A2}/\partial a_i = 0, \qquad (i = 1, 2, \cdots, n) \tag{67}$$

Eq. (65) becomes

$$\sum_{i=1}^{n} w_{i+j} a_j = w_i \tag{68}$$

which can be written in the matrix form

$$\mathbf{A}_n \cdot \mathbf{a} = \mathbf{b}, \tag{69}$$

where $\mathbf{A}_n$ is a n-by-n persymmetric matrix, $\mathbf{a}$ and $\mathbf{b}$ are n-dimensional column vectors.

$$\mathbf{A}_n = \begin{vmatrix} w_2 & w_3 & w_4 & w_{n+1} \\ w_3 & w_4 & & \\ w_4 & & & \\ w_{n+1} & & & w_{2n} \end{vmatrix} \quad \mathbf{a} = \begin{vmatrix} a_1 \\ a_2 \\ a_3 \\ a_n \end{vmatrix} \quad \mathbf{b} = \begin{vmatrix} w_1 \\ w_2 \\ w_3 \\ w_n \end{vmatrix}. \tag{70}$$

Solving for the trial parameters

$$\mathbf{a} = \mathbf{A}_n^{-1} \cdot \mathbf{b} \tag{71}$$

and combining this result with (65) we obtain the new functional

$$j_{Bn}[G_t] = \mathbf{b}^+ \cdot \mathbf{A}_n^{-1} \cdot \mathbf{b}, \tag{72}$$

where $\mathbf{b}^+$ is the n-dimensional row vector

$$\mathbf{b}^+ = |w_1\ w_2 \cdots w_n|. \tag{73}$$

The new functional $j_{Bn}[G_t]$ does not contain any variational parameters. We have derived it from the functional $j_{A2}[\tilde{G}]$ and the n-parameter trial operator (64) by eliminating the variational parameters with the Ritz conditions (67). Moreover in (93) we see that j_{Bn}, unlike j_{A2}, is independent of the normalization of the trial operator.

We now show that $j_{Bn}[G_t]$ is a $2n$th-order generalized variational functional, i.e., a first-order variation of the trial operator G_t about the exact Green's operator G produces a $2n$th-order variational of the functional (72) about the weighted average $(w, Gs) = (w, f)$.

$$j_{Bn}[G + \delta G] = (w, Gs) + \delta^{2n} j_{Bn}. \tag{74}$$

An expression for $\delta^{2n} j_{Bn}$ is also derived.

Let us consider the $n+1$ by $n+1$ persymmetric matrix

$$\mathbf{B}_{n+1} = \begin{vmatrix} w_0 & w_1 & w_2 & w_n \\ w_1 & w_2 & & \\ w_2 & & & \\ w_n & & & w_{2n} \end{vmatrix} = \left|\begin{array}{c|c} w_0 & \mathbf{b}^+ \\ \hline \mathbf{b} & \mathbf{A}_n \end{array}\right|, \tag{75}$$

where

$$w_0 = (w, Gs) \tag{76}$$

is the exact weighted average and the matrices $\mathbf{A}_n$, $\mathbf{b}$, $\mathbf{b}^+$ are given in (70), (73).

In Appendix A we derive the expression

$$\det \mathbf{B}_{n+1}/\det \mathbf{A}_n = (w, Gs) - \mathbf{b}^+ \cdot \mathbf{A}_n^{-1} \cdot \mathbf{b} \tag{77}$$

which relates the functional (72) to the exact weighted average $w_0 = (w, Gs)$ and the persymmetric determinants $\det \mathbf{A}_n$ and $\det \mathbf{B}_{n+1}$.

Let us investigate the effects of a first-order variation of the trial Green's operator G_t about the exact Green's operator G on the cross-diagonal elements w_i of the persymmetric determinant $\mathbf{B}_{n+1}$, i.e., we express G_t as a member of a one-parameter family with small parameter ϵ.

$$G_t = G + \epsilon V, \tag{78}$$

where V is an arbitrary operator. Thus, using the relation $GL = I$ we get

$$\begin{aligned} w_i &= (w, G_t[LG_t]^{i-1}s) \\ &= (w, G[LG_t]^i s) = (w, G[LG + \epsilon LV]^i s) \\ &= (w, G[I + \epsilon LV]^i s) = \left(w, G\sum_{r=0}^{i}\binom{i}{r}\epsilon^r [LV]^r s\right) \\ &= \sum_{r=0}^{i}\binom{i}{r}\epsilon^r (w, G[LV]^r s) = \sum_{r=0}^{i}\binom{i}{r}\epsilon^r c_r, \end{aligned} \tag{79}$$

[13] L. C. Biedenharn and J. M. Blatt, Phys. Rev. **93**, 230 (1954).
[14] T. Kikuta, Progr. Theoret. Phys. (Kyoto) **12**, 10 (1954); **14**, 457 (1955); **15**, 50 (1956); **16**, 231 (1956).
[15] P. Goldhammer and E. Feenberg, Phys. Rev. **101**, 1233 (1956).

where

$$c_r = (w, G[LV]^r s) = (w, V[LV]^{r-1} s). \tag{80}$$

According to the theory of persymmetric determinants,[16] the value of a persymmetric determinant with the cross-diagonal elements

$$w_i = \sum_{r=0}^{i} \binom{i}{r} \epsilon^r c_r \tag{81}$$

is equal to the value of the persymmetric determinant with the cross-diagonal elements $\epsilon^r c_r$.

$$\det \mathbf{B}_{n+1} = \det \begin{vmatrix} c_0 & \epsilon c_1 & \epsilon^2 c_2 & \epsilon^n c_n \\ \epsilon c_1 & \epsilon^2 c_2 & & \\ \epsilon^2 c_2 & & & \\ \epsilon^n c_n & & & \epsilon^{2n} c_{2n} \end{vmatrix}. \tag{82}$$

By noting that

$$\det \begin{vmatrix} c_0 & \epsilon c_1 & \epsilon^n c_n \\ \epsilon c_1 & & \\ \epsilon^n c_n & & \epsilon^{2n} c_{2n} \end{vmatrix} \tag{83}$$

$$= \det \begin{vmatrix} 1 & & \\ & \epsilon & 0 \\ 0 & & \epsilon^n \end{vmatrix} [\det \mathbf{C}_{n+1}] \det \begin{vmatrix} 1 & & \\ & \epsilon & 0 \\ 0 & & \epsilon^n \end{vmatrix},$$

where

$$\mathbf{C}_{n+1} = \begin{vmatrix} c_0 & c_1 & c_2 & c_n \\ c_1 & c_2 & & \\ c_2 & & & \\ c_n & & & c_{2n} \end{vmatrix}, \tag{84}$$

we obtain the important relation

$$\det \mathbf{B}_{n+1} = \epsilon^{n(n+1)} \det \mathbf{C}_{n+1}. \tag{85}$$

Following the same procedure we obtain a similar relation for the other persymmetric determinant $\det \mathbf{A}_n$.

$$\det \mathbf{A}_n = \epsilon^{n(n-1)} \det \mathbf{C}_n + O(\epsilon^{n(n-1)+1}). \tag{86}$$

Consequently, combining (72), (85), and (86) we show that

$$j_{Bn}[G_t] = (w, Gs) + \epsilon^{2n} \det \mathbf{C}_{n+1}/\det \mathbf{C}_n + O(\epsilon^{2n+1}). \tag{87}$$

This functional has the property that the first, second, third, $\cdots$, $(2n-1)$th variations of the functional all vanish leaving only the $(2n)$th and higher variations.

$$\begin{aligned} &\delta^m j_{Bn} = 0 \quad \text{for} \quad m = 1, 2, \cdots, 2n-1, \\ &\delta^{2n} j_{Bn} = \epsilon^{2n} \det \mathbf{C}_{n+1}/\det \mathbf{C}_n. \end{aligned} \tag{88}$$

The variational functional j_{Bn} given in (72) is obtained by inverting the matrix $\mathbf{A}_n$, premultiplying it by the row vector $\mathbf{b}^+$ and then postmultiplying the result by the column vector $\mathbf{b}$. An explicit expression for j_{Bn} can be found from this formalism. In (A7) of Appendix A, if we let

$$p_{00} = 0 \quad p_{ij} = w_{i+j}, \quad (i, j = 1, 2, \cdots, n) \tag{89}$$

then we obtain

$$j_{Bn}[G_t] = -\begin{vmatrix} 0 & w_1 & w_2 & w_n \\ w_1 & w_2 & & \\ w_2 & & & \\ w_n & & & w_{2n} \end{vmatrix} \div \det \mathbf{A}_n. \tag{90}$$

Thus, for example

$$j_{B1} = -\begin{vmatrix} 0 & w_1 \\ w_1 & w_2 \end{vmatrix} \div w_2 = w_1^2/w_2. \tag{91}$$

$$\begin{aligned} j_{B2} &= -\begin{vmatrix} 0 & w_1 & w_2 \\ w_1 & w_2 & w_3 \\ w_2 & w_3 & w_4 \end{vmatrix} \div \begin{vmatrix} w_2 & w_3 \\ w_3 & w_4 \end{vmatrix} \\ &= (w_1^2 w_4 + w_2^3 - 2w_1 w_2 w_3)/(w_2 w_4 - w_3^2). \end{aligned} \tag{92}$$

Using j_{B2} to estimate the flux at the origin for the infinite homogeneous medium problem with $\beta = 0.1$ treated in Sec. 4, we get $j_{B2} = 0.9999994$. This estimate compares very favorably with the exact answer $(s, Gs) = 1.000000$ and is an improvement over the previous estimate obtained from the variational functional j_{A4} which for condition $dj_{A4}/da = 0$ gave $j_{A4} = 0.999998$.

The second-order variational functional j_{B1} is compared with the fourth-order variational functional j_{B2} in Table II.

The variational functional $j_{Bn}[G_t]$ is independent of the normalization of the trial operator G_t. Consider the trial operator aG_t, where a is a normalization constant. Employing a procedure which is similar to that given in (83) with (90), we see that

$$\begin{aligned} j_{Bn}[aG_t] &= -\begin{vmatrix} 0 & aw_1 & a^n w_n \\ aw_1 & & \\ a^n w_n & & a^{2n} w_n \end{vmatrix} \div a^{n(n+1)} \det A_n \\ &= (a^{n(n+1)}/a^{n(n+1)}) j_{Bn}[G_t] = j_{Bn}[G_t]. \end{aligned} \tag{93}$$

[16] T. Muir, *A Treatise on the Theory of Determinants* (Dover Publications, Inc., 1960), pp. 419ff.

TABLE II. Variational estimates for the weighted average (s, Gs) obtained from second- and fourth-order variational functionals.

β	Second-order variational functional j_{B1}	Fourth-order variational functional j_{B2}
0.1	0.9988	0.9999994
0.3	0.991	0.99996
0.5	0.980	0.9998
0.7	0.966	0.9994
0.9	0.951	0.9987
Exact result $(s, Gs) = 1.000000$		

Thus, j_{Bn} is a normalization-independent variational functional. It is a generalization of the functional (15) which uses trial functions.

In Sec. 3 we saw that if we let the trial operator G_t for the perturbed problem

$$[L_o + L_p]G = G[L_o + L_p] = I \tag{94}$$

be the unperturbed Green's operator G_o,

$$L_oG_o = G_oL_o = I, \tag{95}$$

then the nth-order variational functional j_{An} yielded the first n terms of the perturbation series. A variational correction to the perturbation series can be obtained from j_{Bn}.

Appendix A furnishes us with the relation

$$j_{Bn} = p_{00} - \begin{vmatrix} p_{00} & w_1 & w_2 & w_n \\ w_1 & w_2 & & \\ w_2 & & & \\ w_n & & & \end{vmatrix} \div \det \mathbf{A}_n, \tag{96}$$

where p_{00} is arbitrary and as yet unspecified. By employing the theorem on persymmetric determinants used previously to derive (82), we show that the persymmetric determinant in the numerator of the right-hand side of (96) has the equivalent form

$$\begin{vmatrix} p_{00} & w_1 & w_2 & w_n \\ w_1 & w_2 & & \\ w_2 & & & \\ w_n & & & w_{2n} \end{vmatrix} = \begin{vmatrix} d_0 & d_1 & d_2 & d_n \\ d_1 & d_2 & & \\ d_2 & & & \\ d_n & & & d_{2n} \end{vmatrix}, \tag{97}$$

where

$$\begin{aligned} d_0 &= \sum_{i=0}^{2n-1} (-1)^i \binom{2n}{i} w_{2n-i} + p_{00} \\ d_m &= \sum_{i=0}^{2n-m} \binom{2n-m}{i} w_{2n-i}. \end{aligned} \tag{98}$$

If we now set $G_t = G_o$, then

$$\begin{aligned} w_1 &= (w, G_os) = u_1 \\ w_2 &= (w, G_o[L_o + L_p]G_os) = u_1 + u_2 \\ w_3 &= (w, G_o[(L_o + L_p)G_o]^2s) = u_1 + 2u_2 + u_3 \\ w_n &= (w, G_o[(L_o + L_p)G_o]^{n-1}s) = \sum_{i=0}^{n} \binom{n-1}{i-1} u_i, \end{aligned} \tag{99}$$

where $u_i = (w, G_o[L_pG_o]^{i-1}s)$. Next let

$$p_{00} = -\sum_{i=0}^{2n-1} (-1)^i \binom{2n}{i} w_i \tag{100}$$

so that the first entry in the determinant is zero. Then from Eqs. (96) to (100) we can show that j_{Bn} has the alternative form

$$j_{Bn} = \sum_{i=1}^{2n} (-1)^{i+1} u_i - \begin{vmatrix} 0 & d_1 & d_2 & d_n \\ d_1 & d_2 & & \\ d_2 & & & \\ d_n & & & d_{2n} \end{vmatrix} \div \det \mathbf{A}_n, \tag{101}$$

where

$$d_j = \sum_{i=0}^{j-1} \binom{j-1}{i} u_{2n-i}. \tag{102}$$

The first term on the right of (101) is identical to the $(2n)$th-order perturbation series; the second term gives explicitly the variational correction, e.g., for $n = 1$

$$\begin{aligned} j_{B1} &= u_1 - u_2 - \begin{vmatrix} 0 & u_2 \\ u_2 & u_1 + u_2 \end{vmatrix} \div \det \mathbf{A}_1 \\ &= u_1 - u_2 + u_2^2/(u_1 + u_2) \\ &= (w, G_os) - (w, G_oL_pG_os) + u_2^2/(u_1 + u_2). \end{aligned} \tag{103}$$

In (90) we showed that j_{Bn} could be expressed as the ratio of two persymmetric determinants. Another compact expression which involves j_{Bn} can be derived from the persymmetric determinant formalism by setting $p_{00} = j_{Bn}$ in (96). Thus,

$$\begin{vmatrix} j_{Bn} & w_1 & w_2 & w_n \\ w_1 & w_2 & & \\ w_2 & & & \\ w_n & & & w_{2n} \end{vmatrix} = 0. \tag{104}$$

If (a) the source function $s(x)$ and the weighting function $w(x)$ are the same, (b) L is a self-adjoint positive definite operator, and (c) the trial operator G_t is also chosen to be self-adjoint, then j_{Bn} gives

a lower bound to the exact weighted average (s, Gs). This can be shown by considering the derivation of $j_{Bn}[G_t]$ from (63). Since $L^+ = L$ and $G_t^+ = G_t$, the n-parameter trial operator $\tilde{G}_t$ given in (64) is also self-adjoint. This means that

$$j_{A2}[\tilde{G}_t] = (s[2\tilde{G}_t - \tilde{G}_t L \tilde{G}_t]s) \tag{105}$$

and consequently $j_{Bn}[G_t]$ yields a lower bound to the exact weighted average. The variational estimates in Table II are examples of this result.

Finally, we note that j_{Bn} and j_{A2n} are both $(2n)$th-order variational functionals.

$$j_{Bn} = (w, Gs) + \delta^{2n} j_{Bn} \tag{106}$$

$$j_{A2n} = (w, Gs) + \delta^{2n} j_{A2n}. \tag{107}$$

What is the relation between them?

The functional j_{B2} given in (92) can be derived from the quadratic form

$$j(a_1, a_2) = 2 \sum_{i=1}^{2} a_i w_i - \sum_{i=1}^{2} \sum_{j=1}^{2} a_i a_j w_{i+j} \tag{108}$$

$$= 2a_1 w_1 + (2a_2 - a_1^2) w_2 - 2a_1 a_2 w_3 - a_2^2 w_4$$

by subjecting it to the conditions

$$\partial j(a_1, a_2)/\partial a_1 = 0 \qquad \partial j(a_1, a_2)/\partial a_2 = 0. \tag{109}$$

On comparing (108) with j_{A4},

$$j_{A4} = 4aw_1 - 6a^2 w_2 + 4a^3 w_3 - a^4 w_4, \tag{110}$$

we observe that if we restrict the two independent parameters a_1 and a_2 by the condition

$$a_1 = 2a, \qquad a_2 = -a^2 \tag{111}$$

or equivalently

$$a_1^2 = 4a_2, \tag{112}$$

then (108) becomes reduced to (110). Thus, j_{A4} can be considered as a restricted form of j_{B2}. The variational estimates in Tables I and II illustrate this conclusion.

6. APPLICATION

The variational formalism of Sec. 5 is applied to the one-velocity problem of the isotropic point neutron source in an infinite homogeneous isotropic medium. According to neutron transport theory, the flux at position $\mathbf{r}$ which results from a unit isotropic point source at $\mathbf{r}_s$ is given by the integral equation

$$g(\mathbf{r}, \mathbf{r})_s = \frac{c}{4\pi l} \iiint k(\mathbf{r}, \mathbf{r}') g(\mathbf{r}', \mathbf{r}_s)\, dV' + k(\mathbf{r}, \mathbf{r}_s), \tag{113}$$

where $c = l/l_s$, l_s and l are the scattering mean free path and total mean free path, respectively, and

$$k(\mathbf{r}, \mathbf{r}') = (1/|\mathbf{r} - \mathbf{r}'|^2) \exp(-|\mathbf{r} - \mathbf{r}'|/l). \tag{114}$$

Equation (113) can be rewritten in the more convenient operator notation

$$G = (c/4\pi l) KG + K. \tag{115}$$

By rearranging terms, we can formally place (115) in the standard variational form $LG = I$, where $L = K^{-1} - (c/4\pi l) I$. Let us now apply the variational method to the moment problem.

In terms of the Green's function, the pth spatial moment is defined by the integral

$$m_p = \iiint |\mathbf{r} - \mathbf{r}_s|^p\, g(\mathbf{r}, \mathbf{r}_s)\, dV, \tag{116}$$

where the integration extends over all of space. Since the medium is infinite and homogeneous, the moments do not depend on the position of the source, and it is convenient to locate the source at the origin, i.e., set $\mathbf{r}_s = 0$.

$$m_p = \iiint |\mathbf{r}|^p\, g(\mathbf{r}, 0)\, dV = (w, Gs). \tag{117}$$

Here s denotes the point source at the origin and w is the weighting function for the pth moment.

Using the trial Green's function

$$g_t(\mathbf{r}, \mathbf{r}') = \frac{\exp(-|\mathbf{r} - \mathbf{r}'|/\lambda)}{|\mathbf{r} - \mathbf{r}'|}, \tag{118}$$

we estimate the spatial moments with the variational functional j_{Bn}. Since j_{Bn} is a normalization-independent functional, we can omit the normalization constant of (118). The parameter λ determines the rate at which the trial flux decays with distance from the source.

We start by estimating the zeroth spatial moment m_0 with the $n = 1$ variational functional

$$j_{B1} = w_1 (w_2)^{-1} w_1. \tag{119}$$

On substituting

$$w_1 = (w, G_t s) = 4\pi\lambda^2 \tag{120}$$

and

$$w_2 = (w, G_t L G_t s) = 4\pi(1 - c)\lambda^4/l \tag{121}$$

in (119), we secure the variational estimate

$$j_{B1} = 4\pi l/(1 - c) \tag{122}$$

which is identical to the exact zeroth spatial moment

$$m_0 = 4\pi l/(1 - c) \tag{123}$$

given by Case, deHoffmann and Placzek.[17] In this case, the approximate variational method produces the exact answer.

We next estimate the second spatial moment m_2 again employing j_{B1}. Using (118) to calculate w_1 and w_2

$$w_1 = (w, G_t s) = 24\pi\lambda^4 \tag{124}$$

$$w_2 = (w, G_t L G_t s) = w_1 \tfrac{1}{3} l[6(1 - c)(\lambda/l)^2 - 1] \tag{125}$$

we obtain the variational estimate

$$j_{B1} = \frac{72\pi\lambda^4/l}{6(1 - c)(\lambda/l)^2 - 1}. \tag{126}$$

The result (126) depends on the parameter λ. Let us set $\lambda = \lambda_d$, the diffusion length of transport theory which is given implicitly by

$$\frac{1}{c} = \frac{\lambda_d}{2l} \ln \frac{\lambda_d + l}{\lambda_d - l} \tag{127}$$

and compare the result with the exact moment $m_2 = 8\pi l^3/(1 - c)^2$ when the absorption mean free path, $l_a \gg l_s$ (i.e., $\Sigma_a \ll \Sigma_s$, weak absorption case). Here, $(1 - c) = l/l_a = \Sigma_a/\Sigma \ll 1$, and the diffusion length has the expansion

$$(\lambda_d/l)^2 = [1/3(1 - c)][1 + \tfrac{4}{5}(1 - c) + \tfrac{108}{175}(1 - c)^2 + \cdots]. \tag{128}$$

By combining (128) with (126), we find that the $(1 - c)$ term cancels itself so that the exact moment and the leading $(1 - c)^2$ correction term remain.

$$j_{B1} = [8\pi l^3/(1 - c)^2][1 + \tfrac{16}{25}(1 - c)^2 + \cdots] = m_2[1 + O(1 - c)^2]. \tag{129}$$

When scattering dominates over absorption, the variational answer j_{B1} rapidly approaches m_2.

Now let us consider the variational functional j_{B2}, Eq. (92), and from it obtain another estimate of the second spatial moment. We calculate the two additional terms

$$w_3 = (w, G_t[LG_t]^2 s) = w_1(3h^2 - 2h)(3/l)^2$$
$$w_4 = (w, G_t[LG_t]^2 s) = w_1(4h^3 - 3h^2)(3/l)^3, \tag{130}$$

where $h = 3(1 - c)(\lambda/l)^2$. After inserting (124), (125), and (130) into the expression for j_{B2}, we have the variational estimate

$$j_{B2} = 8\pi l^3/(1 - c)^2 \tag{131}$$

which is, in fact, equal to the exact result.

APPENDIX A

Consider the $n + 1$ by $n + 1$ matrix

$$\mathbf{P} = \left|\begin{array}{c|c} p_{00} & \mathbf{q}^+ \\ \hline \mathbf{p} & \mathbf{Q} \end{array}\right|, \tag{A1}$$

where $\mathbf{Q}$ is a nonsingular n-by-n matrix with the elements p_{ij}, $i, j, = 1, 2, \cdots, n$; $\mathbf{p}$ is a column matrix with the elements p_{0j}, $j = 1, 2, \cdots, n$; $\mathbf{q}^+$ is a row matrix with the elements p_{i0}, $i = 1, 2, \cdots, n$. Let $\mathbf{R}$ be the $n + 1$ by $n + 1$ matrix

$$\mathbf{R} = \left|\begin{array}{c|c} 1 & \mathbf{q}^+ \\ \hline -\mathbf{Q}^{-1}\mathbf{p} & \mathbf{Q}^{-1} \end{array}\right|, \tag{A2}$$

where $\mathbf{Q}^{-1}$ is the inverse of $\mathbf{Q}$; $\mathbf{Q}\mathbf{Q}^{-1} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{I}$. Forming the matrix product $\mathbf{PR}$ we get the diagonal matrix

$$\mathbf{PR} = \left|\begin{array}{c|c} PR_{00} & 0 \\ \hline 0 & \mathbf{I} \end{array}\right|, \tag{A3}$$

where $PR_{00} = p_{00} - \mathbf{q}^+ \cdot \mathbf{Q}^{-1} \cdot \mathbf{q}$.

Next we note that

$$\det(\mathbf{Q}\mathbf{Q}^{-1}) = (\det \mathbf{Q})(\det \mathbf{Q}^{-1}) = \det \mathbf{I} = 1,$$

so that

$$\det \mathbf{Q}^{-1} = 1/\det \mathbf{Q}. \tag{A4}$$

From (A3) we get

$$\det(PR) = p_{00} - \mathbf{q}^+ \cdot \mathbf{Q}^{-1} \cdot \mathbf{p} \tag{A5}$$

and from (A2) and (A4)

$$\det(\mathbf{PR}) = (\det \mathbf{P})(\det \mathbf{R}) = (\det \mathbf{P})(\det \mathbf{Q}^{-1}) = (\det \mathbf{P})/\det \mathbf{A}. \tag{A6}$$

Combining (A5) with (A6) we finally have

$$\det \mathbf{P}/\det \mathbf{Q} = p_{00} - \mathbf{q}^+ \cdot \mathbf{Q}^{-1} \cdot \mathbf{p}. \tag{A7}$$

If we set the elements of $\mathbf{P}$ equal to those of $\mathbf{B}_{n+1}$, Eq. (75),

$$p_{ij} = w_{i+j}, \tag{A8}$$

then from (A7) we get the desired result

$$\det \mathbf{B}_{n+1}/\det \mathbf{A}_n = w_0 - \mathbf{b}^+ \cdot \mathbf{A}_n^{-1} \cdot \mathbf{b}. \tag{A9}$$

[17] K. M. Case, F. deHoffmann, and E. Placzek, *Introduction to the Theory of Neutron Diffusion* (Los Alamos Scientific Laboratory, Los Alamos, New Mexico, 1953), Vol. I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Useful Operator in Plasma Kinetic Theory*

CHING-SHENG WU
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received 30 June 1964)

An operator which facilitates the derivation of the plasma kinetic equation is introduced and discussed. The operator determines the first integral of the pair correlation function without the necessity for knowledge of the pair correlation function itself. To break through the lengthy algebraic effort which is usually encountered in solving the truncated BBGKY hierarchy equations, the present operator method is found to be far superior to the singular integral equation technique. The mathematical simplification which can be gained from the use of this operator is demostrated by several examples.

I. INTRODUCTION

THE study of the pair correlation functions from the truncated BBGKY hierarchy[1,2] is currently of considerable interest. This interest has been spurred primarily by the formulation and investigation of the plasma kinetic theory. The unfortunate situation is that, even though we require the theory to be accurate only to first order in the plasma parameter $\epsilon = 1/(\lambda_D^3 n)$(where λ_D is the Debye length and n is particle density), the governing mathematical equations are woefully complicated. Such mathematical complexity has appeared to be one serious drawback to research progress in this particular area of plasma physics.

In the past few years, several authors have applied the singular integral equation technique[3] to obtain the solution of the pair correlation functions for various problems.[4-6] The utilization of such a mathematical method has shown a certain amount of promising success. However, this does not necessarily mean that the algebra involved in obtaining the solution is simple and, in fact, it usually is complicated. Typical detailed discussions may be found in Refs. 5 and 6.

An elegant alternative approach has been suggested recently by Dupree.[7] His proposed scheme is based mainly on the fact that, in many cases, the operator of the pair correlation function equation can be related to the operator to the linearized Vlasov equation if we accept the Bogoliubov adiabatic approximation. Further discussions along the same line have been given by a number of authors.[8-10] Wolff[8] first clarified some of the details involved in such an operational method. Later, Rutherford and Erieman[9] applied a similar method to give a brief discussion on the unstable case.

More recently Rostoker[10] has applied the same method to a study of the test particle problem. Furthermore, Ron[11] has extended the application of such a method to the derivation of the kinetic equations for electrons and phonons in a metal. This literature indicates a growth of interest in this new method. In practice, however, one may find that the complete process of obtaining the solution may still be lengthy and involved, although the basic notion of the method is simple. The algebra involved in such a task often becomes more complicated than one would expect. The discussion given by Wolff,[8] who treated the homogeneous case in great detail, is a good illustration. From such discussion, one can easily imagine how complicated the process could be for a more complex case: for instance, the problem previously studied by Guernsey using the singular integral equation technique.[5] In view of this situation, further improvement of the method seems desirable.

One natural consideration is that, either in the derivation of the kinetic equation or in many other problems, the most useful quantity is not g_{rr} itself,

* The research discussed in this paper was supported by NASA under Contract No. NAS 7-100.

[1] N. H. Bogoliubov, *Problems of a Dynamic Theory in Statistical Physics*, Moscow (1946), translated by E. K. Gora, AFCRC-TR-59-235; or *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962), Vol. 1.

[2] N. Rostoker and M. N. Rosenbluth, Phys. Fluids **3**, 1 (1960).

[3] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

[4] R. L. Guernsey, dissertation, University of Michigan (1960) (unpublished).

[5] R. L. Guernsey, Phys. Fluids **5**, 322 (1962).

[6] C. Oberman, A. Ron, and J. Dawson, Phys. Fluids **5**, 1514 (1962).

[7] T. Dupree, Phys. Fluids **4**, 696 (1961).

[8] P. A. Wolff, Phys. Fluids **5**, 316 (1962).

[9] P. H. Rutherford and E. A. Frieman, Phys. Fluids **6**, 1139 (1963).

[10] N. Rostoker, General Atomic Reports No. GA-4555 and No. GA-4707 (1963) (unpublished).

[11] A. Ron, J. Math. Phys. **4**, 1182 (1963).

but $\sum_r n_r e_r \int d^3v_2 g_{sr}(\mathbf{v}_1, \mathbf{v}_2)$, where g_{sr} is the pair correlation function for particles of the sth and rth species, and n_r and e_r are the number density and charge, respectively, for the rth species particle. However, the procedure adopted in most work is, first, to determine g_{sr}, and then to compute $\sum_r n_r e_r \int d^3v_2 g_{sr}$. In such a case, we might have spent more labor than necessary. Thus, it would be interesting if we could devise a method which determines the quantity $\sum_r n_r e_r \int d^3v_2 g_{sr}$ directly without the knowledge of g_{sr}. In Sec. II, it is shown that this thought indeed leads us to a scheme which provides a great deal of mathematical simplification. As a consequence of the ensuing discussion, we derive a useful operator which can be applied easily to a variety of problems, so that the solution for $\sum_r n_r e_r \int d^3v_2 g_{sr}$ becomes obtainable in a much simpler manner. In Sec. III, we furnish a number of illustrative examples: (1) the computation of high-frequency conductivity, (2) the derivation of the BLGRR (Balescu,[12] Lenard,[13] Guernsey,[4] Rostoker, and Rosenbluth[2]) collision integral, (3) Guernsey's problem[5] for an inhomogeneous plasma, and (4) a study of the kinetic equation with unstable correlation. These examples make evident the usefulness of the new operator, which enables us to save a tremendous amount of mathematical effort.

II. STATEMENT OF THE METHOD

A. The Governing Mathematical Equations

In order to discuss the method of solution, we first list the mathematical equations under consideration. We consider a fully ionized plasma with Coulomb interactions only. The BBGKY hierarchy used to describe such a system can be truncated by an expansion scheme[2] with a small expansion parameter $\epsilon = 1/\lambda_D^3 n$, which represents the inverse of the number of particles in a "Debye cube." To first order in this parameter, the first two members of the hierarchy[1,2] can be written as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_1 + \frac{e_s}{m_s}\mathbf{E}_1\cdot\nabla_{\mathbf{v}_1}\right)\mathcal{F}_s(1, t) = \frac{1}{m_s}\sum_r n_r \int \frac{\partial\phi_{sr}(1, 2)}{\partial\mathbf{r}_1}\cdot\frac{\partial}{\partial\mathbf{v}_1}\mathcal{G}_{sr}(1, 2, t)\, d^3r_2\, d^3v_2, \quad (1)$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_1 + \mathbf{v}_2\cdot\nabla_2 + \frac{e_s}{m_s}\mathbf{E}_1\cdot\nabla_{\mathbf{v}_1} + \frac{e_r}{m_r}\mathbf{E}_2\cdot\nabla_{\mathbf{v}_2}\right) \times \mathcal{G}_{sr}(1, 2, t) - \frac{1}{m_s}\frac{\partial\mathcal{F}_s(1, t)}{\partial\mathbf{v}_1} \cdot \sum_q n_q \int \frac{\partial\phi_{sq}(1, 3)}{\partial\mathbf{r}_1}\mathcal{G}_{rq}(2, 3, t)\, d^3r_3\, d^3v_3 - \frac{1}{m_r}\frac{\partial\mathcal{F}_r(2, t)}{\partial\mathbf{v}_2} \times \sum_q n_q \int \frac{\partial\phi_{rq}(2, 3)}{\partial\mathbf{r}_2}\mathcal{G}_{sq}(1, 3, t)\, d^3r_3\, d^3v_3 = \frac{\partial\phi_{sr}}{\partial\mathbf{r}_1}\left(\frac{\mathcal{F}_r(2, t)}{m_s}\frac{\partial\mathcal{F}_s(1, t)}{\partial\mathbf{v}_1} - \frac{\mathcal{F}_s(1, t)}{m_r}\frac{\partial\mathcal{F}_r(2, t)}{\partial\mathbf{v}_2}\right), \quad (2)$$

and

$$e_s\mathbf{E}_1 = -\sum_r n_r \int \frac{\partial\phi_{sr}(1, 2)}{\partial\mathbf{r}_1}\mathcal{F}_r(2, t)\, d^3r_2\, d^3v_2. \quad (3)$$

Here, $\mathcal{F}(1, t)$ is the one-particle distribution function; $\mathcal{G}(1, 2, t)$ is the pair correlation function; the subscripts s, r, $\cdots$ designate the species of particles; e_s and m_s are the charge and mass of the s type of particles; 1,2, $\cdots$ denote the state variable in phase space [for instance, $(\mathbf{r}_1, \mathbf{v}_1)$, $(\mathbf{r}_2, \mathbf{v}_2)$, $\cdots$] ; $\phi_{sr}(1, 2) = (e_s e_r)/|\mathbf{r}_1 - \mathbf{r}_2|$ is the interparticle Coulomb potential; $\nabla_1 \equiv \partial/\partial\mathbf{r}_1$; and $\nabla_{\mathbf{v}_1} \equiv \partial/\partial\mathbf{v}_1$. For a variety of physical problems, the following model may be justified:

(1) The plasma is supposed to be the sum of two parts: the main body and the small perturbations.

(2) The main body is homogeneous, but not necessarily in an equilibrium state, and the small perturbations give rise to spatial inhomogeneity.

(3) The physical quantities associated with the main body are assumed to vary with a time scale which is long compared with the electron plasma period $1/(\omega_{ps})$; however, those associated with the perturbation part may vary rapidly with a much shorter time scale since, in many cases, high-frequency wave phenomena are involved.

With these considerations, let us split $\mathcal{F}_s(1, t)$ and $\mathcal{G}_{sr}(1, 2, t)$ into two parts:

$$\mathcal{F}_s(1, t) = F_s(\mathbf{v}_1, t) + f_s(\mathbf{r}_1, \mathbf{v}_1, t),$$

$$\mathcal{G}_{sr}(1, 2, t) = G_{sr}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{r}_1 - \mathbf{r}_2, t) + g_{sr}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{r}_1, \mathbf{r}_2, t),$$

where F_s and G_{sr} designate, respectively, the distribution function and correlation function of the main body, and f_s and g_{sr} apply in a similar manner to the perturbed part. We assume, furthermore,

$$F_s \gg f_s,$$

$$G_{sr} \gg g_{sr}.$$

[12] R. Balescu, Phys. Fluids **3**, 52 (1960).
[13] A. Lenard, Ann. Phys. (N. Y.) **10**, 390 (1960).

Thus, Eqs. (1) and (2) may be linearized, and we obtain two sets of equations. The first set describes the behavior of the main body and takes the form

$$\frac{\partial F_s}{\partial t} = \frac{1}{m_s}\sum_r n_r \int \frac{\partial \phi_{sr}(1,2)}{\partial \mathbf{r}_1}\cdot\frac{\partial}{\partial \mathbf{v}_1} \times G_{sr}(1,2,t)\, d^3r_2\, d^3v_2, \tag{4}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_1 + \mathbf{v}_2\cdot\nabla_2\right)G_{sr}(1,2,t) - \frac{1}{m_s}\frac{\partial F_s(1)}{\partial \mathbf{v}_1}\cdot\sum_q n_q \int \frac{\partial \phi_{sq}(1,3)}{\partial \mathbf{r}_1} G_{rq}(2,3)\, d^3r_3\, d^3v_3 - \frac{1}{m_r}\frac{\partial F_r(2)}{\partial \mathbf{v}_2}\cdot\sum_q n_q \int \frac{\partial \phi_{rq}(2,3)}{\partial \mathbf{r}_2} G_{sq}(1,3)\, d^3r_3\, d^3v_3 = \frac{\partial \phi_{sr}}{\partial \mathbf{r}_1}\left(\frac{F_r(2)}{m_s}\frac{\partial F_s(1)}{\partial \mathbf{v}_1} - \frac{F_s(1)}{m_r}\frac{\partial F_r(2)}{\partial \mathbf{v}_2}\right). \tag{5}$$

The second set, which describes the small perturbations due to the wave phenomena and high-frequency processes, can be written as

$$\frac{\partial f_s}{\partial t} + \mathbf{v}_1\cdot\nabla f_s + \frac{e_s}{m_s}\mathbf{E}_1\cdot\frac{\partial F_s}{\partial \mathbf{v}_1} = \frac{1}{m_s}\sum_r n_r \int \frac{\partial \phi_{sr}(1,2)}{\partial \mathbf{r}_1}\cdot\frac{\partial}{\partial \mathbf{v}_1} g_{sr}(1,2,t)\, d^3r_2\, d^3v_2, \tag{6}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_1 + \mathbf{v}_2\cdot\nabla_2\right)g_{sr} - \frac{1}{m_s}\frac{\partial F_s(1)}{\partial \mathbf{v}_1}\cdot\sum_q n_q \int \frac{\partial \phi_{sq}(1,3)}{\partial \mathbf{r}_1} g_{rq}(2,3)\, d^3r_3\, d^3v_3 - \frac{1}{m_r}\frac{\partial F_r(2)}{\partial \mathbf{v}_2}\cdot\sum_q n_q \int \frac{\partial \phi_{rq}(2,3)}{\partial \mathbf{r}_2} g_{sq}(1,3)\, d^3r_3\, d^3v_3 = \frac{1}{m_s}\frac{\partial f_s(1)}{\partial \mathbf{v}_1}\cdot\sum_q n_q \int \frac{\partial \phi_{sq}(1,3)}{\partial \mathbf{r}_1} G_{rq}(2,3)\, d^3r_3\, d^3v_3 + \frac{1}{m_r}\frac{\partial f_r(2)}{\partial \mathbf{v}_2}\cdot\sum_q n_q \int \frac{\partial \phi_{rq}(2,3)}{\partial \mathbf{r}_2} G_{sq}(1,3)\, d^3r_3\, d^3v_3 + \frac{\partial \phi_{sr}}{\partial \mathbf{r}_1}\cdot\left(\frac{1}{m_s}\frac{\partial}{\partial \mathbf{v}_1} - \frac{1}{m_r}\frac{\partial}{\partial \mathbf{v}_2}\right)[f_s(1)F_r(2) + F_s(1)f_r(2)] + \frac{e_s}{m_s}\mathbf{E}_1\cdot\frac{\partial G_{sr}}{\partial \mathbf{v}_1} + \frac{e_r}{m_r}\mathbf{E}_2\cdot\frac{\partial G_{sr}}{\partial \mathbf{v}_2}. \tag{7}$$

Let us now introduce the following Fourier transforms:

$$G_{sr}(\mathbf{k}_1, \mathbf{v}_1, \mathbf{v}_2, t) = \int d^3r_1 e^{-i\mathbf{k}_1\cdot(\mathbf{r}_1-\mathbf{r}_2)} G_{sr}(\mathbf{r}_1 - \mathbf{r}_2, \mathbf{v}_1, \mathbf{v}_2, t), \tag{8}$$

$$f_s(\mathbf{k}_1, \mathbf{v}_1, t) = \int d^3r_1 e^{-i\mathbf{k}_1\cdot\mathbf{r}_1} f_s(\mathbf{r}_1, \mathbf{v}_1, t), \tag{9}$$

$$g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t) = \iint d^3r_1\, d^3r_2 e^{-i\mathbf{k}_1\cdot\mathbf{r}_1 - i\mathbf{k}_2\cdot\mathbf{r}_2} g_{sr}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{v}_1, \mathbf{v}_2, t). \tag{10}$$

Then Eqs. (4), (5), (6), and (7) in Fourier transform space will have the following forms:

$$\frac{\partial F_s}{\partial t} = -\frac{\partial}{\partial \mathbf{v}_1}\frac{i}{(2\pi)^3} \times \int d^3k \frac{4\pi \mathbf{k} e_s}{m_s k^2}\sum_r n_r e_r \int d^3v_2 G_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2, t), \tag{11}$$

$$\left(\frac{\partial}{\partial t} + i\mathbf{k}\cdot\mathbf{v}_1 - i\mathbf{k}\cdot\mathbf{v}_2 - \frac{4\pi e_s i}{m_s k^2}\mathbf{k}\cdot\frac{\partial F_s(1)}{\partial \mathbf{v}_1}\sum_s n_s e_s \int d^3v_1 + \frac{4\pi e_r i}{m_r k^2}\mathbf{k}\cdot\frac{\partial F_r(2)}{\partial \mathbf{v}_2}\sum_r n_r e_r \int d^3v_2\right)G_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2, t) = \frac{4\pi e_s e_r}{k^2} i\mathbf{k}\cdot\left[\frac{F_r(2)}{m_s}\frac{\partial F_s(1)}{\partial \mathbf{v}_1} - \frac{F_s(1)}{m_r}\frac{\partial F_r(2)}{\partial \mathbf{v}_2}\right], \tag{12}$$

$$\left(\frac{\partial}{\partial t} + i\mathbf{k}_1\cdot\mathbf{v}_1 - \frac{4\pi e_s i}{m_s k_1^2}\mathbf{k}_1\cdot\frac{\partial F_s(1)}{\partial \mathbf{v}_1} \times \sum_s n_s e_s \int d^3v_1\right)f_s(\mathbf{k}_1, \mathbf{v}, t) = \frac{\partial}{\partial \mathbf{v}_1}\frac{i}{(2\pi)^3}\int d^3k_2 \frac{4\pi \mathbf{k}_2 e_s}{m_s k_2^2}\sum_r n_r e_r \times \int d^3v_2 g_{sr}(\mathbf{k}_1 - \mathbf{k}_2, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t), \tag{13}$$

$$\left(\frac{\partial}{\partial t} + i\mathbf{k}_1\cdot\mathbf{v}_1 + i\mathbf{k}_2\cdot\mathbf{v}_2 - \frac{4\pi e_s}{m_s k_1^2} i\mathbf{k}_1\cdot\frac{\partial F_s}{\partial \mathbf{v}_1}\sum_s n_s e_s \int d^3v_1 - \frac{4\pi e_r}{m_r k_2^2} i\mathbf{k}_2\cdot\frac{\partial F_r}{\partial \mathbf{v}_2}\sum_r n_r e_r \int d^3v_2\right) \times g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t) = S_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t), \tag{14}$$

$$S_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t) \equiv -\frac{4\pi i e_s e_r}{k_2^2}\mathbf{k}_2\cdot\left[F_r\frac{1}{m_s}\frac{\partial f_s(\mathbf{k}_1+\mathbf{k}_2, \mathbf{v}_1, t)}{\partial \mathbf{v}_1} - \frac{f_s(\mathbf{k}_1+\mathbf{k}_2, \mathbf{v}_1, t)}{m_r}\frac{\partial F_r}{\partial \mathbf{v}_2}\right] + \frac{4\pi i e_s e_r}{k_1^2}\mathbf{k}_1\cdot\left[\frac{f_r(\mathbf{k}_1+\mathbf{k}_2, \mathbf{v}_2, t)}{m_s}\frac{\partial F_s(2)}{\partial \mathbf{v}_1} - \frac{F_s(1)}{m_r}\frac{\partial f_r(\mathbf{k}_1+\mathbf{k}_2, \mathbf{v}_2, t)}{\partial \mathbf{v}_2}\right] + \frac{4\pi\sum n_q e_q i}{|\mathbf{k}_1+\mathbf{k}_2|^2}(\mathbf{k}_1+\mathbf{k}_2)\rho_q(\mathbf{k}_1+\mathbf{k}_2, t) \cdot\left[\frac{e_s}{m_s}\frac{\partial G_{sr}(\mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2)}{\partial \mathbf{v}_1} + \frac{e_r}{m_r}\frac{\partial G_{sr}(\mathbf{k}_1, \mathbf{v}_1, \mathbf{v}_2)}{\partial \mathbf{v}_2}\right] - \frac{1}{m_s}\frac{\partial f_s(\mathbf{k}_1+\mathbf{k}_2, \mathbf{v}_1)}{\partial \mathbf{v}_1} 4\pi e_s \cdot\sum_q n_q e_q \frac{i\mathbf{k}_2}{k_2^2}\int d^3v_3 G_{rq}(\mathbf{k}_2, \mathbf{v}_2, \mathbf{v}_3) - \frac{1}{m_r}\frac{\partial f_r(\mathbf{k}_1+\mathbf{k}_2, \mathbf{v}_2)}{\partial \mathbf{v}_2} 4\pi e_r \cdot\sum_q n_q e_q \frac{i\mathbf{k}_1}{k_1^2}\int d^3v_3 G_{sq}(\mathbf{k}_1, \mathbf{v}_1, \mathbf{v}_3), \tag{15}$$

where

$$\rho_q = \int d^3v f_q(\mathbf{k}_1 + \mathbf{k}_2, \mathbf{v}, t).$$

Since the time scale of $F(\mathbf{v}, t)$ is long compared to $1/(\omega_{pe})$, we may use the adiabatic approximation in the discussion of G_{sr}. However, since $f_s(\mathbf{r}, \mathbf{v}, t)$ and $g_{sr}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{v}_1, \mathbf{v}_2, t)$ are supposed to be highly time-dependent and to have characteristic times of the order of $1/(\omega_{pe})$, which is close to the relaxation time scale of the pair correlation function, the adiabatic approximation cannot be justified. Therefore, in the solution for $g_{sr}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{v}_1, \mathbf{v}_2, t)$, $f_s(\mathbf{r}, \mathbf{v}, t)$ should be treated as time dependent. Finally, it should be remarked that, although these equations are typical in many problems, yet they may not be valid when nonlinear interactions are important. In the subsequent discussions, however, we shall restrict ourselves only to the cases for which Eqs. (11) to (14) are valid.

B. The Method of Solution and Derivation of the Operator

Observing Eqs. (12) and (14), we conclude that, for a variety of problems, we may write the governing equations for the pair correlation function in a typical form: i.e.,

$$\left[\frac{\partial}{\partial t} + H_s(\mathbf{k}_1, \mathbf{v}_1) + H_r(\mathbf{k}_2, \mathbf{v}_2)\right] g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t) = R_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t), \tag{16}$$

where

$$H_l(\mathbf{k}, \mathbf{v}) = i\mathbf{k}\cdot\mathbf{v} - \frac{4\pi e_l}{m_l k^2} i\mathbf{k}\cdot\frac{\partial F_l(\mathbf{v})}{\partial \mathbf{v}} \int d^3v \sum_l n_l e_l.$$

The operator $H_l(\mathbf{k}, \mathbf{v})(l = s, r, \cdots)$ is a linear operator, and R_{sr} is an inhomogeneous forcing term whose explicit form depends upon the problem under study. For example, in Section III, we shall illustrate a simplified calculation of high-frequency conductivity. In that problem, we assume that the plasma is homogeneous, and that the equation can be linearized. The governing equation for the pair correlation function again has the form of Eq. (16), with the inhomogeneous term R_{sr} modified slightly from the right-hand-side terms given in Eqs. (12) and (14). Furthermore, the operator in Eq. (12) is merely a special case ($\mathbf{k}_2 = -\mathbf{k}_1$) of that in Eq. (14); therefore, as far as the discussion of the method is concerned, Eq. (16) is sufficiently general.

Now, according to Eq. (16), we may formally write the solution for g_{sr} as

$$g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t) = e^{-(H_s+H_r)t} g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, 0) + \int_0^t d\tau e^{-(H_s+H_r)\tau} R_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, t - \tau). \tag{17}$$

Here, we have made use of the Bogoliubov adiabatic approximation, and H_s and H_r are thus considered independent of time. Yet the operator $e^{-(H_s+H_r)t}$ so far remains to be determined. Since the operators $H_s(\mathbf{k}_1, \mathbf{v}_1)$ and $H_r(\mathbf{k}_2, \mathbf{v}_2)$ commute, and since the operators $e^{-H_s(\mathbf{k}_1,\mathbf{v}_1)t}$ and $e^{-H_r(\mathbf{k}_2,\mathbf{v}_2)t}$ (as first pointed out by Dupree[7]) can be identified from the solutions to the linearized Vlasov equation, the operator $e^{-(H_s+H_r)t}$ is thus well defined. However, we are not interested in doing this. In the following discussion, we propose a simplified scheme.

To facilitate our treatment, it is found convenient to re-express the operator $e^{-(H_s+H_r)t}$ in a form like that of a propagator, $P_{sr}(\mathbf{v}_1, \mathbf{v}_2|\mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2)$, so that

$$g_{sr} = P_{sr}(\mathbf{v}_1, \mathbf{v}_2 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2) \times g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1', \mathbf{v}_2', t = 0) + \int_0^t d\tau \times P_{sr}(\mathbf{v}_1, \mathbf{v}_2 \mid \mathbf{v}_1', \mathbf{v}_2'; \tau, \mathbf{k}_1, \mathbf{k}_2) \times R_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1', \mathbf{v}_2'; t - \tau), \tag{18}$$

where the operator $P_{sr}(\mathbf{v}_1, \mathbf{v}_2|\mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2)$ can again be written as the product of two operators: i.e.,

$$P_{sr}(\mathbf{v}_1, \mathbf{v}_2 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2) = P_s(\mathbf{v}_1 \mid \mathbf{v}_1'; t, \mathbf{k}_1) P_r(\mathbf{v}_2 \mid \mathbf{v}_2'; t, \mathbf{k}_2). \tag{19}$$

Now, we shall bypass the solution for g_{sr} and make an attempt to compute the quantity $\sum_r n_r e_r \int d^3v_2 g_{sr}$ directly. To do this, we multiply Eq. (18) by $\sum_r n_r e_r$ and integrate with respect to $\mathbf{v}_2$. Thus, we obtain

$$\sum_r n_r e_r \int d^3v_2\, g_{sr} = Q_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2) g_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1', \mathbf{v}_2', t = 0) + \int_0^t d\tau\, Q_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2) R_{sr}(t - \tau, \mathbf{v}_1', \mathbf{v}_2'), \tag{20}$$

where the new operator Q_{sr} is defined by

$$Q_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2) \equiv P_s(\mathbf{v}_1 \mid \mathbf{v}_1'; t, \mathbf{k}_1) \times \sum_r n_r e_r \int d^3v_2\, P_r(\mathbf{v}_2 \mid \mathbf{v}_2'; t, \mathbf{k}_2). \tag{21}$$

As yet, this operator remains to be determined.

Now, let us look at the linearized Vlasov equa-

tion, which takes the form

$$\left(\frac{\partial}{\partial t} + H_s(\mathbf{k}_1, \mathbf{v}_1)\right) f_s(\mathbf{k}_1, \mathbf{v}_1, t) = 0. \tag{22}$$

Since the solution for $f_s(\mathbf{k}_1, \mathbf{v}_1, t)$ can be formally written as

$$f_s(\mathbf{k}_1, \mathbf{v}_1, t) = P_s(\mathbf{v}_1 \mid \mathbf{v}_1'; t, \mathbf{k}_1) f_s(\mathbf{k}_1, \mathbf{v}_1', 0), \tag{23}$$

where $f_s(\mathbf{k}_1, \mathbf{v}_1, 0)$ is the initial condition, it is evident that the operator P_s can be identified from the solutions obtained by Landau[14] or van Kampen[15]; i.e.,[16]

$$P_s(\mathbf{v}_1 \mid \mathbf{v}_1'; t, \mathbf{k}_1) = \frac{1}{2\pi i} \int_{-\infty - i\gamma}^{+\infty - i\gamma} d\tilde{u}_1 \frac{e^{ik_1\tilde{u}_1 t}}{u_1 + \tilde{u}_1} \times \left\{ \int d^3v_1' \, \delta(\mathbf{v}_1 - \mathbf{v}_1') - \frac{D_s(\mathbf{v}_1)}{\epsilon(\tilde{u}_1)} \int \frac{d^3v_1' \sum_s n_s e_s}{u_1' + \tilde{u}_1} \right\}, \tag{24}$$

where

$$D_s(\mathbf{v}_1, \mathbf{k}_1) = -\frac{4\pi e_s \mathbf{k}_1}{m_s k_1^3} \cdot \frac{\partial F_s}{\partial \mathbf{v}_1}, \quad u_1 = \frac{\mathbf{k}_1 \cdot \mathbf{v}_1}{k_1}, \quad \tilde{u}_1 = \frac{\omega_1}{k_1}$$

and[17]

$$\epsilon(\tilde{u}_1, \mathbf{k}_1) = 1 + \sum_s n_s e_s \int d^3v_1 \frac{D_s(\mathbf{v}_1)}{u_1 + \tilde{u}_1}, \quad \tilde{u} \in S_-.$$

It is understood from the usual inversion integral in Laplace transform theory that the path of integration in Eq. (24) is parallel to the real axis and below all the singularities of the integrand in the complex $\tilde{u}_1$ plane. Now, it is instructive to determine the operator $\sum_s n_s e_s \int d^3v_1 P_s(\mathbf{v}_1 | \mathbf{v}_1'; t, \mathbf{k}_1)$. We multiply Eq. (23) by $\sum_s n_s e_s$ and integrate with respect to $\mathbf{v}_1$; i.e.,

$$\sum_s n_s e_s \int d^3v_1 f_s(\mathbf{k}_1, \mathbf{v}_1, t) = \sum_s n_s e_s \int d^3v_1 P_s(\mathbf{v}_1 \mid \mathbf{v}_1'; t, \mathbf{k}_1) f_s(\mathbf{k}_1, \mathbf{v}_1', 0). \tag{25}$$

But, by definition,

$$\sum_s n_s e_s \int d^3v_1 f_s = \rho,$$

where ρ denotes the net perturbed charge density and, from the Landau–Vlasov theory, we know[18]

$$\rho = \frac{1}{2\pi i} \int_{-\infty - i\gamma}^{+\infty - i\gamma} d\tilde{u}_1 \frac{e^{+ik_1\tilde{u}_1 t}}{\epsilon(\tilde{u}_1, \mathbf{k}_1)} \times \int d^3v_1' \frac{\sum_s n_s e_s f_s(\mathbf{k}_1, \mathbf{v}_1', 0)}{(u_1' + \tilde{u}_1)}. \tag{26}$$

Therefore, we can identify the meaning of the operator $\sum_s n_s e_s \int d^3v_1 P_s(\mathbf{v}_1 | \mathbf{v}_1')$ by comparing Eqs. (25) and (26). Hence, we have

$$\sum_s n_s e_s \int d^3v_1 P_s(\mathbf{v}_1 \mid \mathbf{v}_1'; t, \mathbf{k}_1) = \frac{1}{2\pi i} \int_{-\infty - i\gamma}^{+\infty - i\gamma} d\tilde{u}_1 \frac{e^{ik_1\tilde{u}_1 t}}{\epsilon(\tilde{u}_1, \mathbf{k}_1)} \int d^3v_1' \frac{\sum_s n_s e_s}{u_1' + \tilde{u}_1} \tag{27}$$

or, moreover,

$$Q_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2) = \frac{1}{(2\pi i)^2} \int_{-\infty - i\gamma_1}^{+\infty - i\gamma_1} d\tilde{u}_1 \times \int_{-\infty - i\gamma_2}^{+\infty - i\gamma_2} d\tilde{u}_2 \, e^{i(k_1\tilde{u}_1 + k_2\tilde{u}_2)t} \frac{1}{(u_1 + \tilde{u}_1)} \times \left[\int d^3v_1' \, \delta(\mathbf{v}_1 - \mathbf{v}_1') - \frac{D_s(\mathbf{v}_1)}{\epsilon(\tilde{u}_1, \mathbf{k}_1)} \times \int d^3v_1' \frac{\sum_s n_s e_s}{u_1 + \tilde{u}_1} \right] \int d^3v_2' \frac{\sum_r n_r e_r}{\epsilon(\tilde{u}_2, \mathbf{k}_2)(u_2' + \tilde{u}_2)}. \tag{28}$$

Substituting Eq. (28) into Eq. (20), one can compute $\sum n_r e_r \int g_{sr} \, d\mathbf{v}_2$.

C. Representation of the Operator in Laplace Transform Space

If we are dealing with a time-dependent R_{sr}, we usually have to perform the $\mathbf{v}_1'$ and $\mathbf{v}_2'$ integral operations before integrating over τ and, in that case, the situation may be very difficult. However, in view of the convolution-type integral in Eq. (20), the Laplace transform method is evidently useful. If we define the following Laplace transforms,

$$\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2) = \int_0^\infty dt \, e^{-i\omega t} Q_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; t, \mathbf{k}_1, \mathbf{k}_2), \tag{29}$$

$$\tilde{h}_s(\mathbf{v}_1, \omega, \mathbf{k}_1, \mathbf{k}_2) = \int_0^\infty dt \, e^{-i\omega t} h_s(\mathbf{v}_1, t, \mathbf{k}_1, \mathbf{k}_2), \tag{30}$$

$$\tilde{R}_{sr}(\omega) = \int_0^\infty dt \, e^{-i\omega t} R_{sr}(t), \tag{31}$$

where

$$h_s = \sum_r n_r e_r \int d^3v_2 \, g_{sr},$$

then the Laplace transform of Eq. (20) takes the

[14] L. Landau, J. Phys. USSR **10**, 25 (1946).
[15] N. G. van Kampen, Physica **21**, 949 (1955).
[16] To extend Landau's or van Kampen's work to a multispecies plasma is a straightforward operation.
[17] $\epsilon(\tilde{u}, \mathbf{k})$ is first defined for $\tilde{u}$ given in the lower half of the complex plane ($\tilde{u} \in S_-$) and is analytic in the region $-\gamma > \mathrm{Im}\, \tilde{u} > -\infty$, but such definition may be extended to the entire complex plane by analytic continuation.
[18] The same result, of course, may be obtained directly by integrating Eq. (24).

form

$$\tilde{h}_s(\mathbf{v}_1, \omega, \mathbf{k}_1, \mathbf{k}_2) = \tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega) g_{sr}(\mathbf{v}_1', \mathbf{v}_2', \mathbf{k}_1, \mathbf{k}_2, 0) + \tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega) \tilde{R}_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1', \mathbf{v}_2', \omega). \quad (32)$$

The Laplace transform of $Q_{sr}(t)$ defined by Eq. (29) is well defined as long as the condition

$$\operatorname{Im}(\omega - k_1\tilde{u}_1 - k_2\tilde{u}_2) < 0$$

is satisfied. We obtain

$$\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2) = \frac{1}{(2\pi i)^2} \int_{-\infty - i\gamma_1}^{+\infty - i\gamma_1} d\tilde{u}_1 \times \int_{-\infty - i\gamma_2}^{+\infty - i\gamma_2} d\tilde{u}_2 \frac{1}{i(\omega - k_1\tilde{u}_1 - k_2\tilde{u}_2)(u_1 + \tilde{u}_1)} \times \left[\int d^3v_1' \, \delta(\mathbf{v}_1 - \mathbf{v}_1') - \frac{D_s(\mathbf{v}_1, \mathbf{k}_1)}{\epsilon(\tilde{u}_1, \mathbf{k}_1)} \times \int \frac{d^3v_1' \sum_s n_s e_s}{\tilde{u}_1 + u_1'} \right] \int \frac{d^3v_2' \sum_r n_r e_r}{(\tilde{u}_2 + u_2')\epsilon(\tilde{u}_2, \mathbf{k}_2)}. \quad (33)$$

Let us consider the $\tilde{u}_2$ integration first. Since the pole $\tilde{u}_2 = (\omega - k_1\tilde{u}_1)/k_2$ is located below the path of integration, and the function

$$\frac{1}{(u_2' + \tilde{u}_2)\epsilon(\tilde{u}_2, \mathbf{k}_2)}$$

is analytic in the domain $-\gamma_2 \geq \operatorname{Im} \tilde{u}_2 \geq -\infty$, we may close the contour of $\tilde{u}_2$ integration in the lower half-plane. Thus,

$$\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2) = \frac{1}{(2\pi i)k_2} \int_{-\infty - i\gamma_1}^{+\infty - i\gamma_1} d\tilde{u}_1 \frac{1}{i(u_1 + \tilde{u}_1)} \times \left[\int d^3v_1' \, \delta(\mathbf{v}_1 - \mathbf{v}_1') - \frac{D_s(\mathbf{v}_1)}{\epsilon(\tilde{u}_1, \mathbf{k}_1)} \int \frac{d^3v_1' \sum_s n_s e_s}{(\tilde{u}_1 + u_1')} \right] \times \int \frac{d^3v_2' \sum_r n_r e_r}{((\omega - k_1\tilde{u}_1)/k_2 + u_2')\epsilon((\omega - k_1\tilde{u}_1)/k_2, \mathbf{k}_2)}. \quad (34)$$

Since we have considered $\epsilon(\tilde{u}_2, k_2)$ to be analytic in the domain

$$\operatorname{Im}\left(\frac{\omega - k_1\tilde{u}_1}{k_2}\right) < -\gamma_2$$

and have also assumed the condition

$$\operatorname{Im}\omega + k_1\gamma_1 + k_2\gamma_2 < 0$$

we can see that, for $\operatorname{Im}\tilde{u}_1 > -\gamma_1$ (that is, above the path of integration for the $\tilde{u}_1$ integral) the function

$$\epsilon\left(\frac{\omega - k_1\tilde{u}_1}{k_2}, \mathbf{k}_2\right)$$

has no singularity.[19] Making use of this property, one may evaluate the $\tilde{u}_1$ integral for the first term in Eq. (34) by closing the contour in the upper half of the complex plane. The following result is then obtained:

$$\tilde{Q}_{sr} = \frac{1}{ik_2} \int d^3v_1' \frac{\delta(\mathbf{v}_1' - \mathbf{v}_1)}{\epsilon((\omega + k_1u_1)/k_2, \mathbf{k}_2)} \times \int d^3v_2' \frac{\sum_r n_r e_r}{(u_2' + (\omega + k_1u_1)/k_2} - \frac{1}{2\pi i k_2} \times \int_{-\infty - i\gamma_1}^{+\infty - i\gamma_1} d\tilde{u}_1 \frac{D_s(\mathbf{v}_1)}{i(u_1 + \tilde{u}_1)\epsilon(\tilde{u}_1, \mathbf{k}_1)\epsilon((\omega - k_1\tilde{u}_1)/k_2, \mathbf{k}_2)} \times \int \frac{d^3v_1' \sum_s n_s e_s}{(u_1' + \tilde{u}_1)} \int \frac{d^3v_2' \sum_r n_r e_r}{(u_2' + (\omega - k_1\tilde{u}_1)/k_2)}. \quad (35)$$

So far, the operator expressed by Eq. (35) may be applied to quite general situations. For instance, the plasma can be either stable or unstable. For the stable case, since $\epsilon(\tilde{u}_1, k_1)$ has no pole in the lower half of the complex $\tilde{u}_1$ plane, we may deform the path of integration to the real axis: that is, we let $\gamma_1 \to 0_+$. After doing this, we may further consider the limit $\operatorname{Im}\omega \to 0_-$. Then it is convenient to introduce some new notations which are defined as follows:

$$\epsilon^-(\pm\alpha, \mathbf{k}) = \lim_{\operatorname{Im}\alpha \to 0_-} \epsilon(\pm\alpha, \mathbf{k}),$$

$$\epsilon^+(\pm\alpha, \mathbf{k}) = \lim_{\operatorname{Im}\alpha \to 0_+} \epsilon(\pm\alpha, \mathbf{k}),$$

$$\delta_\pm(\alpha) = \frac{\delta(\alpha)}{2} \pm \frac{i}{2\pi} P\frac{1}{\alpha} = \mp\frac{1}{2\pi i} \lim_{\gamma \to 0_+} \frac{1}{\alpha \pm i\gamma},$$

where $\delta(\alpha)$ is the Dirac delta function and P denotes principal value.

In terms of these notations, the operator given by Eq. (35) may be re-expressed as

$$\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2) = \frac{2\pi i}{ik_2} \int d^3v_1' \frac{\delta(\mathbf{v}_1 - \mathbf{v}_1')}{\epsilon^-((\omega + k_1u_1)/k_2, \mathbf{k}_2)} \int d^3v_2' \sum_r n_r e_r \, \delta_-\left(\frac{\omega + k_1u_1}{k_2} + u_2'\right) - \frac{2\pi i}{k_2 i} \int_{-\infty - i0_+}^{+\infty - i0_+} d\tilde{u}_1 \frac{D_s(\mathbf{v}_1)}{(u_1 + \tilde{u}_1)\epsilon^-(\tilde{u}_1, \mathbf{k})\epsilon^-((\omega - k_1\tilde{u}_1)/k_2, \mathbf{k}_2)} \times \int d^3v_1' \sum_s n_s e_s \, \delta_-(u_1' + \tilde{u}_1) \int d^3v_2' \sum_r n_r e_r \, \delta_-\left(u_2' + \frac{\omega - k_1\tilde{u}_1}{k_2}\right). \quad (36)$$

[19] In the discussion above, we have postulated that both k_1 and k_2 are positive, but the result can be stated as valid for the general case.

An alternative form of Eq. (36) can be obtained if we change $\tilde{u}_1$ to $-\tilde{u}_1$. In that case, we have

$$\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2) = \frac{2\pi}{k_2} \int d^3v_1' \frac{\delta(\mathbf{v}_1 - \mathbf{v}_1')}{\epsilon^-((\omega + k_1u_1)/k_2, \mathbf{k}_2)} \int d^3v_2' \sum_r n_r e_r \, \delta_-\left(\frac{\omega + k_1u_1}{k_2} + u_2'\right)$$
$$+ \frac{2\pi}{k_2} \int_{-\infty+i0+}^{+\infty+i0+} d\tilde{u}_1 \frac{D_s(\mathbf{v}_1)}{(\tilde{u} - u_1)\epsilon^+(-\tilde{u}_1, \mathbf{k}_1)\epsilon^-((\omega + k_1\tilde{u}_1)/k_2, \mathbf{k}_2)}$$
$$\times \int d^3v_1' \sum_s n_s e_s \, \delta_+(\tilde{u}_1 - u_1') \int d^3v_2' \sum_r n_r e_r \, \delta_-\left(\frac{\omega + k_1\tilde{u}_1}{k_2} + u_2'\right). \quad (37)$$

For a homogeneous plasma, we may set $k_1 = -k_2 = k$. In this special case, Eq. (37) reduces to the following form:

$$\tilde{Q}_{sr}(\omega, \mathbf{k}, -\mathbf{k}) = \frac{2\pi}{k} \int d^3v_1' \frac{\delta(\mathbf{v}_1 - \mathbf{v}_1')}{\epsilon^-(-\omega/k - u_1, \mathbf{k})} \int d^3v_2' \sum_r n_r e_r \, \delta_-\left(\frac{\omega}{k} + u_1 - u_2'\right)$$
$$+ \frac{2\pi}{k} \int_{-\infty+i0+}^{+\infty+i0+} d\tilde{u}_1 \frac{D_s(\mathbf{v}_1)}{(\tilde{u} - u)\epsilon^+(-\tilde{u}, \mathbf{k})\epsilon^-(-\omega/k - \tilde{u}_1, \mathbf{k})}$$
$$\times \int d^3v_1' \sum_s n_s e_s \, \delta_+(\tilde{u}_1 - u_1') \int d^3v_2' \sum_r n_r e_r \, \delta_-\left(\frac{\omega}{k} + \tilde{u}_1 - u_2'\right). \quad (38)$$

The operator $\tilde{Q}_{sr}$ expressed by Eqs. (35) (general stable or unstable), (36), (37) (inhomogeneous stable), or (38) (homogeneous stable) is easy to apply, since the only operations involved are the $\mathbf{v}_1'$ and $\mathbf{v}_2'$ integrations. This operator will serve as a powerful tool in solving a large number of problems with no need of lengthy algebraic effort. To demonstrate the merit of this operator, it is instructive to study the few examples presented in Sec. III.

III. APPLICATIONS OF THE OPERATOR

A. High-Frequency Conductivity

The first example which we examine is the determination of the conductivity of a fully ionized plasma in the presence of a high-frequency electric field $\mathbf{E}_0e^{i\omega t}$ which is supposed to be uniform in space. By the term *high-frequency*, we mean that the following condition is true:

$$\omega\tau \gg 1,$$

where τ is the collision time (defined as the cumulative 90-deg deflection time). Furthermore, we assume that the perturbation due to the applied field is small, and that the governing equations can be linearized. Hence, we now consider Eqs. (13) and (14), presented in Sec. II, but with the necessary modifications resulting from the presence of an external field. Since the plasma is essentially homogeneous, we set $\mathbf{k}_1 = -\mathbf{k}_2 = \mathbf{k}$. The kinetic equation then takes the form[20]

$$\frac{\partial f_s}{\partial t} + \frac{e_s}{m_s} \mathbf{E}_0e^{i\omega t} \cdot \frac{\partial F_s}{\partial \mathbf{v}_1} = -\frac{\partial}{\partial \mathbf{v}_1} \frac{i}{(2\pi)^3} \int d^3k \frac{4\pi \mathbf{k} e_s}{m_s k^2}$$
$$\times \sum_r n_r e_r \int d^3v_2 g_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2, t \to \infty) \quad (39)$$

and, according to Eq. (20),

$$h_s(t \to \infty) = \sum_r n_r e_r \int d^3v_2 g_{sr}(t \to \infty) = \int_0^\infty d\tau$$
$$\times Q_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \tau, \mathbf{k}, -\mathbf{k}) R_{sr}(\mathbf{k}, \mathbf{v}_1', \mathbf{v}_2', t - \tau). \quad (40)$$

In the present case, R_{sr} can be written as

$$R_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2, t) = \frac{4\pi e_s e_r i\mathbf{k}}{k^2} \cdot \left(\frac{1}{m_s} \frac{\partial}{\partial \mathbf{v}_1} - \frac{1}{m_r} \frac{\partial}{\partial \mathbf{v}_2}\right)[f_r(\mathbf{v}_2, t)F_s(\mathbf{v}_1) + f_s(\mathbf{v}_1, t)F_r(\mathbf{v}_2)]$$
$$+ \frac{1}{m_s} \frac{\partial f_s}{\partial \mathbf{v}_1} \cdot 4\pi e_s \sum_q n_q e_q \frac{i\mathbf{k}}{k^2} \int d^3v_3 G_{rq}(-\mathbf{k}, \mathbf{v}_2, \mathbf{v}_3) - \frac{1}{m_r} \frac{\partial f_r}{\partial \mathbf{v}_2} \cdot 4\pi e_r \sum_q n_q e_q \frac{i\mathbf{k}}{k^2}$$
$$\times \int d^3v_3 G_{sq}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_3) - \mathbf{E}_0e^{i\omega t} \cdot \left(\frac{1}{m_s} \frac{\partial}{\partial \mathbf{v}_1} + \frac{1}{m_r} \frac{\partial}{\partial \mathbf{v}_2}\right) G_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2). \quad (41)$$

[20] We are interested in the behavior of the plasma at large time.

If we require our result to be accurate only to first order in $1/(\lambda_D^3 n)$, we may replace f_s and f_r in Eq. (41) by the collisionless solution of Eq. (39); that is,

$$f_s(\mathbf{v}_1, t) = -\frac{e_s}{m_s}\frac{\mathbf{E}_0 e^{i\omega t}}{i\omega}\cdot\frac{\partial F_s}{\partial \mathbf{v}_1}. \tag{42}$$

Substituting Eq. (42) into (41), we obtain

$$\begin{aligned} R_{sr}(t) = {} & -E_0 e^{i\omega t}\left\{\frac{4\pi e_s e_r}{\omega k^2}\,\mathbf{k}\cdot\left(\frac{1}{m_s}\frac{\partial}{\partial \mathbf{v}_1}-\frac{1}{m_r}\frac{\partial}{\partial \mathbf{v}_2}\right)\hat{\mathbf{k}}_0\cdot\left[\frac{e_s}{m_s}\frac{\partial F_s}{\partial \mathbf{v}_1}F_r(v_2)+\frac{e_r}{m_r}\frac{\partial F_r}{\partial \mathbf{v}_2}F_s(v_1)\right]\right. \\ & + \frac{1}{m_s\omega}\frac{\partial}{\partial \mathbf{v}_1}\left(\frac{e_s}{m_s}\hat{\mathbf{k}}_0\cdot\frac{\partial F_s}{\partial \mathbf{v}_1}\right)\cdot 4\pi e_s\sum_q n_q e_q\frac{\mathbf{k}}{k^2}\int d^3v_3\, G_{rq}(-\mathbf{k}, \mathbf{v}_2, \mathbf{v}_3) - \frac{1}{m_r\omega}\frac{\partial}{\partial \mathbf{v}_2}\left(\frac{e_r}{m_r}\hat{\mathbf{k}}_0\cdot\frac{\partial F_r}{\partial \mathbf{v}_2}\right)\cdot 4\pi e_r \\ & \left.\times\sum_q n_q e_q\frac{\mathbf{k}}{k^2}\int d^3v_3\, G_{sq}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_3) + \hat{\mathbf{k}}_0\left(\frac{e_s}{m_s}\frac{\partial}{\partial \mathbf{v}_1}+\frac{e_r}{m_r}\frac{\partial}{\partial \mathbf{v}_2}\right)G_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2)\right\}, \\ \equiv {} & -E_0 e^{i\omega t}R^0_{sr}, \end{aligned} \tag{43}$$

where $\hat{\mathbf{k}}^0$ is a unit vector parallel to $\mathbf{E}_0$. Therefore,

$$\begin{aligned} h_s(t\to\infty) = {} & -e^{i\omega t}\int_0^\infty d\tau\, e^{-i\omega\tau}Q_{sr} \\ & \times(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \tau, \mathbf{k}, -\mathbf{k})R^0_{sr}(\mathbf{k}, \mathbf{v}_1', \mathbf{v}_2)\mathbf{E}_0. \end{aligned} \tag{44}$$

But

$$\int_0^\infty d\tau e^{-i\omega\tau}Q_{sr}(\tau) = \tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \omega, \mathbf{k}, -\mathbf{k})$$

which is given by Eq. (38). Hence, Eq. (39) may be rewritten as

$$\begin{aligned} \frac{\partial f_s}{\partial t}+\frac{e_s}{m_s}\mathbf{E}_0 e^{i\omega t}\cdot\frac{\partial F_s}{\partial \mathbf{v}_1} = {} & \frac{\partial}{\partial \mathbf{v}_1}\frac{i}{(2\pi)^3}\cdot\int d^3k\,\frac{4\pi e_s\mathbf{k}}{m_s k^2}E_0 e^{i\omega t} \\ & \times\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \omega, \mathbf{k}, -\mathbf{k})R^0_{sr}(\mathbf{k}, \mathbf{v}_1, \mathbf{v}_2). \end{aligned} \tag{45}$$

Now, let us multiply Eq. (45) by $n_s e_s \mathbf{v}_1$, integrate with respect to $\mathbf{v}_1$, and then sum up all components. If we introduce the definition of the current density $\mathbf{J}$ as

$$\mathbf{J} = \sum_s n_s e_s\int d^3v_1\,\mathbf{v}_1 f_s(\mathbf{v}_1, t)$$

then we have

$$\begin{aligned} & \frac{\partial \mathbf{J}}{\partial t}-\sum_s\frac{n_s e_s^2}{m_s}E_0 e^{i\omega t} \\ & = -\frac{i}{(2\pi)^3}\sum_s\int d^3k\,\frac{4\pi n_s e_s^2\mathbf{k}}{m_s k^2}E_0 e^{i\omega t} \\ & \times\int d^3v_1\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \omega)R^0_{sr}(\mathbf{k}, \mathbf{v}_1', \mathbf{v}_2') \end{aligned} \tag{46}$$

or

$$\begin{aligned} \mathbf{J} = {} & -\sum_s\frac{i\omega_s^2}{4\pi\omega}\mathbf{E}_0 e^{i\omega t}-\frac{\mathbf{E}_0 e^{i\omega t}}{(2\pi)^3}\cdot\sum_s\frac{\omega_s^2}{\omega}\int d^3k\,\frac{\mathbf{k}}{k^2} \\ & \times\int d^3v_1\,\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \omega)R^0_{sr}(\mathbf{k}, \mathbf{v}_1', \mathbf{v}_2'), \end{aligned} \tag{47}$$

where

$$\omega_s^2 = 4\pi n_s e_s^2/m_s.$$

If the conductivity σ is defined in the usual way,

$$\mathbf{J} = \sigma\mathbf{E}_0 e^{i\omega t}$$

then we have

$$\begin{aligned} \sigma = {} & -\sum_s\frac{i\omega_s^2}{4\pi\omega}-\frac{1}{(2\pi)^3}\sum_s\frac{\omega_s^2}{\omega}\int d^3k\,\frac{\mathbf{k}\cdot\hat{\mathbf{k}}_0}{k^2}\int d^3v_1 \\ & \times\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \omega)R^0_{sr}(\mathbf{k}, \mathbf{v}_1', \mathbf{v}_2'). \end{aligned} \tag{48}$$

The first part of Eq. (42) yields the usual dominant reactive contribution which is designated as σ_0,

$$\sigma_0 = -\sum_s\frac{i\omega_s^2}{4\pi\omega} \tag{49}$$

and the second term is the high-order correlation contribution σ_1:

$$\begin{aligned} \sigma_1 = {} & -\frac{1}{(2\pi)^3}\sum_s\frac{\omega_s^2}{\omega}\int d^3k\,\frac{\mathbf{k}\cdot\hat{\mathbf{k}}_0}{k^2}\int d^3v_1 \\ & \times\tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2';\, \omega)R^0_{sr}(\mathbf{k}, \mathbf{v}_1', \mathbf{v}_2'). \end{aligned} \tag{50}$$

In the present case, the operator $\tilde{Q}_{sr}(\mathbf{v}_1|\mathbf{v}_1', \mathbf{v}_2';\, \omega)$ takes the form

$$\begin{aligned} \tilde{Q}_{sr}(\omega) = {} & +\frac{2\pi}{k}\int d^3v_1'\,\frac{\delta(\mathbf{v}_1-\mathbf{v}_1')}{\epsilon^-(-(\omega+ku_1)/k, \mathbf{k})}\int d^3v_2'\sum_r n_r e_r\,\delta_-\left(\frac{\omega}{k}+u_1-u_2'\right) \\ & +\frac{2\pi}{k}\int_{0+}d\tilde{u}_1\,\frac{D_s(\mathbf{v}_1)}{(\tilde{u}_1-u_1)\epsilon^+(-\tilde{u}_1, \mathbf{k})\epsilon^-(-\omega/k-\tilde{u}_1, \mathbf{k})} \\ & \times\int d^3v_1'\sum_s n_s e_s\,\delta_+(\tilde{u}_1-u_1')\int d^3v_2'\sum_r n_r e_r\,\delta_-(\omega/k+\tilde{u}_1-u_2'). \end{aligned} \tag{51}$$

Equation (50) gives the general expression of conductivity due to correlation, provided that F_s and G_{sr} are given so that they vary slowly in a time scale which is long compared with the period $1/\omega$. Now, let us study the explicit form of Eq. (45) for a special case in which the following conditions are assumed:

(1) The plasma consists of two species: electrons and ions.

(2) The distribution functions $F_s(\mathbf{v})(s = 1, 2)$ are Maxwellian; i.e.,

$$F_s(v) = \left(\frac{m_s}{2\pi\Theta}\right)^{\frac{3}{2}} \exp\left(-\frac{m_s v^2}{2\Theta}\right).$$

(3) Both electrons and ions are single-charged; i.e., $e_s = eZ_s$, $Z_s = 1$ for ions, and $Z_s = -1$ for electrons.

In kinetic theory, the high-frequency conductivity for this case has been previously studied by Oberman, Ron, and Dawson[6] (hereafter called ORD) by using singular integral equation technique; it should be interesting to see how to obtain the same result by using the method discussed in Sec. II. First of all, it should be pointed out that, in Eq. (50), one can show by straightforward operation, using Eq. (51), that the following relation is true, i.e.,

$$\int d^3k \frac{\mathbf{k}\cdot\hat{\mathbf{k}}_0}{k^2} \int d^3v_1 \tilde{Q}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}, -\mathbf{k}) R^0_{sr}$$

$$= -\int dk \int d^3v_1 \frac{4\pi k^2 e\tilde{Q}_{sr}}{3n(k^2 + k_D^2)\Theta\omega}$$

$$\times [\omega_s^2 F_s(v_1')Z_r F_r(v_2') - \omega_r^2 F_r(v_2')Z_s F_s(v_1')]. \quad (52)$$

Thus,

$$\sigma_1 = -\frac{4\pi e}{(2\pi)^3 3n\Theta\omega^2} \int dk \frac{k^2}{(k^2 + k_D^2)} \sum_s \omega_s^2 \int d^3v_1$$

$$\times \tilde{Q}_{sr}[\omega_s^2 F_s(v_1')Z_r F_r(v_2') - \omega_r^2 F_r(v_2')Z_s F_s(v_1')], \quad (53)$$

where the quantity

$$\tilde{Q}_{sr}[\omega_s^2 F_s(\mathbf{v}_1')Z_r F_r(\mathbf{v}_2') - \omega_r^2 F_r(\mathbf{v}_2')Z_s F_s(\mathbf{v}_1')]$$

can be, of course, written explicitly as follows:

$$\tilde{Q}_{sr}[\omega_s^2 F_s(v_1')Z_r F_r(v_2') - \omega_r^2 F_r(v_2')Z_s F_s(v_1')] = +\frac{2\pi}{k}\left\{\frac{ne}{\epsilon^-[-(w+u), k]}[\omega_s^2 F_s(v_1) \sum_r \bar{F}_{r(-)}(w + u_1)\right.$$

$$- Z_s F_s(\mathbf{v}_1) \sum_r Z_r\omega_r^2 \bar{F}_{r(-)}(w + u_1)] + \int_{0+} d\tilde{u}_1 \frac{n^2e^2 D_s(\mathbf{v}_1)}{(\tilde{u}_1 - u_1)\epsilon^+(-\tilde{u}_1, k)\epsilon^-(-w - \tilde{u}_1, k)}$$

$$\left.\times [\sum_s Z_s\omega_s^2 \bar{F}_{s(+)}(\tilde{u}_1) \sum_r \bar{F}_{r(-)}(w + \tilde{u}_1) - \sum_r Z_r\omega_r^2 \bar{F}_{r(-)}(w + \tilde{u}_1) \sum_s \bar{F}_{s(+)}(\tilde{u}_1)]\right\}, \quad (54)$$

where we have introduced the shorthand notations

$$\bar{F}(u_1) = \int d^3v_1 \, \delta\left(u_1 - \frac{\mathbf{k}\cdot\mathbf{v}_1}{k}\right) F(\mathbf{v}_1),$$

$$\bar{F}_{(\pm)}(\alpha_1) = \int du_1 \, \delta_\pm(\alpha_1 - u_1)\bar{F}(u_1),$$

and

$$w = \omega/k.$$

If we further define

$$J(u) = \sum_s Z_s\omega_s^2 \bar{F}_s(u),$$

$$F(u) = \sum_s \bar{F}_s(u),$$

$$\rho^*(u) = \epsilon^-(-u, k),$$

$$\rho(u) = \epsilon^+(-u, k),$$

the conductivity σ_1 may be expressed as

$$\sigma_1 = -\frac{8\pi^2e^2}{(2\pi)^3 3\Theta\omega^2} \int dk \frac{k}{k_D^2 + k^2} \int du_1 \left\{\frac{J(u_1)J_{(-)}(w + u_1) - \sum_s \omega_s^4 \bar{F}_s(u_1)F_{(-)}(w + u_1)}{\rho^*(w + u_1)}\right.$$

$$\left. - \frac{k_D^2}{2k^2} u_1 J(u_1) \int_{-\infty+i0+}^{+\infty+i0+} d\tilde{u}_1 \frac{1}{(\tilde{u}_1 - u_1)\rho(\tilde{u}_1)\rho^*(w + \tilde{u}_1)} [J_{(+)}(\tilde{u}_1)F_{(-)}(w + \tilde{u}_1) - J_{(-)}(w + \tilde{u}_1)F_{(+)}(\tilde{u}_1)]\right\}. \quad (55)$$

Notice that, in the ORD calculations, the *plus* and *minus* functions are defined as

$$f^{(\pm)} = \int \frac{du' f(u')}{u' - u \mp i\epsilon}.$$

Therefore, we find the following transformation rules:

$$J_{(\mp)}(u) = \mp\frac{1}{2\pi i} J^{(\mp)}(u),$$

$F_{(\mp)}(u) = \mp\frac{1}{2\pi i} F^{(\mp)}(u)$, which enable us to rewrite Eq. (55) in the ORD notations, and we find that

$$\sigma_1 = -\frac{i4\pi e^2}{(2\pi)^3 3\Theta\omega^2}\int dk\,\frac{k}{k^2+k_D^2}\int du_1\left\{\frac{J(u_1)J^{(-)}(w+u_1)-\sum_s \omega_s^4 \bar{F}_s(u_1)F^{(-)}(w+u_1)}{\rho^*(w+u_1)}\right.$$

$$\left.+\frac{k_D^2}{2k^2}\frac{u_1 J(u_1)}{(2\pi i)}\int d\tilde{u}_1\,\frac{1}{(\tilde{u}_1-u,\,+i\epsilon)\rho(\tilde{u}_1)\rho^*(w+\tilde{u}_1)}\left[J^{(+)}(\tilde{u}_1)F^{(-)}(w+\tilde{u}_1)-J^{(-)}(w+\tilde{u}_1)F^{(+)}(\tilde{u}_1)\right]\right\}. \quad (56)$$

Taking account of the fact that we have considered the applied field as being $\mathbf{E}_0 e^{i\omega t}$, rather than $\mathbf{E}_0 e^{-i\omega t}$, we see that Eq. (56) is essentially equivalent to the result obtained by the ORD method, as seen in Eq. (47), Ref. 6.

B. Derivation of the Kinetic Equation for a Homogeneous and Stable Plasma

The second example presented here is the derivation of the well-known BLGRR equation. The major assumptions employed in the usual analyses are stated as follows:

(1) The plasma is homogeneous and stable.
(2) Bogoliubov's adiabatic approximation is valid, and only the asymptotic solution of the pair correlation is needed.

In this case, we designate $\mathbf{k}_1 = -\mathbf{k}_2 = \mathbf{k}$, and the kinetic equation can be written as

$$\frac{\partial F_s}{\partial t} = \frac{1}{(2\pi)^3}\frac{\partial}{\partial \mathbf{v}_1}\int d^3k\,\frac{4\pi e_s \mathbf{k}}{m_s k^2}\,\mathrm{Im}\sum_r n_r e_r \times\int d^3v_2\, G_{sr}(\mathbf{k},\mathbf{v}_1,\mathbf{v}_2,t\to\infty). \quad (57)$$

Again, referring to Eq. (12) in Sec. II,

$$\tilde{R}_{sr}(\mathbf{k},\mathbf{v}_1',\mathbf{v}_2',\omega) = \frac{4\pi e_s e_r i\mathbf{k}}{(i\omega)k^2}\left[\frac{F_r(\mathbf{v}_2')}{m_s}\frac{\partial F_s(\mathbf{v}_1')}{\partial \mathbf{v}_1'}-\frac{F_s(\mathbf{v}_1')}{m_r}\frac{\partial F_r(\mathbf{v}_2')}{\partial \mathbf{v}_2'}\right],$$

$$= -\frac{ik}{(i\omega)}\left[e_r F_r(\mathbf{v}_2')D_s(\mathbf{v}_1')-e_s F_s(\mathbf{v}_1')D_r(\mathbf{v}_2')\right]. \quad (58)$$

Making use of the formula

$$\sum_r n_r e_r\int d^3v_2\, G_{sr}(t\to\infty) = \lim_{i\omega\to 0+}\left\{i\omega\sum_r n_r e_r\int d^3v_2\,\tilde{G}_{sr}(\omega)\right\} \quad (59)$$

and, applying Eqs. (38) and (32), we obtain

$$\sum_r n_r e_r\int d^3v_2\, G_{sr}(t\to\infty) = Q_{sr}^{\infty}(\mathbf{v}_1\mid\mathbf{v}_1',\mathbf{v}_2';t\to\infty,\mathbf{k},-\mathbf{k})R_{sr}(\mathbf{k},\mathbf{v}_1',\mathbf{v}_2'), \quad (60)$$

where

$$\int_0^\infty dt\, Q_{sr}(t) = Q_{sr}^{\infty} = +\frac{2\pi}{k}\int d^3v_1'\,\frac{\delta(\mathbf{v}_1'-\mathbf{v}_1)}{\epsilon^-(-u_1,\mathbf{k})}\int d^3v_2'\sum_r n_r e_r\delta_-(u_1-u_2')$$

$$+\frac{2\pi}{k}\int_{-\infty+i0+}^{+\infty+i0+} d\tilde{u}_1\,\frac{D_s(\mathbf{v}_1,k_1)}{(\tilde{u}_1-u_1)\epsilon^+(-\tilde{u}_1,\mathbf{k}_1)\epsilon^-(-\tilde{u}_1,\mathbf{k})} \times\int d^3v_1'\sum_s n_s e_s\,\delta_+(\tilde{u}_1-u_1') \times\int d^3v_2'\sum_r n_r e_r\,\delta_-(\tilde{u}_1-u_2'). \quad (61)$$

As Lenard[13] first pointed out, in deriving the collision integral one needs only the imaginary part of h_s. Since R_{sr} is a pure imaginary quantity, we need only the real part of Q_{sr}^{∞} in order to obtain Im h_s. Making use of the property

$$\int d^3v_1'\sum_s n_s e_s\,\delta_-(\tilde{u}_1-u_1') \times\int d^3v_2'\sum_r n_r e_r\,\delta_+(\tilde{u}_1-u_2')\tilde{R}_{sr}(\mathbf{v}_1',\mathbf{v}_2') = -\int d^3v_2'\sum_r n_r e_r\,\delta_-(\tilde{u}_1'-u_2') \times\int d^3v_1'\sum_s n_s e_s\,\delta_+(\tilde{u}_1-u_1')\tilde{R}_{sr}(\mathbf{v}_1',\mathbf{v}_2'), \quad (62)$$

we may write down the complex conjugate of Q_{sr}^{∞}, namely $Q_{sr}^{\infty *}$, as follows:

$$Q_{sr}^{\infty *} = +\frac{2\pi}{k}\int d^3v_1'\,\frac{\delta(\mathbf{v}_1'-v_1)}{\epsilon^+(-u_1,\mathbf{k}_1)} \times\int d^3v_2'\sum_r n_r e_r\,\delta_+(u_1-u_2') -\frac{2\pi}{k}\int_{-\infty+i0-}^{+\infty+i0-} d\tilde{u}_1\,\frac{D_s(\mathbf{v}_1,k)}{(\tilde{u}_1-u_1)\epsilon^-(-\tilde{u}_1,\mathbf{k})\epsilon^+(-\tilde{u}_1,\mathbf{k})} \times\int d^3v_1'\sum_s n_s e_s\,\delta_+(\tilde{u}_1-u_1') \times\int d^3v_2'\sum_r n_r e_r\,\delta_-(\tilde{u}_1-u_2'). \quad (63)$$

Therefore,

$$Re\, Q^{\infty}_{sr} = \frac{Q^{\infty}_{s} + Q^{\infty *}_{s}}{2} = +\frac{\pi}{k}\int d^3v_2' \sum_r n_r e_r \times \left[\frac{\delta_+(u_1 - u_2')}{\epsilon^+(-u_1, \mathbf{k})} + \frac{\delta_-(u_1 - u_2')}{\epsilon^-(-u_1, \mathbf{k})}\right]\int d^3v_1'\, \delta(\mathbf{v}_1' - \mathbf{v}_1) - \frac{2\pi^2 i}{k} D_s(\mathbf{v}_1, k)\frac{1}{\epsilon^+(-u_1, \mathbf{k})\epsilon^-(-u_1, \mathbf{k})} \times \int d^3v_1' \sum_s n_s e_s\, \delta_+(u_1 - u_1') \times \int d^3v_2' \sum_r n_r e_r\, \delta_-(u_1 - u_2') = Q^{\mathrm{I}}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2', \mathbf{k}, -\mathbf{k}) + Q^{\mathrm{II}}_{sr}(\mathbf{v}_1 \mid \mathbf{v}_1', \mathbf{v}_2', \mathbf{k}, -\mathbf{k}), \tag{64}$$

where

$$Q^{\mathrm{I}}_{sr} = \frac{\pi}{k\epsilon^+(-u_1)\epsilon^-(-u_1)}\int d^3v_2' \sum_r n_r e_r\, \delta(u_1 - u_2') \times \int d^3v_1'\, \delta(\mathbf{v}_1 - \mathbf{v}_1'),$$

$$Q^{\mathrm{II}}_{sr} = \frac{2\pi^2 i}{k\epsilon^+(-u_1)\epsilon^-(-u_1)} \times \int d^3v_2' \sum_r n_r e_r \Big\{[\delta_-(u_1 - u_2')\, \bar{D}_+(u_1) - \delta_+(u_1 - u_2')\, \bar{D}_-(u_1)]\int d^3v_1'\, \delta(\mathbf{v}_1' - \mathbf{v}_1) - D(\mathbf{v}_1)\, \delta_-(u_1 - u_2')\int d^3v_1'\, \delta_+(u_1 - u_1')\Big\},$$

$$\bar{D}_{\pm}(u_1) = \int d^3v_1'\, \delta_{\pm}(u_1 - u_1') \sum_s n_s e_s\, D_s(\mathbf{v}_1),$$

and

$$\epsilon^{\pm}(-u_1) = 1 \pm 2\pi i\, \bar{D}_{\pm}(u_1).$$

To perform the $\mathbf{v}_1'$ and $\mathbf{v}_2'$ integral operations on $R_{sr}(\mathbf{v}_1', \mathbf{v}_2', k)$ is trivial. One can show readily that Q^{II}_{sr} R_{sr} vanishes. Therefore, we conclude that

$$\mathrm{Im}\sum_r n_r e_r \int G_{sr}\, d^3v_2 = \frac{\pi}{\epsilon^-\left(-\frac{\mathbf{k}\cdot\mathbf{v}_1}{k}\right)\epsilon^+\left(-\frac{\mathbf{k}\cdot\mathbf{v}_1}{k}\right)} \times \int d^3v_2\, \delta(\mathbf{k}\cdot\mathbf{v}_1 - \mathbf{k}\cdot\mathbf{v}_2) \sum_r n_r e_r \frac{4\pi e_s e_r \mathbf{k}}{k^2} \cdot\left[\frac{F_r(\mathbf{v}_2)}{m_s}\frac{\partial F_s(\mathbf{v}_1)}{\partial \mathbf{v}_1} - \frac{F_s(\mathbf{v}_1)}{m_r}\frac{\partial F_r(\mathbf{v}_2)}{\partial \mathbf{v}_2}\right]. \tag{65}$$

Substitution of Eq. (65) into Eq. (57) gives the kinetic equation previously derived independently, with different methods, by Balescu,[12] Lenard,[13] Guernsey,[4] and Rostoker and Rosenbluth.[2]

C. The Guernsey Problem of an Inhomogeneous Plasma

Guernsey[5] has solved Eq. (14) with a Maxwellian distribution function F_s by using the singular integral equation technique. We can see that the solution $\tilde{h}_s = \sum_r n_r e_r \int \tilde{g}_{sr}\, d^3v_2$ is obtainable by the method discussed in this paper. As a matter of fact, the only preparatory work needed is to determine the function $\tilde{R}_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1, \mathbf{v}_2, \omega)$ which, in the present case, is found to be

$$\tilde{R}_{sr} = 4\pi e_s e_r i\Big\{\frac{F_s(v_1)}{(k_1^2 + k_D^2)}\frac{\mathbf{k}_1}{m_r}\cdot\frac{\partial}{\partial \mathbf{v}_2}\tilde{f}_r(\mathbf{v}_2, {}_1\mathbf{k} + \mathbf{k}_2, \omega) + \frac{F_r(v_2)}{(k_2^2 + k_D^2)}\frac{\mathbf{k}_2}{m_s}\cdot\frac{\partial}{\partial \mathbf{v}_1} \times \tilde{f}_s(\mathbf{v}_1, \mathbf{k}_1 + \mathbf{k}_2, \omega) - \frac{\tilde{f}_r(\mathbf{v}_2, \mathbf{k}_1 + \mathbf{k}_2, \omega)}{k_1^2 m_s}\mathbf{k}_1\cdot\frac{\partial F_s(v_1)}{\partial \mathbf{v}_1} - \frac{\tilde{f}_s(\mathbf{v}_1, \mathbf{k}_1 + \mathbf{k}_2, \omega)}{k_2^2 m_r}\mathbf{k}_2\cdot\frac{\partial F_r(v_2)}{\partial \mathbf{v}_2} + \frac{4\pi}{\Theta}\int d^3v_3 \times \sum_q n_q e_q \tilde{f}_q(\mathbf{k}_1 + \mathbf{k}_2, \mathbf{v}_3, \omega)\frac{\mathbf{k}_1 + \mathbf{k}_2}{(\mathbf{k}_1 + \mathbf{k}_2)^2}\cdot\left[\frac{e_s}{m_s(k_2^2 + k_D^2)}\frac{\partial}{\partial \mathbf{v}_1} + \frac{e_r}{m_r(k_1^2 + k_D^2)}\frac{\partial}{\partial \mathbf{v}_2}\right]F_s(v_1)F_r(v_2)\Big\}, \tag{66}$$

where $\Theta = \kappa T$, and κ is the Boltzmann constant.

Since the operator $\tilde{Q}_{sr}(\mathbf{v}_1|\mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2)$ is given by Eq. (37), and we know that

$$\tilde{h}_s(\mathbf{v}_1, \omega, \mathbf{k}_1, \mathbf{k}_2) = \tilde{Q}_{sr}(\mathbf{v}_1, \mathbf{v}_1', \mathbf{v}_2'; \omega, \mathbf{k}_1, \mathbf{k}_2) \times [g_{sr}(\mathbf{v}_1', \mathbf{v}_2; \mathbf{k}_1, \mathbf{k}_2, t = 0) + \tilde{R}_{sr}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{v}_1', \mathbf{v}_2', \omega)], \tag{67}$$

the explicit form of the solution for $\tilde{h}_s$ is therefore immediately obtainable. We omit the final result, since the explicit form is lengthy and not our main concern.

D. The Kinetic Equation for a Weakly Unstable Plasma

So far, we have considered three examples for the stable case. In this example, we consider the derivation of the kinetic equation for a simple unstable

case[21] recently studied by Balescu.[22] If we repeat his assumptions and utilize the operator given by Eq. (35), the solution $\sum_r n_r e_r \int \tilde{G}_{sr}\, d^3v_2$ can be written immediately:

$$\sum_r n_r e_r \int d^3v_2\, \tilde{G}_{sr} = -\Bigg[\frac{1}{ik}\int d^3v_1' \frac{\delta(\mathbf{v}_1' - \mathbf{v}_1')}{\epsilon(-w - u_1)} \times \int d^3v_2' \frac{\sum_r n_r e_r}{(u_2' - w - u_1)} - \frac{1}{(2\pi i)(ik)} \times \int_{-\infty+i\gamma_1}^{+\infty+i\gamma_1} d\tilde{u}\, \frac{D_s(\mathbf{v}_1)}{(\tilde{u} - u_1)\epsilon(-\tilde{u})\epsilon(-w - \tilde{u})} \times \int \frac{d^3v_1' \sum_s n_s e_s}{(u_1' - \tilde{u})} \int \frac{d^3v_2' \sum_r n_r e_r}{(u_2' - w - \tilde{u})}\Bigg] R_{sr},$$

(Im $w < -\gamma$ so that $\epsilon(-w - u_1)$ has no zero)

where

$$w = \omega/k$$

and

$$R_{sr} = \frac{4\pi e_s e_r}{k^2}\, i\mathbf{k}\left[\frac{F_r(2)}{m_s}\frac{\partial F_s(1)}{\partial \mathbf{v}_1} - \frac{F_s(1)}{m_r}\frac{\partial F_r(2)}{\partial \mathbf{v}_2}\right].$$

Equation (68) is essentially in agreement with Eq. (5.3) given by Balescu in Ref. 22. In principle, the kinetic equation is obtained by substituting Eq. (68) into Eq. (11). Of course, a more explicit form of Eq. (68) may be desirable, but such discussion is omitted here because some of the studies have already been done by Balescu.[22]

IV. CONCLUDING REMARKS

In the present communication, a useful operator has been introduced. Such an operator is very convenient to apply and serves as a useful tool for the study of a variety of problems in plasma kinetic theory. To illustrate the possible mathematical simplification resulting from application of such an operator, four examples have been considered. Comparing our analyses with studies based on the singular integral equation technique, one may conclude that the saving of mathematical effort due to the present method is undoubtedly remarkable. Several other advantages relating to the application of this operator may be listed as follows:

(1) The operator enables us to express the integral of the pair correlation function in a rather compact form which is often preferable to the full explicit form (as illustrated in Sec. IIIC).

(2) The present operator method avoids the use of Poincaré–Bertrand transformation formula which is required usually to simplify the results obtained by singular integral equation technique.

(3) The operator method is equally convenient to apply in the stable and the unstable case. Of course, our discussion is mainly concerned with Eqs. (11) to (14). If the instability is strong, the validity of these Equations becomes questionable. (A new theory which was developed recently to treat such a situation, is given by Frieman and Rutherford.[23]) However, further discussion in this regard is beyond the scope of the present communication.

One basic mathematical assumption implicit in our discussion is that the functions R_{sr} behave sufficiently well that the $\mathbf{v}_1'$ and $\mathbf{v}_2'$ integral operations are meaningful. This assumption is apparently acceptable in most physical problems.

Finally, we remark that, in the present discussion, only longitudinal Coulomb interactions have been considered. Further applications of the present operator will be discussed in some forthcoming publications.

ACKNOWLEDGMENT

The author is grateful to Dr. E. H. Klevans for helpful discussions and criticism.

[21] Rutherford and Frieman (Ref. 8) have also given a brief discussion along this line, but they considered only the parts of the G_{sr} solution which are most important at large time.

[22] R. Balescu, J. Math. Phys. 4, 1090 (1963).

[23] E. A. Frieman and R. H. Rutherford, Princeton University Plasma Physics Laboratory Report MATT 179 (1963) (unpublished).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Representation of Fields in a Two-Dimensional Model Theory*

JAN TARSKI
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 28 February 1964)

The two-dimensional model recently suggested by Schroer is examined. The free scalar massless field in two dimensions is discussed in detail. The infrared-myriotic representations of this field which arise in the model are described; it is shown that, up to unitary equivalence, the required representation depends only on the net total charge of the fermions. Discussion is also given of distribution-theoretic aspects of the fields, and in particular, of possible restrictions on test functions.

1. INTRODUCTION

IN the last few years solutions have been obtained for several two-dimensional model field theories.[1,2] These solutions were largely limited to the construction of a complete set of Green's functions. In principle, the Green's functions determine the Hilbert space of states, and the action of the fields.[3] But in practice, it has not been easy to describe the space of states in these theories. A description of the states would be desirable, to complete the solution which was initiated by finding the Green's functions. Furthermore, as the Green's functions for these models show, the states do not constitute a Fock space, and therefore their description could yield some insight into the problem of representation of fields.

In this paper we examine the two-dimensional model field theory, which was suggested recently by Schroer.[4] Schroer's model is one of the simplest examples of coupled fields, but nonetheless, it raises some interesting mathematical questions. One of these concerns the representation of fields. Another deals with distribution-theoretic aspects, such as the imposing of special restrictions on test functions.

The representations of the fields (or, of the canonical commutation relations) which arise in the foregoing models,[1,2,4] are related to the infrared divergences. As is well known, these are due to the fact that a physical state may include an infinite number of very soft massless quanta, e.g. photons, the total energy of which is finite.[5] It has been observed by Friedrichs,[6] and is also well known by now, that such situations are appropriately described by representations of fields, which are inequivalent to the Fock representation. We shall refer to representations of this kind as *infrared representations*. Thus, here the word *infrared* is more restrictive than *myriotic*, which has been introduced[6] for non-Fock representations in general.

In the study of infrared representations, we make use of several results of Friedrichs, and we also find parallels with some of his conclusions. In particular, in the case of scattering of a massless scalar field in four dimensions by a fixed source, the relevant infrared representation depends (up to unitary equivalence) only on the net total charge of the source. In Schroer's model, the conclusion is analogous.

It is convenient to recapitulate Schroer's arguments and conclusions at this point. We consider the field equations

$$\Box\varphi = 0, \tag{1.1a}$$

$$(\partial + M)\psi = ig(\partial\varphi)\psi + (\text{ren.}), \tag{1.1b}$$

which correspond to the interaction Lagrangian density

$$L_{\text{int}} = ig\bar{\psi}(\partial\varphi)\psi + (\text{ren.}). \tag{1.1c}$$

The renormalized solution is

$$\psi(x) = \psi^{(0)}(x)\, :e^{ig\varphi(x)}:, \tag{1.2}$$

where $\psi^{(0)}$ is a free spinor of mass M. (The restriction to two dimensions is needed, in order that the solution be renormalizable.) We assume canonical commutation relations for $\psi^{(0)}$ and φ. The solution (1.2) can thus be interpreted as an operator-dependent gauge transformation applied to a free spinor. Therefore, one would not expect this theory to describe any interesting physical effects, in some two-dimensional world.

Let us consider the field φ. Its two-point function is determined by Eq. (1.1a) and the commutation

* This work represents results obtained at the Courant Institute under the Ford Foundation grant for mathematical physics.

[1] K. Johnson, Nuovo Cimento **20**, 773 (1961), and references given there.

[2] J. Schwinger, Phys. Rev. **128**, 2425 (1962).

[3] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

[4] B. Schroer, Fortschr. Physik **11**, No. 1 (1963), Sec. III.

[5] F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).

[6] K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1953), especially Secs. 13, 14, and 19.

relations, and is

$$\langle\varphi(x)\varphi(y)\rangle_0 = -(4\pi)^{-1} \lim_{\epsilon\to 0+} \log(-\xi^2 + i\epsilon\xi^0) + C, \quad (1.3)$$

where

$$\xi = x - y, \qquad \xi^2 = (\xi^0)^2 - (\xi^1)^2. \quad (1.4)$$

On the other hand, a formal construction of the Fourier transform yields

$$\langle\varphi(x)\varphi(y)\rangle_0 \sim \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{dp^1}{|p^1|} e^{-ip\xi}. \quad (1.5)$$

(Here and in similar expressions, $p^0 = |p^1|$ is understood.) The constant C is related to the ambiguity in the scale of the logarithm, and to the singularity of the invariant measure $dp^1\ |p^1|^{-1}$ at $p^1 = 0$. The important fact is that $\varphi(x)$ is an operator-valued distribution,[7] and if we restrict ourselves to test functions $f \in \mathcal{S}$ which satisfy

$$\int d^2x f(x) = 0 = \tilde{f}(0),$$

where $\tilde{f}$ denotes the Fourier transform,

$$f(x) = \frac{1}{2\pi}\int d^2k e^{ikx}\tilde{f}(k), \quad (1.6)$$

then all the ambiguities in (1.3) and (1.5) are eliminated.[4,8] This subspace of test functions will be denoted by $\mathcal{S}_0$:

$$\mathcal{S}_0 = \{f \in \mathcal{S} : \tilde{f}(0) = 0\}. \quad (1.7)$$

To investigate the ψ quanta, we need the two-point function of $\Upsilon(x) \equiv :e^{ig\varphi(x)}:$,

$$\langle\Upsilon(x)\Upsilon^*(y)\rangle_0 = C'(-\xi^2 + i\epsilon\xi^0)^{-g^2/4\pi}. \quad (1.8)$$

We have here a power-function branch point at $\xi^2 = 0$, and the Fourier transform $Z^{(2)}_{g^2/2\pi}(p)$ shows a similar branch point singularity [cf. Eq. (3.8a)]:

$$Z^{(2)}_{g^2/2\pi} \sim (p^2)^{(g^2/4\pi)-1}.$$

Such branch points are typical of infrared situations.[1,9] This relation also suggests that there is in the theory no discrete state of mass M.

In the context of this model, the representation problem reduces to the following: To describe, and to characterize, the action of (smeared-out) operators $\varphi(f)$ on vectors such as $\Upsilon(h)\ |0\rangle$. We characterize this action by showing that the distributions

$$A^{(\pm)}(p^1) = \varphi^{(\pm)}(p^1) \pm iQH(p^1)$$

have the Fock property,[10] $A^{(-)}(p^1)\tau = 0$ for some vector τ. Here $H(p^1)$ is a function which is singular at $p^1 = 0$, and Q is the net total charge shown by the operators Υ and Υ^*.

With regard to restrictions on test functions, we observe that the condition $\tilde{f}(0) = 0$ does not remove the ambiguity in Eq. (1.8). Furthermore, this condition does not eliminate the infrared behavior, even though the latter is associated with the point $p = 0$. To clarify this puzzling situation, we examine the Wick powers $:\varphi^n(x):$, which occur in the expansion of $\Upsilon(x)$, or their smeared-out forms $:\varphi^n:(f)$.

We ask if the action of $:\varphi^n:(f)$ can be represented on the Fock space generated by operators $\varphi(h)$, where $h \in \mathcal{S}_0$. We find that this is impossible for $n > 2$, and that for $n = 2$, a stronger restriction on test functions is necessary. The remaining alternative is to fix the arbitrary parameters in Eq. (1.3), and to enlarge the Fock space by allowing $\tilde{h}(0) \neq 0$. In this way we conclude that the infrared behavior is inherent in the spinor solution (1.2). (On the other hand, one can avoid the infrared representations by modifying the solution to a bilocal operator. See Sec. 5.)

We should emphasize that our arguments are incomplete in several respects. For example, we do not discuss the details of completion of spaces. Furthermore, we prove the equivalence of representations (cf. above) only for irreducible subrepresentations. We do not consider how to combine these, to yield the desired Hilbert space.

We might note that several interesting aspects of this model, such as asymptotic convergence and renormalization, are beyond the scope of the present investigation. We confine ourselves to the problems which we outlined in the foregoing.

Section 2 contains a discussion of several free scalar massless fields, which were loosely denoted by φ in the foregoing. In Sec. 3 we discuss the Wick powers and the exponential of these fields. The infrared representations are described in Sec. 4. In Sec. 5 we return to the spinor solution, and to the model. Some concluding remarks are made in Sec. 6.

2. FREE SCALAR MASSLESS FIELDS IN TWO DIMENSIONS

The field in question has several special properties, of which we have already noted one: the singularity

[7] A. S. Wightman and L. Gårding, "Fields as Operator-Valued Distributions in Quantum Field Theory" (to be published).

[8] H. J. Borchers, "Three Remarks on Quantum Field Theory" (unpublished manuscript).

[9] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), Sec. 41.

[10] We use the convention of Ref. 9: The annihilation part of B is denoted by $B^{(-)}$.

of the invariant measure at $p = 0$. This section is therefore devoted to a summary of such properties.

Extension of Definition of the Field

We have seen that we can obtain a satisfactory theory of the free scalar field by restricting the test functions. Another possibility is to change the invariant measure

$$d\Omega_P(p) = \tfrac{1}{2}d^2p\,\delta(p^0 - |p^1|)\,|p^1|^{-1} \qquad (2.1)$$

into a suitable distribution. The subscript P refers to positive definiteness, which apparently has to be sacrificed in such an extension. (See below.)

We first consider the rays, $p^0 = p^1 > 0$ and $p^0 = -p^1 > 0$, separately. For the interval $(0, \infty)$, let us set[11,12]

$$(r^{-1})_+ = (d/dr)[\theta(r)\log r] + C\delta(r). \qquad (2.2)$$

This quantity is a distribution which equals r^{-1} for $r > 0$, and which vanishes for $r < 0$. However, its behavior at $r = 0$ expresses the ambiguity of scale for the logarithm. For a test function $\alpha(r)$ we thus have

$$\int_{-\infty}^{\infty} dr(r^{-1})_+\alpha(r) = -\int_0^{\infty} dr \log r\,\alpha'(r) + C\alpha(0).$$

The Fourier transform of $(r^{-1})_+$ is

$$\int_{-\infty}^{\infty} dr\, e^{-irk}(r^{-1})_+ = -\log ik + \Gamma'(1) + C. \qquad (2.3)$$

We now introduce "light cone" coordinates

$$u = p^0 + p^1, \qquad v = p^0 - p^1 \qquad (2.4)$$

and define

$$\omega(p) = \delta(v)(u^{-1})_+ + \delta(u)(v^{-1})_+. \qquad (2.5)$$

Note that Eq. (2.2) implies a contribution $C\delta(u)\delta(v)$ to $\omega(p)$.

Let us determine Lorentz invariance of the two terms in Eq. (2.5). Let

$$\Lambda_\lambda \begin{bmatrix} p^0 \\ p^1 \end{bmatrix} = \begin{bmatrix} \cosh\lambda & \sinh\lambda \\ \sinh\lambda & \cosh\lambda \end{bmatrix} \begin{bmatrix} p^0 \\ p^1 \end{bmatrix}, \qquad (2.6)$$

then

$$\Lambda_\lambda u = e^\lambda u, \qquad \Lambda_\lambda v = e^{-\lambda}v, \qquad (2.7)$$

$$\delta(e^\lambda u)((e^{-\lambda}v)^{-1})_+ = \delta(u)[(v^{-1})_+ - \lambda\delta(v)], \qquad (2.8a)$$

$$\delta(e^{-\lambda}v)((e^\lambda u)^{-1})_+ = \delta(v)[(u^{-1})_+ + \lambda\delta(u)]. \qquad (2.8b)$$

Thus, $\omega(p)$ is invariant, but its two summands separately are not.

We can now define the extended field φ_κ, whose two-point function [cf. Eq. (1.3)],

$$\langle\varphi_\kappa(x)\varphi_\kappa(y)\rangle_0 = -(4\pi)^{-1}\log(-\xi^2 + i\epsilon\xi^0) + C', \qquad (2.9)$$

is the Fourier transform of $\omega(p)$. The arbitrary constants are to depend on the index κ in some definite way, and all test functions in $\mathcal{S}$ are now allowed.

A few remarks about the arbitrary constants are in order. We shall use C and C' throughout this paper to denote such constants. Two such constants, which occur in different parts of the paper, need not be equal, or related in any way. Next, a dilation $r \to ar$, or $p \to a'p$, can, in effect, change C or C' to any given value. Therefore the distributions $(r^{-1})_+$, $\omega(p)$ are not positive. We may also set $C = 0$, without loss of generality. This last possibility suggests that we retain a Lorentz-invariant decomposition, $\varphi_\kappa(x) = \varphi_\kappa^{(+)}(x) + \varphi_\kappa^{(-)}(x)$.

The condition of positive definiteness also requires a comment. As was shown by Wightman,[3] this condition requires that the Fourier transform of the two-point function $\langle\varphi(x)\varphi(y)\rangle_0$ be a positive measure. This is fulfilled by $d\Omega_P(p)$, on the set of two-vectors $p \neq 0$. On the other hand, $\omega(p)$ does not define a measure, and moreover, is not positive. Therefore the field φ_κ requires an indefinite metric. We note, however, that this example of indefinite metric is, apparently, rather different from the other familiar examples.[13] In particular, it is not clear how to construct a subsidiary condition.

We shall denote the field restricted to $\mathcal{S}_0$ by φ_P. Thus,

$$\varphi_P(f) = \varphi_\kappa(f)\,|_{\mathcal{H}_P,\,\mathrm{Fock}} \quad \text{if} \quad f \in \mathcal{S}_0.$$

In the right-hand side we have indicated the restriction to the Fock space of φ_P.

Right and Left Representations and the Group Manifold

The representation of the Poincaré group $\mathcal{P}_2$, which is defined by φ_P, reduces. This corresponds to the factorization

$$\Box = -(\partial_0 - \partial_1)(\partial_0 + \partial_1).$$

This also corresponds to the fact that the measure $d\Omega_P(p)$ breaks into two disconnected, Lorentz-invariant parts. These define the *right* and the *left* representation. On the other hand, $\omega(p)$ cannot be so decomposed, as the relations (2.8) show. Therefore the present discussion applies only to φ_P.

The representation becomes irreducible if one adjoins spatial reflections to $\mathcal{P}_2$, but the reducible

[11] The familiar invariant functions for four dimensions should likewise be expressed as unevaluated derivatives, since the differentiation leads to undefined products (e.g., Ref. 9, p. 148 ff.).

[12] Equations (2.2)–(2.8) are based on unpublished work of Professor A. S. Wightman, and are presented here with his permission.

[13] For instance, E. C. G. Sudarshan, Phys. Rev. **123**, 2183 (1961); and H. J. Schnitzer and E. C. G. Sudarshan, *ibid.*, p. 2193.

case is instructive and useful. We therefore decompose φ_P:

$$\varphi_P(x) = \varphi_r(x) + \varphi_l(x) = \varphi_r(x^0 - x^1) + \varphi_l(x^0 + x^1), \quad (2.10)$$

where

$$(\partial_0 + \partial_1)\varphi_r = 0, \qquad (\partial_0 - \partial_1)\varphi_l = 0.$$

We note a few relations:

$$d\Omega_r(p) = d\Omega_P(p)\theta(p^1), \qquad d\Omega_l(p) = d\Omega_P(p)\theta(-p^1), \quad (2.11)$$

$$\langle \underset{l}{\varphi_r}(x)\underset{l}{\varphi_r}(y)\rangle_0 = -(4\pi)^{-1}\log[i(\xi^0 \mp \xi^1) + \epsilon] \sim \frac{1}{4\pi}\int d\underset{l}{\Omega_r}(p)e^{-ip\xi}. \quad (2.12)$$

Further, with the points x and y understood,

$$\langle\varphi_{r,l}\varphi_\kappa\rangle_0 = \langle\varphi_{r,l}\varphi_P\rangle_0 = \langle\varphi_{r,l}\varphi_{r,l}\rangle_0, \qquad \langle\varphi_r\varphi_l\rangle_0 = 0, \quad \langle\varphi_P\varphi_\kappa\rangle_0 = \langle\varphi_P\varphi_P\rangle_0. \quad (2.13a)$$

These relations mean, in particular, that one can use the appropriate measure $d\Omega_{(\cdot)}(p)$ if one of the two test functions belongs to $\mathcal{S}_0$. Similar relations hold for the commutators:

$$[\varphi_r\varphi_\kappa] = [\varphi_r, \varphi_r], \quad \text{etc.} \quad (2.13b)$$

In discussing the fields φ_r and φ_l, it is sometimes convenient to use functions on the manifold of the homogeneous group.[8] This relates to another peculiarity of the two-dimensional case, that the group manifold has the same dimension as the light cone, i.e., one. From Eqs. (2.6)–(2.7) we have

$$\Lambda_\lambda(p^1, \pm p^1) = e^{\pm\lambda}(p^1, \pm p^1).$$

For the right representation, $p^0 = p^1 > 0$, we may therefore set

$$(p^1, p^1) = e^\lambda(p_a^1, p_a^1), \quad \text{or} \quad \lambda = \log(p^1/p_a^1);$$

and, if $\tilde{f}(0) = 0$,

$$\int d\Omega_r(p)\tilde{f}(p) = \int_0^\infty \frac{dp^1}{p^1} f(p^1) = \int d\lambda f(\lambda). \quad (2.14)$$

The Fock Representation

For completeness, we include here the standard formulas for the Fock representation.[7,14] We first consider φ_r, and we use the λ parametrization.[8] We form the direct sum,

$$\mathfrak{H}_{r,\text{Fock}} = \sum{}^{\oplus}\,\mathfrak{H}_r^{(n)},$$

where $\mathfrak{H}_r^{(0)}$ is the space of complex numbers, and where

$$\mathfrak{H}_r^{(n)} = \{f^{(n)}(\lambda_1 \cdots \lambda_n) : \langle f^{(n)}, f^{(n)}\rangle < \infty \text{ and } f(\lambda_1 \cdots \lambda_n) \text{ symmetric}\},$$

$$\langle f^{(n)}, F^{(n)}\rangle = \int d\lambda_1 \cdots d\lambda_n f^{(n)*}(\lambda_1 \cdots \lambda_n)F^{(n)}(\lambda_1 \cdots \lambda_n).$$

Now, for $h \in \mathfrak{H}_r^{(1)}$, $f^{(n)} \in \mathfrak{H}_r^{(n)}$,

$$\varphi_r^{(+)}(h)f^{(n)} = (n+1)^{-\frac{1}{2}}\sum_{i=1}^{n+1} h(\lambda_i) \times f^{(n)}(\lambda_1 \cdots \lambda_{i-1}, \lambda_{i+1} \cdots \lambda_{n+1}) \in \mathfrak{H}_r^{(n+1)},$$

$$\varphi_r^{(-)}(h)f^{(n)} = n^{\frac{1}{2}}\int d\lambda_n\, h(\lambda_n)f^{(n)}(\lambda_1 \cdots \lambda_n) \in \mathfrak{H}_r^{(n-1)}.$$

Note that we can make the identification

$$\varphi(f_{(x)}) = \tilde{\varphi}(\tilde{f}_{(p)}) = \varphi(f_{(p^1)}) = \varphi(f_{(\lambda)}).$$

These formulas can be immediately adapted to momentum space. Next, the case of φ_l is analogous, and for φ_P, the space is a tensor product:

$$\mathfrak{H}_{P,\text{Fock}} = \mathfrak{H}_{r,\text{Fock}} \otimes \mathfrak{H}_{l,\text{Fock}}. \quad (2.15)$$

However, we omit the discussion of the space $\mathfrak{H}_{\kappa,\text{Fock}}$, which necessarily has a more intricate structure.

Derivative and Bilocal Fields

We observe, finally, that the derivative fields $\partial_\mu\varphi_P$ do not require the restriction of test functions to $\mathcal{S}_0$. If we integrate $\partial_\mu\varphi_P$, we obtain a bilocal field, which has a similar property:

$$\int_\eta^x dy\, \partial_\mu\varphi_P(y) = \varphi_P(x) - \varphi_P(\eta).$$

We may express the foregoing observations by writing

$$[\partial_\mu\varphi]_P \equiv \partial_\mu\varphi_\kappa\,|_{\mathfrak{H}_{P,\text{Fock}}}, \qquad [\varphi(x) - \varphi(\eta)]_P \equiv [\varphi_\kappa(x) - \varphi_\kappa(\eta)]\,|_{\mathfrak{H}_{P,\text{Fock}}}. \quad (2.16)$$

These definitions are to be interpreted as follows. The operators which are obtained by smearing right-hand sides with arbitrary test functions in $\mathcal{S}$ (or in $\mathcal{S}\times\mathcal{S}$), are independent of the arbitrary parameters, and leave the space $\mathfrak{H}_{P,\text{Fock}}$ invariant. We may therefore define restricted operators, and also restricted distributions, i.e., the left-hand sides.

In what follows, we shall gain some additional insight from these fields.

3. WICK POWERS AND THE EXPONENTIAL OF THE SCALAR FIELDS

Wick Powers

We are concerned here with the two-point functions of the fields $:\varphi_{(\cdot)}^n:$ and their Fourier transforms:

[14] J. M. Cook, Trans. Am. Math. Soc. **74**, 222 (1953).

$$ {}^{n}\tilde{W}^{(2)}_{(\cdot)}(p) = \frac{1}{2\pi}\int d^2\xi e^{-ip\xi}\langle :\varphi^n_{(\cdot)}(x): :\varphi^n_{(\cdot)}(y):\rangle_0. \quad (3.1) $$

One has the standard relation,

$$ \langle :\varphi^n_{(\cdot)}(x): :\varphi^n_{(\cdot)}(y):\rangle_0 = n!\,\langle \varphi_{(\cdot)}(x)\varphi_{(\cdot)}(y)\rangle^n_0. \quad (3.2) $$

Let us consider the various fields, starting with φ_r. We have for $n = 1$,

$$ \tilde{W}^{(2)}_r(p) \equiv {}^1\tilde{W}^{(2)}_r(p) = \tfrac{1}{2}\delta(p^0 - p^1)\theta(p^1)(p^1)^{-1}. \quad (3.3) $$

Now let $n = 2$. We find

$$ {}^2\tilde{W}^{(2)}_r(p) = \frac{1}{\pi}\int d^2q \tilde{W}^{(2)}_r(q)\tilde{W}^{(2)}_r(p-q) $$

$$ = (4\pi)^{-1}\delta(p^0 - p^1)\theta(p^1)(p^1)^{-1}(\log p^1 - \lim_{\epsilon\to 0+}\log\epsilon). $$

This divergence cannot be eliminated, and we conclude that $:\varphi^2_r:$ cannot be defined.

The case of φ_l is analogous. For φ_P, if $n = 2$, we can require that the test functions vanish on the light cone. Then we obtain the finite relations,

$$:\varphi^2_P(x): = \varphi^2_P(x) = 2\varphi_r(x)\varphi_l(x), \quad (3.4a) $$

$$ {}^2\tilde{W}^{(2)}_P(p) = 4\tilde{W}^{(2)}_r(p) * \tilde{W}^{(2)}_l(p). \quad (3.4b) $$

However, for $n > 2$, we will always encounter two "right" or two "left" factors, and a consistent definition is not possible.

On the other hand, for φ_κ we have

$$ {}^2\tilde{W}^{(2)}_\kappa(p) = 2\omega(p) * \omega(p) = 2\delta(v)[(u^{-1})_+ * (u^{-1})_+] $$

$$ + 2\delta(u)[(v^{-1})_+ * (v^{-1})_+] + 4(u^{-1})_+ * (v^{-1})_+. \quad (3.5) $$

We recall that distributions can be convoluted[15] (provided that their supports are appropriately restricted), so that $\omega(p) * \omega(p)$ is again a well-defined distribution on $\mathcal{S}$. It is worth noting that, in right-hand side of Eq. (3.5), the last term has support in the entire region $p^0 > 0$, $p^2 > 0$, while the first two terms have their support concentrated on the future light cone.

One can also see that further convolutions, and the fields $:\varphi^n_\kappa:$, are all well defined.

The Exponential

The Wick-ordered exponential is defined formally by the power series. However, the foregoing discussion shows that this definition can be applied only to φ_κ:

$$ \Upsilon(x) \equiv :e^{ig\varphi_\kappa(x)}: = \sum (ig)^n(n!)^{-1}\,:\varphi^n_\kappa(x):. \quad (3.6) $$

This conclusion is confirmed by the form of the two-point function of Υ [cf. Eqs. (1.8), (3.2)]:

$$ \langle\Upsilon(x)\Upsilon^*(y)\rangle_0 = \sum g^{2n}(n!)^{-2}\langle :\varphi^n_\kappa(x): :\varphi^n_\kappa(y):\rangle_0 \quad (3.7a) $$

$$ = C(-\xi^2 + i\epsilon\xi^0)^{-g^2/4\pi}. \quad (3.7b) $$

There is no term in right-hand side which is independent of C, while φ_r, e.g., does not imply any arbitrary constants.

It is also instructive to compare momentum space properties of Υ with those of the Wick powers. The exponential leads to the Riesz distribution,[4,16]

$$ \frac{1}{(2\pi)^2}\int d^2\xi\, e^{-ip\xi}(-\xi^2 + i\epsilon\xi^0)^{-g^2/4\pi} = Z^{(2)}_{g^2/2\pi}(p) $$

$$ = 2^{1-(g^2/2\pi)}\Gamma^{-2}(g^2/4\pi)\theta(p^0)\theta(p^2)(p^2)^{(g^2/4\pi)-1}. \quad (3.8a) $$

(We set $C = 1$.) This distribution defines a positive measure,

$$ d\Omega_{\exp}(p) = d^2p(4\pi)Z^{(2)}_{g^2/2\pi}(p), \quad (3.8b) $$

no part of which is concentrated on the light cone. On the other hand, the expansion (3.7a) yields a series of distributions, each of which is partly or entirely concentrated on the light cone.

The Riesz distribution is analytic in g^2 (as a distribution), but ordinarily, $\int d\Omega_{\exp}(p)\tilde{f}(p)$ cannot be expanded as a series in g^2. In particular, in such a case one cannot use the term-by-term integration which is suggested by Eq. (3.7a).

Arithmetical Properties of Υ

The infrared representations in this model are determined by the exponential $\Upsilon(y)$. For convenience we summarize the arithmetical rules which govern this distribution. We first consider the positive and the negative frequency parts of Υ and of Υ^*:

$$ \Upsilon(y) = \exp[ig\varphi^{(+)}_\kappa(y)]\exp[ig\varphi^{(-)}_\kappa(y)] $$

$$ \equiv \Upsilon^{(+)}(y)\Upsilon^{(-)}(y), \quad (3.9a) $$

$$ \Upsilon^*(y) = :e^{-ig\varphi_\kappa(y)}: = \Upsilon^{*(+)}(y)\Upsilon^{*(-)}(y), \quad (3.9b) $$

$$ \Upsilon^{(-)}(x)\Upsilon^{(+)}(y) $$

$$ = \Upsilon^{(+)}(y)\Upsilon^{(-)}(x)(-\xi^2 + i\epsilon\xi^0)^{+g^2/4\pi}, \quad (3.10a) $$

$$ \Upsilon^{(-)}(x)\Upsilon^{*(+)}(y) $$

$$ = \Upsilon^{*(+)}(y)\Upsilon^{(-)}(x)(-\xi^2 + i\epsilon\xi^0)^{-g^2/4\pi}, \quad (3.10b) $$

$$ \Upsilon(x)\Upsilon^*(y)|_{\mathfrak{X}_P,\,\mathrm{Fock}} $$

$$ = :e^{ig[\varphi(x)-\varphi(y)]_P}:(-\xi^2 + i\epsilon\xi^0)^{-g^2/4\pi}. \quad (3.10c) $$

We set the arbitrary constant $C = 1$. These equations can be deduced formally from the familiar

[15] L. Schwartz, *Théorie des distributions* (Hermann & Cie., Paris, 1957–1959), 2nd ed., Vol. II, p. 102.

[16] L. Schwartz, Ref. 15, Vol. I, p. 49; Vol. II, p. 119; and references given there.

rule for operators A, B, if $[A, B]$ is a c number:

$$e^A e^B = e^B e^A e^{[A,B]} = e^{A+B} e^{\frac{1}{2}[A,B]},$$

but a derivation by power series is more satisfactory, as we have seen.

In particular, in Eq. (3.10c) we assert that the distributions $\Upsilon(x)\Upsilon^*(y)$, when smeared out, leave the space $\mathfrak{H}_{P,\text{Fock}}$ invariant. To see this, we consider (cf. the right-hand side)

$$:[\varphi_\kappa(x) - \varphi_\kappa(y)]^n: .$$

But one can show with the help of the approximation theorem,[17] that such distributions (upon smearing) leave $\mathfrak{H}_{P,\text{Fock}}$ invariant. Our assertion follows.

Further, $[\varphi_r, \varphi_\kappa] = [\varphi_r, \varphi_r]$, etc. [Eqs. (2.13)], and therefore

$$[\varphi_r^{(\pm)}(x), \Upsilon^{(\mp)}(y)] = gD_r^{(\pm)}(x - y)\Upsilon^{(\mp)}(y), \tag{3.11a}$$

$$[\varphi_r(x), \Upsilon(y)] = gD_r(x - y)\Upsilon(y); \tag{3.11b}$$

for Υ^*, replace g by $-g$. Also,

$$[\varphi_r^{(\pm)}(x), \varphi_r^{(\mp)}(y)] = i^{-1}D_r^{(\pm)}(x - y),$$
$$D_r = D_r^{(+)} + D_r^{(-)}.$$

Equations (3.11) remain valid if φ_r is replaced by one of the other φ's, provided the corresponding change of D functions is also made.

4. THE INFRARED REPRESENTATIONS

A Few Immediate Properties

The representations in question correspond to the action of the fields $\varphi_{(\cdot)}$ on vectors of the following kind as cyclic vectors ($\Upsilon^{(\cdot)} = \Upsilon$ or Υ^*):

$$:\Upsilon^{(\cdot)}(f_1) \cdots \Upsilon^{(\cdot)}(f_m): |0\rangle.$$

One can verify that such a vector indeed defines a cylic representation.[18] In case of the fields φ_r, φ_l, and φ_P, positive definiteness of the metric can be established. See remarks in the sequel.

The presence of the spinors $\psi^{(0)}$ and $\bar{\psi}^{(0)}$ in the integrands would not alter the esentials of the present discussion. In particular, the conclusion that different functions f_i yield equivalent representations (see below), would not be altered.

We shall present our results for the field φ_r, but they can be applied directly to the other φ's. For simplicity, we confine ourselves to the case of one exponential Υ. Now, if we consider vector-valued distributions rather than vectors, e.g.,

$$\Upsilon(y)\ |0\rangle \equiv v_y, \tag{4.1}$$

then the representation is described already by the following equation:

$$\varphi_r(x)\ :\varphi_r(x_1) \cdots \varphi_r(x_n)\Upsilon(y):$$
$$= :\varphi_r(x)\varphi_r(x_1) \cdots \varphi_r(x_n)\Upsilon(y):$$
$$- gD_r^{(+)}(x - y)\ :\varphi_r(x_1) \cdots \varphi_r(x_n)\Upsilon(y):$$
$$+ i \sum_{i=1}^{n} D_r^{(+)}(x - x_i)\ :\varphi_r(x_1) \cdots$$
$$\times \hat{\varphi}_r(x_i) \cdots \varphi_r(x_n)\Upsilon(y):, \tag{4.2}$$

where the caret denotes a quantity to be omitted.

Equation (4.2) suggests that we should write, in analogy with the Fock representation,

$$\mathfrak{H}_{r,1} = \sum \mathfrak{H}_{r,1}^{(n)},$$

where $\mathfrak{H}_{r,1}^{(n)}$ is spanned by vectors of the form

$$v(h_1, \cdots, h_n; f) \equiv :\varphi_r(h_1) \cdots \varphi_r(h_n)\Upsilon(f): |0\rangle. \tag{4.3}$$

The function f is now allowed to vary over $\mathcal{S}$. (We do not know whether the representation is still cyclic.) The subscript 1 following $\mathfrak{H}$ indicates one exponential, with positive charge. In contrast to the Fock representation, the subspaces $\mathfrak{H}_{r,1}^{(n)}$ are not orthogonal. Furthermore, we have here

$$\varphi_r^{(+)}(h): \mathfrak{H}_{r,1}^{(n)} \to \mathfrak{H}_{r,1}^{(n+1)}, \tag{4.4a}$$

$$\varphi_r^{(-)}(h): \mathfrak{H}_{r,1}^{(n)} \to \mathfrak{H}_{r,1}^{(n)} + \mathfrak{H}_{r,1}^{(n-1)}. \tag{4.4b}$$

We shall not consider the reducibility of the representations here defined.

Construction of the Representations

We shall now attempt to find formulas to describe the action of smeared-out operators, in analogy to the formulas for the Fock representation in Sec. 2.

Our solution to this problem is only partly successful. The following construction gives the norms of elementary vectors (4.3), but does not apply to their linear combinations.

In particular, this construction does not establish the positive definiteness of the metric in spaces like $\mathfrak{H}_{r,1}$. We note (without proof) that positive definiteness for such spaces can be established as follows. We give the field φ_P a small mass μ. Then positive definiteness is implied by the general theory of the free field. The positive definiteness will be manifestly preserved if one carefully takes the limit $\mu \to 0$.

Let us now look at the norm of the vector $:\varphi_r(h)\Upsilon(f): |0\rangle$. We note the relations

$$\langle\varphi_r(H)\varphi_r(h)\rangle_0 = \pi \int d\Omega_r(p)\tilde{H}(p)\bar{h}(-p), \tag{4.5a}$$

$$\langle\Upsilon^*(F)\Upsilon(f)\rangle_0 = \langle\Upsilon(F)\Upsilon^*(f)\rangle_0$$
$$= \pi \int d\Omega_{\text{exp}}(p)\tilde{F}(p)\bar{f}(-p), \tag{4.5b}$$

and if

[17] L. Schwartz, Ref. 15, Vol. I, p. 108.
[18] H. Araki, J. Math. Phys. 1, 492 (1960).

$$K(y) = F(y)\int d^2x\, H(x)D_r^{(-)}(x-y),$$

then

$$\tilde{K}(p) = \tfrac{1}{2}i\int d\Omega_r(q)\tilde{F}(p-q)\tilde{H}(q)$$
$$\equiv \tfrac{1}{2}i(\tilde{F}\ast_r\tilde{H})(p). \tag{4.6}$$

Now we find

$$||:\varphi_r(h)\Upsilon(f):|0\rangle||^2 = \langle:\Upsilon^*(f^*)\varphi_r(h^*):\varphi_r(h)\Upsilon(f):\rangle_0$$
$$= \pi^2\int d\Omega_{\text{exp}}(\zeta)\,d\Omega_r(\eta)\,|\tilde{f}(-\zeta)|^2\,|\tilde{h}(-\eta)|^2$$
$$+ \tfrac{1}{4}\pi g^2\int d\Omega_{\text{exp}}(\zeta)\,|[\tilde{f}\ast_r(h^*)^{\sim}](-\zeta)|^2. \tag{4.7}$$

We therefore associate

$$:\varphi_r(h)\Upsilon(f):|0\rangle \leftrightarrow (\tilde{f}(-\zeta)\tilde{h}(-\eta),\ \tfrac{1}{2}ig[\tilde{f}\ast_r(h^*)^{\sim}](-\zeta)).$$

For the general vector (4.3), we need the symmetrized products

$$(\tilde{h}_1\circ\cdots\circ\tilde{h}_m)(-\eta_1,\cdots,-\eta_m)$$
$$\equiv (m!)^{-\frac{1}{2}}\sum_{\sigma\in S_m}\tilde{h}_{\sigma(1)}(-\eta_1)\cdots\tilde{h}_{\sigma(m)}(-\eta_m). \tag{4.8}$$

We now define the correspondence

$$v(h_1,\cdots,h_n;f)\leftrightarrow(\tilde{f}(-\zeta)(\tilde{h}_1\circ\cdots\circ\tilde{h}_n)(-\eta_1,\cdots,-\eta_n),$$
$$\tfrac{1}{2}ig\sum_i[\tilde{f}\ast_r(h_i^*)^{\sim}](-\zeta)(\tilde{h}_1\circ\cdots\hat{\tilde{h}_i}\circ\cdots\tilde{h}_n)(-\eta_1,\cdots,-\eta_{n-1}),$$
$$(\tfrac{1}{2}ig)^2\sum_{i<k}[\tilde{f}\ast_r(h_i^*)^{\sim}\ast_r(h_k^*)^{\sim}](-\zeta)(\tilde{h}_1\circ\cdots\hat{\tilde{h}_i}\circ\cdots\hat{\tilde{h}_k}\circ\cdots\tilde{h}_n)(-\eta_1,\cdots,-\eta_{n-2}),\cdots,$$
$$(\tfrac{1}{2}ig)^n[\tilde{f}\ast_r(h_1^*)^{\sim}\ast_r\cdots(h_n^*)^{\sim}](-\zeta)). \tag{4.9}$$

A multiple convolution has the following meaning,

$$(\tilde{f}\ast_r\tilde{H}_1\ast_r\cdots\tilde{H}_m)$$
$$\equiv((\cdots(\tilde{f}\ast_r\tilde{H}_1)\ast_r\cdots)\ast_r\tilde{H}_m), \tag{4.10}$$

and it is symmetric in the $\tilde{H}_j$. The norm of the vector (4.9) is in complete analogy with Eq. (4.7).

One can go further and describe the action of $\varphi_r^{(\pm)}(h_0)$. For $\varphi_r^{(+)}(h_0)$, it suffices to observe that

$$\varphi_r^{(+)}(h_0)v(h_1,\cdots,h_n;f)$$
$$= v(h_0,h_1,\cdots,h_n;f). \tag{4.11a}$$

For $\varphi_r^{(-)}(h_0)$, we have contributions in $\mathfrak{C}_{r,1}^{(n)}$ and in $\mathfrak{C}_{r,1}^{(n-1)}$:

$$\varphi_r^{(-)}(h_0)v(h_1,\cdots,h_n;f)|^{(n)}$$
$$= \tfrac{1}{2}igv(h_1,\cdots,h_n;(\tilde{f}\ast_r\tilde{h}_0)^{\sim}), \tag{4.11b}$$

$$\varphi_r^{(-)}(h_0)v(h_1,\cdots,h_n;f)|^{(n-1)}$$
$$\leftrightarrow\Big(n^{\frac{1}{2}}\tilde{f}(-\zeta)\int d\Omega_r(\eta_n)\tilde{h}_0(\eta_n)$$
$$\times(\tilde{h}_1\circ\cdots\circ\tilde{h}_n)(-\eta_1,\cdots,-\eta_n),\cdots,(\tfrac{1}{2}ig)^{n-1}$$
$$\times\sum_i[\tilde{f}\ast_r(h_1^*)^{\sim}\ast_r\cdots(h_i^*)^{\sim\hat{}}\ast_r\cdots(h_n^*)^{\sim}](-\zeta)$$
$$\times\int d\Omega_r(\eta_1)\tilde{h}_0(\eta_1)\tilde{h}_i(-\eta_1)\Big). \tag{4.11c}$$

We note that for $\tilde{h}_i(\eta)$ only the values on the line $\eta^0=\eta^1$ are relevant, and thus the $\tilde{h}_i(\eta)$ can all be replaced by $h_i(\lambda)$, except in convolution.

Equivalence of Representations

Let us consider for definiteness that representation of φ_r, which is defined by $\Upsilon(f)\,|0\rangle\equiv v_f$ as the cyclic vector. We shall show (for functions f restricted to a dense subset), that there exist operators V_f which leave the cyclic space invariant, and which satisfy

$$A_r^{(-)}(p^1)V_fv_f = 0, \tag{4.12a}$$

$$A_r^{(\pm)}(p^1)\equiv\varphi_r^{(\pm)}(p^1)\pm ig(4\pi p^1)^{-\frac{1}{2}}F(p^1)\theta(p^1) \tag{4.12b}$$

[see Eq. (4.16)]. Here $F(p^1)$ can be any function differentiable at zero and such that

$$F(0)=1 \quad\text{and}\quad \int_0^\infty dp^1\,|F(p^1)|^2<\infty. \tag{4.12c, d}$$

It will follow that the distributions $A_r^{(\pm)}(p^1)$ are associated with a Fock representation, which, in turn, determines the representation of φ_r. The latter representation is therefore independent of the function f. Now, other factors $\Upsilon^{(\cdot)}$ would give similar additive contributions proportional to g, and our previous assertion about equivalence will be established. (But recall the remark in the Introduction about the shortcomings of our analysis.)

As we have already noted, we will make use of the analysis of scattering by a fixed source. We recall that the distributions $b^{(\pm)}(\mathbf{k})$ which arise in this problem are obtained from the free-field distributions $a^{(\pm)}(\mathbf{k})$ by a translation[6,19]

$$b^{(\pm)}(\mathbf{k}) = a^{(\pm)}(\mathbf{k}) - q^{(\pm)}(\mathbf{k}),$$

where $q^{(\pm)}(\mathbf{k})$ are complex-conjugate c-number functions. Let $a^{(\pm)}(\mathbf{k})$ be normalized so that

$$[a^{(-)}(\mathbf{k}),a^{(+)}(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}').$$

[19] See also H. J. Borchers, R. Haag, and B. Schroer, Nuovo Cimento 29, 148 (1963).

Then, this translation can be generated by a unitary transformation if and only if

$$I(q) \equiv \int d^3\mathbf{k}\ |q^{(\pm)}(\mathbf{k})|^2 < \infty. \tag{4.13}$$

In this case, $b^{(\pm)}(\mathbf{k}) = U_q^{-1} a^{(\pm)}(\mathbf{k}) U_q$, where

$$U_q = \exp\,(q^{(+)}, -\tfrac{1}{2}q^{(-)}) \times \exp\,(a^{(+)}, -q^{(-)})\ \exp\,(q^{(+)}, a^{(-)}), \tag{4.14a}$$

$$(B^{(+)}, C^{(-)}) = \int d^3\mathbf{k} B^{(+)}(\mathbf{k}) C^{(-)}(\mathbf{k}). \tag{4.14b}$$

On the other hand, infrared representations arise if $I(q) = \infty$ but

$$\int d^3\mathbf{k}\ |\mathbf{k}|\ |q^{(\pm)}(\mathbf{k})|^2 < \infty. \tag{4.15}$$

This condition guarantees the finiteness of energy.

Let us adapt the foregoing relations to the present problem. We write

$$\varphi_r(x) = \int_0^\infty \frac{dp^1}{(4\pi p^1)^{\frac{1}{2}}} [\varphi_r^{(+)}(p^1)e^{ipx} + \varphi_r^{(-)}(p^1)e^{-ipx}], \tag{4.16}$$

so that

$$[\varphi_r^{(-)}(p^1), \varphi_r^{(+)}(p'^1)] = \delta(p^1 - p'^1)\theta(p^1).$$

These equations were presupposed in Eq. (4.12b). We see that $\int_0 dp^1\ |(p^1)^{-\frac{1}{2}}|^2 = \infty$, so that the translation in Eq. (4.12b) is not a unitary transformation. Moreover, the condition analogous to (4.15) is fulfilled, by virtue of (4.12d).

We shall find it convenient to use distributions, rather than (smeared-out) operators. In the case of the latter, the criteria for equivalence are of course analogous.[20] Vector-valued distributions will also be useful.

Now we find

$$\varphi_r^{(-)}(p^1)v_\nu = ige^{ip\nu}(4\pi p^1)^{-\frac{1}{2}}\theta(p^1)v_\nu, \tag{4.17}$$

$$A_r^{(-)}(p^1)v_\nu = q_\nu^{(-)}(p^1)v_\nu, \tag{4.18}$$

where

$$q_\nu^{(\pm)}(p^1) \equiv \mp ig[e^{\mp ip\nu} - F(p^1)](4\pi p^1)^{-\frac{1}{2}}\theta(p^1). \tag{4.19}$$

The integral $\int dp^1\ |q_\nu^{(\pm)}|^2$ is finite at $p^1 = 0$, but diverges at infinity. This ultraviolet divergence expresses the singular nature of point charges, in analogy with the case of scattering by fixed point sources.[21]

[20] I. E. Segal, Trans. Am. Math. Soc. 88, 12 (1958).
[21] L. Van Hove, Physica 18, 145 (1952).

This divergence should disappear for normalizable states. However, the y dependence of $q_\nu^{(\pm)}$ then becomes a complication. One way to proceed is by a detour to the complex plane, by letting $y \to y + i\eta$.

We therefore consider state vectors of the form

$$:\varphi_r(h_1) \cdots \varphi_r(h_n)\Upsilon^{(+)}(y + i\eta):|0\rangle, \tag{4.20}$$

where η is future timelike. We observe that $\varphi_\kappa^{(+)}(y+i\eta)$ can be expressed as the (convergent) integral

$$\varphi_\kappa^{(+)}(y + i\eta) = \frac{1}{2\pi}\int_{p^0\geq 0} d^2p\tilde{\varphi}_\kappa^{(+)}(p)e^{ip\nu - p\eta}, \tag{4.21}$$

and that

$$||v_{\nu,\eta}||^2 \equiv \langle \Upsilon^{*(-)}(y - i\eta)\Upsilon^{(+)}(y + i\eta)\rangle_0 < \infty, \tag{4.22}$$

etc. Therefore the vectors (4.20) constitute a cyclic subspace of $\mathfrak{K}_{r,1}$. But for this subspace, Eqs. (4.18)–(4.19) already show that the representation is in agreement with the relations (4.12). Note that $q_{\nu,\eta}^{(\pm)}$ should both be taken with a decreasing exponential, and so are square integrable.

The cyclic representation generated by v_f will also be in agreement with (4.12), if v_f can be expressed as an integral

$$v_f = \int d^2y\ H(y)v_{\nu,\eta}. \tag{4.23}$$

This is possible if f is exponentially decreasing in momentum space. But such functions are dense in e.g., $\mathcal{S}$, and the equivalence of representations follows.

It may be worthwhile to give some explicit formulas. By using the analogue of Eq. (4.14a) and by interchanging $A_r^{(-)}$ and $q_{\nu,\eta}^{(-)}$ in the exponents, cf. Eq. (4.18), one can show that

$$U_{q_{\nu,\eta}}v_{\nu,\eta} = :\exp\,(A_r^{(+)}, -A_r^{(-)}): \times \exp\,(q_{\nu,\eta}^{(+)}, \tfrac{1}{2}q_{\nu,\eta}^{(-)})v_{\nu,\eta}, \tag{4.24a}$$

$$A_r^{(-)}(p^1)\ :\exp\,(A_r^{(+)}, -A_r^{(-)}):v_f = 0. \tag{4.24b}$$

This is valid for v_f given by (4.23). We see an operator V_f satisfying Eq. (4.12a), and others can be easily constructed. For example,

$$V_f = :\exp\,(A_r^{(+)}, -A_r^{(-)}): A_r^{(-)}(p^1). \tag{4.25}$$

5. SOME PROPERTIES OF THE MODEL

A summary of the model was included in the introduction, and here we elaborate on certain details. We observe that, unlike the case of the φ's, the familiar properties of free fields remain valid for $\psi^{(0)}$

Remarks on the Field Equations

In Eqs. (1.1) we were quite vague about the renormalization terms. The equation for the renormalized spinor is

$$(\partial + M)\psi(x) = ig\,{:}[\partial\varphi(x)]\psi(x){:}\,. \qquad (5.1)$$

This equation becomes meaningful when it is established that the positive and the negative frequency parts of ψ can be separated. This property of course follows from the solution $\psi^{(0)}\,{:}e^{i\varphi\varphi_\kappa}{:}$.

As we remarked earlier, there should be no scattering in this model. This is suggested by previous studies of similar couplings,[22,23] and by a consideration of Borchers' classes.[24]

The unphysical nature of the field φ_κ, and of the induced indefinite metric, is indicated by the fact, that this field does not really occur in the field equations. While we obtain from $\partial_\mu\psi$ a term proportional to $\partial_\mu\varphi_\kappa$, the latter can be replaced by $[\partial_\mu\varphi]_P$ [Eq. (2.16)]. Moreover, the arithmetical properties of Υ, cf. Sec. 3, do not force one to consider φ_κ at any stage.

One therefore can expect that a physical state, which is defined in terms of the fields φ_P and ψ, cannot make a transition into a ghost state, which would be defined in terms of φ_κ and ψ. Therefore the lack of a subsidiary condition, to eliminate the ghost states of φ_κ, is not a serious deficiency. One can restrict the theory to physical states by an initial condition at a given time, or as $t \to -\infty$.

Hilbert Space and the Action of φ_P

Before discussing the Hilbert space $\mathfrak{H}^{\text{phys}}$ of physical states, two preliminary remarks have to be made. First: The vector $\psi(f)\,|0\rangle$ lies in the tensor product of Hilbert spaces, consisting of vectors $\Upsilon(h_1)\,|0\rangle$ and $\psi^{(0)}(h_2)\,|0\rangle$, respectively. In particular, the representation of φ_P, which is determined by an operator $\psi(f)$, is equivalent to that, which is determined by operators like $\Upsilon(h)$. Second: We can always put all φ's to the left of the ψ's. In this way, the action of φ_P will correspond to the total charge of the given state.

Let us first consider an enlarged space $\mathfrak{H}^*$, where $\psi^{(0)}$ and Υ are not assumed to be related in space–time, so that we can use the construction of Sec. 4. We see that it is convenient to separate the Fock space $\mathfrak{H}_{\text{Fock}}(\psi^{(0)})$ of $\psi^{(0)}$ and $\bar\psi^{(0)}$ into sectors of the same charge. Let $\mathfrak{H}_{\psi^{(0)}}^{[k,l]}$ denote the subspace of states having k $\psi^{(0)}$-quanta, and l $\bar\psi^{(0)}$-quanta. We assert that

$$\begin{aligned}\mathfrak{H}^* = {}& \mathfrak{H}_{P,\text{Fock}} \otimes (1 + \mathfrak{H}_{\psi^{(0)}}^{[1,1]} + \mathfrak{H}_{\psi^{(0)}}^{[2,2]} + \cdots) \\ & + \mathfrak{H}_{P,1} \otimes (\mathfrak{H}_{\psi^{(0)}}^{[1,0]} + \mathfrak{H}_{\psi^{(0)}}^{[2,1]} + \cdots) \\ & + \mathfrak{H}_{P,-1} \otimes (\mathfrak{H}_{\psi^{(0)}}^{[0,1]} + \cdots) + \cdots . \qquad (5.2)\end{aligned}$$

The action of φ_P is here clearly indicated. It is straightforward, at least in principle, to describe the subspace $\mathfrak{H}^{\text{phys}}$ of $\mathfrak{H}^*$, but we forego formal construction.

The construction (5.2) follows from the fact that the operators, which are generated by the distributions in Eqs. (2.16) and (3.10c), leave the spaces $\mathfrak{H}_{P,N}$ invariant. (In particular, the space $\mathfrak{H}_{P,N}$ is determined only by the net total charge N, and not by the number of exponentials.) This is a trivial generalization of the case $N = 0$ considered previously. On the other hand, we have no corresponding statement for spaces like $\mathfrak{H}_{r,1}$ or $\mathfrak{H}_{l,1}$. Therefore, if the free bosons were to correspond to φ_r or to φ_l, the construction would be less simple.

If we were to consider φ_κ rather than φ_P as the field which describes free bosons of the theory, then the space of states would be $\mathfrak{H}_{\kappa,\text{Fock}} \otimes \mathfrak{H}_{\text{Fock}}(\psi^{(0)})$. This is the space of all normalizable states in the theory. Therefore

$$\mathfrak{H}^{\text{phys}} \subset \mathfrak{H}^* \subset \mathfrak{H}_{\kappa,\text{Fock}} \otimes \mathfrak{H}_{\text{Fock}}(\psi^{(0)}) \equiv \mathfrak{H}^{\text{Max}}. \qquad (5.3)$$

The space $\mathfrak{H}^{\text{Max}}$ has states of negative norm, but the two subspaces have a positive definite metric.

Bilocal Solution for ψ

If we do not insist on local fields, then we can construct a bilocal solution to the field equations,

$$\psi_\eta(x) = \psi^{(0)}(x)\,{:}e^{ig[\varphi(x)-\varphi(\eta)]P}{:}\,. \qquad (5.4)$$

Nonlocal solutions of this general form have in fact been considered for other theories.[25,26] This example should serve as a warning, that they may have rather different properties from local solutions.

One difference between ψ and ψ_η, besides that of locality, lies in the representation of φ_P which each defines, i.e., an infrared and the Fock representation, respectively. We can also verify directly that the field ψ_η has a state of mass M, in contrast[4] to ψ. To see this, one can check that the integral

$$\int d^2x\, e^{-ipx}\langle{:}e^{ig[\varphi(x)-\varphi(\eta)]P}{:}\,{:}e^{-ig[\varphi(0)-\varphi(\xi)]P}{:}\rangle_0$$

has a term proportional to $\delta(p)$. But we find that for large $|x^2|$, the vacuum expectation value becomes

[22] F. J. Dyson, Phys. Rev. **73**, 929 (1948).
[23] S. Okubo, Progr. Theoret. Phys. (Kyoto) **11**, 80 (1954); Nuovo Cimento **19**, 574 (1961).
[24] H. J. Borchers, Nuovo Cimento **15**, 784 (1960).
[25] F. L. Scarf and J. Wess, Nuovo Cimento **26**, 150 (1962).
[26] S. Mandelstam, Ann. Phys. (New York) **19**, 1 (1962).

independent of x and nonzero. The conclusion follows.

We are ignoring the question as to how the η dependence should be treated in constructing the states of ψ_η, e.g., by smearing or by allowing distributions in η. But the foregoing conclusions should be valid in either case.

6. CONCLUSION

The present work can be described as an application of elementary field-theoretic methods to a particularly simple model theory, in which fields are coupled. The operator solution to this theory was examined, and its meaning was made quite precise. Physically more interesting problems, such as asymptotic convergence, were not considered. However, this work can be regarded as laying the groundwork for further study.

The conclusion which appears the most interesting is the relation (5.3). It is tempting to speculate that in more realistic theories, where divergences and myriotic representations arise, a similar approach might be fruitful; i.e., it might be possible to enlarge the definition of the field, and then to identify the physical space, as a subset of the enlarged Fock space. In this way one could try to gain an orientation into the myriotic representations, which, most likely, abound in any realistic field theory. (Such an imbedding of spaces of course is not novel.[21] But in our case, the enlarged space has a natural Fock structure and a natural metric.)

An approach of this kind may perhaps be applicable to both infrared and ultraviolet divergences. In fact, in two-dimensional theories there is much similarity between them.[27] The constants C and C', which were introduced at various places, originate as infinities, in integrating $|p^1|^{-1}$ at $p^1 = 0$. Only through an artifice they were given finite, albeit arbitrary values. This is characteristic of renormalization in the ultraviolet, rather than of, e.g., soft photon effects. It might be appropriate to say, to stress this point, that this model requires an infrared renormalization.

[27] This similarity has been emphasized by H. J. Borchers (private communication).

We make two further observations, which are of more limited interest.

One of these concerns the equivalence of infrared representations for the same charge. We have already mentioned the analogy in this respect between Schroer's model and the fixed-source problem in four dimensions.[6] What is surprising about this is that the two models deal with different dimensions and in each case the conclusion is valid only for the dimension in question. Indeed, an exponential such as in Schroer's model leads to unrenormalizable theories in higher dimensions, while for the scattering problem, the crucial quantity is the integral

$$\int d^n\mathbf{k}\ |\mathbf{k}|^{-3}\ |\tilde{\jmath}(\mathbf{k})|^2.$$

Only for $n = 3$ is the part which diverges at $\mathbf{k} = \mathbf{0}$, proportional to $|\tilde{\jmath}(\mathbf{0})|^2$.

The second observation reflects on the remark about interchange of limiting operations in Sec. 3. There, the nonvalidity of interchange occurs together with T, and the latter leads to infrared representations. For contrast, the Fock exponential $:e^{ig[\varphi(x)-\varphi(\eta)]_P}:$ apparently allows such an interchange (Sec. 5). It seems that also in other instances, the nonvalidity of interchange of limiting operatoins can be related to the occurrence of inequivalent representations of fields. For example, in many-body problems, some of the difficulties with the limit $V \to \infty$ may well be of this kind.[28]

ACKNOWLEDGMENTS

I would like to express my appreciation to Professor A. S. Wightman for suggesting this problem, and for a number of very helpful discussions. I have also benefited from discussions with Dr. H. J. Borchers, Mr. A. M. Jaffe, Dr. R. F. Streater, and Professor K. Symanzik. I would like to express my thanks to Professor B. Zumino and to the Courant Institute for hospitality. Some of this work was carried out at the Summer Institute at the University of Wisconsin, and I thank Professor R. G. Sachs for hospitality there.

[28] H. Araki and E. J. Woods, J. Math. Phys. 4, 637 (1963).

On the Representations of the Semisimple Lie Groups. III. The Explicit Conjugation Operation for SU_n*†

G. E. Baird‡ and L. C. Biedenharn
Department of Physics, Duke University, Durham, North Carolina
(Received 21 April 1964; final manuscript received 24 July 1964)

The analog of the SU_2 $(1-j)$ symbol is defined and discussed in detail for SU_n. An appropriate generalization to SU_n of the Condon–Shortley phase convention is explicitly given.

I. INTRODUCTION AND SUMMARY

THE present series of papers[1] is concerned with the explicit constructive determination of the representation of the semisimple Lie groups by an extension of the Racah–Wigner angular momentum calculus developed for the group SU_2. We have discussed in I the physical motivation underlying this program.

As discussed earlier, the program to be followed has been laid out in detail by the work of Wigner[2] and Racah[3] and consists of essentially three problems: (a) the determination of invariant operators ("Casimir invariants") that uniquely specify the irreducible representations, (b) the determination of sufficient "labeling operators" to uniquely specify the states of an irreducible representation, and (c) the determination of explicit Wigner coefficients by solution of the problem of simple reducibility.[4,5] In I, a more complete discussion of these problems was given and solution for general Casimir operators (I_n) *for the unitary groups* was constructed.[6,7] In II, a solution of the labeling for the unitary groups—based on Weyl's branching theorem—was discussed; using these results, an explicit determination of the matrices of the generators of the unitary groups SU_n was given.

The problem of determining explicit Wigner coefficients involves the ancillary problem of determining the SU_n analog of the particular Wigner coefficient known as the $(1-j)$ symbol (for the SU_2 group). Alternatively this determination of the "$SU_n(1-j)$ symbol" may be viewed as a discussion of the *conjugation operation for* SU_n. Phase conventions naturally play an important part in this discussion.

Because our motivation is primarily toward physical applications, we have attempted in the present paper to discuss the conjugation operation in complete and explicit detail.[8] Let us note that our SU_n phase conventions reduce to those of Condon and Shortley for $n = 2$.

An appendix details an interesting algebraic treatment of the SU_n generators using a method suggested by Fano.

In the paper immediately following, the conjuga-

* Supported in part by the U. S. Army Research Office (Durham) and the National Science Foundation.

† This paper is excerpted in part from the Ph.D. thesis of G. E. Baird (submitted to Duke University, April 1964).

‡ Present address: Northrop Space Laboratories, Huntsville, Alabama.

[1] L. C. Biedenharn, J. Math. Phys. 4, 436 (1963); G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963); throughout we refer to these works as I and II, respectively.

[2] E. P. Wigner, lecture notes, Princeton University, Princeton, New Jersey, 1955 (unpublished); "On the Matrices which Reduce the Kronecker Product of Representations of Simply Reducible Groups" in *Selected Papers on the Quantum Theory of Angular Momentum* (Academic Press Inc., New York, to be published).

[3] G. Racah, lecture notes, The Institute for Advanced Study, Princeton, New Jersey, 1951 (unpublished).

[4] E. P. Wigner, Am. J. Math. **63**, 57 (1941).

[5] Actually, problems (b) and (c) are not distinct, for in both the problem is to uniquely specify the states of the irreducible representations of a subgroup contained in the uniquely labeled states of the irreducible representation of the larger group. Conditions under which such a problem is soluble have been discussed by E. P. Wigner, "On the Condition that the Irreducible Representation of a Group, Considered as Representations of a Subgroup, Do Not Contain any Representation of the Subgroup More than Once" (unpublished).

[6] References to the literature on this problem will be found in I and II. We should like to call attention to an earlier solution of the problem (in very different terms, however!) by I. M. Gelfand in his paper entitled "The Center of an Infinitesimal Group Ring" [Mat. Sb. **26**, 103 (1950)]. We are indebted to Dr. J. P. Louck not only for calling this (and other of Gelfand's papers) to our attention, but for the courtesy of supplying translations as well.

We would also like to call attention to the work of M. Moshinsky [J. Math. Phys. **4**, 1128 (1963) and references cited there] which derives many of the results contained in II, Sec. II.

[7] For other recent work on this problem see: M. Umezawa, Nucl. Phys. 57, 65 (1964); M. Micu, "Construction of Invariants for Simple Lie Groups" (preprint).

[8] There is a precedent in the literature for this course: as noted by Condon and Shortley in their introduction of the standard (Condon and Shortley) phase some three decades ago, "it has been too little recognized that a matrix is not fully useful unless the relative phases of the states to which the components refer are in some way specified."

tion operation is applied to the primary task of the determination of the SU_n Wigner coefficients.

II. RÉSUMÉ OF THE BASIS STATES

In the sections to follow we make continual use of the Gelfand basis vectors for the irreducible representations of SU_n. Although these basis vectors were discussed in II, for the convenience of the reader, and for clarity, we briefly discuss their properties again.

Any irreducible representation of SU_n may be defined by its *Young symmetry pattern* which is characterized by the partition $(m_{1,n}m_{2,n}m_{3,n} \cdots m_{n-1,n}m_{n,n} = 0)$, $m_{n,n}$ being zero because of the unimodular restriction.

To obtain the *states* of an irreducible representation, one considers all possible *lexical Young tableaux*, which are obtained by filling the *Young pattern* with the integers one to n with repetitions but in lexical order. These states are designated as Weyl basis states. This general Weyl basis tableau may be written in the form

First row: m_{11} 1's, followed by
$(m_{12} - m_{11})$ 2's, $\cdots$, $(m_{1n} - m_{1,n-1})$ n's;
Second row: m_{22} 2's, followed by (1)
$(m_{23} - m_{22})$ 3's, $\cdots$, $(m_{2,n} - m_{2,n-1})$ n's;
kth row: $m_{k,k}$ k's, followed by
$(m_{k,k+1} - m_{k,k})(k+1)$'s, $\cdots$, $(m_{k,n} - m_{k,n-1})$ n's,
where the $m_{i,j}$ are integers.

We associate the Gelfand basis state (m) with this Weyl basis tableau, whereby, by (m), we denote the triangular pattern

$$(m) \equiv \begin{bmatrix} m_{1,n} & m_{2,n} & \cdots & m_{n-1,n} & m_{n,n} \\ & m_{1,n-1} & \cdots & & m_{n-1,n-1} \\ & & \cdots & & \\ & & & m_{11} & \end{bmatrix}. \quad (2)$$

The requirement that the Weyl basis tableau be *lexical* leads to the condition on the $m_{i,j}$ that

$$m_{i,j+1} \geq m_{i,j} \geq m_{i+1,j+1}. \quad (3)$$

In terms of the triangular pattern the $m_{i,j}$'s lie *between* the integers directly above as is implied by the notational scheme. It is often useful to distinguish the state (m) and the basis vector $|(m)\rangle$ associated with the state.

In what follows we shall speak of a *maximal* state $(m)_{\max}$. In terms of the triangular pattern this will mean that the values of all of the $m_{i,j}$'s will be *as great as possible* [that is, the $m_{i,j}$'s move as far to the *left* as allowed by Eq. (3)]. This implies that

$$(m)_{\max} \equiv \begin{bmatrix} m_{1,n} & m_{2,n} & \cdots & m_{n-1,n} & m_{n,n} \\ & m_{1,n} & m_{2,n} & & m_{n-1,n} \\ & & \cdots & & \\ & & m_{1,n} & m_{2,n} & \\ & & & m_{1,n} & \end{bmatrix}. \quad (4)$$

Likewise, a *minimal* state $(m)_{\min}$ will be one where the $m_{i,j}$'s are as small as possible—the $m_{i,j}$'s go as far to the *right* as allowed.

III. THE CONJUGATION OPERATION

One may approach the operation of conjugation in two different but equivalent ways: (1) by viewing the operation with respect to the representation as a whole (in which case the conjugation operation is complex conjugation and carries a character into the complex conjugate character) and (2) by focusing attention on the states of the representation (in which case, the operation appears most simply as the "metric" operation of raising and lowering indices).

Taking the second point of view, let us define conjugation as that operation by which the states of two distinct representations are coupled to produce an invariant. Let the states of one representation be denoted by the basis vectors $|(m)\rangle$; the states of the second representation by $|(m')\rangle$. We then seek coefficients $\langle m', m\rangle$ such that

$$I \equiv \sum_{m,m'} \langle m', m\rangle \, |(m)\rangle \, |(m')\rangle \quad (5)$$

is invariant under the group generators

$$X_A^{\text{tot}} \equiv X_A + X'_A. \quad (6)$$

(There are two distinct points of view one may take with regard to the operators on the coupled system. One may either postulate that the two systems are in *different spaces*, and thus $[X_A, X'_B] = 0$, all A, B—i.e., kinematically independent systems, or, we may let the two representations be instances of the *same* system. For the second point of view to be equivalent to the first, the generators must be explicitly assumed to have the *derivative property*.[9])

Specializing to the diagonal operators H_i, one sees that if I is invariant under H_i then $H_i \cdot I = 0 \Rightarrow \sum_{m,m'} \langle m', m\rangle(m_i + m'_i) \, |(m)\rangle \, |(m')\rangle = 0$. Thus the matrix $\langle m', m\rangle$ is diagonal, that is,

$$\langle m', m\rangle = f(m', m)\delta_{m'_i}^{-m_i}. \quad (7)$$

It is useful in deriving this result to recall (II;

[9] L. C. Biedenharn, *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience Publishers, Inc., New York, 1963), Vol. 5, p. 386.

p. 1462) that

$$H_i \rightarrow m_i \equiv \frac{1}{i}\sum_{j=1}^{i} m_{j,i} - \frac{1}{i+1}\sum_{j=1}^{i+1} m_{j,i+1}. \quad (8)$$

It follows that the two representations must have the same dimension. Moreover, it follows—since the diagonal quantum numbers of the two representations are the *negatives* of each other—that the *characters* of the two representations are complex conjugates. This shows not only that the conjugation operation is equivalent to complex conjugation (and hence the two views mentioned above are equivalent), but that the conjugation operation associates to each representation a unique conjugate representation, and to each state $|(m)\rangle$ a unique conjugate state $|(\tilde{m})\rangle$. We denote this conjugation operation by

$$\mathcal{K}\ |(m)\rangle \rightleftarrows |(\tilde{m})\rangle. \quad (9)$$

In order to obtain the function $f(m', m)$ in Eq. (7) it is necessary now to use the nondiagonal generators. The action of the generators on the states is completely characterized by the matrix elements that have been obtained in II. To simplify the discussion let us note that these matrix elements obey the rule

$$\langle(b)|\, E_{i,j}\ |(a)\rangle = (-)^{j-i+1}\langle(\tilde{a})|\, E_{i,j}\ |(\tilde{b})\rangle, \quad (10)$$

where the quantum numbers of the conjugate basis are given by

$$\tilde{m}_{ij} \equiv m_{1,n} - m_{j-i+1,j}. \quad (11)$$

This result is most easily demonstrated by direct examination of the generator matrix elements [given in Eqs. (61) and (62) of II]. (It should be noted that by this definition, the conjugate of a maximal state is a minimal state.)

Let us now define the function $\rho(n)$ by

$$\rho(n) \equiv \sum_{j=1}^{n}\sum_{i=1}^{j} m_{ij}, \quad (12)$$

that is, $\rho(n)$ is just the sum of all of the state labels in the bottom n rows of the basis state (m). Using this function we may now define the *phase* associated with the SU_n conjugation operation. The operation of conjugation is specified by

$$\mathcal{K}\ |(m)\rangle = (-)^{\delta((m))}\ |(\tilde{m})\rangle \quad (13)$$

where

$$\delta((m)) = \rho(n-1) - \rho(n-1)_{\max}, \quad (14)$$

and $\rho(n)_{\max}$ is $\rho(n)$ evaluated for $(m)_{\max}$. [An *over-all* phase is arbitrary in the operation of conjugation, since the above determination fixes only the relative phase. We have chosen the phase δ, as given above, to be a simple form which agrees with the time-reversal operation in angular momentum, in that the maximal state has $(-)^{\delta} = +1$.]

The invariant I is then specified by the function

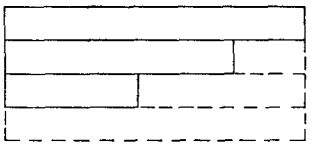


FIG. 1. A Young pattern (for a Weyl basis tableau) and its conjugate pattern.

$\langle m', m\rangle$ where

$$\langle m', m\rangle = \delta^{(\tilde{m})}_{(m')}\cdot(-)^{\delta((m))}[\dim(m)]^{-\frac{1}{2}}, \quad (15)$$

and dim (m) is the Weyl dimension formula. The over-all phase and normalization of this result are arbitrary, but we have chosen this form to agree with the SU_2 paradigm. [In the SU_2 model the corresponding coefficient is the "one-j symbol" $C^{j\ j\ 0}_{m\ m'\ 0} = (-)^{j-m}\cdot\delta^{-m}_{m'}\cdot[2j+1]^{-\frac{1}{2}}$. The operation of conjugation in SU_2—note that all representations are self-conjugate—is often given a metric interpretation. The "one-j symbol" can be considered as the operation of raising–lowering an index, i.e., $\sum_\mu \Psi^\mu C^{j\ j\ 0}_{\mu\ \mu'\ 0} \rightleftarrows \Psi^j_{\mu'}$. This is the so-called "Herring metric."][10]

In terms of the Weyl basis tableaux, the operation of conjugation assumes a diagrammatic interpretation. The operation of conjugation associates with a given Weyl basis tableau (a given state, that is) a unique new Weyl basis tableau, the new tableau having rows of lengths $m_{1,n}-m_{n,n}$; $m_{1,n}-m_{n-1,n}$; $m_{1,n}-m_{n-2,n}$; $\cdots$; $m_{1,n} - m_{2,n}$. This new Weyl basis tableau (conjugate to the original basis tableau) is obtained by filling in the new Young pattern (shown below beneath the original Young pattern in Fig. 1 and indicated by the dotted lines) with the integers one to n in such a way that the new tableau is *lexical* from the *bottom right corner*. In filling in the new tableau *in the position shown*, one uses the rule that *no column in the new tableau may contain any numeral in the same column of the original Weyl basis tableau.* In this way all of the integers, one through n, will occur once in every column of the above rectangle (consisting of the original pattern and its conjugate—the latter pattern being inverted). As an example of the above rules, Fig. 2, shows an SU_4 Weyl basis tableau, its conjugate, and both tableaux put together in the form of a rectangle (as in Fig. 1).

A *phase* associated with the new tableau is also fixed by this procedure. This "pattern phase", ω, is the product of the permutation signatures of all of the columns of the completed rectangle. This pattern phase ω is not the same as δ but differs by

[10] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*, translated by J. J. Griffin (Academic Press Inc., New York, 1959); see p. 293, Footnote 9 of this reference.

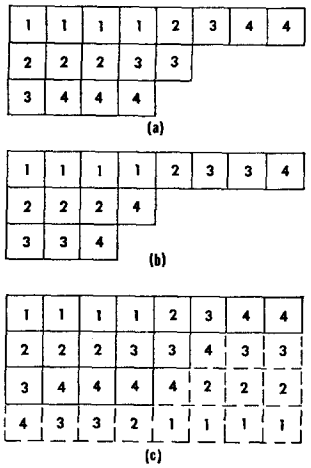


FIG. 2. A Weyl basis tableau (in SU_4), its conjugate Weyl basis tableau, and the two combined into a rectangular array.

only a fixed phase. In SU_n for the phase ω one gets

$$\omega \equiv \rho(n-1) + \begin{cases} m_{1,n} \text{ if } n \equiv 2 \text{ or } 3 \pmod 4, \\ 0 \quad \text{otherwise.} \end{cases} \tag{16}$$

IV. THE QUESTION OF PHASE CONVENTIONS

The question of a convenient phase convention to adopt for a systematic discussion of the unitary groups is an annoying one, but of considerable practical importance. It is helpful, before specifying the phase conventions to be followed, to re-examine the angular momentum paradigm.

It will be recalled that the Condon and Shortley[11] (Wigner) phase convention for the spherical harmonics $Y_\ell^m(\Theta, \varphi)$ differed by the phase $(-)^m$ (for $m > 0$) from the earlier definitions used in the mathematical literature (and adopted by Bethe[12]). The Condon–Shortley phase is, at first glance, a rather curious one; how does it come about that it is a useful choice?

The answer—which we generalize at once—is that this phase choice results from the properties of the generators $\mathbf{J} = \{E_{1,2}, E_{2,1}, H_1\}$, [which are Hermitian and obey the time-reversal rule $J^T = -J$] with the convention that the matrix elements of $E_{1,2}$ (the raising operator) be real and positive.

The generators themselves play two different roles: (1) operators and (2) the carrier space of the adjoint representation. The latter role establishes a mapping of the generators onto the basis states of the adjoint representation $X_A \rightarrow |[X_A]\rangle$. Using now the commutator relations one sees that

(a) $[H_1, E_{1,2}] = (2)^{-\frac{1}{2}} E_{1,2}$

$$\rightleftarrows E_{1,2}\,|[H_1]\rangle = -(2)^{-\frac{1}{2}}\,|[E_{1,2}]\rangle, \tag{17}$$

(b) $\langle(m')|\, E_{1,2}\,|(m)\rangle =$ positive, real number.

The relations given in Eq. (17) imply that the states associated with $E_{1,2}$ and H_1 differ in relative phase by a minus sign. In a similar way, the commutator $[E_{1,2}, E_{2,1}] = (2)^{-\frac{1}{2}} H_1$—viewed as a generator relation on states—shows that the states associated with H_1 and $E_{2,1}$ have the *same* relative phase. Thus one finds that the states of the adjoint representation have the relative phases

$$\begin{aligned} |\ell = 1, m = 1\rangle &\rightleftarrows -E_{1,2}, \\ |\ell = 1, m = 0\rangle &\rightleftarrows H_1, \\ |\ell = 1, m = -1\rangle &\rightleftarrows E_{2,1}, \end{aligned} \tag{18}$$

which is precisely the phase $(-)^m$ mentioned earlier. The convention that the Wigner coefficient $C^{j_1\, j_2\, j}_{j_1\, j-\, j_1\, j}$ have positive real phase then establishes this convention in general.

{Let us note that this mapping of operators onto states does not include the conjugation operation. Thus, for example,

$$\mathcal{K} X_A \mathcal{K}^{-1} = -X_{-A} \tag{19}$$

may be associated with

$$\mathcal{K}\,|(X_A)\rangle = -|(X_{-A})\rangle, \tag{20}$$

but this need not follow. This extra freedom [a result of the fact that Eq. (19) is "bilinear" in $\mathcal{K}$ while Eq. (20) is "linear"] is quite important [as will be clear in IV (following paper)]. Let us note, too, that the choice expressed by Eqs. (19) and (20) is often very convenient.[13] It leads to $i^\ell Y_\ell^m = |\ell, m\rangle$, but it is not the conventional choice.}

The generalization to all unitary groups is now self-evident: the raising operators and the maximal Wigner coefficients are defined to have positive, real phase. By means of the mapping: generators $\rightleftarrows$ basis vectors of the adjoint representation, one establishes the relative phase of the various states. Finally, the over-all phase is established by adopting in general the choice equivalent to Eq. (18). *This establishes a single consistent convention for the unitary groups as a whole.*

This phase convention has already been employed in the matrix elements of the generators given in II, p. 1466. It is quite essential to note, however, that these matrix elements require a further phase and

[11] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, Cambridge, England, 1935).

[12] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One and Two Electron Atoms* (Academic Press Inc., New York, 1957).

[13] L. C. Biedenharn and M. E. Rose, Rev. Mod. Phys. 25, 729 (1953).

normalization before they may be identified as Wigner coefficients (recall the similar situation in SU_2). This is explicitly given below.

Let us proceed to establish this convention explicitly. First consider the commutator

$$[H_i, E_{n-1,n}] = m_i(n-1, n)E_{n-1,n} \tag{21}$$

where the general value of $m_i(n-1, n)$ is

$$m_i(n-1, n) = [-1/(n+1)]\delta_i^{n-2} + [1/(n-1)]\delta_i^{n-1}. \tag{22}$$

In Table I, some values of $m_i(n-1, n)$ are tabulated so that the *structure* of this result will be clearer.

From consideration of Table I, it may easily be seen that the phase for the H_i's must be $(-)^{i+1}$ (except for, perhaps, an over-all phase), that is, the H_i's *alternate in sign* when considered as states. To see this, consider the commutator

$$[E_{n-1,n}, H_i] = -m_i(n-1, n)E_{n-1,n}. \tag{23}$$

Considered in its significance as a relation concerning *states*, Eq. (23) shows that the relative phase of the various states associated with the H_i is just the sign of $(-m_i)$. (The matrix elements of $E_{n-1,n}$ are assumed positive.) Looking at a fixed row of Table I, shows that (since the m_i alternate in sign) the H_i must have the phase $(-)^{i+1}$.

Next we shall show that $E_{i-2,i-1}$ and $E_{i-1,i}$, when considered as states, differ in phase by -1. To see this consider the commutators

$$[E_{i-2,i-1}, E_{i-1,i}] = -[E_{i-1,i}, E_{i-2,i-1}] = E_{i-2,i}, \tag{24}$$

which implies that

$$(-)^{\delta_1}E_{i-1,i}\,|[E_{i-2,i-1}]\rangle = -(-)^{\delta_2}E_{i-2,i-1}\,|[E_{i-1,i}]\rangle = -(-)^{\delta_3}\,|[E_{i-2,i}]\rangle. \tag{25}$$

{Here the $(-)^{\delta}$ are the phases associated with $E_{ij} \to (-)^{\delta}\,|[E_{ij}]\rangle$.}

From the same commutator (Eq. 24), one easily sees now that $|[E_{i-2,i}]\rangle$ also has the phase $(-)$ relative to $|[E_{i-2,i-1}]\rangle$. By commuting $E_{i-2,i}$ with $E_{i-3,i-2}$ to get $E_{i-3,i}$ one establishes a recursive procedure for showing that the phase associated with $|[E_{i,j}]\rangle$ is $(-)^{i+1}$ for $i < j$.

TABLE I. Values of $m_i(n-1, n)$ for $n < 7$.

	H_1	H_2	H_3	H_4	H_5	⋯
$E_{1,2}$	1	0	0	0	0	⋯
$E_{2,3}$	$-\frac{1}{2}$	$\frac{1}{2}$	0	0	0	⋯
$E_{3,4}$	0	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	⋯
$E_{4,5}$	0	0	$-\frac{1}{4}$	$\frac{1}{4}$	0	⋯
$E_{5,6}$	0	0	0	$-\frac{1}{5}$	$\frac{1}{5}$	⋯
⋯	⋯	⋯	⋯	⋯	⋯	⋯

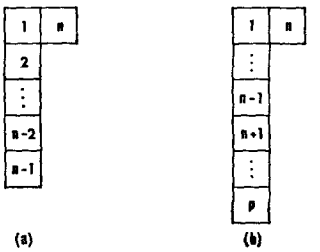


FIG. 3. Weyl basis tableaux for the generator H_{n-1}.

Lastly, by considering the commutator $[E_{i-1,i}, E_{i,i-1}] = -[E_{i,i-1}, E_{i-1,i}]$, one sees that $E_{i-1,i}$ and $E_{i,i-1}$ have opposite phases as states; then, exactly as above, it may be established that the phase of $|[E_{i,j}]\rangle$ in general is $(-)^{i+1}$ (for $i \neq j$). This establishes a consistent phase convention for any generator onto an equivalent state.

To summarize, the phase convention

$$E_{i,j} \to (-)^{i+1}\,|[E_{i,j}]\rangle; \quad H_i \to (-)^{i+1}\,|[H_i]\rangle \tag{26}$$

has been shown to be a consistent generalization for all SU_n equivalent to the Condon–Shortley–Wigner phase for SU_2.

V. THE EXPLICIT ASSOCIATION: GENERATORS → $|(m)\rangle$

In Sec. IV it was shown how to establish the phases of the generators when they are mapped onto associated states; in the present section we give the explicit state of the adjoint representation onto which each generator maps.

We use the commutators of the *diagonal* generators H_i with an *arbitrary* generator X_A to define this mapping. The following equivalence

$$[H_i, X_A] = m_i(A)X_A \to H_i\,|[X_A]\rangle = \ell_i(A)\,|[X_A]\rangle \tag{27}$$

is used. We impose the condition

$$m\ell(A) = \ell_i(A) \tag{28}$$

upon the mapping. A mapping of the generators of SU_n *onto* the states of the adjoint representation (the $[2\,1\,1\cdots1\,0]$ representation) satisfying the above condition is now given. This mapping is a one-to-one mapping. (It is a useful abbreviation to employ the numerical convention that $\dot{1}$, $\dot{2}$, etc. denotes a *repeated* numeral. Thus the adjoint representation is denoted $[2\,\dot{1}\,0]$.)

Consider first the diagonal generator H_{n-1} which is added in going from SU_{n-1} to SU_n. This generator will map onto Weyl basis tableaux shown in Fig. 3. Fig. 3(a) shows the tableau for SU_n and Fig. 3(b) shows how this tableau is extended to obtain the tableau for H_{n-1} in SU_p where $p > n$. One sees that

to convert the SU_n tableau for H_i to a tableau in SU_p one merely adds the integers $n+1, n+2, \cdots, p-1, p$ in lexical order in the $p-n$ additional rows of the new tableau. This method of changing an SU_n Weyl basis tableau to an SU_p Weyl basis tableau is completely general. Thus, the state of the adjoint representation corresponding to a generator is known, for all SU_p, from the pattern found from its first occurrence in SU_n, $p > n$.

The Weyl basis tableau shown in Fig. 3(b) has the (triangular pattern) basis state [obtained by Eqs. (1) and (2)].

$$|[H_{n-1}]\rangle = (-)^n \left|\begin{array}{ccccccc} 2 & 1 & 1 & 1 & 1 & \cdots & 1\ 1\ 1\ 1\ 0 \\ 2 & 1 & 1 & 1 & 1 & \cdots & 1\ 1\ 1\ 1\ 0 \\ & & & & \cdots & & \\ 2 & 1 & 1 & 1 & & \cdots & 1\ 1\ 1\ 0 \\ 1 & 1 & 1 & 1 & & \cdots & 1\ 1\ 1\ 1 \\ & & & & \cdots & & \\ & & & & & 1 & 1 \\ & & & & & & 1 \end{array}\right| \begin{array}{l} p\text{th row} \\ \\ n\text{th row} \\ (n-1)\text{th row}, \end{array} \tag{29}$$

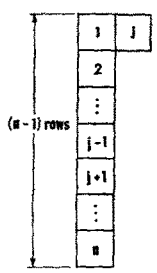


FIG. 4. The Weyl basis tableau for $E_{i,j} (i < j)$ in SU_n.

where the rows of the state vector are numbered from the bottom—the bottom row being the first row and the top row being the pth row.

The Weyl basis tableau for $E_{i,j}$ $(i < j)$ is shown in Fig. 4.

To obtain the conjugate Weyl basis tableau $E_{j,i}$ one applies the conjugation operation of Sec. III. The Weyl basis tableau of Fig. 4 has the equivalent basis vector

$$|[E_{i,j}]\rangle = (-)^{j+1} \left|\begin{array}{ccccccc} 2 & 1 & 1 & 1 & 1 & \cdots & 1\ 1\ 1\ 0 \\ 2 & 1 & 1 & 1 & 1 & \cdots & 1\ 1\ 1\ 0 \\ & & & & \cdots & & \\ 2 & 1 & 1 & 1 & & \cdots & 1\ 1\ 0 \\ 2 & 1 & 1 & 1 & & \cdots & 1\ 1\ 1 \\ & & & & \cdots & & \\ 2 & 1 & 1 & & & \cdots & 1\ 1 \\ 1 & 1 & 1 & & & \cdots & 1\ 1 \\ & & & & \cdots & & \\ & & & & & 1 & 1 \\ & & & & & & 1 \end{array}\right| \begin{array}{l} j\text{th row} \\ \\ i\text{th row}. \end{array} \tag{30}$$

To find the Gelfand basis vector for $E_{i,j}$ where $i > j$, one merely uses the conjugate basis state to that of Eq. (30) [conjugating the state of Eq. (30) according to Eq. (11)]. Thus we have established a mapping from the generators onto the states. It is easily verified that the above mapping does, indeed, satisfy the requirement of Eq. (28). This is accomplished by using Eq. (8), the commutators of the H_i, and the fact that

$$H_i \to M_i = [i(i+1)]^{-1} \sum_{k=1}^{i} e_{kk} - (i+1)^{-1} e_{i+1,i+1}, \tag{31}$$

where $\ell_i(m)$ is the result of H_i operating upon the state $|(m)\rangle$ as defined in Eq. (27). It should be noted that the *phases* associated with the mapping of the generators onto states have been included in Eqs. (29) and (30).

APPENDIX A: ALGEBRAIC TREATMENT OF THE SU_n GENERATORS

Shortly after the completion of I, Dr. U. Fano called to our attention a most elegant and economical proof that the symmetric coupling coefficients $[A\ B^C]$ had the properties proved in the appendix of I. The algebraic method that Fano proposed is very valuable in that it suggests many further applica-

tions—particularly to the task of discussing the generalized Wigner and Racah coefficients—and discussion of these aspects were postponed until the present paper. In the meantime, an independent treatment, similar in content to Fano's method, has been published by Klein,[14] and we accordingly only sketch the method and indicate briefly possible further developments here.

The essential idea is to introduce for the *fundamental* (n-dimensional) representation of SU_n the explicit generator matrices Σ_A, which are but a generalization of the Pauli spin matrices familiar from SU_2. A second, kinematically independent, set of generators X_A is then introduced and in addition the generators $X_A^{\text{tot}} \equiv \Sigma_A + X_A$—which function analogously to the total angular momentum operator—are introduced.

It follows immediately that the scalar operator

$$\mathbf{\Sigma}\cdot\mathbf{X} \equiv \sum_A (\Sigma_A X_{-A}) \tag{A1}$$

is invariant under $\{X^{\text{tot}}\}$. Moreover, one easily sees that one may reduce the product $(\mathbf{\Sigma}\cdot\mathbf{X})^2$ by separating into symmetric and antisymmetric parts, i.e.,

$$(\mathbf{\Sigma}\cdot\mathbf{X})^2 = \sum_{A,B} (\Sigma_A \Sigma_B X_A X_B)$$
$$= \tfrac{1}{2} \sum_{A,B} \{[\Sigma_A, \Sigma_B]_+ X_A X_B + [\Sigma_A, \Sigma_B]_- X_A X_B\}. \tag{A2}$$

Introducing the definitions of the symmetric and antisymmetric coupling coefficients [I, Eqs. (13) and (15)]

$$[\Sigma_A, \Sigma_B]_- \equiv \sum_C (AB^C)\Sigma_C, \tag{A3}$$

$$[\Sigma_A, \Sigma_B]_+ \equiv n^{-2}\delta_A^B + \sum_C [AB^C]\Sigma_C, \tag{A4}$$

one finds that

$$2(\mathbf{\Sigma}\cdot\mathbf{X})^2 + \tfrac{1}{2}\mathbf{\Sigma}\cdot\mathbf{X} = (n^2)^{-1}\mathbf{X}\cdot\mathbf{X}$$
$$+ \sum_{ABC} [AB^C]\Sigma_C X_{-A} X_{-B} = n^{-2}\mathbf{X}\cdot\mathbf{X} + \mathbf{\Sigma}\cdot\mathbf{X}^{(2)}, \tag{A5}$$

where

$$X_C^{(2)} \equiv \sum_{AB} [AB^C] X_A X_B, \tag{A6}$$

and the first of the two useful general relations

$$\sum_{AB} (AC^B)(BD^A) = \delta_C^D, \tag{A7}$$

$$\sum_{AB} [AC^B][BD^A] = \delta_C^D (n^2 - 4)/n^2, \tag{A8}$$

has been used.

It follows immediately from (A5) that $\mathbf{\Sigma}\cdot\mathbf{X}^{(2)}$ is invariant under $\{X^{\text{tot}}\}$, and hence $X_A^{(2)}$ transforms under $\{X_A\}$ as the adjoint representation. This proves that the symmetric coupling coefficients have the properties claimed in I (since the null situation occurs only for SU_2 as is easily shown).

[14] A. Klein, J. Math. Phys. 4, 1283 (1963).

The relation (A5) may next be generalized [using (A7) and (A8)] to yield the reduction

$$(\mathbf{\Sigma}\cdot\mathbf{X})(\mathbf{\Sigma}\cdot\mathbf{X}^{(k)}) = (2n^2)^{-1} I_{k+1}(\mathbf{X})$$
$$+ \tfrac{1}{2}\mathbf{\Sigma}\cdot\mathbf{X}^{(k+1)} - \tfrac{1}{4}\mathbf{\Sigma}\cdot\mathbf{X}^{(k)}, \tag{A9}$$

where $I_{k+1}(\mathbf{X})$ denotes the $(k+1)$th invariant operator formed from $\{X_A\}$.

It is also convenient to introduce the invariants of the total generators $\{X_A^{\text{tot}}\}$,

$$I_2(\mathbf{X}^{\text{tot}}) = I_2(\mathbf{X}) + I_2(\mathbf{\Sigma}) + 2\mathbf{\Sigma}\cdot\mathbf{X}, \tag{A10}$$

$$I_3(\mathbf{X}^{\text{tot}}) = I_3(\mathbf{X}) + I_3(\mathbf{\Sigma}) + 3\mathbf{\Sigma}\cdot\mathbf{X}^{(2)}$$
$$+ [3(n^2 - 4)/2n^2]\mathbf{\Sigma}\cdot\mathbf{X}. \tag{A11}$$

It is easily seen from these relations that (1) the symmetrically coupled generators $\mathbf{X}^{(k)}$ obey the product law $\mathbf{X} \circledS \mathbf{X}^{(k)} = \mathbf{X}^{(k+1)}$, where $\circledS$ denotes a symmetric vector coupling; (2) the symmetrically coupled generators *terminate*, (that is, $\mathbf{X}^{(n)}$ is expressible in terms of lower $\mathbf{X}^{(k)}$); and (3) the invariant $\mathbf{\Sigma}\cdot\mathbf{X}$ obeys an nth-order polynomial equation. We illustrate these results for the SU_3 group.

From the SU_2 paradigm one knows that the operator $-(\boldsymbol{\delta}\cdot\mathbf{L} + 1)$ plays a special and familiar role in discussing the angular momentum operator $\mathbf{J} = \mathbf{L} + \tfrac{1}{2}\boldsymbol{\delta}$. In a certain sense $\boldsymbol{\delta}\cdot\mathbf{L} + 1$ provides a linear (operator) factorization of the total angular momentum operator $\mathbf{J}^2$—a result exploited in the operator treatment of the spin-$\frac{1}{2}$ problem using the Pauli-spherical eigenfunctions χ_κ^μ of Ref. 13. A similar operator $-(6\mathbf{\Sigma}\cdot\mathbf{X} + 1) \equiv Q$ plays an analogous role for the SU_3 group. The polynomial equation satisfied by Q is

$$Q^3 + 6I_3(\mathbf{X}^{\text{tot}}) = (3I_2(\mathbf{X}^{\text{tot}}) + 1)Q, \tag{A12}$$

from which one shows that Q may be brought to the diagonal matrix form

$$Q = \frac{1}{3}\begin{bmatrix} p+q+3 & 0 & 0 \\ 0 & q-2p-3 & 0 \\ 0 & 0 & p-2q \end{bmatrix} \equiv \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}. \tag{A13}$$

(This result is clearly related to the standard treatments of the unitary group by Weyl; it is similar to the techniques of Gelfand and of Klein.) It is useful to note that this diagonal form for Q has a close connection to Weyl's dimension formula: in fact, $\dim [p\ q\ 0] = \tfrac{1}{2}(a-c)(a-b)(c-b)$.

The general relation (A9) now implies the special results:

$$\mathbf{\Sigma}\cdot\mathbf{X}^{(2)} = \tfrac{1}{18}Q(Q + \tfrac{1}{2}) - \tfrac{1}{9}(I_2 + \tfrac{1}{4})$$
$$\mathbf{\Sigma}\cdot\mathbf{X}^{(3)} = -(I_2 + \tfrac{1}{4})(Q + 1), \tag{A14}$$

the latter relation implying the desired result

$$X_A^{(3)} = [(4I_2 + 1)/(36)]X_A. \qquad \text{(A15)}$$

We do not pursue Fano's technique further, but it is useful to note, however, that the method is extremely valuable for the explicit construction of *projection operators* (and hence also invariant operators) for the SU_n groups. *The fact that the fundamental Wigner coefficients are completely known for the general case can be very nicely combined with this algebraic technique in the construction of these projection operators.*

Note added in proof: The review paper of de Swart [Rev. Mod. Phys. **35**, 916 (1963)] discusses the phase question very carefully for the octet model (the SU_3/C_3 group) and arrives at a different phase convention than used in III above. (References to the earlier literature may also be found in his paper.) The convention used by de Swart is based on: (1) E_{12} and E_{13} are postulated to have only positive real matrix elements (in contrast to E_{12} and E_{23} used in III above) and (2) $I_z = Y = 0$ states are to have positive real phase under conjugation (in contrast to the use of the maximal state for this in III above). The resulting phase convention of de Swart is quite convenient for the octet model since conjugation then takes a state with charge Q into $(-)^Q \times$ (state of charge $-Q$).

There are two principal reasons not to adopt de Swart's phase convention as a basis for the general SU_n group [aside from the obvious point that the phase $(-)^Q$ is specifically tailored to the octet model]. Firstly, one recognizes from II that the basic operators are the $E_{i+1,i}$, (since, roughly speaking, the complexity of an operator increases with distance from the "diagonal"). Hence a natural phase convention would be based on these. Secondly, unless the phase conventions are patterned on the $E_{i+1,i}$ (postulated to have positive real matrix elements), the simple relationship between conjugation and the "pattern phase" of the Weyl tableaux is lost. We feel that this latter point is especially persuasive in indicating the choice made in III on the vexed question of phase conventions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

On the Representations of the Semisimple Lie Groups. IV. A Canonical Classification for Tensor Operators in SU_3*†

G. E. BAIRD‡ AND L. C. BIEDENHARN
Department of Physics, Duke University, Durham, North Carolina
(Received 21 April 1964; final manuscript received 24 July 1964)

It is shown that the multiplicity structure of the general SU_n operators may be put in a one-to-one correspondence with the multiplicity structure of the corresponding states. This result allows a convenient labeling scheme to be devised for the general SU_n *Wigner operator* and leads in a natural way to the concept of a *reduced Wigner operator*. The problem of multiplicity in tensor operators is shown to have a canonical resolution in the conjugation classification which is discussed in detail for the SU_3 case.

I. INTRODUCTION AND SUMMARY

THE present series of papers[1] is concerned with the explicit constructive determination of the semisimple Lie groups by an extension of the Racah–Wigner angular momemtum calculus developed for the group SU_2.

It is the purpose of the present paper to discuss in detail a canonical solution to the problem of simple reducibility with particular attention to the unitary group SU_3.

The problem of simple reducibility arises when one seeks to determine explicit matrices ("Wigner coefficients") which effect the decomposition of the direct product (inner Kronecker product) of two irreducible representations of a group G. If in every such direct product the irreducible representations occur either once or not at all, then the Wigner coefficients are uniquely (to within phases) specified by the group.

* Supported in part by the U. S. Army Research Office (Durham) and the National Science Foundation.
† This paper is excerpted in part from the Ph.D. thesis of G. E. Baird (submitted to Duke University, April, 1964).
‡ Present address: Northrop Space Laboratories, Huntsville, Alabama.
[1] L. C. Biedenharn, J. Math. Phys. 4, 436 (1963); G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963); 5, 1723 (1964). We refer to these works as I, II, and III, respectively.

the latter relation implying the desired result

$$X_A^{(3)} = [(4I_2 + 1)/(36)]X_A. \qquad \text{(A15)}$$

We do not pursue Fano's technique further, but it is useful to note, however, that the method is extremely valuable for the explicit construction of *projection operators* (and hence also invariant operators) for the SU_n groups. *The fact that the fundamental Wigner coefficients are completely known for the general case can be very nicely combined with this algebraic technique in the construction of these projection operators.*

Note added in proof: The review paper of de Swart [Rev. Mod. Phys. **35**, 916 (1963)] discusses the phase question very carefully for the octet model (the SU_3/C_3 group) and arrives at a different phase convention than used in III above. (References to the earlier literature may also be found in his paper.) The convention used by de Swart is based on: (1) E_{12} and E_{13} are postulated to have only positive real matrix elements (in contrast to E_{12} and E_{23} used in III above) and (2) $I_z = Y = 0$ states are to have positive real phase under conjugation (in contrast to the use of the maximal state for this in III above). The resulting phase convention of de Swart is quite convenient for the octet model since conjugation then takes a state with charge Q into $(-)^Q \times$ (state of charge $-Q$).

There are two principal reasons not to adopt de Swart's phase convention as a basis for the general SU_n group [aside from the obvious point that the phase $(-)^Q$ is specifically tailored to the octet model]. Firstly, one recognizes from II that the basic operators are the $E_{i+1,i}$, (since, roughly speaking, the complexity of an operator increases with distance from the "diagonal"). Hence a natural phase convention would be based on these. Secondly, unless the phase conventions are patterned on the $E_{i+1,i}$ (postulated to have positive real matrix elements), the simple relationship between conjugation and the "pattern phase" of the Weyl tableaux is lost. We feel that this latter point is especially persuasive in indicating the choice made in III on the vexed question of phase conventions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

On the Representations of the Semisimple Lie Groups. IV. A Canonical Classification for Tensor Operators in SU_3*†

G. E. BAIRD‡ AND L. C. BIEDENHARN
Department of Physics, Duke University, Durham, North Carolina
(Received 21 April 1964; final manuscript received 24 July 1964)

It is shown that the multiplicity structure of the general SU_n operators may be put in a one-to-one correspondence with the multiplicity structure of the corresponding states. This result allows a convenient labeling scheme to be devised for the general SU_n *Wigner operator* and leads in a natural way to the concept of a *reduced Wigner operator*. The problem of multiplicity in tensor operators is shown to have a canonical resolution in the conjugation classification which is discussed in detail for the SU_3 case.

I. INTRODUCTION AND SUMMARY

THE present series of papers[1] is concerned with the explicit constructive determination of the semisimple Lie groups by an extension of the Racah–Wigner angular momemtum calculus developed for the group SU_2.

It is the purpose of the present paper to discuss in detail a canonical solution to the problem of simple reducibility with particular attention to the unitary group SU_3.

The problem of simple reducibility arises when one seeks to determine explicit matrices ("Wigner coefficients") which effect the decomposition of the direct product (inner Kronecker product) of two irreducible representations of a group G. If in every such direct product the irreducible representations occur either once or not at all, then the Wigner coefficients are uniquely (to within phases) specified by the group.

* Supported in part by the U. S. Army Research Office (Durham) and the National Science Foundation.
† This paper is excerpted in part from the Ph.D. thesis of G. E. Baird (submitted to Duke University, April, 1964).
‡ Present address: Northrop Space Laboratories, Huntsville, Alabama.
[1] L. C. Biedenharn, J. Math. Phys. 4, 436 (1963); G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963); 5, 1723 (1964). We refer to these works as I, II, and III, respectively.

The problem of simple reducibility was originally posed by Wigner[2] in his categorization of the special properties of the angular momentum group that underlie the vector addition coefficients. As defined by Wigner, a simply reducible group possesses two properties: (a) all classes are ambivalent (every element is equivalent to its inverse) and (b) it is multiplicity free (the Kronecker product of any two irreducible representations of the group contains no representations more than once). As discussed by Mackey[3] and by Sharp,[4] the ambivalence requirement may be largely removed; the essential condition is that the group be multiplicity free.

The importance of the problem of constructing explicit Wigner coefficients for a group may be seen in a more physical way. The general operators on a group may be classified by the mapping: operators $\rightarrow$ states. This is the generalization[5] of the familiar *tensor operator classification* well known from Wigner's and Racah's work on the angular momentum calculus. If a group is multiplicity free the tensor operator classification is complete; one obtains then a generalization of the Wigner–Eckart result, in which the Wigner coefficients play the role of unit tensor operators and cover the space of admissible operator structures on the group.[6]

But if a given representation occurs, say, *twice*, in a reduction of the inner Kronecker product, one now has the freedom to arbitrarily "mix" the occurrences—the Wigner coefficients now have an unwanted ambiguity. Clearly there are any number of ways to define "Wigner coefficients" in such a case; suitable operators that accomplish the splitting can be constructed almost at will.

The problem of constructing Wigner coefficients for a non-multiplicity-free group is then not merely to *remove* the ambiguity in the definition of the Wigner coefficients, but to remove the ambiguity in a canonical way—preferably by an appeal to properties of the group itself.[7] It is the purpose of the present paper to demonstrate explicitly such a canonical resolution of the multiplicity problem, by classifying tensor operators—having identical transformation properties—via the *conjugation operation*,[8] discussed in III. (The conjugation operation serves to determine the SU_n equivalent of Wigner's "$1-j$ symbol" for the SU_2 group.)

In the language of the Wigner–Eckart theorem, one may now assert this multiplicity free-tensor operator classification designates a complete set of orthonormal operators for the unitary group SU_3; the unit tensor operators classified via conjugation span the space of admissible operator structures in SU_3.

In order to discuss the multiplicity problem for the unitary groups, it is necessary first to determine explicitly the extent of the multiplicity. This is accomplished by the general result: *the multiplicity structure of the tensor operators may be put in one-to-one correspondence with the multiplicity structure of the states*; there exists therefore a *second* operator mapping, in addition to the tensor operator mapping itself (cf. Sec III). This result is not completely new; it is clearly contained (though not this explicitly) in Weyl's discussions of the Kronecker product reduction.[9] It is, however, only possible to discuss this result as a one-to-one mapping in terms of tensor operator structures; this is the essential element introduced in the present work.

By means of this second operator mapping, a convenient notational convention for the general unit operators in SU_n is now possible (Sec. IV). As a necessary preliminary to these results, a detailed discussion of the basis states, the general phase convention, and the explicit generator mapping onto states has been given in III. This is particularly important for understanding the conjugation operation.[10]

The conjugation classification is introduced by means of a detailed treatment of the tensor operators in the SU_3 group. This classification is discussed in several equivalent formulations; in particular it is shown that the SU_3 conjugation classification is itself equivalent to a mapping into the generators of SU_4.

A final section discusses very briefly the generalization of the conjugation classification to SU_n.

---

[2] E. P. Wigner, Am. J. Math. **63**, 57 (1941).

[3] G. W. Mackey, Pacific J. Math. **8**, 503 (1958).

[4] W. T. Sharp, "Racah Algebra and the Contraction of Groups," CRT-935 (AECL 1098), September 1960.

[5] A. P. Stone, Proc. Cambridge Phil. Soc. **57**, 460, 469 (1961); an excellent bibliography citing the extensive literature of the tensor operator classification may be found in these papers.

[6] (a) G. Racah, lecture notes, Institute for Advanced Study, Princeton, New Jersey, 1951 (unpublished); (b) L. C. Biedenharn, *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience Publishers, New York, 1963), Vol. V.

[7] Earlier discussions of the simple reducibility problem have not always emphasized this point. See: J. Ginibre, J. Math. Phys. **4**, 720 (1963); M. Moshinsky, J. Math. Phys. **4**, 1128 (1963); B. Diu, Nuovo Cimento **28**, 466 (1963); G. F. Koster, Phys. Rev. **109**, 227 (1958).

[8] A preliminary discussion was given by L. C. Biedenharn, Phys. Letters **3**, 254 (1963).

[9] H. Weyl, lecture notes, The Institute for Advanced Study, Princeton, New Jersey (1934) (unpublished).

[10] I. M. Shumushkevich, Doklady Akad. Nauk SSSR **103**, 235 (1955); A. J. Macfarlane, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **5**, 576 (1964).

The present work is for the most part formal, and seeks to delimit, and classify, the type of operator structures on the general unitary groups. Specific results—that is, explicit Wigner and Racah coefficients—are given subsequently. Let us note however that II presented the fundamental Wigner coefficients for the general unitary group, and that other special cases of the Wigner coefficients have been discussed by Jucys *et al.*,[11] Moshinsky,[12] Lurie and Macfarlane, [13] Hecht,[14] Konuma and Wada,[15] Gasiorowicz,[16] Gabriel,[17] among others.

II. CONCEPT OF A TENSOR OPERATOR

The introduction of the concept of a tensor operator into quantum mechanics stems from early work of Pauli and Güttinger,[18] and of Wigner.[19] The original work concerned *vector* operators, and for this the prototype operator is the angular momentum $\mathbf{J}$. For diagonal matrix elements this leads to the familiar result of the vector model $\mathbf{V} \rightarrow \mathbf{V}\cdot\mathbf{J}\ (\mathbf{J}^2)^{-1}\mathbf{J}$ (diagonal matrix elements only). The generalization to the complete vector operator (all matrix elements) is the essential content of Wigner's determination of the ($J = 1$) vector coupling coefficients. The ultimate generalization of these ideas, for angular momentum, is the content of the famous Wigner–Eckart result; for this, the Racah definition of the general tensor operator[20] in terms of the commutation properties under the generators $\mathbf{J}$ is especially valuable. We may summarize these results as direct analogs to the vector model result: the most general tensor operator can be expressed as a linear combination (coefficients = "reduced matrix elements") of the *unit* tensor operators ("Wigner operators"). In other words, the Wigner operators are a canonical basis for the space of SU_2 tensor operators.

We may most conveniently abstract the essential elements of these fundamental ideas if we note that the introduction of a tensor operator expresses a mapping: operators $\rightarrow$ states. This, however, is but the familiar result of vector spaces: *the operators on a vector space are themselves a vector space.* It is natural then to classify this vector space of operators by the same canonical classification used to classify the original vector space. Thus we achieve the mapping stated.

Let us indicate this mapping more precisely. (We use for definiteness SU_n as the underlying group, but this is inessential and the results are general.) The generators of the group are the operators, $\{X_A\}$; using the results of Papers I and II, we classify the states of the representations by the canonical (Gelfand) labels (m), i.e., the orthonormal states are $|(m)\rangle$. Then the operator mapping is expressed by ($\mathcal{O}$ being a generic operator)

$$\mathcal{O} \rightarrow |(m)\rangle$$
$$[X_A, \mathcal{O}] \rightarrow X_A\,|\,(m)\rangle. \tag{1}$$

From this mapping we may now label the tensor operators as $\mathcal{O}((m))$. Taking matrix elements, one obtains Racah's definition of the general tensor operator,

$$[X_A, \mathcal{O}((m))] = \sum_{(m')} \langle (m')\ |X_A|\ (m)\rangle \mathcal{O}((m')). \tag{2}$$

It is essential to note that this definition of the general (SU_n) tensor operator is fully defined, since the matrix elements of the generators have been completely and explicitly determined in II. (Note, too, that if one wishes to rewrite this definition in terms of Wigner coefficients, it is necessary that the phase and normalization conventions of III be taken into account.)

It is useful to put this (infinitesimal) definition of the tensor operators into integral form. Let $U(\theta_i)$ be the general unitary transformation, where

$$U(\theta_i) \equiv \exp\,[i \sum_A k_A(\theta_i) X_{-A}]. \tag{3}$$

Then directly from the mapping, (or equivalently from the Campbell–Baker–Hausdorff rule) one finds

$$U^{-1}\mathcal{O}((m))U = \sum_{(m')} \langle (m')\ |U|\ (m)\rangle \mathcal{O}((m')). \tag{4}$$

Note that in obtaining this result, Eq. (4), one uses the multiple commutator mapping

$$(X)^k \rightarrow [X, [X, [X, \cdots]]]\ (k \text{ times})]. \tag{5}$$

This shows that tensor operators transform under finite "rotations" via the general transformation matrices in exactly the same way as the states (to which the operators are associated) transform.

[11] A. A. Bandzaitis and A. P. Jucys, Liet. Fiz. Rin. **2**, 5 (1962); K. Zukauskas, R. Karazija, A. A. Bandzaitis and A. P. Jucys, *ibid.* **3**, 377 (1963); V. Vanagas and A. P. Jucys, *ibid.* **2**, 199 (1962).

[12] M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962).

[13] D. Lurie and A. J. Macfarlane, J. Math. Phys. **5**, 565 (1964).

[14] K. T. Hecht, Bull. Am. Phys. Soc. **8**, 57 (1963).

[15] M. Konuma, K. Shima, and M. Wada, "Simple Lie Algebras of Rank 3 and Symmetries of Elementary Particles in the Strong Interaction" [Progr. Theoret. Phys. (Kyoto) Suppl.] (to be published).

[16] S. Gasiorowicz, "A Simple Graphical Method in the Analysis of SU_3," ANL-6729.

[17] J. Gabriel, "New Methods for Reduction of Group Representations Using an Extension of Schur's Lemma", J. Math. Phys. **5**, 494 (1964).

[18] P. Güttinger and W. Pauli, Z. Physik **67**, 743 (1931).

[19] E. P. Wigner, *Gruppentheorie und Ihre Anwendung auf die Quantenmechanik der Atomspektren* (Friedrick Vieweg und Sohn, Braunschweig, Germany, 1931), 1st ed.

[20] G. Racah, Phys. Rev. **62**, 438 (1942); the generalized version is discussed by Racah, Ref. 6, p. 91.

The mapping: operator onto states, has been discussed by Schwinger in a different, but equivalent context in his treatment of quantum mechanics as a unitary geometry.[21] Schwinger points out there exist two essential ways in which the vector space of operators is distinguished from the vector space[22] of the states: *the adjoint operation and the multiplication of elements are defined in the same space.* Both distinctions are of essential importance to the present work. A particularly interesting remark of Schwinger's concerns the Hermitian operators: *the geometry of the Hermitian operators is Euclidian.* The results to follow are illustrations of Schwinger's observations.

The generalization to tensor operators of arbitrary unitary groups thus proceeds smoothly with no new ideas required. Where then is the difficulty? The difficulty occurs when one seeks results equivalent to the Wigner–Eckart theorem: the tensor operator mapping is not one-to-one, for there exist, in the general case, *several* tensor operators that map into the same state. This is the "multiplicity problem," whose canonical solution[23] is discussed in Sec. III below.

The original definition of the Wigner coefficients ("matrix elements of the Wigner operator") emphasized a quite different aspect: Wigner introduced his matrices as coefficients of vector coupling.[24] This is an equally fundamental, and fully equivalent property possessed by tensor operators. We show that: *the matrix elements of a tensor operator are generalized coupling coefficients for the group.* It is not difficult to prove this assertion, but this simplicity should not blind one to the significance of the result.

Let $\mathcal{O}((m))$ be a tensor generator, and consider the matrix elements: $\langle(m'')|\ \mathcal{O}((m))\ |(m')\rangle$ where $|(m')\rangle$ and $|(m'')\rangle$ are basis states in SU_n. Let us now consider two (kinematically) independent SU_n systems 1 and 2 whose generators are $\{X_A^{(1)}\}$ and $\{X_A^{(2)}\}$ and whose states are $|(a)\rangle$ and $|(b)\rangle$. Consider next the direct product representation whose bra vectors are defined by

$$\langle(c)| \equiv \sum_{(a),(b)} \langle(c)|\ \mathcal{O}((a))\ |(b)\rangle\langle(a)|\ \langle(b)|. \tag{6}$$

Under the combined generators $X_A^{(T)} \equiv X_A^{(1)} + X_A^{(2)}$, the right-hand side transforms as

$$\begin{aligned}\Sigma(\cdots)X_A^{(T)} &= \sum_{(a'),(b')} \langle(a')|\ \langle(b')| \\ &\times \{ \sum_{(a),(b)} \langle(c)|\ \mathcal{O}((a))\ |(b)\rangle\cdot[\langle(a)|\ X_A^{(1)}\ |(a')\rangle\delta_{(b)}^{(b')} \\ &+ \langle(b)|\ X_A^{(2)}\ |(b')\rangle\delta_{(a)}^{(a')}]\}. \end{aligned} \tag{7}$$

The terms in curly brackets, $\{\cdots\}$, can be simplified using the defining tensor operator property

$$[X_A, \mathcal{O}((m))] \equiv \sum_{(m')} \langle(m')|\ X_A\ |(m)\rangle\mathcal{O}((m')), \tag{8}$$

and the fact that the matrices of the generators are the same for all realizations of the abstract group. Thus one finds that

$$\begin{aligned}\{\cdots\} &= \langle(c)|\ [X_A, \mathcal{O}((a'))]\ |(b')\rangle \\ &\quad + \langle(c)|\ \mathcal{O}((a'))X_A\ |(b')\rangle \\ &= \langle(c)|\ X_A\mathcal{O}((a'))\ |(b')\rangle \\ &= \sum_{(c')} \langle(c)|\ X_A\ |(c')\rangle\ \langle(c')|\ \mathcal{O}((a'))\ |(b')\rangle. \end{aligned} \tag{9}$$

Introducing this result into Eq. (7) it follows that

$$\begin{aligned}\langle(c)|\ X_A^{(T)} &= \sum_{(a'),(b'),(c')} \langle(a')|\ \langle(b')|\ \langle(c')|\ \mathcal{O}((a'))\ |(b')\rangle \\ &\quad \times \langle(c)|\ X_A\ |(c')\rangle, \\ &= \sum_{(c')} \langle(c)|\ X_A\ |(c')\rangle\ \langle(c')|. \end{aligned} \tag{10}$$

The direct-product states therefore transform properly under the generators $\{X_A^{(T)}\}$, and thus one concludes that they form a representation of SU_n given by the appropriate labels in the pattern (c).

For convenience we have carried out the demonstration using bra vectors, but the use of ket vectors is equally direct, employing matrix elements of the result that $X^\dagger = X_{-A}$.

By means of the tensor operator mapping, one might expect that precisely the same coupling property would hold for the tensor operators themselves. This is correct if one of the following two interpretations holds:

(1) The operators $\mathcal{O}((a))$ and $\mathcal{O}((b))$ act on kinematically independent systems, or, (2) the operators $\mathcal{O}((a))$ and $\mathcal{O}((b))$ act on the *same* system, but the realization of the generators $\{X_A\}$ has the derivative property [i.e., $X(fg) = (Xf)g + f(Xg)$].

If either of these interpretations is valid, then we may conclude (by a repetition of the previous argument) that the coupled operator, $\mathcal{O}((c))$,

$$\mathcal{O}((c)) \equiv \sum_{(a),(b)} \langle(c)|\ (a)\ |(b)\rangle\mathcal{O}((a))\mathcal{O}((b)) \tag{11}$$

is a tensor operator as specified by the labels (c).

[21] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **45**, 1542 (1959); **46**, 257, 570, 833, 1401 (1960); **47**, 1075 (1961); **48**, 603 (1962).

[22] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **46**, 257 (1960); see especially p. 264.

[23] This solution to the multiplicity problem was sketched briefly in Ref. 8.

[24] Ref. 6 (b), p. 341 discusses several other useful and rather distinct ways to regard the Wigner coefficients—an indication of the importance of these coefficients.

For brevity, this technique for coupling tensor operators will be designated as

$$\mathcal{O}(c) = [\mathcal{O}((a)) \otimes \mathcal{O}((b))]_{(c)}. \tag{12}$$

Owing to the multiplicity problem this coupling procedure is, at the moment, ambiguous to the extent that the tensor operators of a specific type are not unique. We remedy this defect below, and with this in mind reserve the $\otimes$ notation for the canonically defined, normalized, coupling coefficients.

III. MULTIPLICITY PROBLEM FOR TENSOR OPERATORS

It is essential to note that the classification of operators by the tensor operator mapping is not necessarily exhaustive or complete, that is, there exists the possibility that two independent operators may possess the same tensor-operator labels. It is not difficult to see, by examples, that this situation—the existence of multiplicity—is, in fact, the general one. For the general unimodular unitary group, SU_n, there exist $(n-1)$ independent invariant operators; it follows that there exist $(n-1)$ independent tensor operators classified by their transformation properties as belonging to the adjoint representation. These operators are designated as $\{X_A^{(k)}\}$, and result by coupling k times the operators $\{X_A^{(1)}\}$ corresponding to the generators, using the symmetric coupling coefficients defined in I. (The $\{X_A^{(k)}\}$ have been discussed more fully in III, Appendix I.) It is clear from this example that for the unitary groups the problem of multiplicity in the tensor operator classification certainly exists for $n > 2$.

It is the purpose of the present section to examine the multiplicity problem for the general SU_n group, and to determine, in particular, the precise extent of the multiplicity (the number of transformationally equivalent operators belonging to a given representation labeling). We demonstrate a surprisingly simple general result for the SU_n group: *the operator multiplicity labeling problem may be mapped in a one-to-one fashion onto the state labeling problem.* With the *structure* of the multiplicity problem thus determined, the solution lies at hand immediately, in the form of *an additional operator labeling using a Gelfand state labeling pattern.* A unique and complete operator labeling scheme thus exists, and the generalization of the Racah–Wigner calculus developed for SU_2 can then be shown explicitly for all SU_n, upon demonstrating a canonical basis for the second labeling pattern.

Demonstration of the Assertion

Let us proceed now to the demonstration of the result: *the operator multiplicity labeling problem for SU_n may be mapped in a one-to-one fashion onto the state labeling problem.* Consider the group SU_n and assume that the operator, $\mathcal{O}$, has been classified by its transformation properties under $\{X_A\}$, to belong to the state labeled by (m). That is,

$$[X_A, \mathcal{O}((m))] = \sum_{(m')} \langle (m')|\, X_A\, |(m)\rangle \mathcal{O}((m')), \tag{13}$$

where (m) denotes the representation labels $m_{1,n}$, $m_{2,n}$, $\cdots$, $m_{n,n} = 0$, and the subgroup labels $m_{i,j}$, $1 \leq i, j \leq n-1$; similarly (m') denotes the same representation labels, but a general set of subgroup labels.

The problem now is to determine the matrix elements of the operator $\mathcal{O}((m))$, i.e., $\langle (b)|\, \mathcal{O}((m))\, |(a)\rangle$, where $|(a)\rangle$, $|(b)\rangle$ denote arbitrary states in SU_n. The problem may be simplified by noting that, from Cartan's fundamental result,[25] any state $|(a)\rangle$ may be obtained from the maximal state $|(a)_{\max}\rangle$. Thus one may write

$$|(a)\rangle = F_-\, |(a)_{\max}\rangle, \tag{14}$$

where the F_- is a definite polynomial in the generators whose typical term may be put in the normal form, $(E_{n,i})^a (E_{n,j})^b \cdots (E_{2,1})^m (H)^n \cdots$, which consists only of *lowering* operators and diagonal operators, ordered from the left by decreasing index n. (Similarly the state $|(b)\rangle$ may be written as: $|(b)\rangle = G_-\, |(b)_{\max}\rangle$ where G_- is a lowering polynomial operator.)

We want to show first that the general matrix element $\langle (b)|\, \mathcal{O}((m))\, |(a)\rangle$ is a linear function of the matrix elements $\langle (b)|\, \mathcal{O}((m))\, |(a)_{\max}\rangle$. To do this, one notes that

$$F_- = a_1 + X_- F_-^{(1)}, \tag{15}$$

where $F_-^{(1)}$ is of lower degree and thus

$$\begin{aligned}
&\langle (b)|\, \mathcal{O}((m))\, |(a)\rangle \\
&= \langle (b)|\, \mathcal{O}((m)) F_-\, |(a)_{\max}\rangle \\
&= a_1 \langle (b)|\, \mathcal{O}((m))\, |(a)_{\max}\rangle \\
&\quad + \sum_{(b')} \langle (b)|\, X_-\, |(b')\rangle \langle (b')|\, \mathcal{O}((m)) F_-^{(1)}\, |(a)_{\max}\rangle \\
&\quad - \sum_{(m')} \langle (m')|\, X_-\, |(m)\rangle \langle (b)|\, \mathcal{O}((m')) F_-^{(1)}\, |(a)_{\max}\rangle.
\end{aligned} \tag{16}$$

The matrix elements $\langle \cdots |\, X_-\, | \cdots \rangle$ are generator matrix elements and thus completely known.

Repeating the process, the degree of $F_-^{(1)}$ is successively lowered to zero, and the desired result follows: the general matrix element $\langle (b)|\, \mathcal{O}((m))\, |(a)\rangle$

[25] See Refs. 1 and 6 for a discussion of this result.

can be determined from the special matrix elements having (a) replaced by $(a)_{\max}$.

Next one introduces the form $G_- |(b)_{\max}\rangle$ for the general state and obtains

$$\langle(b)| \mathcal{O}((m)) |(a)_{\max}\rangle = \langle(b)_{\max}| G_-^\dagger \mathcal{O}((m)) |(a)_{\max}\rangle, \quad (17)$$

and uses the fact that the Hermitian adjoint of a lowering polynomial is a raising polynomial. Just as before, we may commute the operators X_+ through $\mathcal{O}((m))$, where they annihilate against $|(a)_{\max}\rangle$. That is, we use successively, (the existence of H_i operators in $G_-^\dagger$ is no restriction),

$$G_-^\dagger = b_1 + F_+^{(1)} X_+; \qquad X_+ |(a)_{\max}\rangle = 0 \quad (18)$$

and

$$[X_+, \mathcal{O}((m))] = \sum_{(m')} \langle(m')| X_+ |(m)\rangle \mathcal{O}((m')) \quad (19)$$

to successively reduce the matrix element $\langle(b)| \mathcal{O}((m)) |(a)_{\max}\rangle$ to a linear combination of the further specialized matrix elements

$$\langle(b)_{\max}| \mathcal{O}((m)) |(a)_{\max}\rangle.$$

In summary, we may determine the most general matrix element $\langle(b)| \mathcal{O}((m)) |(a)\rangle$ from a linear combination of the matrix elements

$$\langle(b)_{\max}| \mathcal{O}((m)) |(a)_{\max}\rangle,$$

where the coefficients in the linear combination can be calculated, in principle, from knowledge of the generator matrix elements alone.

This has the immediate consequence that the maximum number of independent matrix elements is dim $[(m)]$, but in fact, the number is far less. This limitation results from the diagonal operators H_i—$(n-1)$ in number—which imply that

$$[m_i((b)_{\max}) - m_i((a)_{\max}) - m_i((m))] \times \langle(b)_{\max}| \mathcal{O}((m)) |(a)_{\max}\rangle = 0 \quad (20)$$

for $i = 1, 2, \cdots, n-1$. Thus, the existence of a nonvanishing matrix element $\langle(b)_{\max}| \mathcal{O}((m)) |(a)_{\max}\rangle$ requires that, for fixed (m), a system of $(n-1)$ linear equations hold for the $(n-1)$ differences,

$$\Delta_i \equiv b_{i,n} - a_{i,n}. \quad (21)$$

More explicitly,

$$\sum_{i=1}^{j} \Delta_i - j\,\Delta_j = (j+1) \sum_{i=1}^{j} m_{ij} - (j) \sum_{i=1}^{j+1} m_{i,j+1}. \quad (22)$$

Given a set of state labels (m) for the operator $\mathcal{O}((m))$ these equations determine definite values for the $(n-1)$ Δ's and a nonvanishing matrix element $\langle(a+\Delta)_{\max}| \mathcal{O}((m)) |(a)_{\max}\rangle$ may be assigned to the operator $\mathcal{O}((m))$. [The existence of at least one such operator follows from the fact that the operator corresponding to the maximal state (m) exists for maximal (Δ), and is unique and nonvanishing, (essentially the Cartan result again). This operator may be lowered to yield nonmaximal Δ's.]

It is essential to note now, that although the state labels (m) uniquely assign the set $\{\Delta_i\}$, the converse is not true. It is easy to see that the number of different sets (m), corresponding to a fixed set $\{\Delta_i\}$ is precisely the degeneracy of the states $|(m)\rangle$ corresponding to fixed diagonal quantum numbers $H_i \to m_i((m))$. This follows from the fact [which is easily verified by use of III, Eq. (8) and Eq. (22) above] that

$$\Delta_k = -(k-1)m_{k-1} + \sum_{i=k}^{n-2} m_i + n m_{n-1}, \quad (23)$$

and thus the sets Δ_i and m_i are equivalent, in the sense that one implies the other.

This proves the desired result, since we may now *define* (*for purposes of discussion*) the general tensor operator by the labeling scheme: $\mathcal{O}_{(m)}^{(M)}$, where the Gelfand labels correspond to:

(1) the transformation properties of the operator $\mathcal{O}$ under the generators $\{X_A\}$ which determine (m) in accord with Eq. (2), and,

(2) the labels (M) which correspond to the set of matrix elements

$$\langle((a)_{\max} + \Delta((M)))| \mathcal{O}_{(M)}^{(M)} |(a)_{\max}\rangle$$

with all other matrix elements defined to vanish.

This set of matrix elements is a complete and unique specification of the tensor operator denoted by the symbol $\mathcal{O}_{(m)}^{(M)}$. Let us note that this definition of the general tensor operator allows for the possibility that *certain* of the matrix elements

$$\langle((a)_{\max} + \Delta((M)))| \mathcal{O}_{(m)}^{(M)} |(a)_{\max}\rangle$$

may vanish for particular choices of the states $(a)_{\max}$. (As an elementary example, consider the SU_2 tensor operator

$$\mathcal{O}\begin{pmatrix} 2j & 0 & 0 \\ 2j & 0 & 0 \\ & j+m & \end{pmatrix}$$

which corresponds to $\Delta_1 = -j$, i.e., $j_f - j_i = -j$. Clearly this operator vanishes for an initial state having $j_i < j$.)

We must now show that such a vanishing cannot occur for *all* values of $(a)_{\max}$. In other words, we must now prove that the general tensor operator actually possesses the maximum dimensionality.

Dimensionality Considerations

It has been shown above that the number of independent, nonidentically-vanishing fundamental

matrix elements of the operator $\mathcal{O}((m))$ is equal to or less than the dimension of (m). We now prove that the number of such matrix elements is exactly equal to dim $[(m)]$. To do this we must first prove a lemma:

If the matrix element $\langle (b)_{\max}| \mathcal{O}^{(M)}_{(M)} |(a)_{\max}\rangle \neq 0$ *for* $(a)_{\max}$, $(b)_{\max}$, *then the matrix element*

$$\langle ((b) + (\Lambda))_{\max}| \mathcal{O}^{(M)}_{(M)} |((a) + (\Lambda))_{\max}\rangle \neq 0,$$

where $[\Lambda]$ *has* $\Lambda_1 \geq \Lambda_2 \geq \Lambda_3 \geq \cdots \geq 0$.

Proof of the above lemma requires itself an ancillary result: an operator $\mathcal{O}^{(M)\min}_{(m)}$ whose upper state label (labeling the multiplicity) is minimal when operating upon an *maximal* state $|(a)_{\max}\rangle$, gives a nonzero result only for a *maximal* final state $|(b)_{\max}\rangle$, and a *minimal* set of tensor operator labels $(m)_{\min}$. That is,

$$\langle (b)| \mathcal{O}^{(M)\min}_{(m)} |(a)_{\max}\rangle = \delta^{(b)\max}_{(b)} \delta^{(M)\min}_{(m)} \langle (b)_{\max}| \mathcal{O}^{(M)\min}_{(M)\min} |(a)_{\max}\rangle. \quad (24)$$

The demonstration of this ancillary result is a straightforward application of III, Eq. (8) and Eq. (23) above, and is easily verified. With this result in hand, let us now prove the lemma.

First observe that the operator, $\mathfrak{D}$, defined by

$$\mathfrak{D} \equiv \sum_{(\lambda)} (-)^{\delta((\lambda))} \mathcal{O}^{(\Lambda)\max}_{(\lambda)} \mathcal{O}^{(M)}_{(m)} \mathcal{O}^{(\tilde{\Lambda})\max}_{(\tilde{\lambda})} \quad (25)$$

transforms under the generators, $\{X_A\}$, precisely like the operator $\mathcal{O}^{(M)}_{(m)}$ itself. This proves that the operator $\mathfrak{D}$ possesses the lower pattern (m). Next let us take matrix elements of the operator $\mathfrak{D}((m))$ between the *maximal* states $(e)_{\max}$ and $(f)_{\max}$.

$$\langle (e)_{\max}| \mathfrak{D}((m)) |(f)_{\max}\rangle = \sum_{(\lambda),(c),(d)} \langle (e)_{\max}| \mathcal{O}^{(\Lambda)\max}_{(\lambda)} |(c)\rangle \times \langle (c)| \mathcal{O}^{(M)}_{(m)} |(d)\rangle \langle (d)| \mathcal{O}^{(\tilde{\Lambda})\min}_{(\tilde{\lambda})} |(f)_{\max}\rangle. \quad (26)$$

We now show that the sum, Eq. (26), reduces to a *single term*. By our ancillary result above, the state (d) must be a maximal state, and moreover, by the same result, $(\tilde{\lambda})$ must be minimal. This removes the sums over both (λ) and (d), and results in

$$\langle (e)_{\max}| \mathfrak{D}((m)) |(f)_{\max}\rangle = \sum_{(c)} \langle (e)_{\max}| \mathcal{O}^{(\Lambda)\max}_{(\Lambda)\max} |(c)\rangle \times \langle (c)| \mathcal{O}^{(M)}_{(m)} |(d)_{\max}\rangle \langle (d)_{\max}| \mathcal{O}^{(\tilde{\Lambda})\min}_{(\tilde{\Lambda})\min} |(f)_{\max}\rangle. \quad (27)$$

Next we note that the matrix element

$$\langle (e)_{\max}| \mathcal{O}^{(\Lambda)\max}_{(\Lambda)\max} |(c)\rangle$$

vanishes unless the state (c) is *maximal*. [The proof of this assertion follows from III, Eq. (8) and Eq. (23) above, as earlier for our ancillary result.] Thus the last remaining summation, over (c), is removed, and Eq. (26) reduces to a single term as asserted.

Since, however, the operator $\mathcal{O}^{(M)}_{(m)}$ taken between maximal states vanishes (by definition) unless $(m) = (M)$, we see that the operator $\mathfrak{D}((m))$, taken between maximal states, also vanishes except for $(m) = (M)$. This proves that the operator $\mathfrak{D}$ is nothing else but a (nonzero) multiple of the original operator $\mathcal{O}^{(M)}_{(m)}$.

The result demonstrated implies, moreover, that the matrix element $\langle (e)_{\max}| \mathfrak{D}^{(M)}_{(M)} |(f)_{\max}\rangle$ does not vanish. This is true for the matrix element of $\mathcal{O}^{(M)}_{(M)}$ by hypothesis, and is true also for the matrix elements of $\mathcal{O}^{(\Lambda)\max}_{(\Lambda)\max}$ and $\mathcal{O}^{(\tilde{\Lambda})\min}_{(\tilde{\Lambda})\min}$ [Since these latter operators are unique (multiplicity 1), and the matrix elements in question are nonvanishing by Cartan's result cited earlier].

This, however, proves the lemma for if we let $(e)_{\max} = (b + \Lambda)_{\max}$ and $(f)_{\max} = (a + \Lambda)_{\max}$, then the result, Eq. (27), shows that the matrix element $\langle (b + \Lambda)_{\max} |\mathcal{O}^{(M)}_{(M)} |(a + \Lambda)_{\max}\rangle$ is nonvanishing if $\langle (b)_{\max} |\mathcal{O}^{(M)}_{(M)} |(a)_{\max}\rangle$ is nonvanishing.

Now we proceed to prove the dimensionality of the general operator $\mathcal{O}^{(M)}_{(m)}$ equals dim $[(m)]$. In demonstrating this, an inductive approach is employed. We thus assume that the result is true for $\mathcal{O}((m))$, that is, there exist dim $[(m)]$ independent, nonidentically-vanishing matrix elements of $\mathcal{O}((m))$.

Consider the direct product operator

$$\mathcal{O}^{[1\dot{0}]} \otimes \mathcal{O}((m)),$$

and matrix elements of this product operator. By construction (Paper II) we know that there exist n independent nonzero matrix elements of the $\mathcal{O}^{[1\dot{0}]}$ operator. Similarly, by assumption, there exist dim $[(m)]$ independent nonzero matrix elements of the $\mathcal{O}((m))$ operator. We wish to prove now that this implies that there exist $n \times$ dim $[(m)]$ independent nonzero matrix elements for the product operator. To see this, take matrix elements of

$$\mathcal{O}^{[1\dot{0}]} \otimes \mathcal{O}^{(M)}_{(M)}$$

between *maximal* states. That is,

$$\langle (a)_{\max}| \mathcal{O}^{[1\dot{0}]} \otimes \mathcal{O}^{(M)}_{(M)} |(b)_{\max}\rangle = \sum_{(c)} \langle (a)_{\max}| \mathcal{O}^{[1\dot{0}]} |(c)\rangle \langle (c)| \mathcal{O}^{(M)}_{(M)} |(b)_{\max}\rangle. \quad (28)$$

Once again the sum reduces to a single term. This results from the vanishing of the matrix element $\langle (c)| \mathcal{O}^{(M)}_{(M)} |(b)_{\max}\rangle$, for all (c) except $(c) = (c)_{\max}$. [Application of III, Eq. (8) and Eq. (23) above demonstrates this result, which is a particular property of operators with identical upper and lower pattern labels operating on maximal states.]

The resultant product of matrix elements is

again nonvanishing, provided a suitable intermediate state $(c)_{\max}$ may be found. That this is always possible for $\mathcal{O}^{(M)}_{(M)}$ is the content of the lemma proved above. (The matrix elements of $\mathcal{O}^{[10]}$ do not vanish between lexical maximal states.) This establishes that there exist $n \times \dim [(m)]$ independent nonzero matrix elements for the product operator.

On the other hand, the product operator may be reduced [using the Wigner coefficients of II] into operators belonging to representations

$$\begin{aligned}&[m_{1,n}+1, m_{2,n}, \cdots, 0]\\ &\quad \dot{+} [m_{1,n}, m_{2,n}+1, m_{3,n}, \cdots, 0] \dot{+} \cdots\\ &\quad \dot{+} [m_{1,n}, m_{2,n}, \cdots, m_{n-1,n}+1, 0]\\ &\quad \dot{+} [m_{1,n}-1, m_{2,n}-1, \cdots, m_{n-1,n}-1, 0],\end{aligned}$$

[either once or zero]. The *maximum* number of independent nonzero matrix elements for these operators is given by $\sum_{(m')} \dim [(m')]$, by the previous section. But this maximum number is precisely $n \times \dim [(m)]$. Therefore we conclude *that every operator in the direct product possesses precisely* $\dim [(m)]$ *independent nonidentically-vanishing matrix elements.* Since all tensor operators are obtainable by reduction of the direct products (using only the Wigner coefficients given in II, in principle) we have therefore proved the desired result in general.

Two points in this proof should be noted: (a) it is *not* necessary for n representations to occur in the reduction of the direct product, for the proof requires consideration only of those representation that actually do occur; (b) it should be noted that we do *not* need to reduce the direct product into operators properly labeled by *two* patterns. (The tensor operator pattern itself is, of course, defined completely by the $[1\dot{0}]$ Wigner coefficients.)

Symmetry Vanishings[26]

The familiar SU_2 example shows that the (SU_2) operator $\mathcal{O}^{(M)}_{(m)}$ does not vanish in general, but that matrix elements of this operator may vanish for special cases. We must now examine this point critically for the SU_n case.

It has already been noted that there exist n operators belonging to the representation label $[1\dot{0}]$ in SU_n, and that there exist n conjugate operators belonging to the conjugate representation label $[\dot{1}0]$. These operators are uniquely labeled by the n state labels belonging to $[1\dot{0}]$ and to $[\dot{1}0]$ and, in turn, these n upper patterns determine the n Δ values: $(1\dot{0})$, $(01\dot{0})$, $\cdots$. Moreover, there exists an initial state for which *all* of the matrix elements are nonvanishing (this can be seen explicitly from II, or by the results of the previous section). Thus there exists a fixed representation $[M]$ such that

$$\begin{aligned}[m_{1,n}, m_{2,n}, \cdots, 0] \otimes [1\dot{0}] = [m_{1,n}+1, \cdots, 0] \dot{+} \cdots\\ \dot{+} [m_{1,n}-1, m_{2,n}-1, \cdots, 0] \qquad (29)\end{aligned}$$

and every representation on the right actually occurs. In other words, one has the result that

$$n \times \dim [(M)] = \sum_{\text{all } n\Delta \text{ values}} \dim [(M+\Delta)]. \qquad (30)$$

(The condition on (M) is simply that the pattern $[M+\Delta]$ be lexical.)

This algebraic result must, however, remain valid even for representations $[m]$ that lead to nonlexical patterns $[m+\Delta]$. In order for this not to lead to a contradiction one sees that this requires the algebraic extension of the meaning of Eq. (30) to remain valid for the sum extended over either (a) the *allowed* representations only or (b) *all* representations $[m+\Delta]$. This in turn requires that $\sum_{(m')} \dim [(m')] = 0$ where the sum is restricted to the *unallowed* representations. Since the dimensions of the unallowed representations do not vanish in general (an example easily shows this), it follows that the algebraic extension of Eq. (30) consists in the extension of the dimension formula to *negative* dimensions.

This is quite reasonable, as the SU_2 paradigm shows. Consider the case $\mathbf{j}_1 \times \mathbf{j}_2$, when $j_2 > j_1$. If we apply the mapping result (operator $\rightarrow$ states) to the $\times \mathbf{j}_1$, it follows that the $2j_1+1$ terms j_2+j_1, $j_2+j_1-1, \cdots, j_2-j_1$ occur in the Kronecker product. This is as it should be; but suppose we apply the mapping result to $\times \mathbf{j}_2$. There the mapping result (operators $\rightarrow$ states) shows that the *general* result contains $2j_2+1$ terms, but specific matrix elements may nonetheless vanish. Comparing the dimension result Eq. (30) in the two cases, one sees that the unallowed representations j_2-j_1-1, $j_2-j_1-2, \cdots, j_1-j_2$ obey the requirement $\sum \dim [(m')] = 0$.

The SU_2 example is quite instructive for it shows that the unallowed representations may be divided into three types: those of positive nonzero dimension, those of negative nonzero dimension, and those of zero dimension. The nonzero dimensions occur in pairs, related by the association: $j \rightleftarrows -j-1$, while the zero-dimension case is self-associated. This provides the rule by which the unallowed representations are identified in general. Of course, one hardly needs to approach

[26] We would like to thank Professor C. N. Yang for the favor of an enlightening conversation on this subject.

the problem for SU_2 in this complicated way! Nevertheless this example illustrates the principle underlying the general (SU_n) case, as is discussed explicitly below.

The existence of matrix elements that vanish for special cases [and the existence of rules for the determination of such cases] are a small price to pay for the insight that the mapping result (operators $\rightarrow$ states) yields into the structure of the general operator on SU_n.

The determination of the conditions on $(a)_{\max}$ under which the operator matrix element

$$\langle (a + \Delta)_{\max} | \, \mathcal{O}^{(M)}_{(m)} \, | (a)_{\max} \rangle$$

vanishes is equivalent to the determination of the unallowed states in the Kronecker product decomposition of $[M] \otimes [a]$. The previous section established this decomposition (with multiplicity) in general. It follows that the algebraic extension of the dimension formula involves simply the extension to negative dimensions. [A complete discussion of this problem would determine not only the conditions on $(a)_{\max}$ for which matrix elements have symmetry vanishings, but also the specific operator $\mathcal{O}^{(M)}_{(m)}$ connected with this vanishing. We shall, however, defer this point until after a *canonical* meaning has been established for upper patterns.]

Let us turn to the problem of identifying the "unallowed states," or more precisely the vanishing matrix elements, to be associated with the operator $\mathcal{O}^{(M)}_{(m)}$. The essential point has been illustrated in the SU_2 paradigm. The problem is to exhibit explicitly the association between unallowed states of positive and negative dimensions. Since the representations of negative dimension violate the lexicality requirement they are easily identified in the collection of representations obtained from the mapping result; the two associated representations (positive $\rightleftarrows$ negative dimension) are thus identified as unallowed. Similarly, the self-associated representations are also easily identified. Let us detail these results explicitly. Consider the Weyl dimension formula

$$\dim [(m)] = \prod_{i,j} \frac{(m_{i,n} - m_{i+j,n} + j)}{j}. \tag{31}$$

Consider now the group of linear transformations on the $m_{i,n} = \sum_j A_{i,j} m_{j,n}$ which (1) leave dim $[(m)]$ invariant and (2) carry dim $[(m)]$ into $\pm$ dim $[(m)]$. The group corresponding to (1) is already familiar: it is isomorphic to the group S_n, which occurred already as the symmetry group of the vector diagram (see I). The group corresponding to (1) and (2) is also familiar from I; it is the *complete* symmetry group of the vector diagram corresponding to adjunction of the outer automorphism of conjugation. [Note, however, the realization of these groups differs: conjugation is a central reflection of the vector diagram, whereas it leaves dim $[(m)]$ invariant, for (m) unimodular.] This latter group is generated by the operations R_{ij} and $\mathcal{K}$

$$R_{ij}: m_{i,n} + n - i \rightarrow m_{j,n} + n - j, \qquad \text{all other } m_{i,n} \text{ unchanged}, \tag{32}$$

$$\mathcal{K}: m_{i,n} \rightarrow m_{n,n} + m_{1,n} - m_{n+1-i,n}. \tag{33}$$

Consider now a representation that arises from $\mathcal{O}^{(M)}_{(m)}$. The representations of negative dimension (but nonzero) have the characteristic that the $[m_{i,n}]$ are nonlexical. The operations of the group are now employed to transform this nonlexical pattern into a lexical pattern, thereby associating to each pattern of negative (nonzero) dimension the desired pattern of positive dimension.

Note added in proof: J. J. De Swart, Rev. Mod. Phys. **35**, 916 (1963), employs (without proof) a diagrammatic technique (attributed to D. R. Speiser, Proceedings of the Istanbul International Summer School, 1962) for reducing the Kronecker product in the octet model. The diagrammatic method is essentially equivalent to the algebraic method discussed above, with negative regions in the weight space (vector diagram) playing the role of the Weyl formula for negative dimension.

IV. A NOTATION FOR THE CANONICAL TENSOR OPERATORS

One of the immediate consequences of the operator mapping proven in the previous section is that it allows one to devise notations which uniquely label the general tensor operator. It follows from the mapping result that a *complete* characterization of the general operator in an SU_n space is given by two Gelfand patterns: one pattern designates completely the tensor operator (transformation) properties, the second pattern designating both the $\Delta m_{i,n}$ induced by the operator and some resolution of the multiplicity problem. We discuss shortly a canonical resolution of the multiplicity problem, but it is convenient at this point to introduce a notation to facilitate discussion of the general tensor operator. It is to be explicitly understood that this notation is to refer exclusively to the canonical resolution, as described later.

The first point to note is that the use of two Gelfand patterns is slightly redundant, since these patterns share a common row, the row $m_{i,n}$. This row is the longest in the Gelfand scheme and is

already distinguished by the fact that the $m_{i,n}$ are the Young symmetry pattern labels.

The second point to note is that the inequalities implicit in the Gelfand pattern *are not affected by inverting the diagram.*

We propose to exploit both points by assigning to operators a diamond-shaped pattern composed of two Gelfand patterns—one inverted—joined by a common row (the Young pattern labels). To make the operator $\rightleftarrows$ matrix element relation notationally obvious we designate the diamond shaped pattern—the operator pattern—by the bra-ket notation.

The SU_3 example makes the general result clear. Consider the general SU_3 operator; such an operator requires nine integers arranged in the pattern

$$\begin{array}{ccccc} & & \cdot & & \\ & \cdot & & \cdot & \\ \cdot & & \cdot & & \cdot \\ & \cdot & & \cdot & \\ & & \cdot & & \end{array}$$

The Young symmetry labels for this operator are the middle row; call these labels A_{13}, A_{23}, $A_{33} = 0$. The operator is referred to collectively as $\langle A \rangle = \langle A_{13}A_{23}0 \rangle$. The upper pattern labels are α_{12}, α_{22}, α_{11}. The lower pattern labels (the analogs to magnetic quantum numbers) are: a_{12}, a_{22}, a_{11}. We may refer to the operator $\langle A \rangle$ in various more or less specific ways:

$$\langle A \rangle = \langle A_{13}A_{23}A_{33} \rangle = \langle \begin{matrix} & \alpha_{11} & \\ \alpha_{12} & & \alpha_{22} \\ A_{13} & A_{23} & A_{33} \\ a_{12} & & a_{22} \\ & a_{11} & \end{matrix} \rangle \tag{34}$$

and further combinations suppressing upper or lower pattern labels. It is convenient to designate the upper pattern

$$\begin{bmatrix} & \alpha_{11} & \\ \alpha_{12} & & \alpha_{22} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}$$

by (α) and the lower pattern

$$\begin{bmatrix} A_{13} & A_{23} & A_{33} \\ & a_{12} & a_{22} \\ & & a_{11} \end{bmatrix}$$

by (a).

One of the genuine difficulties in dealing with the SU_n system is the lengthy and extensive detail required in the notation. Gelfand has given a valuable aid in introducing the "Gelfand pattern." The "double Gelfand patterns" seem similarly to recommend themselves. One point should be noted. The notation $\langle A \rangle$ makes it clear that the more useful object to discuss is not the Wigner *coefficient*, but rather the *operator* (the Wigner operator) defined by the matrix elements. As a familiar example let us note that in SU_2 the Wigner operator reads

$$\langle A \rangle = \left\langle \begin{matrix} & J + \Delta J & \\ 2J & & 0 \\ & J + M & \end{matrix} \right\rangle, \tag{35}$$

which specifies fully a specific component of the Wigner coefficient, but for arbitrary j_{initial}.

The Concept of a Reduced Tensor Operator

The essential content of the results obtained above is to make explicit the "coordinatization" of the (dim $[A]$)^2 space of the canonical tensor operators; the notation is designed to make this content obvious.

A particular merit of the notation is that this "geometrical" view, which it implies, suggests further extensions to consider. Such, for example, is the idea of a *reduced tensor operator.* In geometrical terms, a reduced tensor operator is the projection of a given (arbitrary) tensor operator in SU_n onto the space of the canonical tensor operators in the SU_n or *lower unitary spaces* (SU_m, $m < n$). This is but a generalization of the ideas implicit in the Wigner–Eckart theorem.

The simplest application is that of the reduced matrix element which appears in the Wigner–Eckart theorem. For the present language, the reduced matrix element for a tensor operator $\mathcal{O}((m))$ in SU_n is a particular matrix element of the reduced matrix element *operator*, $\mathcal{O}((m))\cdot\langle A \rangle$ where $\langle A \rangle$ is a canonical tensor operator in SU_n, and the inner ("dot") product is carried out using the conjugation operation. Alternatively one may view this result as the expansion of an arbitrary tensor operator in SU_n,

$$\mathcal{O}((m)) = \sum_{(A)} (\mathcal{O}((m))\cdot\langle A \rangle)\langle A \rangle. \tag{36}$$

It is clear from this equation that one should distinguish between the two possible orders (reduced matrix element operator on the left, or on the right), since this affects the (SU_n) representation that actually enters into the specific matrix elements. In practice, this is often unnecessary since the operator notation partially distinguishes the initial (rhs) representations from the final (lhs), and this distinction is complete for the reduced operator decomposition of the $\langle A \rangle$ itself. The conjugation properties—established in the final section—specify the phase properties of the reduced matrix element operator (for Hermitian operators).

The extension of the reduced operator concept to the canonical tensor operators themselves is

clearly the next step. Consider for example the operator $\langle A_n \rangle$, where the n specifies that $\langle A \rangle$ is a canonical tensor operator in U_n. The inner product (in U_{n-1}) of this operator with a canonical tensor operator $\langle A_{n-1} \rangle$ in U_{n-1} defines an U_{n-1} invariant operator. *This is precisely the reduced Wigner operator*, where matrix elements for the $\langle \dot{1}0 \rangle$ and $\langle 1\dot{0} \rangle$ operators were introduced explicitly in II. By this remark one accomplishes, by iteration, the complete reduction of the canonical tensor operators. Note that a specific *order* is to be specified in the reduction.

A special notation is indicated to describe the reduced tensor operators. Such a notation was introduced in II, for the special case of the fundamental Wigner operators, and can be modified to fit the general case. Let us designate the reduced canonical tensor operator by the notation

$$\begin{pmatrix} (\alpha): n \\ (\beta): n-1 \end{pmatrix} \equiv \left\langle \begin{matrix} \alpha \\ A \\ a \end{matrix} \right\rangle \cdot \left\langle \begin{matrix} \beta \\ B \\ b \end{matrix} \right\rangle , \tag{37}$$

where the operator $\langle A \rangle$ is in U_n, the operator $\langle B \rangle$ is in U_{n-1} and the dot product implies a conjugation of $\langle B \rangle$.

We must justify this notation; among other things, for example, it suppresses some labels in a not-too-obvious manner. Let us note first, that the notation "$(\alpha) : n$" designates the upper (U_n) Gelfand pattern of $\langle A \rangle$, including the labels $[A]$. Similarly "$(\beta) : n - 1$" designates the upper (U_{n-1}) Gelfand pattern of $\langle B \rangle$, including the labels $[B]$. Thus only the labels "(a)" and "(b)" on the right-hand side are suppressed. In order for the dot product to be meaningful, however, the labels $a_{i,n-1}$ and $b_{i,n-1} = B_{i,n-1}$ must agree identically. The remaining labels $a_{i,j}$ and $b_{i,j}$ ($j \leq n - 2$) are summed over in the dot product. Thus the reduced operator notation is in fact a *complete specification.*

Comparing this notation to that used in Eq. (60) of II, one sees that the patterns (α) and (β) furnish a complete and unambiguous designation of the *change* in the representation labels $[m_{i,n}] \rightarrow [m'_{i,n}]$ and $[m_{i,n-1}] \rightarrow [m'_{i,n-1}]$ induced by the tensor operator. This is a direct generalization of the notation of II, which concerned only the special changes (± 1, in one index only) associated with the fundamental Wigner operator.

The complete decomposition of the canonical U_n tensor operator is thus given by iterating the basic decomposition:

$$\left\langle \begin{matrix} \alpha \\ A_n \\ a \end{matrix} \right\rangle = \sum_{(\beta)} \begin{pmatrix} (\alpha): n \\ (\beta): n-1 \end{pmatrix} \left\langle \begin{matrix} \beta \\ B \\ a \end{matrix} \right\rangle , \tag{38}$$

where the Gelfand pattern given by B and a on the right specifies the U_{n-1} pattern (a). It is not very valuable to try to indicate the complete result of this iteration. [Comparing this result with Eq. (60) in II shows that the "U_k Wigner coefficient" of II is just the matrix element of the tensor operator $\langle A_k \rangle$ evaluated between U_k representations where the U_{k-1} labels are *unchanged.* It is not essential therefore to distinguish these matrix elements from the usual matrix elements of the general Wigner operator.]

Let us note again that the phase properties of the reduced canonical tensor operators follows from the general conjugation and Hermiticity properties of the canonical tensor operators themselves (see Sec. VI).

The Racah Operator

Although the present paper is already unwieldy in size, we cannot refrain from noting here that the canonical tensor operator notation suggests of itself the idea of the "Racah operator." Consider two tensor operators in U_n, $\langle A \rangle$ and $\langle B \rangle$. It is clearly suggested that one employ the Wigner coefficients themselves to define a new operator, e.g., $\langle A \rangle \otimes \langle B \rangle \equiv \langle C \rangle$. To make the specification more complete, one must indicate just which Wigner coefficient is used in the coupling. This, however, necessarily fails to be a complete designation, since a complete upper pattern is required for the operator $\langle C \rangle$. It is clear that the construction of new operators by this cross product does not define a unique tensor operator but rather a *sum* of such operators. The coefficients in this sum are just the *Racah operators*

$$R \equiv \langle A \rangle \overset{(\delta)}{\otimes} \langle B \rangle \cdot \langle C \rangle ,$$

where (δ) specifies the specific Wigner operator used in the coupling. From this sketch of the Racah operator one sees at once that the Racah operator plays a role as a form of "coupling coefficient for upper patterns." In a succeeding paper, we hope to discuss such ideas more precisely, and in more detail.

V. THE CANONICAL TENSOR OPERATORS FOR SU_3

It has been demonstrated in the previous sections that the operators in SU_n may be uniquely categorized by two mappings onto the states of SU_n, each mapping introducing a Gelfand pattern. The operators in SU_n are thus in one-to-one correspondence in the double Gelfand patterns—the diamond-shaped "operator patterns" introduced above.

Although this is a very satisfactory categorizing of all possible SU_n operator structures, there is one

feature which is undesirable: unlike the tensor operator mapping (lower Gelfand pattern) which was unique, the assignment of the upper Gelfand patterns is *arbitrary* to a unitary transformation (orthogonal transformation for Hermitian operators) within the set of operators belonging to a given Young pattern $[A]$ and corresponding to the same Δ value. This is the familiar multiplicity problem (the more difficult part of the simple reducibility problem[2]).

By a canonical resolution of the multiplicity problem we mean a unique classification scheme based upon some basic unifying principle, sufficiently fundamental or desirable that the appellation canonical is fitting. For example, consider the finite group A_4. This group is not multiplicity free[4]; however A_4 is a subgroup of group S_4 which *is* multiplicity free. It is clear that there exists a canonical resolution of the multiplicity problem by embedding A_4 in S_4. It does not appear to be possible to resolve the multiplicity problem in the general case by an embedding in a larger multiplicity free group, although we do not claim to have examined this question in detail. As an indication of the possible difficulties that may arise in such an inquiry, even in the finite groups, consider the group A_5. This group is not multiplicity free, yet there does exist a multiplicity-free embedding, *an embedding in the continuous group* R_3.[27] Whether there exists a *finite* embedding for A_5 is not known to us.

It is the thesis of the present paper to assert that for the unitary groups there does indeed exist a canonical resolution of the multiplicity problem, based upon the operator transformation properties under *conjugation*. The fundamental nature of this classification principle is sufficiently clear as to warrant the claim that the resolution is canonical. In order to verify this claim, and to demonstrate the classification by conjugation, it is simplest to proceed inductively from the simplest non-multiplicity-free unitary group, SU_3, to which we now turn.

The operator structures $\langle 100\rangle$ and $\langle 110\rangle$ in SU_3 are multiplicity free and have been given explicitly in II. For these operators the upper patterns are uniquely determined by the changes induced in the SU_3 Young pattern labels, that is, upper patterns $\rightleftarrows$ $\Delta \equiv (\Delta p\ \Delta q\ 0)$. The first occurrence of multiplicity greater than one is for the operators $\langle 210\rangle$; there exist two operators,

$$\begin{matrix} 1 & \\ 2\ 0 & \\ \langle 2\ 1\ 0\rangle & \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & \\ 1\ 1 & \\ \langle 2\ 1\ 0\rangle & \end{matrix}$$

corresponding to $\Delta = (000)$.

There is, however, *a single pre-eminent operator structure in any Lie group*: these are the generators $\{X_A\}$, which necessarily have the property that the $\{X_A\}$ leave the representation labels unchanged. Thus one of the two operators

$$\begin{matrix} 1 & \\ 2\ 0 & \\ \langle 2\ 1\ 0\rangle & \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & \\ 1\ 1 & \\ \langle 2\ 1\ 0\rangle & \end{matrix}$$

must be identified as the generators $\{X_A\}$. The basic step in I was, however, the explicit construction for all SU_n of the independent operators belonging to the adjoint representation, i.e., the operators $\langle 2\dot{1}0\rangle$. For SU_3 there were two such operators $\{X_A\}$ and $\{X_A^{(2)}\}$, the latter being the *symmetrically* coupled generators. By construction, these operators were not orthogonal; in fact the invariants were defined by: $I_n \equiv X_A \cdot X_A^{(n-1)}$. It is easy to orthogonalize the operators but the prescription is unique only if we use the principle that the generators are the pre-eminent operator structure and must be preserved in all classifications.[28] With this it follows that the orthogonal SU_3 operators $\langle 210\rangle$ having $\Delta = (000)$ are $\{X_A\}$ and $\{X_A^{(2)}\} - (I_3/I_2)\{X_A^{(1)}\}$.

Which of these operators is to be identified with

$$\begin{matrix} 1 & \\ 1\ 1 & \\ \langle 2\ 1\ 0\rangle ? & \end{matrix}$$

To answer this, let us note that under conjugation the generators reverse, that is $\mathcal{K}X_A\mathcal{K}^{-1} = -X_{-A}$, whereas the $\{X_A^{(2)}\}$ do not: $\mathcal{K}X_A^{(2)}\mathcal{K}^{-1} = +X_{-A}^{(2)}$. This does not of itself choose between the two upper patterns, but we may extend the idea of conjugation to *all* the $\langle 210\rangle$ operators and then divide the operators according to their sign $(\pm)$ under conjugation. The result (this requires construction of the remaining operators $\langle 210\rangle$, see below) is quite simple: there is but *one* operator,

$$\begin{matrix} 1 & \\ 1\ 1 & \\ \langle 2\ 1\ 0\rangle & \end{matrix}$$

which reverses under conjugation; that is:

[27] This could mean that for finite groups in general it might be necessary to embed in continuous groups. This example is due Professor W. T. Sharp.

[28] Let us note that if we consider the group specified by the additional restriction to self-conjugate representations only ($I_3 \to 0$) then the conjugation classification for the operator structure becomes a unique and complete specification. The tensor operators are then the carrier space of a group locally isomorphic to $SU_2 \times SU_3$.

$$\begin{matrix} & & 1 & & \\ & 1 & & 1 & \\ \langle 2 & & 1 & & 0\rangle \end{matrix} \equiv \{X_A\}.$$

It is convenient to collect together all the operators with the same conjugation properties. This divides the $\langle 210\rangle$ operators into two sets: (a) the operator

$$\begin{matrix} & & 1 & & \\ & 1 & & 1 & \\ \langle 2 & & 1 & & 0\rangle \\ & \cdot & & \cdot & \\ & & \cdot & & \end{matrix}$$

which reverses under conjugation and (b) the remaining seven operators

$$\begin{matrix} & & \cdot & & \\ & \cdot & & \cdot & \\ \langle 2 & & 1 & & 0\rangle \end{matrix}$$

which do not reverse. The set (a) is labeled as $c = 1$ the set (b) as $c = 2$.

This result is quite easy to see, but the really interesting fact is that by this single step *all remaining operators is SU_3 are uniquely classified.*

To prove this assertion, consider the classification of upper patterns for "self-conjugate" operators, that is, the classification of those tensor operators whose Young pattern labels are self-conjugate. For SU_3, the self-conjugate Young patterns are all of the form $[2k\ k\ 0]$, $k = 0, 1, \cdots$. The operators $\langle 210\rangle$ belong to this category and have been classified; the next operators of this type are $\langle 420\rangle$. Knowledge of the operator $\langle 210\rangle$ is equivalent (see Sec. II) to the possession of coupling coefficients, which may be used to couple two $\langle 210\rangle$ operators, themselves, to the maximal system $\langle 420\rangle$. These coupling coefficients, being maximal, are unique, and necessarily symmetric for the case $[210] \otimes [210] \rightarrow [420]$. The two $\langle 210\rangle$ operators to be coupled, however, may be chosen in several different ways, each choice defining a different $\langle 420\rangle$ operator. In terms of the two sets of $\langle 210\rangle$ operators ($c = 1$ and $c = 2$) there are but *three* choices: (a) $c = 4 (c_1 = c_2 = 2)$ (b) $c = 3 (c_1 = 2$ or 1, $c_2 = 1$ or 2—symmetry of the coupling coefficients), and (c) $c = 2 (c_1 = c_2 = 1)$. This divides the 27 independent $\langle 420\rangle$ tensor operators into three sets: $c = 4$, possessing 19 members; $c = 3$ possessing 7 members; $c = 2$ possessing 1 member. The conjugation properties of these three sets are specified by $(-)^c$, with $c = 4$, 3, or 2.

It is not difficult to see now that this procedure is sufficient to determine all the self-conjugate operators uniquely. At each step the operators so classified provide coupling coefficients to determine the next step. The result is a classification which assigns to the general self-conjugate operator $\langle 2k\ k\ 0\rangle$ the conjugation labels $c = 2k, 2k-1, \cdots, k$. These conjugation labels specify that each set possesses the conjugation property $(-)^c$. To the set c, *which we may label generically as*

$$\left\langle \begin{matrix} c \\ 2k\ k \end{matrix} \right\rangle$$

there belong

$$\dim\left(\left\langle \begin{matrix} c \\ 2k\ k \end{matrix} \right\rangle\right) = 3(c-k)(c-k+1)+1 \qquad (39)$$

tensor operators. (As a check, note that

$$\sum_c \dim\left(\left\langle \begin{matrix} c \\ 2k\ k \end{matrix} \right\rangle\right) = \dim [2k\ k\ 0].)$$

The generalization to all SU_3 operators is now immediate. The operators $\langle l00\rangle$ and $\langle ll0\rangle$ are multiplicity free and therefore uniquely known. These operators thus define coupling coefficients by which the tensor operators $\langle 2k+l\ k\ 0\rangle \equiv \langle l00\rangle \otimes \langle 2k\ k\ 0\rangle$ and $\langle 2k+l\ k+l\ 0\rangle \equiv \langle ll0\rangle \otimes \langle 2k\ k\ 0\rangle$ may be uniquely determined. The multiplicity of the operators is precisely the multiplicity of the operators $\langle 2k\ k\ 0\rangle$, (as is shown directly from the operator mapping result of Sec. III), and thus the canonical classification via conjugation is complete, as asserted.[29]

A Reformulation of the Canonical Classification

It is essential for the purpose of generalization to formulate the above classification in various equivalent, but differently appearing, ways—in order that the essence of the conjugation classification become apparent.

The essential point in the classification scheme is that the various tensor operators, belonging to given Young pattern labels $\langle A\rangle$, are grouped together into sets possessing the same conjugation properties. Let us examine these sets now more closely.

Consider the self-conjugate operators, $\langle 2k\ k\ 0\rangle$. There were $k+1$ sets labeled by $c = k, \cdots, 2k$. The classification into these sets is equivalent to the statement that the upper Gelfand patterns are classified into sets of U_2 labels. There is a convenient geometrical way to present this result. The decomposition of $[2k\ k\ 0]$ into U_2 labels fills out a rectangle, whose vertices are given by the points: $(2k, k)$, $(2k, 0)$, $(k, 0)$, (k, k). The conjugation classification groups these into hook-shaped patterns. Fig. 1 shown for [420]—makes this classification quite obvious.

The essential characteristic of this decomposition

[29] This result was first given in a brief (and therefore cryptic!) note; L. C. Biedenharn, Phys. Letters 3, 254 (1963).

into sets is now clear. *Each set contains every allowed* $\Delta \equiv (\Delta p, \Delta q, 0)$ *precisely once or not at all.* Moreover, each set contains those Δ values, of all Gelfand patterns belonging to the representation $[2k-c\ k\ c]$. (These assertions are clear by inspection of the diagram.)

This suggests a notational way in which to make the *general* SU_3 conjugation classification more or less obvious. Consider the upper pattern labels belonging to the operator $\langle p\ q\ 0\rangle$. The complete operator (all upper patterns) is broken into subsystems classified by conjugation by the process

$$\langle p\ q\ 0\rangle \supset (p\ q\ 0) \dotplus (p-1\ q\ 1) \dotplus \cdots \dotplus \begin{cases} (p-q\ q\ \ q), & p-q \geq q \\ (q\ \ q\ p-q), & p-q \leq q. \end{cases} \tag{40}$$

The subsystems $(p-c\ q\ c)$ are just the operators having the conjugation property $(-)^c$. Such a subsystem contains the U_2 upper pattern labels given by

$$(p-c\ q\ c) = [p-c\ q] \dotplus [p-c\ q-1] \dotplus \cdots \dotplus [p-c\ c] \dotplus [p-c-1\ c] \dotplus \cdots \dotplus [q\ c]. \tag{41}$$

For given (upper pattern) U_2 labels, all (upper pattern) U_1 labels allowed by the Gelfand inequalities belong to the same operator set specified by c.

It is helpful to note that the decomposition of $(p-c\ q\ c)$ into U_2 labels is affected by *systematically lowering the labels, beginning from the right,* until a given label is minimal, then proceeding to the next label on the left.

This systematic decomposition procedure shows directly that:

(1) The Δ values associated with a given operator $\langle p\ q\ 0\rangle$ belonging to the label c are precisely those Δ values that occur for the representation $[p-c\ q\ c]$, *but with multiplicity one*;

(2) The sets of operators labeled by c together span the entire space of the upper Gelfand patterns of the operator $\langle p\ q\ 0\rangle$ with the correct multiplicity for each Δ value.

An Algebraic Formulation

The preceding discussion has shown, diagrammatically, the significance of the conjugation classification. Let us approach the same problem algebraically.

The conjugation classification has been induced from the fundamental distinction that exists between $\{X_A\}$ and $\{X_A^{(2)}\}$ under conjugation. This

[420]:	[40] [30] [20]	[41] [31] [21]	[42] [32] [22]
c=4:	[40] [30] [20]	[41]	[4]
c=3:		[31] [21]	[32]
c=2:			[22]

FIG. 1. The decomposition of the [4 2 0] patterns into sets of U_2 patterns labeled by the conjugation parameter c.

distinction may be expressed in terms of properties of the matrix elements under conjugation. Consider first the matrix elements

$$\langle (m')|\ \begin{matrix}1\\1\ 1\\(a)\end{matrix}\langle 2\ 1\ 0\rangle\ |(m)\rangle,$$

which are simply the matrix elements of the $\{X_A\}$ *properly phased* (see III). These matrix elements are real, and under conjugation one finds that

$$\langle (m')|\ \begin{matrix}1\\1\ 1\\(a)\end{matrix}\langle 2\ 1\ 0\rangle\ |(m)\rangle = (-)^{\delta(m)+\delta(m')+\delta(a)+1}\langle (\tilde{m}')|\ \begin{matrix}1\\1\ 1\\(\tilde{a})\end{matrix}\langle 2\ 1\ 0\rangle\ |(\tilde{m})\rangle. \tag{42}$$

(Note that a separate orthogonalization of the operators

$$\begin{matrix}1\\2\ 0\\\langle 2\ 1\ 0\rangle\end{matrix} \quad \text{and} \quad \begin{matrix}1\\1\ 1\\\langle 2\ 1\ 0\rangle\end{matrix}$$

is *not* necessary for matrix elements taken between self-conjugate states. For such matrix elements, $I_3 = 0$ and hence $X^{(1)} \cdot X^{(2)} = 0$. The conjugation classification is thus partly orthogonalized by the conjugation properties themselves.)

Thus it follows that under conjugation the operator

$$\begin{matrix}1\\1\ 1\\\langle 2\ 1\ 0\rangle\\(a)\end{matrix}$$

obeys the rule

$$\mathcal{K}\begin{matrix}1\\1\ 1\\\langle 2\ 1\ 0\rangle\\(a)\end{matrix}\mathcal{K}^{-1} = (-)^{\delta(a)+1}\begin{matrix}1\\1\ 1\\\langle 2\ 1\ 0\rangle\\(\tilde{a})\end{matrix}. \tag{43}$$

Let us now generalize this operation step by step. For the operator

$$\begin{matrix} & 1 & \\ & 2\ \ 0 & \\ \langle & 2\ \ 1\ \ 0 & \rangle \\ & (a) & \end{matrix}$$

one finds that

$$\begin{matrix} & 1 & \\ & 2\ \ 0 & \\ \mathcal{K}\langle & 2\ \ 1\ \ 0 & \rangle\mathcal{K}^{-1} \end{matrix} = (-)^{\delta(a)} \begin{matrix} & 1 & \\ & 2\ \ 0 & \\ \langle & 2\ \ 1\ \ 0 & \rangle, \\ & (\tilde{a}) & \end{matrix} \tag{44}$$

since—by the conjugation classification—the operator

$$\begin{matrix} & 1 & \\ & 2\ \ 0 & \\ \langle & 2\ \ 1\ \ 0 & \rangle \\ & \cdot\ \ \cdot & \end{matrix}$$

does not reverse ($(-)^2$). It is clear that in the general case the operator $\langle p\ q\ 0\rangle$ belonging to conjugation label c possesses the additional phase $(-)^c$ under conjugation.

Consider now the general matrix element

$$\langle (m')|\ \begin{matrix} (\alpha) \\ \langle p\ \ q\ \ 0 \rangle \\ (a) \end{matrix}\ |(m)\rangle.$$

This matrix element is real, and in consequence one finds that

$$\langle (m')|\ \begin{matrix} (\alpha) \\ \langle p\ \ q\ \ 0 \rangle \\ (a) \end{matrix}\ |(m)\rangle = (-)^{\delta(m')+\delta(m)} \langle (\tilde{m}')|\ \begin{matrix} (\alpha) \\ \mathcal{K}\langle p\ \ q\ \ 0 \rangle\mathcal{K}^{-1} \\ (a) \end{matrix}\ |(\tilde{m})\rangle. \tag{45}$$

By definition of the significance of the upper pattern, the labels $m'_{i,3}$ are given by: $m'_{i,3} = m_{i,3} + \Delta_i$. Moreover, to be nonvanishing the matrix elements on both sides of Eq. (45) must satisfy the restrictions imposed by the H_i operators. From this it follows that under conjugation the lower operator pattern

$$(a) = \begin{pmatrix} p & q & 0 \\ a_{12} & a_{22} & \\ a_{11} & & \end{pmatrix}$$

becomes the conjugate pattern

$$(\tilde{a}) \equiv \begin{pmatrix} p & p-q & 0 \\ p-a_{22} & p-a_{12} & \\ p-a_{11} & & \end{pmatrix}$$

and the operator is multiplied by the phase $(-)^{\delta(a)}$. Taking account of the conjugation classification, it follows that

$$\mathcal{K}\left\langle \begin{matrix} \alpha_{11} & & \\ \alpha_{12} & \alpha_{22} & \\ p & q & 0 \\ a_{12} & a_{22} & \\ a_{11} & & \end{matrix} \right\rangle \mathcal{K}^{-1} = (-)^{c+\delta(a)} \left\langle \begin{matrix} \bar{\alpha}_{11} & & \\ \bar{\alpha}_{12} & \bar{\alpha}_{22} & \\ p & q & 0 \\ a_{12} & a_{22} & \\ a_{11} & & \end{matrix} \right\rangle, \tag{46}$$

where

$$c \equiv \begin{cases} \alpha_{12} & \text{if } \alpha_{12} + \alpha_{22} \geq p, \\ p - \alpha_{22} & \text{if } \alpha_{12} + \alpha_{22} \leq p, \end{cases} \tag{47}$$

with $\delta = q + a_{12} + a_{22} + a_{11}$, [by III Eq. (14)], and $\bar{\alpha}_{12}$, $\bar{\alpha}_{22}$, $\bar{\alpha}_{11}$ are yet to be defined.

The conditions which determine the effect of conjugation on the upper patterns are obtained from the requirements that $\tilde{m}'_{i,3} = \tilde{m}_{i,3} + \Delta_i(\bar{\alpha})$ in order that $(\bar{\alpha})$ be a meaningful upper pattern. Since $\tilde{m}'_{i,3} = m'_{1,3} - m_{4-i,3}$ and moreover $m'_{i,3} = m_{i,3} + \Delta_i$, one sees that the requirement becomes

$$\Delta_i(\bar{\alpha}) = \Delta_1(\alpha) - \Delta_{4-i}(\alpha) \equiv \tilde{\Delta}_i(\alpha). \tag{48}$$

The significant fact is that these equations do not possess a unique solution, except for the case in which the Δ_i imply a unique upper pattern. *The ambiguity in the conjugation operation for upper patterns is resolved by the assignment of the conjugation labels c.* Alternatively, one may view this result as establishing the conjugation classification itself in an algebraic fashion.

The result is a complete definition of the conjugation operation on upper patterns,

$$(\alpha) \equiv \begin{pmatrix} & \alpha_{11} & \\ \alpha_{12} & & \alpha_{22} \\ A_{13} & A_{23} & 0 \end{pmatrix}, \tag{49a}$$

$\mathcal{K} : (\alpha) \rightarrow (\bar{\alpha})$, where

$$(\bar{\alpha}) \equiv \begin{pmatrix} & \bar{\alpha}_{11} & \\ \bar{\alpha}_{12} & & \bar{\alpha}_{22} \\ A_{13} & A_{23} & 0 \end{pmatrix} \tag{49b}$$

and

$$\bar{\alpha}_{11} = \alpha_{12} + \alpha_{22} - A_{13}, \tag{49c}$$

with

$$\left.\begin{aligned} \bar{\alpha}_{12} &= \alpha_{12} \\ \bar{\alpha}_{22} &= A_{13} - A_{23} + \alpha_{11} - \alpha_{12} \end{aligned}\right\} \text{ if } \alpha_{11} \geq A_{23} \tag{49d}$$

or

$$\left.\begin{aligned} \bar{\alpha}_{12} &= \alpha_{11} + \alpha_{12} - A_{23} \\ \bar{\alpha}_{22} &= A_{13} - \alpha_{12} \end{aligned}\right\} \text{ if } \alpha_{11} \leq A_{23}. \tag{49e}$$

This fully determines the behavior of the general SU_3 operator under conjugation, and, of course, the explicit transformation properties of the matrix elements as well.

Let us complete this algebraic discussion by giving the one remaining operator property ex-

plicitly: behavior under Hermitian conjugation. We have the general operator result

$$\langle A\rangle^{\dagger} = \begin{matrix}(\alpha)\\(a)\end{matrix}\langle A\rangle^{\dagger} = (\text{dim op})^{\frac{1}{2}}\begin{matrix}(\tilde{\alpha})\\(\tilde{a})\end{matrix}\langle \tilde{A}\rangle(\text{dim op})^{-\frac{1}{2}} \tag{50}$$

where "dim op" is dimension operator giving the dimension (according to the Weyl dimension formula) of a state upon which it acts.

It would be convenient to eliminate the square-root factors, but this weakens the analogy to the SU_2 Wigner coefficients, and is probably not really useful.

A Mapping of the SU_3 Operators into SU_4 Systems

Looked at in a slightly different way, the results obtained above are in essence but a mapping of the SU_3 operators into simpler operator structures in SU_4. It is the purpose of the present section to discuss this remark in some detail.

The mapping into SU_4 is already clear in principle from the fact that the fundamental SU_3 Wigner operators—$\langle 100\rangle$ and $\langle 110\rangle$—were themselves obtained by subgroup decomposition from the generators in SU_4. It is also clear, in principle, that every SU_3 operator—regardless of the manner in which one chooses to resolve the multiplicity problem—is obtainable by composition from the fundamental operators, using the $\langle 100\rangle$ and $\langle 110\rangle$ operator matrix elements as coupling coefficients. Thus it is possible to expect that one may, in principle, map the general SU_3 operator into a suitable product of SU_4 *generators*. This is, in fact, true for the SU_3 case, though the general case is not quite as simple.

In order to be precise let us note that the mapping discussed here is typified by the $\langle 100\rangle$ and $\langle 110\rangle$ cases. The matrix elements of the generators SU_4 are given by $\langle (m')\ |X_A^{(SU_4)}|\ (m)\rangle$ where $m'_{i,4} = m_{i,4}$. Viewed in SU_3 these same matrix elements differ only by an over-all factor (reduced matrix element) from matrix elements of $\langle 100\rangle$, $\langle 110\rangle$, $X_A^{(SU_3)}$, and $\langle 000\rangle$—depending on the particular $X_A^{(SU_4)}$ in question. We express this mapping symbolically as

$$\langle 100\rangle \rightleftarrows E_{i4},$$
$$\langle 110\rangle \rightleftarrows E_{4i}, \quad \text{where} \quad i = 1, 2, 3. \tag{51}$$

If we now confine our attention to the *maximal* (lower pattern) states of these operators, it follows that the Young pattern labels *add*. That is, the product $E_{14}E_{43}$ corresponds to a tensor operator in SU_3 with the labels $\langle 210\rangle$.

The conjugation classification now results immediately. Under conjugation the product becomes

$$\mathcal{K}(E_{14}E_{43})\mathcal{K}^{-1} = E_{41}E_{34}. \tag{52}$$

For the operator $\langle 210\rangle$ belonging to $c = 2$, the desired conjugation properties are, for example,

$$\mathcal{K}\langle \begin{matrix}2\\2\ 1\\2\end{matrix}\ \begin{matrix}1\ 0\rangle\\0\end{matrix}\mathcal{K}^{-1} = +\langle \begin{matrix}2\\2\ 1\\1\ 0\end{matrix}\ \begin{matrix}1\ 0\rangle.\\0\end{matrix} \tag{53}$$

We achieve the desired mapping by using the anti-commutator, i.e.,

$$\langle 210\rangle_{c=2} \rightleftarrows E_{14}E_{43} + E_{43}E_{14}. \tag{54}$$

The commutator $[E_{14}, E_{43}] = E_{13}$ shows that the remaining $\langle 210\rangle$ operator belonging to $c = 1$ is simply E_{13} itself. [Clearly the operators not having maximal (lower) patterns are obtained by using the SU_3 generators to lower the quantum numbers. This operation is of no concern.]

It is quite immediate now to obtain the general SU_3 operator. For the self-conjugate operators $\langle 2k\ k\ 0\rangle$ one sees that there are $k + 1$ different operators specified by the label $c = 2k, 2k-1, \cdots, k$. These are

$$\langle 2k\ k\ 0\rangle_c \rightleftarrows (E_{14}E_{43} + E_{43}E_{14})^{c-k}(E_{13})^{2k-c}. \tag{55}$$

The general operator $\langle p\ q\ 0\rangle$ is obtained by multiplying one of these $(q + 1)$ operators

$$\langle 2q\ q\ 0\rangle \quad \text{by} \quad (E_{14})^{p-2q} (\text{if } p \geq 2q)$$
$$\text{or} \quad \langle 2(p-q)\ p-q\ 0\rangle \quad \text{by} \quad (E_{43})^{2q-p} (\text{if } p \leq q). \tag{56}$$

It should be noted that in order to convert this mapping into means of calculating the matrix elements two further conditions must be imposed: (1) the normalization must be made explicit and (2) the orthogonality must be more explicit. This latter step is the same as that introduced by the association:

$$\langle \begin{matrix}1\\1\ 1\end{matrix}\ \begin{matrix}0\rangle\\\cdot\end{matrix} \propto X_A, \quad \langle \begin{matrix}1\\2\ 1\end{matrix}\ \begin{matrix}0\rangle\\\cdot\end{matrix} \propto [X_A^{(2)} - (I_3/I_2)X_A], \tag{57}$$

as discussed earlier in this section.

The mapping of the SU_3 operator into the generators of SU_4 provides a clear view of the significance of the conjugation classification; this mapping is rather similar to the mapping of the states into the boson operators used so profitably in II. The fact that E_{14} and E_{43} do not commute obscures the parallel, however.

VI. GENERALIZATION OF THE CONJUGATION CLASSIFICATION

It is the purpose of this concluding section to indicate briefly how the conjugation classification

may be extended in principle to the general unitary group U_n in order to obtain a canonical resolution of the tensor operator multiplicity problem. As is seen below, this generalization is inherently very complicated and the usefulness of any detailed general treatment is questionable. We therefore merely sketch the essential ideas in the generalization—without attempting a complete proof.

The essence of the conjugation classification—as abstracted from the SU_3 example—is the grouping together of the various upper (operator) patterns—belonging to given tensor operators $\langle A\rangle$—into *sets*, each set containing a given Δ value either once or not at all. The defining property of these sets is that under upper-pattern conjugation, a given set transforms (with specified phases) into itself—for self-conjugate operators—or into a conjugate set (of the same dimensionality), for non-self-conjugate operators. Rather than dealing directly with upper pattern conjugation (operation $\langle A\rangle \rightarrow \mathcal{K}^{-1}\langle A\rangle\mathcal{K}$), it is somewhat easier to use the combined operation of Hermitian conjugation followed by conjugation. This operation, which we call associated conjugation,[30] $\mathcal{C}$, has the desirable property of carrying the sets (of both self-conjugate and non-self-conjugate operators) into themselves.

The difficulty in this conjugation classification is that, at present, we have no *explicit* general result detailing the effect of associated conjugation upon an arbitrary admissible upper pattern, rather this transformation must be treated iteratively from lower unitary groups.

The operator set corresponding to the *maximal* Δ is clearly of multiplicity one, and the effect of associated conjugation may be easily given here for the general case. In what follows we use a slightly different set of Δ's than that used earlier. It will, henceforth, be required that

$$\sum_{i=1}^{n} \Delta_i = \sum_{i=1}^{n} m_{i,n}. \tag{58}$$

With this requirement and Eq. (22) one arrives at

$$\Delta_i = \sum_{j=1}^{i} m_{j,i} - \sum_{j=1}^{i-1} m_{j,i-1}. \tag{59}$$

These new (nonunimodular) Δ's will be used in all of the following work.

The unique operator set corresponding to maximal Δ is specified as follows:

(a) the upper pattern of highest Δ is maximal and has $\Delta = (m_{1,n}m_{2,n} \cdots m_{n,n})$;

(b) the remaining upper patterns of the set have the Δ values:

$$(m_{1,n}m_{2,n} \cdots m_{n-1,n} - 1,\ m_{n,n} + 1),$$
$$\cdots (m_{1,n}m_{2,n} \cdots m_{n,n}m_{n-1,n}),$$
$$\cdots (m_{1,n}m_{2,n} \cdots m_{n-1,n}\ m_{n,n}m_{n-2,n})$$
$$\cdots (m_{2,n}m_{3,n} \cdots m_{n,n}m_{1n}).$$

(c) The upper patterns corresponding to these Δ's are derived from the maximal pattern by first decreasing $a_{n-1,n-1}$ from $m_{n-1,n}$ to $m_{n,n}$; then decreasing $a_{n-2,n-2} = a_{n-2,n-1}$ from $m_{n-2,n}$ to $m_{n-1,n}$; $\cdots$; finally decreasing $a_{11} = a_{12} = a_{13} \cdots = a_{1,n-1}$ from $m_{1,n}$ to $m_{2,n}$.

The operation [in (b) above] whereby the highest Δ value is stepwise decreased to the lowest Δ value is of particular importance, since it characterizes a basic property of the sets. This operation—which we call a "constrained shift"— is succinctly categorized by noting that the ith symbol in Δ is systematically lowered to its minimal value $(=\Delta_{i+1})$ after the $(i-1)$th symbol has been lowered to its minimum, beginning with the $(n-1)$th symbol. It is clear that by this definition *the set of Δ's obtained from a constrained shift realizes all Δ values belonging to the representation* $[\Delta_1 \cdots \Delta_n]$ *once and only once.*

A given set of upper patterns belonging to the operator $\langle A\rangle$ is classified by the *highest upper pattern of the set* (the pattern having the highest Δ value; the Δ values being ranked by the same process that ranked weights). A given set of upper patterns has one pattern corresponding to each Δ obtained by a constrained shift from the highest Δ.

These definitions now suffice to detail the iterative method by which the operator sets in U_n are completely categorized. Let us suppose that the U_n tensor operator $[A_1 \cdots A_n]$ has been classified into operator sets, each set being denoted by a highest upper pattern. Consider now an arbitrary U_{n+1} tensor operator $[A'_1 \cdots A'_{n+1}]$. All upper patterns belonging to this operator may be classified by upper patterns (formerly in U_n) having the longest row $[a_{1n} \cdots a_{nn}]$. All upper patterns of $[a_{1n} \cdots a_{nn}]$ may next be classified into sets, characterized by a highest U_n pattern and corresponding (lexical) Δ values (containing n symbols).

Since the row $[a_{1,n} \cdots a_{n,n}]$ implies a value for Δ_{n+1}, the Δ values may all be increased to $(n+1)$ symbols. Beginning with the Δ values having $\Delta_{n+1} = m_{n,n}$ all lower Δ values corresponding to a constrained shift are assigned to this highest Δ value and highest pattern. This systematically

[30] One might with justice call this operation "transposition" but its effect on upper patterns is hardly a transposition, and this designation would prove confusing.

identifies by explicit construction all upper patterns belonging to a given set (characterized by a highest upper pattern). Since the Δ values do not uniquely characterize the upper patterns, it is necessary to note that the rules—induced to designate a highest upper pattern in U_n—are required to make this assignment explicit. (It is this feature which prevents a fully explicit characterization of all U_n upper patterns at present.) We have not succeeded in obtaining a complete proof by this direct process; the *existence* of the conjugation classification is, however, guaranteed by the fact that this classification is a mapping of the operators of U_n into the operators of $U_{(n+[\frac{1}{2}n])}$ where $[\frac{1}{2}n]$ is the largest integer in $\frac{1}{2}n$, and can thus be seen to exist in general. We have carried out the conjugation classification for SU_4 in complete detail, but, as noted earlier, the results are rather complicated and—aside from their existence—need not be discussed further here.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Applications of the Lorentz Transformation Properties of Canonical Spin Tensors

A. CHAKRABARTI
Centre de Physique Théorique de l'Ecole Polytechnique, Paris, France
(Received 1 June 1964)

Some applications of the Lorentz transformations of relativistic spin tensors in the canonical representation are discussed. The problem of precession of polarization is discussed in Sec. 2. It is shown that the kinematical equation, obtained quite simply, already contains the "Thomas factor." In Sec. 3, applications to polarization analysis of decay products are considered. The canonical form of S-matrix elements are used and multipole parameters for successive decays of the type $a \to b + c$ are obtained in an arbitrary frame in a relatively simple way. The exact relativistic way in which the multipole parameters depend, in an arbitrary frame, on the particle momenta are discussed for decays of the type $a \to b + c + d$. While the canonical representation is used mainly, the corresponding technique in the spinor representation is discussed.

1. INTRODUCTION

IN a series of previous articles[1] we have considered in detail many aspects of the canonical representation for particles of nonzero rest mass.

The fundamental feature of this representation is that, considering a particle of spin S, mass m ($\neq 0$) and 4-momentum p, the effect of any transformation Λ of the homogeneous Lorentz group corresponds to the "Wigner rotation"

$$\Lambda_{(p')} \cdot \Lambda \cdot \Lambda_{(p)}{}^{-1} = \begin{vmatrix} 1 & \\ - & \end{vmatrix} \begin{vmatrix} & \\ & R_w \end{vmatrix}, \qquad (1.1)$$

where

$$\Lambda \cdot p = p', \quad \Lambda_{(p)} \cdot p = (m, \mathbf{0})$$

and R_w is a 3×3 rotation matrix.

The state vectors are transformed by $\mathfrak{D}^{(s)}(R_w)$ and the expectation value of spin $\langle \mathbf{s} \rangle$ or, equivalently, the polarization vector $\mathbf{P}\,[=(1/s)\langle \mathbf{s} \rangle]$ is transformed to

$$\mathbf{P}' = R_w \cdot \mathbf{P}.$$

When Λ is a pure rotation, R_w simply coincides with Λ. When Λ is a pure Lorentz transformation, we have, when a 4-velocity u is imparted,

$$R_w = I + a^{-1}(bA + A \cdot A), \qquad (1.2)$$

where I is the 3×3 unit matrix and

$$a = (u^\circ + 1)(p^\circ + m)(u^\circ p^\circ + \mathbf{u} \cdot \mathbf{p} + m),$$

$$b = (u^\circ + 1)(p^\circ + m) + \mathbf{u} \cdot \mathbf{p},$$

$$A = (\mathbf{u} \otimes \mathbf{p} - \mathbf{p} \otimes \mathbf{u}).$$

Noting that

$$A \cdot \mathbf{P} = (\mathbf{u} \times \mathbf{p}) \times \mathbf{P},$$

and putting

$$\hat{n} = (\mathbf{p} \times \mathbf{u})/|\mathbf{p} \times \mathbf{u}|,$$

[1] A. Chakrabarti, J. Math. Phys. 4, 1215, 1223 (1963); 5, 922 (1964).

identifies by explicit construction all upper patterns belonging to a given set (characterized by a highest upper pattern). Since the Δ values do not uniquely characterize the upper patterns, it is necessary to note that the rules—induced to designate a highest upper pattern in U_n—are required to make this assignment explicit. (It is this feature which prevents a fully explicit characterization of all U_n upper patterns at present.) We have not succeeded in obtaining a complete proof by this direct process; the *existence* of the conjugation classification is, however, guaranteed by the fact that this classification is a mapping of the operators of U_n into the operators of $U_{(n+[\frac{1}{2}n])}$ where $[\frac{1}{2}n]$ is the largest integer in $\frac{1}{2}n$, and can thus be seen to exist in general. We have carried out the conjugation classification for SU_4 in complete detail, but, as noted earlier, the results are rather complicated and—aside from their existence—need not be discussed further here.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Applications of the Lorentz Transformation Properties of Canonical Spin Tensors

A. CHAKRABARTI
Centre de Physique Théorique de l'Ecole Polytechnique, Paris, France
(Received 1 June 1964)

Some applications of the Lorentz transformations of relativistic spin tensors in the canonical representation are discussed. The problem of precession of polarization is discussed in Sec. 2. It is shown that the kinematical equation, obtained quite simply, already contains the "Thomas factor." In Sec. 3, applications to polarization analysis of decay products are considered. The canonical form of S-matrix elements are used and multipole parameters for successive decays of the type $a \to b + c$ are obtained in an arbitrary frame in a relatively simple way. The exact relativistic way in which the multipole parameters depend, in an arbitrary frame, on the particle momenta are discussed for decays of the type $a \to b + c + d$. While the canonical representation is used mainly, the corresponding technique in the spinor representation is discussed.

1. INTRODUCTION

IN a series of previous articles[1] we have considered in detail many aspects of the canonical representation for particles of nonzero rest mass.

The fundamental feature of this representation is that, considering a particle of spin S, mass m ($\neq 0$) and 4-momentum p, the effect of any transformation Λ of the homogeneous Lorentz group corresponds to the "Wigner rotation"

$$\Lambda_{(p')} \cdot \Lambda \cdot \Lambda_{(p)}{}^{-1} = \begin{vmatrix} 1 & \\ - & \end{vmatrix} \begin{vmatrix} & \\ & R_w \end{vmatrix}, \qquad (1.1)$$

where

$$\Lambda \cdot p = p', \quad \Lambda_{(p)} \cdot p = (m, \mathbf{0})$$

and R_w is a 3×3 rotation matrix.

The state vectors are transformed by $\mathfrak{D}^{(s)}(R_w)$ and the expectation value of spin $\langle \mathbf{s} \rangle$ or, equivalently, the polarization vector $\mathbf{P}\,[=(1/s)\langle \mathbf{s} \rangle]$ is transformed to

$$\mathbf{P}' = R_w \cdot \mathbf{P}.$$

When Λ is a pure rotation, R_w simply coincides with Λ. When Λ is a pure Lorentz transformation, we have, when a 4-velocity u is imparted,

$$R_w = I + a^{-1}(bA + A \cdot A), \qquad (1.2)$$

where I is the 3×3 unit matrix and

$$a = (u^\circ + 1)(p^\circ + m)(u^\circ p^\circ + \mathbf{u} \cdot \mathbf{p} + m),$$

$$b = (u^\circ + 1)(p^\circ + m) + \mathbf{u} \cdot \mathbf{p},$$

$$A = (\mathbf{u} \otimes \mathbf{p} - \mathbf{p} \otimes \mathbf{u}).$$

Noting that

$$A \cdot \mathbf{P} = (\mathbf{u} \times \mathbf{p}) \times \mathbf{P},$$

and putting

$$\hat{n} = (\mathbf{p} \times \mathbf{u})/|\mathbf{p} \times \mathbf{u}|,$$

[1] A. Chakrabarti, J. Math. Phys. 4, 1215, 1223 (1963); 5, 922 (1964).

we have

$$\mathbf{P}' = R_w \cdot \mathbf{P} = \cos\omega \mathbf{P} + (1 - \cos\omega)(\hat{n}\cdot\mathbf{P})\hat{n} + \sin\omega \hat{n} \times \mathbf{P}, \quad (1.3)$$

where

$$\cos\omega = 1 - (1/a)(\mathbf{u}\times\mathbf{p})^2,$$
$$\sin\omega = (b/a)\,|\mathbf{u}\times\mathbf{p}|.$$

Thus $\mathbf{P}$ is seen to undergo a rotation ω around the axis $(\mathbf{p}\times\mathbf{u})$.

The most detailed discussion of this rotation has been given by Ritus.[2] It may be verified that ω is always less than the angle α turned through by the momentum as a consequence of Λ and that $\omega \to \alpha$ as $m \to 0$. This limiting tendency corresponds to the well known fact that for $m = 0$, the parallelism of spin and momentum is a Lorentz-invariant concept.

Let us now consider the "multipole parameters" $t_L^M (0 \leq L \leq 2S, -L \leq M \leq L)$, which, together, give a complete description of the state of polarization of the particle. t_L^M is the expectation value of the operator which is constructed in terms of S_x, S_y, S_z (properly symmetrized because of noncommutativity) exactly as the spherical harmonic y_L^M is constructed out of x, y, z. Thus, by construction, if $\langle \mathbf{S} \rangle$ undergoes a rotation R then the irreducible tensorial set $[t_L]$, is transformed by the matrix $\mathfrak{D}^{(L)}(R)$.

So far as pure spatial rotations are concerned, this property is widely utilized in analyzing phenomena involving particle polarizations, the whole technique being quite well known,[3] (Chaps. 18, 19 in particular).

But it is evident from the foregoing discussion of the transformation properties of $\langle \mathbf{S} \rangle$, that for pure Lorentz transformation also we have, in the canonical representation, a simple rule for the set $[t_L]$. Symbolically, we can write

$$[t_L]' = \mathfrak{D}^{(L)}(R_w)\cdot[t_L]. \quad (1.4)$$

The purpose of this paper is to indicate the insights and advantages to be gained by exploiting systematically the above-mentioned fundamental property.

To this end we discuss two different types of applications. The first is the phenomenon of precession of polarization under acceleration. It will be shown that the famous "Thomas factor" is very simply and naturally included in our formalism. This will be just another evidence of the profound physical significance of the way in which spin and momentum are interconnected in our formalism.

The second type of applications will be to polarization analysis. The case of decays of particles of arbitrary spins will be studied in some detail in Sec. 3. Here we briefly point out, as an example, an immediate use of (1.4).

Let ρ, the density matrix of a particle (of spin j and momentum p) produced in a certain process, be parametrized as (following closely the notation of Ref. 4),

$$\rho = (\mathrm{tr}\ \rho)[(2j+1)^{-1} \sum_{L,M} (2L+1) t_L^{M*} T_L^M], \quad (1.5)$$

where

$$(T_L^M)mm' = C(jLj; m'M),$$
$$t_L^{M*} = (-)^M t_L^{-M}, \quad (1.5')$$

and

$$t_L^M = (1/\mathrm{tr}\ \rho)\{\mathrm{tr}\ (\rho T_L^M)\} = \langle T_L^M \rangle.$$

Let the experimental setup used to carry out the necessary polarization analysis be represented mathematically by the "efficiency matrix" E, parametrized as

$$E = N \sum_{L,M} \epsilon_L^{M*} T_L^M, \quad (1.6)$$

where N is a normalization factor. (We may suppose, more generally, E to be any operator whose mean value is to be determined.) It is natural to suppose that the definitions of the irreducible sets $[\epsilon_L]$ are bound to the laboratory frame. Let us now suppose that (1.5) gives ρ in some other frame, in which it has a particularly simple form, say, the center-of-mass frame for the scattering or decay process in which the particle is obtained. The required mean value is then directly given, in an obvious notation, by

$$\frac{1}{\mathrm{tr}\ \rho_{\mathrm{lab}}} \mathrm{tr}\ (\rho_{\mathrm{lab}}\cdot E) = \frac{1}{\mathrm{tr}\ \rho_{\mathrm{lab}}} \sum_L [(t_L)_{\mathrm{lab}}]\cdot[\epsilon_L], \quad (1.7)$$

$$= \frac{1}{\mathrm{tr}\ \rho} \sum_L \{\mathfrak{D}^{(L)}(R_w)\cdot[t_L]\}\cdot[\epsilon_L], \quad (1.8)$$

where R_w is the Wigner rotation corresponding to the transformation from the initial frame (for ρ) to the laboratory frame.

The formula (1.8), which may also be supposed to include any desired adjustment of the difference in the orientations of the two frames, generalizes

[2] V. I. Ritus, Zh. Eksperim. i. Teor. Fiz. **40**, 352 (1961) [Eng. transl.: Soviet Phys.—JETP **13**, 240 (1961)].

[3] U. Fano and G. Racah, *Irreducible Tensorial Sets*. (Academic Press Inc., New York, 1959).

[4] H. Joos, Fortsch. Phys. **10**, 65 (1962).

the corresponding formula (19.1) of Fano–Racah for a pure rotation only.

2. PRECESSION OF POLARIZATION

We have already discussed the nature of the rotation undergone by the polarization vector under a Lorentz transformation. This makes it evident that if we consider a continuous series of Lorentz transformations, the polarization will undergo a precessional motion. It is, however, useful and instructive to derive the consequent equation of motion from the "infinitesimal" point of view.

In the canonical representation the infinitesimal operators of the homogeneous Lorentz group are[1]:

$$\mathbf{N} = -iP^0\frac{\partial}{\partial \mathbf{P}} - \frac{\mathbf{P}\times\mathbf{S}}{P^0+m}, \qquad \mathbf{M} = -i\mathbf{P}\times\frac{\partial}{\partial \mathbf{P}} + \mathbf{S}, \tag{2.1}$$

where

$$P^0 = (\mathbf{P}^2 + m^2)^{\frac{1}{2}}.$$

The form of $\mathbf{M}$ indicates, of course, that as usual $\langle\mathbf{S}\rangle$ transforms as a vector under a spatial rotation.

Let us now consider a pure Lorentz transformation of an infinitesimal velocity ($\tanh\chi$) parallel to the unit vector $\hat{n}$.

We have

$$[i\mathbf{N}\cdot\hat{n}\chi,\ \mathbf{S}] = -[i\chi/(P^0+m)][\hat{n}\cdot(\mathbf{P}\times\mathbf{S}),\ \mathbf{S}]$$

$$= -(\chi/P^0+m)(\hat{n}\times\mathbf{P})\times\mathbf{S} \tag{2.2}$$

$$= -\chi[\gamma/(\gamma+1)](\hat{n}\times\mathbf{v})\times\mathbf{S}, \tag{2.3}$$

where

$$\mathbf{v} = \mathbf{P}/P^0 \quad\text{and}\quad \gamma = (1-\mathbf{v}^2)^{-\frac{1}{2}}.$$

Since

$$[i\mathbf{N}\cdot\hat{n}\chi,\ \mathbf{v}] = \chi\{\hat{n} - (\mathbf{v}\cdot\hat{n})\mathbf{v}\}, \tag{2.4}$$

writing $\delta\langle\mathbf{S}\rangle$ for the change in $\langle\mathbf{S}\rangle$, we have from (2.3) and (2.4),

$$\delta\langle\mathbf{S}\rangle = [\gamma/(\gamma+1)](\mathbf{v}\times\delta\mathbf{v})\times\langle\mathbf{S}\rangle. \tag{2.5}$$

Now, considering successive Lorentz transformations and passing to the limit, we have

$$\langle\dot{\mathbf{S}}\rangle = [\gamma/(\gamma+1)](\mathbf{v}\times\dot{\mathbf{v}})\times\langle\mathbf{S}\rangle \tag{2.6}$$

Thus we have obtained the required equation of motion giving the precession of $\langle\mathbf{S}\rangle$, *which is due to the acceleration only*. The factor $\gamma/(\gamma+1)$, which $\to\frac{1}{2}$ as $\mathbf{v}\to 0$, supplies the famous Thomas factor, which is thus seen to be built into our formalism from the very beginning. (This aspect is discussed in more detail in Appendix B.)

Let us now consider the multipole parameters in general, i.e., $t_L^M = \langle T_L^M\rangle$. We note that for any vector $\mathbf{a}$, in the representation in which S_3 is diagonal (which we always adapt implictly), we have

$$i[\mathbf{a}\cdot\mathbf{S},\ T_L^M] = iMa_zT_L^M + \tfrac{1}{2}i\{L(L+1)-M(M+1)\}^{\frac{1}{2}}(a_x-ia_y)T_L^{M+1} + \tfrac{1}{2}i\{L(L+1)+M(M-1)\}^{\frac{1}{2}}(a_x+ia_y)T_L^{M-1}. \tag{2.7}$$

To obtain $\delta\langle T_L^M\rangle$ under an infinitesimal Lorentz transformation, we have now only to put [see Eqs. (2.1)–(2.4)]

$$\mathbf{a} = -[\chi/(P^0+m)](\hat{n}\times\mathbf{P}), \qquad = [\gamma/(\gamma+1)](\mathbf{v}\times\delta\mathbf{v}). \tag{2.8}$$

The geometrical interpretation is, of course, not evident for $[t_L]$ in general as it is for the vector.

We may note however, that when the z axis is chosen to be parallel to $(\mathbf{v}\times\delta\mathbf{v})$, we have

$$\delta t_L^M = iMat_L^M, \tag{2.9}$$

where now

$$a = [\gamma/(\gamma+1)]\ |\mathbf{v}\times\delta\mathbf{v}|. \tag{2.10}$$

In the special case when all the successive velocities lie in one plane, which may be chosen as the xy plane, we have,

$$\dot{t}_L^M = iMa(t)t_L^M \tag{2.11}$$

or

$$(t_L^M)_\tau = \exp\left(iM\int_0^\tau a(\tau')\,d\tau'\right)(t_L^M)_{\tau=0}, \tag{2.12}$$

where the parameter τ denotes the time.

Since from (1.5), we have

$$\rho_{m_1m_2} = (2j+1)^{-1}\sum(-)^M t_L^{-M}C(jLj;\ m_2M), \tag{2.13}$$

corresponding to (2.12) we can write

$$(\rho_{m_1m_2})_\tau = \exp\left(i(m_1-m_2)\int_0^\tau a(\tau')\,d\tau'\right)(\rho_{m_1m_2})_0. \tag{2.14}$$

In particular, the diagonal elements of ρ do not vary.

3. POLARIZATION ANALYSIS IN PARTICLE DISINTEGRATIONS

We now proceed to discuss some applications of (1.4) to decay phenomena. We consider particles of nonzero rest mass only.

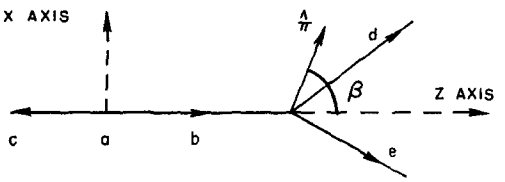


FIG. 1. Successive two-particle decay.

Let us consider the process

$$a \to b + c + d + \cdots, \tag{3.1}$$

where the spins of a and b are j_a, and j_b, respectively. For the sake of simplicity we put $j_c = j_d = \cdots = 0$.

In terms of the relevant S-matrix elements, and ρ^a, the density matrix of a, that of b is given by

$$\rho^b = S\rho^a S^\dagger, \tag{3.2}$$

where the particles states are all supposed to be characterized by the respective eigenvalues of momentum and spin projection.

A. Notation for S-Matrix Elements

Our technique is to evaluate ρ^b always in a "standard" frame of reference, namely the rest system of a with the z axis parallel to the momentum of $b[\mathbf{p}_a = 0; (p_b)_x = (p_b)_y = 0]$. Then we pass over to any other frame, as required [examples are given in Subsections 3B, 3C], by applying (1.4).

This is considerably simpler than evaluating the S-matrix elements directly in an arbitrary frame or (equivalently) transforming the relatively simple S-matrix elements of the "standard" frame and carrying out all the necessary summations and rearrangements of terms to obtain the final result.

In conformity with our canonical definition of spin, we will use the canonical S-matrix elements,[1,5,6] since the internal angular momenta of the product particles must be combined "canonically" to obtain the spin of the decaying particle in its rest frame.

(i) For decay into two particles:

$$a \to b + c;$$

in the "standard" frame, we have (with $\mathbf{p}_a = 0$ and $\mathbf{p}_b \parallel z$ axis):

$$\langle p_b, p_c, j_b, \sigma_b \,|S|\, p_a, j_a, \sigma_a\rangle = \sum_l C(j_b, l, j_a; \sigma_b 0) S_l, \tag{3.3}$$

where

$$S_l = 2m_a^{\frac{1}{2}}\lambda^{-\frac{1}{4}}(m_a^2, m_b^2, m_c^2)\,\delta(p_a - p_b - p_c)\,\delta(\sigma_a - \sigma_b) \times \left(\frac{2l+1}{4\pi}\right)^{\frac{1}{2}} \langle (j_b, l); j_a \,|S|\, j_a\rangle. \tag{3.4}$$

In the above formula m_a, m_b, m_c are the respective masses; j_a, j_b, 0 are the respective spins,

$$\lambda(a, b, c) = a^2 + b^2 + c^2 - 2(bc + ca + ab),$$

and $|(j_b, l); j_a\rangle$ denotes the state vector of total angular momentum j_a obtained by coupling j_b and l. For the sake of convenience we have put for the "reduced" S-matrix element S_l the entire expression on the right-hand side of (3.4) and not only $\langle (j_i, l); j_a|S|j_a\rangle$.

(ii) For decay into three particles:

$$a \to b + c + d,$$

of which c, d have both zero spins, we have (with $\mathbf{p}_a = 0$, $\mathbf{p}_b \parallel z$ axis):

$$\langle p_b, p_c, p_d; j_b\sigma_b \,|S|\, p_a, j_a, \sigma_a\rangle = \sum_{l'l''l} C(l'l''l; m'0) C(lj_bj_a; m'\sigma_b) \times y_{l'm'}(\hat{\pi}) S_{(l'l'')l}. \tag{3.5}$$

In the above formula,

$$\pi = \Lambda_{(p_c+p_d)} \cdot p_c, \qquad \hat{\pi} = \pi/|\pi|. \tag{3.6}$$

[The "standard" frame may be supposed to be so oriented as to have $\hat{\pi}$ in the z–x plane when $y_{lm}(\hat{\pi})$ is real and $y_{lm}(\hat{\pi}) = (-)^m y_{l-m}(\hat{\pi})$.]

$$S_{(l'l'')l} = 2m_{cd}^2\lambda^{-\frac{1}{4}}(m_c^2, m_d^2, m_{cd}^2) 2m_a^2\lambda^{-\frac{1}{4}}(m_b^2, m_{cd}^2, m_a^2) \times \delta(p_a - p_b - p_c - p_d)\left(\frac{2l''+1}{4\pi}\right)^{\frac{1}{2}} \times \langle (l'l'')l; (l, j_b)j_a \,|S|\, j_a\rangle \tag{3.7}$$

with

$$m_{cd}^2 = (p_c + p_d)^2, \qquad m' = (\sigma_a - \sigma_b).$$

B. Successive Decays into Two Particles

Let us consider the events (Fig. 1)

$$a \to b + c$$

followed by

$$b \to d + e,$$

where

$$j_a = j, \qquad j_l = \tfrac{1}{2} = j_d; \qquad j_c = 0 = j_e$$

[$\Xi_{13}^* \to \Xi^- + \pi^\circ$; $\Xi^- \to \Lambda + \pi^-$ is an example with $j = \frac{3}{2}$].

[5] A. J. Macfarlane, Rev. Mod. Phys. 34, 41 (1962).
[6] N. Byers and S. Fenster, Phys. Rev. Letters 11, 52 (1963).

With $\mathbf{p}_a = 0 = (\mathbf{p}_b)_x = (\mathbf{p}_b)_y$, the multipole parameters of b (t_a's) are given in terms of t_a's in the general case (i.e., without putting $j_b = \frac{1}{2}$) as

$$t_{bL}^{M} = (-)^{M}(2j_a + 1)^{-1} \sum_{L} (2L + 1) t_{aL}^{M} g_{LL}^{M}, \qquad (3.8)$$

where

$$g_{LL}^{M} = \sum_{l,l'} S_l S_{l'}^{*} \sum_{\sigma_b} c(j_a L j_a; \sigma_b' - M) C(j_b \tilde{L} j_b; \sigma_b M)$$

$$\times C(j_b l j_a; \sigma_b 0) C(j_b l' j_a; \sigma_b' 0)(\sigma_b' = \sigma_b + M). \qquad (3.9)$$

Similarly the t_d's are given by an exactly similar formula in terms of the t_b's in *the rest frame of b with the momentum of d parallel to the z axis.*

Thus the next step is to pass over from the "standard" frome for the event $a \to c + d$, to that for the event $b \to d + e$.

If the plane of the momenta $\mathbf{p}_b$, $\mathbf{p}_d$, $\mathbf{p}_e$ in the first frame is taken as the zx plane, this involves (i) a pure Lorentz transformation $\Lambda_{(p_b)}$ [where $\Lambda_{(p_b)} \cdot p_b = (m_b, \mathbf{0})$] and (ii) a rotation with the Euler angles $(0, \beta, 0)$ where

$$\cos\beta = \mathbf{p}_b, \boldsymbol{\pi}_d / |\mathbf{p}_b| \cdot |\boldsymbol{\pi}_d| , \qquad (3.10)$$

with

$$\boldsymbol{\pi}_d = \Lambda_{(p_b)} \cdot p_d$$

or

$$\boldsymbol{\pi}_d = \left(\mathbf{p}_d - \frac{p_b \cdot p_d + m_b p_d^0}{m_b(m_b + p_b^0)} \mathbf{p}_b \right). \qquad (3.11)$$

Now from (1.4) we find that (i) $\Lambda_{(p_b)}$ leaves the t_b's unaffected and (ii) $R(0, \beta, 0)$ simply rotates the polarization $\mathbf{p}_b$ about y axis through an angle β.

$$\left[\text{for } j = \tfrac{1}{2}, \mathbf{P} = \sqrt{3} \left\{ \frac{1}{\sqrt{2}} (t_1^{-1} - t_1^{1}), \frac{i}{\sqrt{2}} (t_1^{-1} + t_1^{1}), t_1^{0} \right\} \right].$$

Thus denoting by S' the S matrix for the event $b \to d + e$, we have

$$\text{tr } \rho_d = \text{tr } \rho_b \left[(|S_0'|^2 + \tfrac{1}{3} |S_1'|^2) + \frac{1}{\sqrt{3}} (S_0' S_1'^{*} + S_0'^{*} S_1')(\cos\beta P_b^3 + \sin\beta P_b^1) \right],$$

$$P_d^3 = \frac{\text{tr } \rho_b}{\text{tr } \rho_d} \left[\frac{1}{\sqrt{3}} (S_0' S_1'^{*} + S_0'^{*} S_1') + (|S_0'|^2 + \tfrac{1}{3} |S_1'|^2)(\cos\beta P_b^3) \right], \qquad (3.12)$$

$$P_d^1 = \frac{\text{tr } \rho_b}{\text{tr } \rho_d} \left[(|S_0'|^2 - \tfrac{1}{3} |S_1'|^2)(\cos\beta P_b^1 - \sin\beta P_b^3) + \frac{i}{\sqrt{3}} (S_0' S_1'^{*} - S_0'^{*} S_1') P_b^2 \right],$$

$$P_d^2 = \frac{\text{tr } \rho_b}{\text{tr } \rho_d} \left[(|S_0'|^2 - \tfrac{1}{3} |S_1'|^2) P_b^2 - \frac{i}{\sqrt{3}} (S_0' S_1'^{*} - S_0'^{*} S_1')(\cos\beta P_b^1 - \sin\beta P_b^3) \right].$$

tr ρ_d corresponds to the differential cross section when $\mathbf{P}_d$ is not observed. The explicit expressions for tr ρ_b and $\mathbf{P}_b$ in terms of S and ρ_a are given in Appendix C (i). The above formulas are all exact without approximations. Additional simplifications may be introduced in particular cases by neglecting, according to the circumstances, certain parameters of ρ_a, S, S'.

C. Decays into Three Particles

As has already been emphasized the fundamental physical fact brought out most simply in the canonical representation is the dependence of the internal degrees of freedom denoted by "spin" on the external ones of momenta. Taking the relatively simple but already interesting case of the decay

$$a \to b + c + d$$

we now propose to show how our technique helps us to extract the exact dependence of the multipole parameters of the product particles on the momenta involved. As before [A (ii)], we suppose that $j_c = j_d = 0$. (When more than one product particle has spin, the irreducible components into which the direct product of the corresponding density matrices may be decomposed, no longer have any very simple transformation properties under pure Lorentz transformations, several different momenta being involved. However the case we consider is quite often realized in practice in decays through strong interactions and hence is of considerable interest.)

Using (3.5) we can write in the "standard" frame,

$$(\rho_b)_{\sigma_b \sigma_b'} = \sum_{\zeta} (\rho_a)_{\sigma_a \sigma_a'} y_{LM}(\hat{\pi}) g_\zeta, \qquad (3.13)$$

where the summation index ζ stands for the set $(\sigma_a \sigma_a'; l', l'', l; \tilde{l}', \tilde{l}'', \tilde{l}; L)$ and

$$g_\zeta = S_{(l'l'')l} S_{(\tilde{l}'\tilde{l}'')\tilde{l}}^{*} \cdot (-)^{\tilde{m}'} \left[\frac{(2l' + 1)(2\tilde{l}' + 1)}{4\pi(2L + 1)} \right]^{\frac{1}{2}} \times (\tilde{l}' 0 l' 0 \mid L0)(\tilde{l}' - \tilde{m}' l' m' \mid LM) \times (l' m' l'' 0 \mid l m')(\tilde{l}' \tilde{m}' \tilde{l}'' 0 \mid \tilde{l} \tilde{m}') \times (l m' j_b \sigma_b \mid j_a \sigma_a)(\tilde{l} \tilde{m}' j_b \sigma_b' \mid j_a \sigma_a') \qquad (3.14)$$

with

$$m' = \sigma_a - \sigma_b, \qquad \tilde{m}' = \sigma_a' - \sigma_b', \qquad M = m' - \tilde{m}'.$$

Using (1.5), (1.5′) along with (3.13) and (3.14), we obtain the $[t_b]$'s directly in terms of the $[t_a]$'s.

For practical purposes an approximation has to be made, usually by neglecting all the reduced S-matrix elements $(S_{(l'l'')l})$ except those which correspond to the lowest few angular momenta eigenvalues.

Let us consider the simplest example, namely,

$$j_a = \tfrac{1}{2} = j_b,$$

$$(\text{example: } Y_{01}^{*} \to \Lambda + \pi + \pi),$$

keeping, as an example of approximate results, only $S_{(00)0}$ and $S_{(10)1}$ and neglecting all the rest.

This gives

$$\mathbf{P}_b = \frac{1}{\operatorname{tr} \rho_b}\left[(\eta_1 + \eta_2)\mathbf{P}_a - \eta_3\hat{\pi} + \eta_4\mathbf{P}_a \times \hat{\pi} + 2\eta_2\mathbf{P}_a \times \hat{\pi} \times \hat{\pi}\right] \tag{3.15}$$

and

$$\operatorname{tr} \rho_b = [(\eta_1 + \eta_2) - \eta_3\mathbf{P}_a, \hat{\pi}], \tag{3.16}$$

where

$$\eta_1 = \frac{1}{4\pi} S_{(00)0}S_{(00)0}^{*}, \quad \eta_2 = \frac{1}{4\pi} S_{(10)1}S_{(10)1}^{*},$$

$$\eta_3 = \frac{1}{4\pi}(S_{(00)0}S_{(10)1}^{*} + S_{(00)0}^{*}S_{(10)1}),$$

$$\eta_4 = \frac{1}{4\pi i}(S_{(00)0}S_{(10)1}^{*} - S_{(00)0}^{*}S_{(10)1}).$$

[If, in addition, we also take into account $S_{(01)1}$ and $S_{(11)1}$ then (along with $\mathbf{P}_a$ and $\hat{\pi})\hat{k}$, the unit vector parallel to the z axis, also enters into the right-hand side of (3.15) and (3.16).]

Now if we pass over to another (arbitrary) frame of reference $\mathbf{P}_b$ will undergo a rotation R_W corresponding to the pure Lorentz transformation or tation (or a combination of both) that is used. We see that $\mathbf{P}_b' = R_W\mathbf{P}_b$ is exactly by the same equation as above (3.15), (3.16) provided we replace $\mathbf{P}_a$ and $\hat{\pi}$ by $\mathbf{P}_a' = R_W \cdot \mathbf{P}_a$ and $\hat{\pi}' = R_W \cdot \hat{\pi}$, respectively.

In using this compact expression it should be noted that $\mathbf{P}_a'$ does not represent the polarization of a in the new frame if the latter is obtained by a pure Lorentz transformation (when in fact $\mathbf{P}_a$ simply remains invariant) while for a pure rotation it indeed does so.

The corresponding expressions for the case $j_a = \frac{3}{2}$, $j_b = \frac{1}{2}$ are given in Appendix C (ii).

Generalizations to the cases where there are a greater number of product particles or to successive decays may be carried out, though the corresponding expressions become more and more complicated.

What is to be particularly noted in the preceding formulas is the way in which the polarization parameters depend not only on the direction but also on the *magnitudes* of the momenta involved.

Even in the "standard" frame, the magnitudes are directly involved since $\hat{\pi}$ [in (3.5), (3.6)] is obtained through a Lorentz transformation. Magnitudes are again brought in (through R_W) as we transform to any other frame. Our technique permits us to extract this dependence (a fundamental relativistic feature) in the most convenient manner.

4. CONCLUSION

In this paper we have discussed some applications of the Lorentz transformation properties of the canonical spin tensors. Two principal types of applications have been discussed namely, to the problem of precession of spin and to that of polarization analysis of decay products. In the precession problem we have mainly tried to show how simply and directly our formalism leads to the kinematical equation (2.6) which already includes the "Thomas factor" (as discussed more fully in Appendix B).

So far as applications to polarization analysis are concerned we have carried through the calculations up to the construction of the corresponding density matrices and multipole parameters. Once these parameters have been obtained they may be used to obtain information on the possibly unknown quantum numbers of the particles involved on the reduced S-matrix elements characterizing the interaction process. These calculations may, however, carried out as usual.[6,7] No specially new technique is involved once the multipole parameters are obtained as indicated in Sec. 4. We have used the canonical representation. The corresponding picture in the spinor representation is discussed in Appendix A.

APPENDIX A: THE CORRESPONDING FORMALISM IN SPINOR REPRESENTATION

In this paper we have throughout used the canonical representation (1). Here we indicate the equivalent technique in the spinor representation.

To this end, we first consider some aspects of the interrelations of these two types of representations.

The infinitesimal operators of the homogeneous Lorentz group are, for the two types of nonequivalent spinor representations $[(s,0), (0,s)]$,

$$\mathbf{N} = -iP^0(\partial/\partial\mathbf{P}) \mp i\mathbf{S}$$
$$\mathbf{M} = -i\mathbf{P} \times \partial/\partial\mathbf{P} + \mathbf{S} \tag{A1}$$

[7] M. Amadello and R. Gatto, Phys. Rev. 133, B531 (1964).

where $\mathbf{S}$ is the usual $(2S+1)\times(2S+1)$ Hermitian spin matrix. [These are to be compared with the operators (2.1) for the canonical representation.] The wavefunctions of both types are supposed to satisfy the KG equation. To obtain a representation invariant under space reflection, the above two may be combined as follows:
let

$$\Gamma^0 = \begin{vmatrix} 0 & I_{2S+1} \\ I_{2S+1} & 0 \end{vmatrix}, \quad \mathbf{\Sigma} = \begin{vmatrix} \mathbf{S} & 0 \\ 0 & \mathbf{S} \end{vmatrix}, \quad \mathbf{A} = \begin{vmatrix} \mathbf{S} & 0 \\ 0 & -\mathbf{S} \end{vmatrix}, \tag{A2}$$

where I_{2S+1} is the $(2S+1)\times(2S+1)$ unit matrix.

We note that

$$[\Gamma^0, \mathbf{\Sigma}]_- = 0, \qquad [\Gamma^0, \mathbf{A}]_+ = 0. \tag{A2'}$$

Let Ω_s, χ_s be the wavefunctions transforming according to $(s,0)$ and $(0,s)$, respectively. Now, coupling together the two types of spinors we obtain a wavefunction

$$\begin{vmatrix} \chi_s \\ \Omega_s \end{vmatrix} \tag{A3}$$

for which the intrinsic parts of N and M are now given by $i\mathbf{A}$ and $\mathbf{\Sigma}$, respectively.

In Ref. 1(a), we showed that the Dirac equation is transformed to the form

$$(\gamma^0(p^2)^{\frac{1}{2}} - m)\psi(p) = 0, \tag{A4}$$

in the canonical representation and then may directly be decomposed into 2 two-component equations. The operator for this transformation is obtained from the spinor matrix corresponding to the pure Lorentz transformation to the rest system (1).

[This matrix was denoted in Ref. 1 by $Q(p, \Lambda_{k\leftarrow p})$. This is a notational confusion. The correct notation is $Q(k, \Lambda_{k\leftarrow p})$ in view of the equation

$$(U(\Lambda_{k\leftarrow p})\varphi)(k) = Q(k, \Lambda_{k\leftarrow p})\varphi(p).$$

None of the results derived in (1) need, of course, be changed.]

The spinor wave equation for the general case may be obtained directly by putting the question: "what equation is transformed to

$$(\Gamma^0(p^2)^{\frac{1}{2}} - m)\psi(p) = 0 \tag{A5}$$

in the canonical representation?"

The intrinsic part of N now being given by $i\mathbf{A}$, we have to transform [in order to obtain (A5) and (2.1) with $\mathbf{S}$ replaced by $\mathbf{\Sigma}$] by the operator

$$\exp(-\mathbf{A}\cdot\hat{P}\theta),$$

$$\text{where } \hat{P} = \mathbf{P}/|\mathbf{P}|, \qquad \tanh\theta = |\mathbf{P}|/P^0. \tag{A6}$$

Thus the required spinor wave equation for arbitrary spin is

$$(\exp(\mathbf{A}\cdot\hat{P}\theta)\Gamma^0\exp(-\mathbf{A}\cdot\hat{P}\theta)(p^2)^{\frac{1}{2}} - m)\psi_S = 0 \tag{A7}$$

(where ψ_S denotes the spinor wavefunction) or

$$[(P^2)^{\frac{1}{2}}\Gamma^0\exp(-2\mathbf{A}\cdot\hat{P}\theta) - m]\psi_S = 0. \tag{A8}$$

For $S = \frac{1}{2}$ this reduces to the familiar Dirac equation. For $S = 1$, we have

$$\Gamma^0[m^2 + 2(P^0\mathbf{A}\cdot\mathbf{P} - (\mathbf{A}\cdot\mathbf{P})^2]\psi_S = m^2\psi_S. \tag{A9}$$

In fact (A8) gives just the spinor wave equations recently given in an explicit form by Weinberg.[8]

Moreover, we see that [putting $(p^2)^{\frac{1}{2}} = m$ in (A5) for positive mass values, which are, of course, only to be retained] the canonical wavefunction

$$\psi_c = \begin{vmatrix} \chi_c \\ \Omega_c \end{vmatrix}$$

must satisfy $\chi_c = \Omega_c = \varphi$, say. Thus the coupling in the spinor representation must be of the form

$$\psi_s = \exp(\mathbf{A}\cdot\hat{P}\theta)\begin{vmatrix} \varphi \\ \varphi \end{vmatrix}. \tag{A10}$$

Diagonalizing Γ^0, the above solutions give for $S = \frac{1}{2}$, the familiar "large" and "small" components of the Dirac wavefunctions. Thus we see that in terms of the canonical representation the coupling necessary for the "doubled" spinor representation has an extremely simple significance.

We now note that since Γ^0 commutes with $\mathbf{\Sigma}$, the projection operator for the eigenvalue $(s-r)$ of $\mathbf{\Sigma}\cdot\hat{n} = \Sigma_n$ (where $\hat{n}$ is any unit vector) may directly be defined on the manifold of the solutions of (A5) as

$$(\pi_{s-r})_c = \frac{(\Sigma_n - s)(\Sigma_n - s + 1)\cdots(\Sigma_n - s + r - 1)(\Sigma_n - s + r + 1)\cdots(\Sigma_n + s)}{(-)^r r!\,(2s - r)!}. \tag{A11}$$

The corresponding projection operator in the spinor representation is given by

$$(\pi_{s-r})_s = \exp(\mathbf{A}\cdot\mathbf{P}\theta)(\pi_{s-r})_c\exp(-\mathbf{A}\cdot\mathbf{P}\theta). \tag{A12}$$

This, of course, is the same as replacing $\mathbf{\Sigma}$ by

$$\exp(\mathbf{A}\cdot\mathbf{P}\theta)\mathbf{\Sigma}\exp(-\mathbf{A}\cdot\mathbf{P}\theta). \tag{A13}$$

[8] S. Weinberg, Phys. Rev. **133**, B1318 (1964).

This definition of spin operator, though defined in the general case in (1) was discussed in detail only for $S = \frac{1}{2}$. Utilizing this definition, the applications discussed in this paper may be carried through exactly similarly in the spinor representation. The corresponding calculations [the commutation on the lhs of (2.2), to take only one example] if carried out directly in the spinor formalism, are usually much more tedious, though the final result must be equivalent.

The definition (A13) is often indirectly worked into the theory by stating that the eigenvalues of the spin to be taken as those obtained after transforming to the rest system. This involves defining at first the "spin operator" as $\mathbf{\Sigma}$ and then (since it proves inadequate) "reinterpretating" it in the above way. This gives equivalent results though in (A13) not change of reference system is involved [see Footnote 3 of Ref. 1(a)]. This roundabout way seems quite unnecessary and is probably one of the factors preventing up to now a systematic use of the Lorentz transformation properties of the canonical spin tensors.

APPENDIX B

In Sec. 2 we have derived the formula

$$\langle \dot{S} \rangle = \frac{\gamma}{\gamma + 1} (\mathbf{v} \times \dot{\mathbf{v}}) \times \langle \mathbf{S} \rangle. \tag{B1}$$

Here, we propose to show that the factor $\gamma/(\gamma + 1)$ corresponds to the Thomas factor of correction for the precession of polarization in an electromagnetic field.

We do not claim to provide a rigorous demonstration. Our aim is rather to indicate the essential link in the simplest possible fashion.

If we consider the acceleration to be that due to an external field the above equation gives in general only a part of the time variation of $\langle \mathbf{S} \rangle$, namely that directly due to the acceleration induced. The effect of such forces as magnetic fields which contribute directly to the precession is not included. They are to be added to give the total effect. Also in (2.1) $P_j \equiv i(\partial/\partial xj)$ $(j = 1, 2, 3)$ corresponds only to the kinetic part of the momentum which is, however, only needed—the precession being a purely kinematical effect.

Let us first consider only an electric field $\mathbf{E}(\mathbf{H} = 0)$. Supposing the trajectory of the particle to be the one given by the classical Lorentz force, i.e., putting

$$\frac{d}{dt}(\gamma \mathbf{v}) = -\frac{e}{m} \mathbf{E} \tag{B2}$$

we have

$$\gamma(\mathbf{v} \times \dot{\mathbf{v}}) = -\frac{e}{m} \mathbf{v} \times \mathbf{E}. \tag{B3}$$

Thus in this case,

$$\langle \dot{S} \rangle = -\frac{e}{m} \frac{1}{\gamma + 1} \mathbf{v} \times \mathbf{E}. \tag{B4}$$

This is exactly the result of Thomas[9] for $\mathbf{H} = 0$.

If there is a magnetic field as well, accepting the result of Larmor's theorem that the effect of a relatively weak $\mathbf{H}$ in presence of an electric field is just to add (so far as the "normal" part of the magnetic dipole moment is concerned) an angular velocity $(e/m\gamma)\mathbf{H}$, we have

$$\langle \dot{S} \rangle = \frac{e}{m} \left(\frac{\mathbf{H}}{\gamma} - \frac{1}{\gamma + 1} \mathbf{v} \times \mathbf{E} \right) \times \langle \mathbf{S} \rangle. \tag{B5}$$

This is the complete formula of Thomas[9] (for $g = 2$).

Bacry[10] derived the Thomas equation from the covariant BMT equation[11] by replacing the 4-vector σ (satisfying $p.\sigma = 0$) of BMT by $(0, \sigma') = L_p \cdot \sigma$. The equivalent of our canonical $\mathbf{S}$ in the spinor representation is $\Lambda_{(p)} \cdot \omega$ where $\omega = -P \cdot M^*$ is the Bargmann–Wigner 4-vector satisfying $\omega \cdot P = 0$. Hence, the final agreement of the results is only to be expected.

It has been shown by Fradkin and Good[12] and by Rubinow and Keller,[13] that considering explicitly the Dirac equation in presence of an electromagnetic field and a 4-vector definition of the polarization, one can derive, making all the required approximations, the BMT equation. Passing to the three-component definition with the help of $\Lambda_{(P)}$ we obtain again the Thomas equation.[12] This is certainly a relatively rigorous approach. However, not only are such calculations rather complicated (even for spin $\frac{1}{2}$, which case only has been treated), but it obscures the simple and fundamental relation of the result to the Lorentz transformation properties of the polarization vector.

APPENDIX C

(i) In the "Standard" frame for the event

$$a \rightarrow b + c$$

[9] L. Thomas, Phil. Mag. **3**, 1 (1927).
[10] H. Bacry, thesis, Faculté des Sciences, Marseille, 1963 (unpublished).
[11] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Letters **2**, 435 (1959).
[12] D. M. Fradkin and R. H. Good, Jr., Rev. Mod. Phys. **33**, 343 (1961).
[13] S. I. Rubinow and J. B. Keller, Phys. Rev. **131**, 2789 (1963).

[see Sec. 3B], we have putting

$$\eta_1 \equiv \left(\frac{2j+1}{4j}\right)^{\frac{1}{2}} S_{j-\frac{1}{2}}, \qquad \eta_2 = \left(\frac{2j+1}{4(j+1)}\right)^{\frac{1}{2}} S_{j+\frac{1}{2}}$$

and

$$\begin{aligned}
&\rho \equiv \rho_a,\\
\operatorname{tr} \rho_b &= \{(\rho_{\frac{1}{2}\frac{1}{2}} + \rho_{-\frac{1}{2}-\frac{1}{2}})(|\eta_1|^2 + |\eta_2|^2)\\
&\quad + (\rho_{\frac{1}{2}\frac{1}{2}} - \rho_{-\frac{1}{2}-\frac{1}{2}})(\eta_1\eta_2^* + \eta_1^*\eta_2)\},\\
P_b^3 &= \frac{1}{\operatorname{tr} \rho_b}\{(\rho_{\frac{1}{2}\frac{1}{2}} - \rho_{-\frac{1}{2}-\frac{1}{2}})(|\eta_1|^2 + |\eta_2|^2)\\
&\quad + (\rho_{\frac{1}{2}\frac{1}{2}} + \rho_{-\frac{1}{2}-\frac{1}{2}})(\eta_1\eta_2^* + \eta_1^*\eta_2)\},\\
P_b^1 &= \frac{1}{\operatorname{tr} \rho_b}\{(\rho_{\frac{1}{2}-\frac{1}{2}} + \rho_{-\frac{1}{2}\frac{1}{2}})(|\eta_1|^2 - |\eta_2|^2)\\
&\quad - (\rho_{\frac{1}{2}-\frac{1}{2}} - \rho_{-\frac{1}{2}\frac{1}{2}})(\eta_1\eta_2^* - \eta_1^*\eta_2)\},\\
P_b^2 &= \frac{1}{\operatorname{tr} \rho_b}\{(\rho_{\frac{1}{2}-\frac{1}{2}} - \rho_{-\frac{1}{2}\frac{1}{2}})(|\eta_1|^2 - |\eta_2|^2)\\
&\quad - (\rho_{\frac{1}{2}-\frac{1}{2}} + \rho_{-\frac{1}{2}\frac{1}{2}})(\eta_1\eta_2^* - \eta_1^*\eta_2)\},
\end{aligned} \tag{C1}$$

where

$$\begin{aligned}
\rho_{\pm\frac{1}{2}\pm\frac{1}{2}} &= (2j+1)^{-1}\sum_L (2L+1) t_L^0 C(jLj; \pm\tfrac{1}{2}0),\\
\rho_{\pm\frac{1}{2}\mp\frac{1}{2}} &= -(2j+1)^{-1}\\
&\quad \times \sum_L (2L+1) t_L^{\mp 1} C(jLj; \mp\tfrac{1}{2} \pm 1).
\end{aligned} \tag{C2}$$

(ii) Let the event

$$a \to b + c + d,$$

(where $j_a = \frac{3}{2}$, $j_b = \frac{1}{2}$, $j_c = 0 = j_d$) be considered in its "standard" frame. For the sake of simplicity we suppose that only the diagonal elements of ρ_a are nonzero (which corresponds to a cylindrical symmetry of the production process, giving a, about the z axis) and that all the reduced S-matrix elements except $\alpha = S_{(10)1}$ and $\beta = S_{(01)1}$ may be neglected.

Denoting $(\rho_a)_{mm}$ by ρ_m and supposing it to be normalized to give $\operatorname{tr} \rho = 1$, we have (writing simply y_{lm} for $y_{lm(\phi)}$)

$$\begin{aligned}
(4\pi)^{\frac{1}{2}}(\rho^b_{\frac{1}{2}\frac{1}{2}} \pm \rho^b_{-\frac{1}{2}-\frac{1}{2}}) &= \alpha\alpha^*(y_{00} - 5^{-\frac{1}{2}} y_{20})\\
&\quad \times \{(\rho_{\frac{3}{2}} \pm \rho_{-\frac{3}{2}}) \pm \tfrac{1}{3}(\rho_{\frac{1}{2}} \pm \rho_{-\frac{1}{2}})\} + \tfrac{2}{3}\{(\alpha\alpha^* + \beta\beta^*)y_{00}\\
&\quad + (\alpha\beta^* + \alpha^*\beta)y_{10} + 2\cdot 5^{-\frac{1}{2}}\alpha\alpha^* y_{20}\}(\rho_{\frac{1}{2}} \pm \rho_{-\frac{1}{2}}),
\end{aligned} \tag{C3}$$

$$\begin{aligned}
(4\pi)^{\frac{1}{2}}(\rho^b_{\frac{1}{2}-\frac{1}{2}} + \rho^b_{-\frac{1}{2}\frac{1}{2}}) &= \tfrac{1}{3}\sqrt{2}\,\{(\alpha\beta + \alpha^*\beta)y_{11}\\
&\quad + 2(\tfrac{3}{5})^{\frac{1}{2}}\alpha\alpha^* y_{21}\}(\rho_{\frac{1}{2}} - \rho_{-\frac{1}{2}}),\\
(4\pi)^{\frac{1}{2}}(\rho^b_{\frac{1}{2}-\frac{1}{2}} - p^b_{-\frac{1}{2}\frac{1}{2}}) &= \tfrac{1}{3}\sqrt{2}\,(\alpha\beta^* - \alpha^*\beta)y_{1}(\rho_{\frac{1}{2}} + \rho_{-\frac{1}{2}}),
\end{aligned}$$

with

$$\begin{aligned}
\operatorname{tr} \rho &= (\rho_{\frac{3}{2}} + \rho_{-\frac{3}{2}}) + (\rho_{\frac{1}{2}} + \rho_{-\frac{1}{2}}) = 1,\\
(\rho_{\frac{3}{2}} + \rho_{-\frac{3}{2}}) &= \tfrac{1}{2}(1 + 5^{\frac{1}{2}} t_2^0),\\
(\rho_{\frac{1}{2}} + \rho_{-\frac{1}{2}}) &= \tfrac{1}{2}(1 - 5^{\frac{1}{2}} t_2^0),\\
(\rho_{\frac{3}{2}} - \rho_{-\frac{3}{2}}) &= \frac{1}{2(5^{\frac{1}{2}})}(3\sqrt{3}\, t_1^0 + 7^{\frac{1}{2}} t_3^0),\\
(\rho_{\frac{1}{2}} - \rho_{-\frac{1}{2}}) &= \frac{3}{2(5^{\frac{1}{2}})}\left(3^{-\frac{1}{2}} t_1^0 + 7^{\frac{1}{2}} t_3^0\right).
\end{aligned}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Reduction of the N-Particle Variational Problem*

CLAUDE GARROD† AND JEROME K. PERCUS
Courant Institute of Mathematical Sciences
New York University, New York, New York
(Received 9 June 1964)

A variational method is presented which is applicable to N-particle boson or fermion systems with two-body interactions. For these systems the energy may be expressed in terms of the two-particle density matrix: $\Gamma(1, 2 \mid 1', 2') = (\Psi \mid a_2^+ a_1^+ a_{1'} a_{2'} \mid \Psi)$. In order to have the variational equation: $\delta E/\delta\Gamma = 0$ yield the correct ground-state density matrix one must restrict Γ to the set of density matrices which are actually derivable from N-particle boson (or fermion) systems. Subsidiary conditions are presented which are necessary and sufficient to insure that Γ is so derivable. These conditions are of a form which render them unsuited for practical application. However the following necessary (but not sufficient) conditions are shown by some applications to yield good results: It is proven that if $\Gamma(1, 2 \mid 1', 2')$ and $\gamma(1 \mid 1')$ are the two-particle and one-particle density matrices of an N-particle system [normalized by $\mathrm{tr}\,\Gamma = N(N-1)$ and $\mathrm{tr}\gamma = N$] then the associated operator: $G(1, 2 \mid 1', 2') = \delta(1 - 1')\gamma(2 \mid 2') + \sigma\,\Gamma(1', 2 \mid 1, 2') - \gamma(2 \mid 1)\gamma(1' \mid 2')$ is a nonnegative operator. [Here σ is $+1$ or -1 for bosons or fermions respectively.]

1. INTRODUCTION

We present in this paper a method of calculating ground-state properties of many-particle systems without using many-particle wavefunctions. It is a variational method in which the trial function is the one-particle or two-particle density matrix. These density matrices (which we define in due course) are functions of two or four variables respectively, and our approximate method, if carried through, yields a lower bound to the ground-state energy. There are a number of calculations presented here in which the results are exact but these are admittedly quite special cases.

In order to understand the motivation behind the method, let us review the many-particle variational calculation as it is usually done. We begin with a knowledge of (or a private conviction about) the mutual interactions of the particles and the external potential in which they move. That is, we begin with a Hamiltonian of the form:

$$H_N = \sum_{i=1}^{N} \left[\frac{\hbar^2}{2m} \nabla_i^2 + V(\mathbf{r}_i) \right] + \sum_{\substack{i,j=1 \\ i<j}}^{N} U(\mathbf{r}_i - \mathbf{r}_j). \tag{1.1}$$

We must then introduce a function of N variables which is symmetric with respect to the interchange of any pair of variables if the particles are bosons or antisymmetric if the particles are fermions. Among some set of such functions (which may be a rather restricted set due to the difficulty of manipulating functions of many variables) we find the one which gives the smallest expectation value of H_N. That is, we find the minimum of

$$E[\psi] = \frac{\int \psi^*(1, 2, \cdots, N) H_N \psi(1, 2, \cdots, N)\, d^3r_1 \cdots d^3r_N}{\int |\psi(1, 2, \cdots, N)|^2} \tag{1.2}$$

or more concisely,

$$E[\psi] = (\psi, H_N\psi)/(\psi, \psi). \tag{1.3}$$

The minimum value of E we thus obtain is an upper bound for the true ground-state energy of an N-particle system with the above interactions.

We now take the function, call it $\psi_0(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$, which gave the minimum as an approximation to the ground-state wavefunction. Our justification for doing this may be rather scant since the essential advantage of the variational method is that first-order errors in the trial function produce only second-order errors in the energy. Hence a fairly poor approximation to the wavefunction of the ground-state may produce a surprisingly good value for the energy. (Incidentally, the variational method we present will not have this property.) Using ψ_0,

* Supported in part by U. S. Atomic Energy Commission under Contract AT(30-1)-1480.
† Present address: University of California at Davis, California.

we must now calculate the quantities for the system which are of physical significance, that is, those properties which are more than theoretically measurable, such as

(1) the density

$$\rho(\mathbf{r}) = N \int |\psi(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N)|^2 \, d^3r_2 \cdots d^3r_N, \quad (1.4)$$

(2) the two-particle correlation function

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \times \int |\psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \cdots, \mathbf{r}_N)|^2 \, d^3r_3 \cdots d^3r_N, \quad (1.5)$$

(3) the momentum density

$$\rho(\mathbf{k}) = \frac{N}{(2\pi)^3} \int e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \psi^*(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N) \times \psi(\mathbf{r}', \mathbf{r}_2, \cdots, \mathbf{r}_N) \, d^3r \, d^3r' \, d^3r_2 \cdots d^3r_N \quad (1.6)$$

(where $\mathbf{p} = \hbar\mathbf{k}$).

We can calculate all of the above, and in general any quantity which depends only upon the coordinates or velocity of at most one or two particles at a time, if we first calculate the following "density matrices":

(1) the single-particle density matrix

$$\gamma(\mathbf{r} \mid \mathbf{r}') = N \int \psi^*(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N) \times \psi(\mathbf{r}', \mathbf{r}_2, \cdots, \mathbf{r}_N) \, d^3r_2 \cdots d^3r_N, \quad (1.7)$$

(2) the two-particle density matrix

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = N(N-1) \int \psi^*(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \cdots, \mathbf{r}_N) \times \psi(\mathbf{r}_1', \mathbf{r}_2', \mathbf{r}_3, \cdots, \mathbf{r}_N) \, d^3r_3 \cdots d^3r_N. \quad (1.8)$$

The particle density and momentum density are now given respectively by

$$\rho(\mathbf{r}) = \gamma(\mathbf{r} \mid \mathbf{r}) \quad (1.9)$$

and

$$\rho(\mathbf{k}) = \frac{1}{(2\pi)^3} \int e^{i\mathbf{k}\cdot\mathbf{r}} \gamma(\mathbf{r} \mid \mathbf{r}') e^{-i\mathbf{k}\cdot\mathbf{r}'} \, d^3r \, d^3r'. \quad (1.10)$$

More generally, the average number of particles in any normalized single-particle state $\phi^*(\mathbf{r})$ is given by:

$$\rho_\phi = (\phi, \gamma\phi) = \int \phi^*(\mathbf{r}) \gamma(\mathbf{r} \mid \mathbf{r}') \phi(\mathbf{r}') \, d^3r \, d^3r'. \quad (1.11)$$

The above formulas (1.9) and (1.10) are special cases in which

$$\phi = \delta(\mathbf{r} - \mathbf{r}_0) \quad \text{and} \quad \phi = [1/(2\pi)^{\frac{3}{2}}] e^{-i\mathbf{k}\cdot\mathbf{r}}.$$

Similarly the average number of pairs in an arbitrary two-particle state, $g^*(\mathbf{r}_1, \mathbf{r}_2)$, is

$$\rho_g = (g, \Gamma g). \quad (1.12)$$

Looking back, we see that we began with functions of one or two variables—namely the external potential and the two-particle interaction—and the final results of our calculation are other functions of a small number of variables. The many-particle wavefunctions appear only in an intermediate stage in going from the one set to the other. If we look at the definition of $E[\psi]$ we notice:

$$E[\psi] = \frac{\hbar^2}{2M} \int k^2 \rho(\mathbf{k}) \, d^3k + \int V(\mathbf{r}) \rho(\mathbf{r}) \, d^3r + \frac{1}{2} \int U(\mathbf{r}_1 - \mathbf{r}_2) \rho(\mathbf{r}_1, \mathbf{r}_2) \, d^3r_1 \, d^3r_2. \quad (1.13)$$

That is, the energy which we are to minimize only involves functions which can be immediately calculated from the density matrices. The question has occurred to many people—Why not avoid all reference to wavefunctions and use the density matrices themselves as variational functions? Because you get preposterous results—that's why. We show this by an example.

Let us calculate the ground-state density matrix for a system of one-dimensional hard-core bosons, that is, particles with an interaction potential

$$U(\mathbf{r}_1 - \mathbf{r}_2) = \begin{cases} 0 & \text{if } |\mathbf{r}_1 - \mathbf{r}_2| > d \\ \infty & \text{if } |\mathbf{r}_1 - \mathbf{r}_2| < d \end{cases}, \quad (1.14)$$

where d is the diameter of the "hard core". In making this calculation we are going to be a little bit smart and take into account all of the properties of the density matrices which are obvious from their definition. These properties are:

(1) symmetry

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \Gamma(\mathbf{r}_2, \mathbf{r}_1 \mid \mathbf{r}_1', \mathbf{r}_2') = \Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_2', \mathbf{r}_1'); \quad (1.15)$$

(2) relationship between γ and Γ

$$\gamma(\mathbf{r} \mid \mathbf{r}') = \frac{1}{N-1} \int \Gamma(\mathbf{r}, \mathbf{r}_2 \mid \mathbf{r}', \mathbf{r}_2) \, d^3r_2; \quad (1.16)$$

(3) trace

$$\text{tr}\,(\Gamma) = \int \Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1, \mathbf{r}_2) \, d^3r_1 \, d^3r_2 = N(N-1); \quad (1.17)$$

(4) nonnegativity of Γ as an operator

$$(g, \Gamma g) \geq 0 \quad \text{for } \textit{any} \quad g(\mathbf{r}_1, \mathbf{r}_2); \tag{1.18}$$

(5) Hermiticity

$$\Gamma(\mathbf{r}_1', \mathbf{r}_2' \mid \mathbf{r}_1, \mathbf{r}_2) = \Gamma^*(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2'). \tag{1.19}$$

We choose our trial Γ's from a set which satisfy Conditions (1), (3), (4), and (5). We then use Condition (2) to derive γ from Γ. We assume that the number of particles, N, is very large, and put these particles in a large one-dimensional "box" of length $L = N/\rho_0$. We neglect edge corrections in calculating energy integrals. It is convenient to introduce the variables:

$$x = r_1 - r_2 \quad \text{and} \quad R = \tfrac{1}{2}(r_1 + r_2). \tag{1.20}$$

In order to avoid an infinite value for the potential energy we must then have

$$\rho(r_1, r_2) = 0 \quad \text{for} \quad |x| \leq d. \tag{1.21}$$

We choose as our trial density matrices a one-parameter set:

$$\Gamma_\beta(r_1, r_2 \mid r_1', r_2') = 0 \quad \text{if} \quad |x| \text{ or } |x'| < d \tag{1.22}$$

otherwise:

$$\Gamma_\beta(r_1, r_2 \mid r_1', r_2') = C(1 - e^{-\beta(|x|-d)})(1 - e^{-\beta(|x'|-d)}), \tag{1.23}$$

where C is a positive normalization constant. That Γ_β satisfies Conditions (1), (4), and (5) is trivial to verify. To fix C we evaluate

$$\operatorname{tr} \Gamma_\beta = 4C \int_0^{L/2} dR \int_d^{L-2R} (1 - e^{-\beta(x-d)})^2 \, dx, \tag{1.24}$$

where the factor of 4 results from our having taken half the range of integration for both variables. Evaluating the integral we obtain

$$C = \frac{N(N-1)}{L(L - 3/\beta)} \tag{1.25}$$

or, for N, L very large:

$$C = \rho_0^2 \tag{1.26}$$

Now the Hamiltonian for our system is

$$H_N = \frac{-\hbar^2}{2M} \sum_{i=1}^{N} \frac{\partial^2}{\partial r_i^2} + \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} U(r_i - r_j). \tag{1.27}$$

But if we use the fact that

$$\frac{\partial^2}{\partial r_i^2} + \frac{\partial^2}{\partial r_j^2} = \frac{1}{2} \frac{\partial^2}{\partial R_{ij}^2} + 2 \frac{\partial^2}{\partial x_{ij}^2}, \tag{1.28}$$

we may rearrange the kinetic energy part in the form:

$$\sum_{i=1}^{N} \frac{\partial^2}{\partial r_i^2} = \frac{1}{N-1} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \left(\frac{\partial^2}{\partial r_i^2} + \frac{\partial^2}{\partial r_j^2} \right)$$

$$= \frac{1}{N-1} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \left(\frac{1}{2} \frac{\partial^2}{\partial R_{ij}^2} + 2 \frac{\partial^2}{\partial x_{ij}^2} \right). \tag{1.29}$$

Since the density matrix has been constructed to make the expectation value of U equal zero we need only evaluate the kinetic energy. Taking note of the definition of Γ [Eq. (1.8)] and the fact that, due to symmetry, we need only calculate the kinetic energy of one pair of particles and multiply by the number of pairs ($N(N-1)$) we see that the energy expectation value we obtain with our trial density matrix is

$$E = -\frac{4C}{(N-1)} \frac{\hbar^2}{M} \int_0^{L/2} dR \times \int_d^{L-2R} dx \, (1 - e^{-\beta(x-d)}) \left[\frac{\partial^2}{\partial x^2} (1 - e^{-\beta(x-d)}) \right]$$

$$= C\hbar^2 \beta L/(N-1)M. \tag{1.30}$$

Using the value of Eq. (1.25) for C we get

$$E = \frac{\hbar^2 \beta}{2M} \frac{N}{L - 3/\beta} \to \frac{\hbar^2 \rho_0 \beta}{2M} \quad \text{for} \quad N, L \text{ large.}$$

There are two things to notice about the result we have obtained. The first is that the total energy not only remains constant as N increases, but can be made arbitrarily small by choosing β small. The more disastrous but also more illuminating fact is that our trial density matrices behave perfectly well for d, the hard-core diameter larger than L/N. But one cannot possibly fit N hard-core particles in a length less than Nd! Clearly the density matrix which we would obtain in the case $d > L/N$ could never be derived from any N-particle wavefunction whatsoever according to Eq. (1.8). Hence if we want to avoid nonsensical results we must add the subsidiary condition that we only use, as trial density matrices, those which are derivable from some symmetric (or antisymmetric for fermions) N-particle wavefunction; that is, that Eq. (1.8) have at least one symmetric solution.

One way we could be sure that this subsidiary condition was satisfied would be to start with some set of symmetric N-particle wavefunctions and actually derive the corresponding density matrices. But it is just the use of N-particle functions that the introduction of density matrices was intended to avoid. Therefore, rather than take the more direct approach we try to solve the following problem

which will allow us to avoid the use of N-particle functions entirely.

Problem: Describe criteria by which one may determine whether any given function of four variables is in fact the two-particle density matrix of some system. That is, find necessary and sufficient conditions on the function, $\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2')$, for Eq. (1.8) to have at least one symmetric solution.

Certainly a partial solution to this problem is that the conditions given by Eqs. (1.15) through (1.19) are necessary. That they are not sufficient has been amply illustrated by the results of our sample calculation in which they were all utilized. Of course, by changing the word "symmetric" to "antisymmetric" one obtains the corresponding problem for fermions. There is also a similar problem in classical physics. For a classical system of N identical particles, the Hamiltonian function

$$H_N = \sum_{i=1}^{N} \frac{p_i^2}{2M} + V(\mathbf{r}_1, \cdots, \mathbf{r}_N)$$

is symmetric under the interchange of the coordinates of any two particles. Hence the equilibrium density:

$$\rho(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \frac{\exp\left(-V(\mathbf{r}_1, \cdots, \mathbf{r}_N)/kT\right)}{\int \exp\left(-V/kT\right) d^3r_1 \cdots d^3r_N}$$

is a completely symmetric, positive function. If we now calculate the two-particle density function for such a system:

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \int \rho(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)\, d^3r_3 \cdots d^3r_N,$$

we obtain a positive symmetric function of two variables. But not all positive symmetric functions of two variables can be derived in this way from positive symmetric functions of N variables, as can be seen by looking at the density function associated with the erroneous two-particle density matrix derived above. The problem is then to describe the set of two-particle functions which can be derived in this way. We present a partial solution to this problem.

In order to make the calculational method clear to those people who are unfamiliar with Fock spaces we organize the paper in the following way. In Sec. 2 we present a short and incomplete summary of earlier work on the problem. In Sec. 3 we investigate the problem of determining whether a given two-particle density function represents any real N-particle system. Necessary and sufficient conditions are presented for such representability. The conditions are of a form which make them of little value for practical application. A set of necessary (but not sufficient) conditions which are simple and yet strong enough to be sufficient in many special cases is also presented. Some applications of these necessary conditions are made. In Sec. 4 our attention turns to quantum-mechanical systems of bosons. Here we present conditions which must be satisfied by any two-particle density matrix. Using these conditions as constraints on our set of trial density matrices a simple calculation is made. In Sec. 5 the same is done for fermion systems. There, the fermion system to which we apply our variational method is one with the BCS pairing-force Hamiltonian. We also discuss the relationship of our variational method to the familiar Hartree–Fock method. Section 6 contains necessary and sufficient conditions for a given two-particle density matrix to be derivable from an N-particle system. Both bosons and fermions are treated. Again the necessary and sufficient conditions are of a form which makes them of little use in practical application. Section 7 contains the derivation of some of the necessary conditions used in the earlier sections. It also contains a generalization of the method for Hamiltonians which do not conserve particle number.

2. PREVIOUS INVESTIGATIONS

The quantum mechanical version of the problem has been with us for many years. What we judge to be the earliest reference to it is in the well-known book by von Neumann.[1] Since he never introduced the requirement that the many-particle functions satisfy Bose–Einstein or Fermi–Dirac statistics, he was led to ask something equivalent to the following question: Given a function of four variables Γ under what conditions can we be sure there exists some N-particle system from which this may be derived as the two-particle density matrix? His solution is that Γ must be a Hermitian, nonnegative operator of tr $N(N-1)$. In other words, if we neglect statistics, our Conditions (3), (4), and (5) are sufficient.

Since the ground state of a system of bosons is identical to the ground state of a system of particles which have the same interactions but no symmetry restrictions, one is tempted to calculate the ground-state properties of a system of bosons by solving the corresponding problem for an identical system without statistics using the von Neumann result.

[1] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955) (German edition—Springer-Verlag, Berlin, 1932).

The technique is invalid because, as soon as the requirement of symmetry is relaxed, one must introduce a different two-particle density matrix, $\Gamma_{ij}(\mathbf{r}_i\mathbf{r}_j \mid \mathbf{r}_i', \mathbf{r}_j')$, for each pair of particles. One is then led to the question: Given a set of functions, $\Gamma_{ij}(\mathbf{r}_i, \mathbf{r}_j \mid \mathbf{r}_i', \mathbf{r}_j')$, $(i, j \leq N)$ how can one determine whether these are the two-particle density matrices associated with any N-particle function? This problem is a generalization of the problem considered in this article.

From the time of von Neumann's investigation until 1955, when the papers of Löwdin[2] and Mayer[3] appeared, the only use made of density matrices was in statistical physics and polarization phenomena. Besides giving a full account of the physical interpretation and formal properties of one- and two-particle density matrices, Löwdin proved that the most efficient sequence of single-particle functions for a "configuration-interaction" calculation of the helium ground state is the sequence of eigenfunctions of the single-particle density matrix.

Mayer's paper reported a variational calculation of the ground-state density matrices for an electron gas. He varied the two-particle density matrix directly, taking into account our Conditions (1) through (5) and the Pauli exclusion principle. His results seemed very good. However in a later paper Tredgold[4] pointed out a mathematical error in Mayer's work and for the first time clearly showed that some unknown subsidiary conditions on the density matrices used in any variational calculation were necessary to obtain sensible results. Miyuno and Iyuyama[5] at about the same time also wrote an article with essentially the same content.

Shortly after this Ayres[6] took up the search for the required subsidiary conditions. Using the Pauli exclusion principle he obtained some asymptotic conditions on the one- and two-particle density matrices. In 1959, Bopp[7] utilized a clever method involving the two-particle density matrix to derive lower bounds for the energies of certain atoms and ions. A basic error in the method was later corrected by Coleman[8] to whom the reader is referred for a complete discussion.

In 1961, two papers[9,10] were published concerning the properties of density matrices. The first, by Carlson and Keller, proved an interesting theorem relating the p-particle density matrix[11] and the $(N - p)$-particle density matrix for an N-particle system, namely that the nonzero eigenvalues and their associated multiplicities were the same for both density matrices. The second article, by Yamada, used geometric analysis to derive certain conditions on the two-particle correlation function. His results is discussed in the body of the article where a more general form of his conditions is derived.

We also discuss the important papers of Yang[12] and Coleman[8] in a topic-by-topic fashion as we touch upon the same or related results. For a general review of the theory and applications of density matrices the reader is referred to the review articles of ter Haar[13] and McWeeney.[14] Most of the results reported here were presented in a research report of the Courant Institute of Mathematical Sciences.[15]

3. CLASSICAL SYSTEMS

A classical statistical ensemble of N-particle systems is described by any function of N variables which is nowhere negative and has an integral equal to one. If the particles are identical, then in all physical calculations we may restrict ourselves to symmetric ensembles. These are described by all functions with the properties:

(i) $$\rho(\mathbf{r}_1, \cdots, \mathbf{r}_i, \cdots, \mathbf{r}_j, \cdots, \mathbf{r}_N) = \rho(\mathbf{r}_1, \cdots, \mathbf{r}_j, \cdots, \mathbf{r}_i, \cdots, \mathbf{r}_N) \qquad (3.1)$$

for all i and j;

(ii) $$\rho(\mathbf{r}_1, \cdots, \mathbf{r}_N) \geq 0 \text{ for all values of } \mathbf{r}_1 \text{ to } \mathbf{r}_N; \qquad (3.2)$$

(iii) $$\int \rho(\mathbf{r}_1, \cdots, \mathbf{r}_N)\, d^3r_1 \cdots d^3r_N = 1. \qquad (3.3)$$

The question we consider is the following:

Given a function, $\rho(\mathbf{r}_1, \mathbf{r}_2)$, how may we determine if there exists at least one symmetric ensemble for which this function represents the two-particle

[2] P.-O. Löwdin, Phys. Rev. **97**, 1474, (1955).
[3] J. E. Mayer, Phys. Rev. **100**, 1579, (1955).
[4] R. H. Tredgold, Phys. Rev. **105**, 1421, (1957).
[5] Y. Mizuno and T. Isuyama, Progr. Theor. Phys. Kyoto **18**, 33, (1957).
[6] R. U. Ayres, Phys. Rev. **111**, 1453, (1958).
[7] F. Bopp, Z. Physik **156**, 1421 (1957).
[8] A. J. Coleman, Rev. Mod. Phys. **35**, 668, (1963).
[9] B. C. Carlson and J. M. Keller, Phys. Rev., **121**, 659 (1961).
[10] M. Yamada, Progr. Theoret. Phys. Kyoto **25**, 579 (1961).

[11] The p-particle density matrix is defined:
$$D_p(1, \cdots, p|1', \cdots, p') = \times \binom{p}{N} \int \psi^*(1, \cdots, N)\psi(1' \cdots, p', p+1, \cdots, N) \times dx_{p+1} \cdots dx_N.$$
[12] C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962).
[13] D. ter Haar, Rept. Progr. Phys. **24**, 304 (1961).
[14] R. McWeeney, Rev. Mod. Phys. **32**, 335 (1960).
[15] C. Garrod and J. K. Percus, AEC Research Report NYO-10, 431, Courant Institute of Mathematical Sciences (1963).

density, where the two-particle density is defined in terms of the ensemble density by

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \times \int \rho(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \cdots, \mathbf{r}_N)\, d^3r_3 \cdots d^3r_N. \tag{3.4}$$

In Sec. 6 we prove the following: If, for every integrable function $v(\mathbf{r}_1, \mathbf{r}_2)$, the inequality

$$\int \rho(\mathbf{r}_1, \mathbf{r}_2) v(\mathbf{r}_1, \mathbf{r}_2)\, d^3\mathbf{r}_1\, d^3\mathbf{r}_2 \geq \min \left\{ \sum_{\substack{i,j=1 \\ i \neq j}}^{N} v(\mathbf{R}_i, \mathbf{R}_j) \right\} \tag{3.5}$$

is satisfied, where the minimum on the right is taken over all sets of N spatial points $(\mathbf{R}_1, \cdots, \mathbf{R}_N)$ within the system volume, then there does exist at least one N-particle ensemble which has $\rho(\mathbf{r}_1, \mathbf{r}_2)$ as two-particle density.

In practice the minimum on the right-hand side of (3.5) is quite impossible to calculate which makes these necessary and sufficient conditions, in their present form, of little practical value. Since the right-hand side of (3.5) may be interpreted as the minimum value for an N-particle system of the two-particle potential $v(\mathbf{r}_1, \mathbf{r}_2)$ the conditions stated above suffer from the further difficulty that they involve just those quantities which we would ordinarily be interested in calculating. A more useful set of necessary although not sufficient conditions, the justification of which we leave to the end of this section, is the following:

If $\rho(\mathbf{r}_1, \mathbf{r}_2)$ is the two-particle density of any symmetric N-particle ensemble then:

(a) $\rho(\mathbf{r}_1, \mathbf{r}_2)$ is symmetric,

(b) $\rho(\mathbf{r}_1, \mathbf{r}_2) \geq 0$ for all values of $\mathbf{r}_1$ and $\mathbf{r}_2$. (3.6)

(c) $$\int \rho(\mathbf{r}_1, \mathbf{r}_2)\, d^3r_1\, d^3r_2 = N(N-1). \tag{3.7}$$

(d) First extract the function

$$\rho(\mathbf{r}) = \frac{1}{N-1} \int \rho(\mathbf{r}, \mathbf{r}_2)\, d^3r_2, \tag{3.8}$$

then construct the function

$$K(\mathbf{r}, \mathbf{r}') = \rho(\mathbf{r}, \mathbf{r}') + \delta(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}) - \rho(\mathbf{r})\rho(\mathbf{r}'). \tag{3.9}$$

Condition (4) is now that

$$\int f(\mathbf{r}) K(\mathbf{r}, \mathbf{r}') f(\mathbf{r}')\, d^3r\, d^3r' \geq 0 \tag{3.10}$$

for every real function $f(\mathbf{r})$.

Certainly the first three conditions are trivial. Condition (d) states that, if we consider $K(\mathbf{r}, \mathbf{r}')$ as the kernel of an integral operator, then it has no negative eigenvalues. In order to see how this condition may bring out the N-particle structure of the density let us apply it to a simple minimization problem for which we can easily determine the true minimum. Consider N one-dimensional particles with the following interparticle interactions:

$$U(x_i - x_j) = \begin{cases} \infty & \text{if } |x_i - x_j| \leq d \\ 0 & \text{if } |x_i - x_j| > d \end{cases}. \tag{3.11}$$

The particles are all in an external potential:

$$V(x) = \infty \qquad x < 0 \tag{3.12}$$

$$V(x) = -e^{-x} \qquad x > 0. \tag{3.13}$$

What is the minimum value of the total potential energy?

In terms of the one- and two-particle densities the energy may be written:

$$E = \int_{-\infty}^{\infty} V(x)\rho(x)\, dx + \frac{1}{2} \iint_{-\infty}^{\infty} U(x_1 - x_2)\rho(x_1, x_2)\, dx_1\, dx_2. \tag{3.14}$$

If the energy is not to be infinite we must have:

$$\rho(x) = 0 \quad \text{for} \quad x < 0 \tag{3.15}$$

$$\rho(x_1, x_2) = 0 \quad \text{for} \quad |x_1 - x_2| < d. \tag{3.16}$$

Let us now apply Condition (d) using a test function $f(x)$ which is zero everywhere, except in the interval from x_0 to $x_0 + d$, where $f(x)$ equals one. Then the product $f(x_1)\rho(x_1, x_2)f(x_2)$ is always zero. Hence Condition (d) gives

$$\int_{x_0}^{x_0+d} \rho(x)\, dx - \left(\int_{x_0}^{x_0+d} \rho(x)\, dx \right)^2 \geq 0 \tag{3.17}$$

But in order for this to be true we must have:

$$\int_{x_0}^{x_0+d} \rho(x)\, dx \leq 1. \tag{3.18}$$

Thus we obtain the expected result that the average number of particles in any interval of length d cannot be larger than one. Now let us look at the interval, $0 \leq x \leq d$. Our condition states that we cannot have more than one particle within this interval. Certainly we obtain the largest negative contribution to the energy if we have exactly one particle at the origin. Looking at the second interval, $d \leq x \leq 2d$ we see we get a minimum of E if we concentrate exactly one particle at $x = d$. Going on, interval by interval, each time applying Eq. (3.18), we obtain the one-particle density:

$$\rho(x) = \sum_{n=0}^{N-1} \delta(x - nd). \tag{3.19}$$

Thus the density which yields the energy minimum clearly exhibits the N-particle structure of the system although the N-particle density function was never introduced.

Perhaps the most common class of systems encountered in statistical physics is that of systems with uniform density. If $\rho(\mathbf{r})$ is constant then the two-particle density, $\rho(\mathbf{r}_1, \mathbf{r}_2)$ depends only upon the coordinate $r = |\mathbf{r}_1 - \mathbf{r}_2|$ (if the system is also isotropic). We write this dependence in the form:

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \rho_0[\rho_0 - g(r)] \tag{3.20}$$

where ρ_0 is the constant value of $\rho(\mathbf{r})$. The kernel needed in Condition (d) is then

$$K(\mathbf{r}, \mathbf{r}') = \rho_0\delta(\mathbf{r} - \mathbf{r}') - \rho_0 g(|\mathbf{r} - \mathbf{r}'|). \tag{3.21}$$

We must demand that the eigenvalues of this kernel be nonnegative. But these eigenvalues are not hard to compute since the normalized eigenfunctions may be shown to be:

$$F_{\mathbf{k}}(\mathbf{r}) = V^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad \text{where } V \text{ is the volume of the system.} \tag{3.22}$$

Operating with the kernel K upon these functions we obtain

$$\int K(\mathbf{r}, \mathbf{r}')F_{\mathbf{k}}(\mathbf{r}')\, d^3r' = \rho_0 V^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}} - \rho_0 V^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}} \int g(|\mathbf{r} - \mathbf{r}'|) e^{i\mathbf{k}\cdot(\mathbf{r}'-\mathbf{r})}\, d^3r'. \tag{3.23}$$

In the second integral we change variables from $\mathbf{r}'$ to $\mathbf{r}' - \mathbf{r}$. Neglecting a surface correction we obtain

$$\int K(\mathbf{r}, \mathbf{r}')F_{\mathbf{k}}(\mathbf{r}')\, d^3r' = \rho_0(1 - g(k))F_{\mathbf{k}}(\mathbf{r}), \tag{3.24}$$

where

$$g(k) = \int g(r) e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3r = \frac{4\pi}{k}\int_0^\infty r g(r) \sin(kr)\, dr \tag{3.25}$$

Thus the eigenvalues of K are just

$$\lambda_k = 1 - g(k). \tag{3.26}$$

These are nonnegative if and only if:

$$g(k) \leq 1. \tag{3.27}$$

As an illustration of the use of this condition let us consider an infinite system of unit density which has the two-particle distribution

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = 0 \quad \text{if} \quad |\mathbf{r}_1 - \mathbf{r}_2| = r < d \tag{3.28}$$

$$= 1 \quad \text{if} \quad r > d. \tag{3.29}$$

The Fourier transform of this function is

$$g(k) = (4\pi/k^3)[\sin(kd) - kd\cos(kd)]. \tag{3.30}$$

The maximum of this function occurs at $k = 0$. Our condition on $g(k)$ is thus

$$g(0) = \tfrac{4}{3}\pi d^3 \leq 1, \tag{3.31}$$

which simply states that the "correlation hole" cannot have a volume larger than a mean particle volume.

It might seem that this could have been derived trivially from the normalization condition [Eq. (3.7)]. However the normalization condition may always be circumvented simply by adding an infinitesmal but very-long-range positive correlation after the hole. That is by choosing a distribution of the form:

$$\begin{aligned} \rho(\mathbf{r}_1, \mathbf{r}_2) &= 0 && r < d \\ \rho(\mathbf{r}_1, \mathbf{r}_2) &= 1 + \epsilon_1 e^{-\epsilon_2 r} && r > d, \end{aligned} \tag{3.32}$$

we could satisfy the normalization condition with d larger than the value allowed by Eq. (3.31). We could do this with ϵ_1 arbitrarily small by making ϵ_2 appropriately small. The distribution above would not satisfy the condition that $g(k) \leq 1$ if ϵ_1 is chosen very small (although it would satisfy the normalization condition) since for $k\epsilon_2 \ll 1$ the term involving ϵ_1 and ϵ_2 makes a negligible contribution to $g(k)$.

We now derive Condition (4), the consequences of which we have been analyzing in this section. Any symmetric N-particle distribution may be written as a sum, with positive coefficients, of distributions which each describe the situation of having N particles at definite positions,[16] $\lambda_1, \cdots, \lambda_N$. That is, take the set $C(\Lambda)$ of all distinct sets of N points $\Lambda = (\lambda_1, \cdots, \lambda_N)$. Then the N-particle distribution may be written:

$$\rho(x_1, \cdots, x_N) = \sum_{C(\Lambda)} \alpha_\Lambda (N!)^{-1} \sum_{\text{perm}} \delta(x_1 - \lambda_1) \times \cdots \times \delta(x_N - \lambda_N), \tag{3.33}$$

where the second sum is over all the permutations of the x_i and

$$\sum_{C(\Lambda)} \alpha_\Lambda = 1. \tag{3.34}$$

One can now calculate the one- and two-particle distributions in terms of these "Λ" distributions

[16] We treat the case of one-dimensional particles. The generalization is trivial.

$$\rho(x) = \sum_{C(\Lambda)} \alpha_\Lambda \sum_i \delta(x - \lambda_i) \tag{3.35}$$

$$\rho(x, y) = \sum_{C(\Lambda)} \alpha_\Lambda [\sum_{i,j} \delta(x - \lambda_i)\, \delta(y - \lambda_j) - \delta(x - y) \sum_i \delta(y - \lambda_i)]. \tag{3.36}$$

We are now in a position to exhibit $K(x, y)$ explicitly as a sum, with positive coefficients, of manifestly nonnegative operators. Using Eqs. (3.34), (3.35), and (3.36) it is not difficult to verify that:

$$K(x, y) = \tfrac{1}{2} \sum_{C(\Lambda), C(\Lambda')} \alpha_\Lambda \alpha_{\Lambda'} \times \{\sum_i [\delta(x - \lambda_i) - \delta(x - \lambda_i')]\} \times \{\sum_i [\delta(y - \lambda_i) - \delta(y - \lambda_i')]\}. \tag{3.37}$$

Since the matrix $\alpha_\Lambda \alpha_{\Lambda'}$ is nonnegative and the operator following it is a nonnegative operator we obtain Condition (d) of this section.

4. QUANTUM SYSTEMS—BOSONS

In the introduction it was shown that the expectation value of any N-particle Hamiltonian which does not contain more than two-body forces could be expressed in terms of the one- and two-particle density matrices. Therefore, if we had some way of guaranteeing that the density matrices we were admitting into our variational calculations really were derivable from some physically sensible N-particle boson system (which might include not only states described by wavefunctions but also statistical ensembles of such states of the type used in quantum statistical mechanics) then we could calculate the ground-state properties of an N-particle system with a given Hamiltonian by making a variational calculation using the density matrices as variational functions.

We have already presented some of the conditions which must be satisfied by any pair of one- and two-particle density matrices which represent an N-particle system [See Sec. 1, Eqs. (1.15)–(1.19)].

We have shown by a test calculation that these conditions are far from sufficient to ensure that Γ is derivable from an N-particle system. We present another condition which must be satisfied by any legitimate two-particle density matrix. The condition is closely related to the nontrivial condition which was presented for classical systems. It is

(6) nonnegativity of G as an operator:

First construct the function

$$G(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \Gamma(\mathbf{r}_1', \mathbf{r}_2 \mid \mathbf{r}_1, \mathbf{r}_2') + \delta(\mathbf{r}_1 - \mathbf{r}_1')\gamma(\mathbf{r}_2 \mid \mathbf{r}_2') - \gamma^*(\mathbf{r}_1 \mid \mathbf{r}_2)\gamma(\mathbf{r}_1' \mid \mathbf{r}_2'). \tag{4.1}$$

Then, if we choose any function $g(\mathbf{r}_1, \mathbf{r}_2)$ (not necessarily symmetric) the quantity

$$(g, Gg) = \int g(\mathbf{r}_1, \mathbf{r}_2) G(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') \times g(\mathbf{r}_1', \mathbf{r}_2')\, d^3r_1\, d^3r_2\, d^3r_1'\, d^3r_2' \geq 0. \tag{4.2}$$

That is, if we treat G as the kernel of an integral operator which operates on functions of two variables then G has no negative eigenvalues. In order for Condition (6) to make sense, G must be explicitly Hermitian or else it might have imaginary eigenvalues. That $G(\mathbf{r}_1', \mathbf{r}_2' \mid \mathbf{r}_1, \mathbf{r}_2)$ is in fact equal to $G^*(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2')$ follows easily from the properties of Γ and γ. The reader should note, however that $G(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2')$ is not symmetric under interchange of $\mathbf{r}_1$ and $\mathbf{r}_2$ or $\mathbf{r}_1'$ and $\mathbf{r}_2'$. That is why it is necessary to use nonsymmetric $g(\mathbf{r}_1, \mathbf{r}_2)$ in the statement of Condition (6). The justification for this condition is presented in Sec. 7 where it appears as a special case of a much more general (and more complicated) condition. The circumstances under which Condition (6) is sufficient to insure that γ and Γ are the legitimate offspring of a proper N-particle system have not been determined.

In order to present our variational method in a form which will be the same in all representations, it is necessary to write both the kinetic energy and the potential in the form of continuous matrix (or integral) operators. Thus we define

$$T(\mathbf{r} \mid \mathbf{r}') = -(\hbar^2/2m)\nabla^2 \delta(\mathbf{r} - \mathbf{r}'), \tag{4.3}$$

$$V(\mathbf{r} \mid \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')V(\mathbf{r}), \tag{4.4}$$

and

$$U(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \delta(\mathbf{r}_1 - \mathbf{r}_1')\delta(\mathbf{r}_2 - \mathbf{r}_2')U(\mathbf{r}_1 - \mathbf{r}_2). \tag{4.5}$$

This will allow us to transform all formulas very easily into representations (such as the momentum representation) in which the potential energy operator and the interaction operator are not diagonal. Then the product of two one-particle operators (operators with only one primed and one unprimed index) is another one-particle operator calculated as a matrix product. That is

$$C = AB \tag{4.6}$$

means

$$C(\mathbf{r} \mid \mathbf{r}') = \int d^3r'' A(\mathbf{r} \mid \mathbf{r}'')B(\mathbf{r}'' \mid \mathbf{r}') \tag{4.7}$$

if A and B are one-particle operators or

$$C(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \int d^3\mathbf{r}_1'' \, d^3\mathbf{r}_2'' A(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1'', \mathbf{r}_2'') \times B(\mathbf{r}_1'', \mathbf{r}_2'' \mid \mathbf{r}_1', \mathbf{r}_2') \quad (4.8)$$

if A and B are two-particle operators.

The trace is calculated as for matrices:

$$\text{tr } \Gamma = \int d^3r_1 \, d^3r_2 \, \Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1, \mathbf{r}_2) \quad (4.9)$$

$$\text{tr } \gamma = \int d^3r \, \gamma(\mathbf{r} \mid \mathbf{r}). \quad (4.10)$$

The reader may easily verify that the energy is now given by

$$E[\Gamma] = \text{tr } (T\gamma) + \text{tr } (V\gamma) + \tfrac{1}{2} \text{tr } (U\Gamma). \quad (4.11)$$

Since the Conditions (1) through (6) are all necessary for a legitimate density matrix, the set of two-particle operators which obey these conditions contain all possible legitimate two-particle density matrices as a subset. Thus the minimum of $E[\Gamma]$ over the class of Γ which obey these conditions must be smaller than or equal to the true N-particle ground-state energy. In each case we consider (for bosons, and in the next section, for fermions) the minimum we obtain is the exact ground-state energy. These are admittedly very special Hamiltonians; however in each case it is clear that if Condition (6) were relaxed we would obtain nonsensical results. A numerical check on the method for systems of a few particles in a small number of possible states but with quite general interactions is being carried out now on the IBM 7094 at the Courant Institute.

The following case simply illustrates the type of physical information contained in Condition (6). Consider a system of N bosons whose Hamiltonian consists of the usual kinetic energy operator plus a two-particle interaction which in momentum representation has the very simple form:

$$U(\mathbf{k}_1, \mathbf{k}_2 \mid \mathbf{k}_1', \mathbf{k}_2') = U(\mathbf{k}_1)\delta(\mathbf{k}_1 - \mathbf{k}_2) \times \delta(\mathbf{k}_1 - \mathbf{k}_1')\delta(\mathbf{k}_2 - \mathbf{k}_2'), \quad (4.12)$$

where $U(\mathbf{k}) \geq 0$. That is, any two particles in the same momentum state make a positive contribution to the energy but remain in the same momentum state after interaction. The energy can then be expressed in terms of the one- and two-particle momentum density functions:

$$\rho(\mathbf{k}) = \gamma(\mathbf{k} \mid \mathbf{k}) \quad (4.13)$$

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \Gamma(\mathbf{k}_1, \mathbf{k}_2 \mid \mathbf{k}_1, \mathbf{k}_2) \quad (4.14)$$

$$E = (\hbar^2/2m) \sum_k k^2\rho(\mathbf{k}) + \tfrac{1}{2} \sum_k U(\mathbf{k})\rho(\mathbf{k}, \mathbf{k}). \quad (4.15)$$

The normalization conditions are

$$\sum \rho(\mathbf{k}) = N \quad (4.16)$$

and

$$\sum_{\mathbf{k}'} \rho(\mathbf{k}, \mathbf{k}') = (N - 1)\rho(\mathbf{k}). \quad (4.17)$$

These restrictions may be satisfied while still keeping E arbitrarily small by choosing $\rho(\mathbf{k})$ very large for small $|\mathbf{k}|$ but keeping $\rho(\mathbf{k}, \mathbf{k}) = 0$. We would then generate momentum densities which describe the physically absurd situation of having many particles in the momentum state $\mathbf{k}$ for small $|\mathbf{k}|$ but no pairs in the state $(\mathbf{k}, \mathbf{k})$. In order to avoid this we must use the condition that the operator G of Eq. (4.1) have nonnegative eigenvalues. If an operator has nonnegative eigenvalues then all its diagonal elements must be nonnegative.Hence,

$$G(\mathbf{k}_1, \mathbf{k}_2 \mid \mathbf{k}_1, \mathbf{k}_2) = \rho(\mathbf{k}_1, \mathbf{k}_2) + \rho(\mathbf{k}_2) - |\gamma(\mathbf{k}_1 \mid \mathbf{k}_2)|^2 \geq 0. \quad (4.18)$$

For our purposes we need only consider the above equation for $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}$. It is then

$$\rho(\mathbf{k}, \mathbf{k}) + \rho(\mathbf{k}) \geq \rho^2(\mathbf{k}). \quad (4.19)$$

In order to use this relation most directly let us rewrite the energy as

$$E = \sum_k [(\hbar^2/2m)k^2 - \tfrac{1}{2}U(\mathbf{k}) - \lambda]\rho(\mathbf{k}) + \tfrac{1}{2} \sum U(\mathbf{k})[\rho(\mathbf{k}, \mathbf{k}) + \rho(\mathbf{k})], \quad (4.20)$$

where the Lagrange parameter, λ, has been inserted to take care of the normalization condition. For a given density $\rho(\mathbf{k})$ the two-particle momentum distribution which minimizes the second sum is certainly the one which just satisfies (4.19) as an equality. We may therefore set

$$\rho(\mathbf{k}, \mathbf{k}) + \rho(\mathbf{k}) = \rho^2(\mathbf{k}). \quad (4.21)$$

If we then set $\partial E/\partial\rho(\mathbf{k}) = 0$ we obtain

$$(\hbar^2/2m)k^2 - \tfrac{1}{2}U(\mathbf{k}) - \lambda + U(\mathbf{k})\rho(\mathbf{k}) = 0 \quad (4.22)$$

or:

$$\rho(\mathbf{k}) = \frac{\lambda + \tfrac{1}{2}U(\mathbf{k}) - (\hbar^2/2m)k^2}{U(\mathbf{k})}. \quad (4.23)$$

We henceforth assume that $U(\mathbf{k})$ depends only upon $k = |\mathbf{k}|$. Two special cases deserve particular attention. If $U(k) = \infty$, then $\rho(k, k) = 0$ and Eq. (4.21) tells us that $\rho(k)$ is equal to zero or one. If $U(k_0) = 0$ then (4.23) gives $\rho(k_0) = \infty$ unless $\lambda = (\hbar^2/2m)k_0^2$, in which case $\rho(k_0)$ is not determined by Eq. (4.28) but must be determined separately

by application of the normalization condition as in the case of Bose–Einstein condensation. If $k_0 = 0$ then clearly we obtain as ground state that state in which all particles have momentum zero. As an example of a system midway between these two extremes let us consider the case in which $U(k)$ is equal to a constant, U_0. Then

$$\rho(k) = \frac{\lambda + \frac{1}{2}U_0 - (\hbar^2/2m)k^2}{U_0} \quad \text{for} \quad \lambda + \tfrac{1}{2}U_0 \geq (\hbar^2/2m)k^2$$

and

$$\rho(k) = 0 \quad \text{for} \quad \lambda + \tfrac{1}{2}U_0 < (\hbar^2/2m)k^2.$$

The value of λ is then determined by the normalization condition. Written for a large system with the sums over $\mathbf{k}$ approximated by integrals this is:

$$\frac{V}{2\pi^2}\int_0^{k_m}\left(\frac{\lambda}{U_0} + \frac{1}{2} - \frac{\hbar^2}{2mU_0}k^2\right)k^2\,dk = N,$$

where $\lambda + \frac{1}{2}U_0 = (\hbar^2/2m)k_m^2$. Thus k_m is given by $k_m^5 = (30\pi^2 mU_0/\hbar^2)N/V$. Since k_m remains finite as $N \to \infty$ for fixed density it is easy to verify that the total energy is proportional to N.

5. FERMIONS

For fermions the subsidiary conditions which must be satisfied by the density matrices are of a type similar to those imposed upon boson density matrices. They are however more numerous and more stringent. Again we leave the derivation of these restrictions to Sec. 7. These subsidiary conditions have not been shown to be sufficient to guarantee that the density matrices γ and Γ are actually derivable from some N-particle fermion system. Therefore an exact variational calculation utilizing them can only yield a proven lower bound to the ground-state energy.

The conditions to be imposed upon γ and Γ may be separated into two classes. First there are those restrictions which are obvious from the definition of the density matrices. They are:

(1) antisymmetry

$$\begin{aligned}\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') &= -\Gamma(\mathbf{r}_2, \mathbf{r}_1 \mid \mathbf{r}_1', \mathbf{r}_2') \\ &= -\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_2', \mathbf{r}_1')\end{aligned} \tag{5.1}$$

(2) relationship between γ and Γ

$$\gamma(\mathbf{r} \mid \mathbf{r}') = \frac{1}{N-1}\int \Gamma(\mathbf{r}, \mathbf{r}_2 \mid \mathbf{r}', \mathbf{r}_2)\,d^3r_2; \tag{5.2}$$

(3) trace

$$\operatorname{tr}(\Gamma) = N(N-1); \tag{5.3}$$

(4) nonnegativity of Γ as an operator

$$(g, \Gamma g) \geq 0 \quad \text{for } \textit{any} \quad g(\mathbf{r}_1, \mathbf{r}_2); \tag{5.4}$$

(5) Hermiticity

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \Gamma^*(\mathbf{r}_1', \mathbf{r}_2' \mid \mathbf{r}_1, \mathbf{r}_2). \tag{5.5}$$

In addition to these obvious conditions we show (Sec. 7) that three operators, to be defined below must have nonnegative eigenvalues. These additional conditions are:

(6) exclusion principle

$$g(\mathbf{r} \mid \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') - \gamma^*(\mathbf{r} \mid \mathbf{r}') \tag{5.6}$$

is a nonnegative operator;

(7) nonnegativity of G as an operator where

$$\begin{aligned}G(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') &= -\Gamma(\mathbf{r}_1', \mathbf{r}_2 \mid \mathbf{r}_1, \mathbf{r}_2') \\ &+ \delta(\mathbf{r}_1 - \mathbf{r}_1')\gamma(\mathbf{r}_2 \mid \mathbf{r}_2') - \gamma^*(\mathbf{r}_1 \mid \mathbf{r}_2)\gamma(\mathbf{r}_1' \mid \mathbf{r}_2');\end{aligned} \tag{5.7}$$

(8) nonnegativity of Q as an operator where

$$\begin{aligned}Q(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') &= \Gamma(\mathbf{r}_1', \mathbf{r}_2' \mid \mathbf{r}_2, \mathbf{r}_2) \\ &- \delta(\mathbf{r}_2 - \mathbf{r}_2')\gamma(\mathbf{r}_1 \mid \mathbf{r}_1') + \delta(\mathbf{r}_1 - \mathbf{r}_2')\gamma(\mathbf{r}_1' \mid \mathbf{r}_2) \\ &+ \delta(\mathbf{r}_2 - \mathbf{r}_1')\gamma(\mathbf{r}_2' \mid \mathbf{r}_1) - \delta(\mathbf{r}_1 - \mathbf{r}_1')\gamma(\mathbf{r}_2' \mid \mathbf{r}_2) \\ &+ \delta(\mathbf{r}_1 - \mathbf{r}_1')\delta(\mathbf{r}_2 - \mathbf{r}_2') - \delta(\mathbf{r}_1 - \mathbf{r}_2')\delta(\mathbf{r}_2 - \mathbf{r}_1').\end{aligned} \tag{5.8}$$

The first question to be answered is why we call Condition (6) the exclusion principle. Remembering that the average number of particles in any normalized single-particle state $\phi^*(\mathbf{r})$ is given by $(\phi, \gamma\phi)$ we see that Condition (6) requires that

$$(\phi, \phi) - (\phi, \gamma\phi) \geq 0 \tag{5.9}$$

or

$$(\phi, \gamma\phi) \leq 1. \tag{5.10}$$

This is simply a statement of the condition that no more than one particle ever occupy a single state, which is commonly called the exclusion principle. It will be proven later [Sec. (6)] that the exclusion principle plus the conditions on γ which can be derived directly from its definition are sufficient to ensure that there exist some N-particle fermion system from which γ may be derived. This fact was proven earlier by Coleman.[8] Actually the exclusion principle need not be postulated separately; it is contained in Condition (7) as is shown immediately.

In interpreting Conditions (7) and (8) it is convenient to work in a representation in which the variables are discrete (such as, for finite systems, the momentum representation) because the singular Dirac delta functions are then replaced by finite

Kronecker delta functions. In such a representation the operator G is written:

$$G(n_1, n_2 \mid m_1, m_2) = -\Gamma(m_1, n_2 \mid n_1, m_2) + \delta(n_1, m_1)\gamma(n_2 \mid m_2) - \gamma^*(n_1 \mid n_2)\gamma(m_1 \mid m_2). \tag{5.11}$$

The diagonal elements of G, which must be nonnegative if G is to have nonnegative eigenvalues, are

$$-\rho(n_1, n_2) + \rho(n_2) - |\gamma(n_1 \mid n_2)|^2 \geq 0. \tag{5.12}$$

Since $\rho(n, n) = 0$ by antisymmetry we obtain from (5.12), by choosing $n_1 = n_2$, the exclusion principle in the form:

$$\rho(n)[1 - \rho(n)] \geq 0. \tag{5.13}$$

The function $x(1 - x)$ is a concave downward parabola which cuts the axis at $x = 0$ and $x = 1$. Hence (5.13) is equivalent to the condition that

$$0 \leq \rho(n) \leq 1. \tag{5.14}$$

The operator Q in a discrete representation is written:

$$Q(n_1, n_2 \mid m_1, m_2) = \Gamma(m_1, m_2 \mid n_1, n_2) - \delta(n_2, m_2)\gamma(n_1 \mid m_1) + \delta(n_1, m_2)\gamma(m_1 \mid n_2) + \delta(n_2, m_1)\gamma(m_2 \mid n_1) - \delta(n_1, m_1)\gamma(m_2 \mid n_2) + \delta(n_1, m_1)\delta(n_2, m_2) - \delta(n_1, m_2)\delta(n_2, m_1). \tag{5.15}$$

The requirement that the diagonal elements of Q be nonnegative yields the inequalities:

$$\rho(n_1, n_2) \geq \rho(n_1) + \rho(n_2) - 1 \quad (n_1 \neq n_2). \tag{5.16}$$

This expresses the quite reasonable idea that if single-particle states n_1 and n_2 are filled then there must be a pair in the pair state (n_1, n_2). Restrictions of this type are important when the states n_1 and n_2 are of low kinetic energy while the pair state (n_1, n_2) has a large positive interaction energy. If we ignore this type of restrictions when making a variational calculation on such a system using density matrices then the set of density matrices which yield the energy minimum will exhibit the phenomenon of "overcorrelation." That is, the correlations expressed by the two-particle distribution function will be inconsistent with the single-particle density.

Up to now we have only used the simplest possible consequence of the fact that the operators we have been considering have nonnegative eigenvalues. That is, we have said that if the operator $G(n_1, n_2 \mid m_1, m_2)$ has nonnegative eigenvalues then it must have nonnegative diagonal elements. Thus $G(n_1, n_2 \mid n_1, n_2) \geq 0$. But there are many more conditions on the elements of G which are implied by the fact that all eigenvalues of G are nonnegative. These further conditions involve determinants made up of elements of G. The determinantal conditions are as follows: Choose s distinct diagonal elements,

$$G(n_1, m_1 \mid n_1, m_1), \quad G(n_2, m_2 \mid n_2, m_2), \cdots, G(n_s, m_s \mid n_s, m_s).$$

Consider the determinant whose $(i\text{–}j)$th element is $G(n_i, m_i \mid n_j, m_j)$. This determinant is nonnegative.

To illustrate the condition let us take $s = 2$. The determinantal condition is then:

$$G(n_1, m_1 \mid n_1, m_1)G(n_2, m_2 \mid n_2, m_2) - G(n_1, m_1 \mid n_2, m_2)G(n_2, m_2 \mid n_1, m_1) \geq 0. \tag{5.17}$$

If we now write G in terms of Γ and γ we get

$$\{-\rho(n_1, m_1) + \rho(m_1) - [\gamma(n_1 \mid m_1)]^2\} \times \{-\rho(n_2, m_2) + \rho(m_2) - [\gamma(n_2 \mid m_2)]^2\} \geq [-\Gamma(n_2, m_1 \mid n_1, m_2) + \delta(n_1, n_2)\gamma(m_1 \mid m_2) - \gamma^*(n_1 \mid m_1)\gamma(n_2 \mid m_2)]^2. \tag{5.18}$$

Choosing $n_1 = m_1$ and $n_2 = m_2$ $(n_1 \neq n_2)$ in (5.18) we obtain

$$|\rho(n_1, n_2) - \rho(n_1)\rho(n_2)|^2 \leq \rho(n_1)[1 - \rho(n_1)]\rho(n_2)[1 - \rho(n_2)]. \tag{5.19}$$

This equation, together with Eq. (5.16), gives us both upper and lower bounds on $\rho(n_1, n_2)$ as functions of $\rho(n_1)$ and $\rho(n_2)$.

Both Coleman and Yang have shown that the largest possible eigenvalue of Γ is N. We derive this using the nonnegativity condition on G. Restating the problem we want to obtain an upper bound to

$$\Lambda = \sum F^*(n_1, n_2)\Gamma(n_1, n_2 \mid m_1, m_2)F(m_1, m_2), \tag{5.20}$$

where F is a normalized, antisymmetric function.

$$F(n, m) = -F(m, n) \tag{5.21}$$

and

$$\sum |F(n, m)|^2 = 1. \tag{5.22}$$

Without loss in generality we may assume F is real. Then the matrix:

$$M_{nm} = iF(n, m) \tag{5.23}$$

is a Hermitian matrix and can thus be diagonalized in terms of its eigenvectors $v_i(n)$.

$$iF(n, m) = \sum_i \lambda_i v_i(n) v_i^*(m). \tag{5.24}$$

Equation (5.22) can now be written

$$\sum_i \lambda_i^2 = 1. \tag{5.25}$$

Expressing Λ in terms of the eigenvalues and eigenvectors of iF we obtain

$$\Lambda = \sum \lambda_i \lambda_j v_i^*(n_1) v_j(n_2) \Gamma(n_1, n_2 \mid m_1, m_2) v_j(m_1) v_i^*(m_2). \tag{5.26}$$

If we define a matrix $\tilde{\Gamma}$ by

$$\tilde{\Gamma}_{ij} = \sum v_i^*(n_1) v_j(n_2) \Gamma(n_1, n_2 \mid m_1, m_2) v_j(m_1) v_i^*(m_2), \tag{5.27}$$

then

$$\Lambda = \sum \lambda_i \tilde{\Gamma}_{ij} \lambda_j. \tag{5.28}$$

But, since the λ_i satisfy the normalization condition (5.25),

$$\Lambda \leq \text{max. eigenvalue of } \tilde{\Gamma}. \tag{5.29}$$

The fact that Γ is nonnegative implies that $\Lambda \geq 0$ for any choice of the λ_i. But then Eq. (5.28) tells us that $\tilde{\Gamma}$ is a nonnegative matrix. For any nonnegative matrix: max. eigenvalue $\leq$ trace = sum of eigenvalues. Thus,

$$\Lambda \leq \sum_i \tilde{\Gamma}_{ii}. \tag{5.30}$$

Since G is a nonnegative operator:

$$\sum_{\substack{n_1, n_2 \\ m_1, m_2}} v_i(n_1) v_i(n_2) G(n_1, n_2 \mid m_1, m_2) v_i^*(m_1) v_i^*(m_2) \geq 0. \tag{5.31}$$

Using the definition of G, Eq. (5.31) becomes

$$-\tilde{\Gamma}_{ii} + (v_i, \gamma v_i) - [\sum_{n_1, n_2} v_i(n_1) v_i(n_2) \gamma(n_1 \mid n_2)]^2 \geq 0. \tag{5.32}$$

We may combine Eqs. (5.30) and (5.32) to obtain

$$\Lambda \leq N - \sum_i [\sum_{n_1, n_2} v_i(n_1) v_i(n_2) \gamma(n_1 \mid n_2)]^2 \leq N, \tag{5.33}$$

where we have used the relation:

$$\sum_i (v_i, \gamma v_i) = \operatorname{tr} \gamma = N. \tag{5.34}$$

Yang has obtained a more detailed limit for a system with only a finite number of available single-particle states. Using a more detailed argument it is possible to reproduce his result by the methods presented here.

Due to the exclusion principle it is impossible in a fermion system for all the particles or even a finite fraction of them to "condense" into the lowest-energy single-particle state as the temperature goes to zero. On the other hand it is possible to have a type of fermion condensation into pair states. There are certain similarities between boson condensation (which is generally considered responsible for superfluidity) and fermion condensation (usually attributed to superconductors); however there are also important differences. As we have mentioned, fermion condensation is a condensation into pair states rather than single-particle states.

Even the pair-state condensation for a fermion system is not as extreme as that of a boson system. For a condensed boson system the most heavily occupied pair state has in it an average number of pairs proportional to N^2 where N is the number of particles in the system. For a condensed fermion system the upper bound we have obtained above shows that out of the total of $\frac{1}{2}N(N-1)$ distinct pairs of particles no more than N pairs can simultaneously occupy the same pair state. It can also be shown (see Coleman or Yang) that it is only certain very special types of pair states which can attain anything approaching the maximal occupation. For example it is quite easy to prove [using Eq. (5.12)] that the maximum number of pairs in the state, $F(1, 2) = (1/\sqrt{2})\,[f(1)g(2) - f(2)g(1)]$, (where f and g are any two normalized single-particle states) is one.

The most noticeable feature of those special pair states which are capable of bearing the weight of fermion condensation is that when they are expanded in terms of any set of single-particle states whatever, they have a very large number of very small expansion coefficients. (Large in comparison with N, small in comparison with $1/N$.) For more details concerning the comparison of boson and fermion condensation the reader is referred to the paper of Yang.

The next application we make of the nonnegativity restrictions involving the operator G is to the problem of finding the ground-state energy of what is usually called the reduced BCS Hamiltonian,[17] which currently is the best "model" for superconductivity which has proven solvable. In order to write the Hamiltonian in a simple form, we number all the single-particle momentum states of the electrons which have spin $+\frac{1}{2}$ by a sequence of positive integers. The states of opposite spin and momentum are then given the corresponding negative integer.

[17] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

The Hamiltonian can then be written in the form:

$$H = \sum_{n>0} 2T_n a_n^+ a_n - \sum_{n,m>0} V_n V_m a_{-n}^+ a_n^+ a_m a_{-m}. \quad (5.35)$$

Adding a Lagrange parameter which will be adjusted to give the correct particle number, we may write the expectation of H as

$$E - \lambda N = 2 \sum_{n>0} (T_n - \lambda)\rho(n) - \sum_{n,m>0} V_n V_m \Gamma(n, -n \mid m, -m), \quad (5.36)$$

where $\rho(n) = \gamma(n \mid n)$.

We assume that all $V_n > 0$ although we could easily modify the calculation if this were not the case. We use a condition which is somewhat weaker than Condition (7) of Eq. (5.7), namely that the operator

$$\tilde{G}(n_1, n_2 \mid n_1', n_2') = -\Gamma(n_1', n_2 \mid n_1, n_2') + \delta(n_1, n_1')\gamma(n_2 \mid n_2') \quad (5.37)$$

be nonnegative. The determinantal inequality which then corresponds to Eq. (5.18) is:

$$[\rho(n_2) - \rho(n_1, n_2)][\rho(n_2') - \rho(n_1', n_2')] \geq [-\Gamma(n_1', n_2 \mid n_1, n_2') + \rho(n_1, n_1')\gamma(n_2 \mid n_2')]^2. \quad (5.38)$$

If we choose $n_1 = m$, $n_2 = n$, $n_1' = -n$, and $n_2' = -m$, we obtain an inequality involving the elements of Γ we are interested in. It is

$$[\rho(n) - \rho(m, n)][\rho(-m) - \rho(-n, -m)] \geq [\Gamma(n, -n \mid m, -m)]^2. \quad (5.39)$$

If we now demand that, in Eq. (5.36), the value of Γ satisfy the above inequality the minimum is given by

$$E - \lambda N = 2 \sum_{n>0} (T_n - \lambda)\rho(n) - \sum_{n,m>0} V_n V_m \times \{[\rho(n) - \rho(m, n)][\rho(m) - \rho(n, m)]\}^{\frac{1}{2}}, \quad (5.40)$$

where we have used the fact that since the Hamiltonian is unchanged by the transformation $n \to -n$ the ground-state densities $\rho(n)$ and $\rho(n, m)$ will have the same symmetry.

We now have two possible paths, only one of which leads to the correct result. Firstly we might use Eq. (5.19) to express $\rho(n, m)$ in terms of $\rho(n)$ and $\rho(m)$ thereby obtaining an expression for the energy in terms of the one-particle distribution $\rho(n)$ alone. Variation with respect to $\rho(n)$ would then yield a lower bound to the energy. The lower bound we obtain this way is not bad, falling below the true value by an amount approximately equal to $\frac{1}{2} \sum_{n,m>0} V_n V_m q_n q_m (q_n - q_m)^2$ where, $q_n = [\rho_n(1-\rho_n)]^{\frac{1}{2}}$. The reason our lower bound falls below the true ground-state energy is not hard to find. We have repeatedly replaced conditions of operator nonnegativity by weaker determinantal inequalities. Weakening any significant restrictions on our trial density matrices leads to a lowering of our energy minimum.

If, instead of using Eq. (5.19) to relate $\rho(n, m)$ to $\rho(n)$ and $\rho(m)$ we use the fact (known from the BCS solution) that in the ground-state there exist no correlations between particles with the same spin, we get the relation:

$$\rho(n, m) = \rho(n)\rho(m)[1 - \delta(n, m)]. \quad (5.41)$$

Using this, our expression for the energy becomes:

$$E - \lambda N = 2 \sum_{n>0} (T_n - \lambda)\rho(n) - \left(\sum_{n>0} V_n \{\rho(n)[1 - \rho(n)]\}^{\frac{1}{2}}\right)^2 + \sum_{n>0} V_n^2 [\rho(n)]^2. \quad (5.42)$$

The last term may be neglected for large systems since it tends to a constant while the other terms are proportional to N for large N at fixed density. Setting to zero the derivative of the first two terms with respect to $\rho(n)$ we obtain:

$$T_n - \lambda = \frac{\frac{1}{2} - \rho(n)}{\{\rho(n)[1 - \rho(n)]\}^{\frac{1}{2}}} V_n \times \left(\sum_{m>0} V_m \{\rho(m)[1 - \rho(m)]\}^{\frac{1}{2}}\right). \quad (5.43)$$

By comparing the above with the equation obtained by BCS the reader may easily verify that the ground-state distribution we now obtain is the correct one. For the solution and interpretation of the above equation the reader is referred to any of the introductory articles on superconductivity.

We now explore the connection between the variational method presented here and what is certainly the most common and important variational technique used in many-particle calculations on fermion systems, namely the Hartree–Fock Approximation. In the HF approximation one assumes that the N-particle wavefunction may be represented by a single N-by-N Slater determinant. Using this assumption one can then derive the following connection between the two-particle density matrix Γ and the one-particle density matrix γ:

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \gamma(\mathbf{r}_1 \mid \mathbf{r}_1')\gamma(\mathbf{r}_2 \mid \mathbf{r}_2') - \gamma(\mathbf{r}_1 \mid \mathbf{r}_2')\gamma(\mathbf{r}_2 \mid \mathbf{r}_1'). \quad (5.44)$$

The one-particle density matrix γ may be easily expressed in terms of the N orthonormal single-particle functions, $\phi_1(\mathbf{r}), \cdots, \phi_N(\mathbf{r})$, which make up the Slater determinant. The expression is

$$\gamma(\mathbf{r}' \mid \mathbf{r}) = \sum_{n=1}^{N} \phi_n(\mathbf{r})\phi_n^*(\mathbf{r}'). \tag{5.45}$$

Since γ satisfies the relation

$$\gamma(\mathbf{r} \mid \mathbf{r}') = \int d^3r''\gamma(\mathbf{r} \mid \mathbf{r}'')\gamma(\mathbf{r}'' \mid \mathbf{r}'), \tag{5.46}$$

it is simple to verify that the necessary relation between Γ and γ, namely:

$$\gamma(\mathbf{r} \mid \mathbf{r}') = \frac{1}{N-1}\int d^3r''\Gamma(\mathbf{r}, \mathbf{r}'' \mid \mathbf{r}', \mathbf{r}'') \tag{5.47}$$

is satisfied. Of course, since the HF density matrix is a valid two-particle density matrix it satisfies all of the necessary conditions listed at the beginning of this section. In terms of the ϕ_n the operator G is

$$G(\mathbf{r}_1', \mathbf{r}_2' \mid \mathbf{r}_1, \mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_1') \sum_{n=1}^{N} \phi_n(\mathbf{r}_2)\phi_n^*(\mathbf{r}_2') - \sum_{n,m=1}^{N} \phi_m^*(\mathbf{r}_1)\phi_m(\mathbf{r}_1')\phi_n(\mathbf{r}_2)\phi_n^*(\mathbf{r}_2'). \tag{5.48}$$

Using the completeness relation for a complete set of functions $\phi_1, \phi_2, \cdots$, which include our functions as the first N functions

$$\delta(\mathbf{r}_1 - \mathbf{r}_1') = \sum_{m=1}^{\infty} \phi_m^*(\mathbf{r}_1)\phi_m(\mathbf{r}_1'), \tag{5.49}$$

we see that G may be written

$$G(\mathbf{r}_1', \mathbf{r}_2' \mid \mathbf{r}_1, \mathbf{r}_2) = \sum_{m=N+1}^{\infty} \phi_m^*(\mathbf{r}_1)\phi_m(\mathbf{r}_1') \sum_{n=1}^{N} \phi_n(\mathbf{r}_2)\phi_n^*(\mathbf{r}_2'), \tag{5.50}$$

which shows that G is clearly a nonnegative operator. So far we have shown only that the HF density matrices satisfy our necessary conditions. We now show that the assumption of Eq. (5.44) relating Γ and γ when combined with our restrictions on the density matrices yield the HF theory. Thus we assume only

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2 \mid \mathbf{r}_1', \mathbf{r}_2') = \gamma(\mathbf{r}_1 \mid \mathbf{r}_1')\gamma(\mathbf{r}_2 \mid \mathbf{r}_2') - \gamma(\mathbf{r}_1 \mid \mathbf{r}_2')\gamma(\mathbf{r}_2 \mid \mathbf{r}_1'). \tag{5.44}$$

Since γ must be Hermitian it may be diagonalized in the form:

$$\gamma(\mathbf{r}' \mid \mathbf{r}) = \sum_{n=1}^{\infty} C_n\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}'),$$

with real C_n. Since γ is nonnegative, the C_n are greater or equal to zero. The trace of γ must equal N; thus: $\sum C_n = N$.

The relation between Γ and γ is

$$\gamma(\mathbf{r} \mid \mathbf{r}') = \frac{1}{N-1}\int d^3r''\Gamma(\mathbf{r}, \mathbf{r}'' \mid \mathbf{r}', \mathbf{r}''),$$

which yields the following relation among the C_n:

$$\sum_n C_n\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}') = [1/(N-1)] \sum_n (NC_n - C_n^2)\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')$$

which, since the ϕ_n are orthonormal gives:

$$C_n = 1 \quad \text{or} \quad C_n = 0.$$

Thus we obtain exactly N filled single-particle states which is equivalent to assuming that the wavefunction is a single N-by-N Slater determinant.

6. NECESSARY AND SUFFICIENT CONDITIONS

We now present conditions on the one- and two-particle density matrices of a boson (or fermion) system which are sufficient to insure that the given density matrices are actually derivable from at least one physically sensible many-particle system. First we must describe precisely what we mean by a physically sensible many-particle system. We treat boson and fermion systems simultaneously by introducing a dichotomic variable σ which must be assigned the value $+1$ if the particles under discussion are bosons and -1 if they are fermions.

What we accept as a physically sensible many-particle system is a statistical ensemble of the type used in quantum statistical mechanics. Any many-particle system which is in a unique quantum state may be represented by a vector $|\psi\rangle$ in an appropriate Fock space, S_F. However there are many-particle systems which cannot be considered to be in any definite quantum state. The simplest case of this is any many-particle subsystem of a larger system from which it is not dynamically isolated, for instance, the particles in a particular cubic centimeter of a larger volume of gas. This subsystem could interchange both particles and energy with the larger system; thus even if the larger system (the complete gas) is isolated and in some fixed quantum state, no quantum state can be ascribed to the subsystem.

Thus the "physical state" of the subsystem cannot be described mathematically by a "state vector" but must be described instead by an operator on the Fock space called the "density operator" of the many-particle subsystem. (For more details see von

Neumann.[1]) The physical interpretation of the density operator, w, is that the probability of finding the subsystem in any particular quantum state, $|\psi\rangle$, is given by

$$P_\psi = \langle\psi| w |\psi\rangle. \tag{6.1}$$

Since P_ψ must be nonnegative we see that w can have no negative eigenvalues. Since a measurement must always find the subsystem in some one of a complete set of states we obtain the normalization:

$$\sum_{n=1}^{\infty} \langle n| w |n\rangle = \operatorname{tr} w = 1, \tag{6.2}$$

where the sequence of states $|n\rangle$ is any complete orthonormal set.

Any operator which satisfies these two conditions of nonnegativity and normalization may be sensibly interpreted as a many-particle density operator. Thus when we speak of a physically reasonable many-particle system we mean any system or subsystem which may be described by a nonnegative, normalized density operator. In terms of the density operator w, we may now define our one- and two-particle density matrices, γ and Γ respectively, by

$$\gamma(n \mid m) = \operatorname{tr} (a_n^+ a_m w) \tag{6.3}$$

and

$$\Gamma(k, l \mid m, n) = \operatorname{tr} (a_l^+ a_k^+ a_m a_n w), \tag{6.4}$$

where a_n^+ and a_n are single-particle creation and annihilation operators. To see that this definition agrees with our original definition [Eqs. (1.7) and (1.8)] whenever the many-particle system is one which contains exactly N particles in some N-particle quantum state $\psi(\mathbf{r}_1, \cdots, \mathbf{r}_N)$, we need only realize that the density operator w in such a case would be

$$w = |\Psi\rangle\langle\Psi|. \tag{6.5}$$

Eqs. (6.3) and (6.4) for γ and Γ if expressed in the position representation would then have the forms

$$\gamma(x \mid x') = \langle\Psi| \psi^+(x)\psi(x') |\Psi\rangle \tag{6.6}$$

and

$$\Gamma(x_1, x_2 \mid x_1', x_2') = \langle\Psi| \psi^+(x_2)\psi^+(x_1)\psi(x_1')\psi(x_2') |\Psi\rangle \tag{6.7}$$

which if converted to "first quantized" form yield exactly the desired expressions.

We have yet a little more preliminary work before we can come to the basic theorem of this section. The set of real single-particle operators $F(n \mid m)$ which have the property that $\sum_{n,m} |F(n \mid m)|^2 < \infty$ forms a Hilbert space H_1 with scalar product

$$(F, G) = \sum_{n,m} F(n \mid m) G(n \mid m). \tag{6.7}$$

Similarly the set of real two-particle operators $F(n_1, n_2 \mid m_1, m_2)$ which have the properties

$$\sum_{\substack{n_1, n_2 \\ m_1, m_2}} |F(n_1, n_2 \mid m_1, m_2)|^2 < \infty \tag{6.8}$$

and

$$F(n_2, n_1 \mid m_1, m_2) = \sigma F(n_1, n_2 \mid m_1, m_2) \tag{6.9}$$

form a Hilbert space H_2 with scalar product

$$(F, G) = \sum_{\substack{n_1, n_2 \\ m_1, m_2}} F(n_1, n_2 \mid m_1, m_2) G(n_1, n_2 \mid m_1, m_2). \tag{6.10}$$

The Hilbert space we need is a bit more complicated than either of these. It is in fact the direct product of H_1, H_2, and the space R of real numbers. That is, we consider a set of elements

$$\xi = (F_0, F_1(n \mid m), F_2(n_1, n_2 \mid m_1, m_2)), \tag{6.11}$$

where F_0 is a real number, $F_1(n \mid m)$ is an element of H_1, and $F_2(n_1, n_2 \mid m_1, m_2)$ is an element of H_2. The set of such elements forms a Hilbert space:

$$\mathfrak{K} = R \otimes H_1 \otimes H_2. \tag{6.12}$$

Given another element ζ in $\mathfrak{K}$, where:

$$\zeta = (G_0, G_1(n \mid m), G_2(n_1, n_2 \mid m_1, m_2)) \tag{6.13}$$

we define our scalar product as

$$(\xi, \zeta) = F_0 G_0 + (F_1, G_1) + (F_2, G_2). \tag{6.14}$$

We now want to consider the following question: Given an element, $\xi = (F_0, F_1, F_2)$, of $\mathfrak{K}$ how may we determine if there exists any nonnegative operator w on the Fock space S_F for which

$$F_0 = \operatorname{tr} (w), \tag{6.15}$$

$$F_1(n \mid m) = \operatorname{tr} (a_n^+ a_m w), \tag{6.16}$$

and

$$F_2(n_1, n_2 \mid m_1, m_2) = \operatorname{tr} (a_{n_2}^+ a_{n_1}^+ a_{m_1} a_{m_2} w)? \tag{6.17}$$

If we could answer this question, we could quite trivially calculate the set of density matrices which were derivable from normalized nonnegative w by dividing F_1 and F_2 by F_0. Let us give the set of elements of $\mathfrak{K}$ which can be derived from some nonnegative w the name Q. Thus, $\xi \in Q$ implies that there exists at least one operator w on S_F which has nonnegative eigenvalues and which satisfies Eqs. (6.15) through (6.17).

The only property of the set Q which we use is that Q is a "convex cone of vectors." That is, for any pair of vectors ξ' and ξ'' which are both in Q and any pair of positive real numbers, α' and α'' the vector

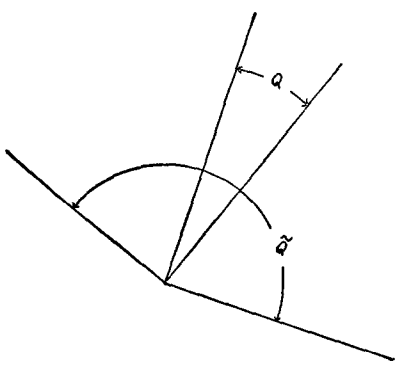
$$\xi = \alpha'\xi' + \alpha''\xi'' \tag{6.18}$$

is also in Q. This property is not difficult to verify, for suppose the equations (6.15) through (6.17) for ξ' are satisfied by some nonnegative operator w' and the same equations for ξ'' are satisfied by another nonnegative operator w''. Then these equations for ξ are satisfied by the nonnegative operator

$$w = \alpha'w' + \alpha''w''. \tag{6.19}$$

The reason for calling such a set of vectors a convex cone is that any set in a three-dimensional space which had this property would form an infinite cone with apex at the origin and which if intersected with any plane would yield a convex plane figure.

We must now define another convex cone of vectors in which we call the "polar of Q" and write $\tilde{Q}$, by the following: A vector ξ is an element of $\tilde{Q}$ if and only if it has a nonnegative scalar product with every vector in Q. For a two-dimensional convex cone the picture of $\tilde{Q}$ is the following:



There is an established theorem on convex cones (a proof of which is presented in Appendix A) which states that the polar of $\tilde{Q}$ is Q. That is:

$$Q = \tilde{\tilde{Q}}. \tag{6.20}$$

Thus if we could describe the set $\tilde{Q}$ we would have a method of determining whether any given vector was derivable from a nonnegative operator w by Eqs. (6.15) through (6.17).

A physical interpretation of this theorem is not hard to find. If a vector, $\zeta = (F_0, F_1, F_2)$, is an element of $\tilde{Q}$ then, for every $\xi = (1, \gamma, \Gamma)$ which is derivable from a real many-particle ensemble, that is for every normalized ξ in Q,

$$(\zeta, \xi) = F_0 + \sum_{n,m} F_1(m \mid n)\gamma(n \mid m) + \sum_{\substack{n_1,n_2 \\ m_1,m_2}} F_2(m_1, m_2 \mid n_1, n_2)\Gamma(n_1, n_2 \mid m_1m_2) \geq 0. \tag{6.21}$$

If we now consider $F_1(m \mid n)$ and $F_2(m_1, m_2 \mid n_1, n_2)$ as the one-and two-particle parts of some many-particle Hamiltonian:

$$H = \sum F_1(m \mid n)a_m^+a_n + \sum F_2(m_1, m_2 \mid n_1, n_2)a_{m_2}^+a_{m_1}^+a_{n_1}a_{n_2}$$

whose minimum energy is $-E_0$, then Eq. (6.21) says that $F_0 \geq E_0$. Thus the elements of $\tilde{Q}$ may be constructed by taking F_1 and F_2 to be any one- and two-particle operators and taking F_0 to be any real number larger than the absolute value of the ground-state energy of the corresponding Hamiltonian. Therefore if we are given any pair of operators $\gamma(n \mid m)$ and $\Gamma(n_1, n_2 \mid m_1, m_2)$ which are not derivable from any many-particle ensemble, there will always exist at least one Hamiltonian which is capable of detecting this fact in the sense that γ and Γ will yield an energy for this Hamiltonian which is lower than its minimum energy.

Offhand, the above result seems of very little value since we would already need to know the ground-state energy of almost all Hamiltonian operators in order to apply these conditions. Actually we see that by using a fairly simple class of Hamiltonians whose minimum energies are trivial to calculate, we may derive all the restrictions used in the earlier sections of this paper.

We are now in a position to prove that if the eigenvalues of γ all lie between zero and one (inclusively) then γ is a legitimate single-particle density matrix for a fermion system. Consider any single-particle Hamiltonian operator, $H(x \mid x')$, whose eigenfunctions and eigenvalues are

$$f_n(x) \qquad n = 1, 2, \cdots$$

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \cdots .$$

The ground-state energy of an N-particle system with Hamiltonian H is:

$$E_0(n) = \sum_{n=1}^{N} \lambda_n,$$

since the N-particle fermion ground state is just that state in which the N lowest energy levels are occupied while all others are empty. If γ has eigenvalues between zero and one, then

$$\operatorname{tr}(\gamma H) = \sum_{n=1} \lambda_n(f_n, \gamma f_n) \geq E_0.$$

Thus γ can never yield an energy expectation value which is lower than the N-particle ground-state energy. But this is just the condition which is necessary and sufficient to insure that γ is derivable from some N-particle fermion system.

7. NECESSARY CONDITIONS

Certainly if we choose a Hamiltonian which is the product of an operator A and its Hermitian adjoint A^+, that is

$$H = A^+A, \tag{7.1}$$

then the minimum value of the energy, E, must be greater than zero. Thus

$$E = \text{tr}\,(Hw) \geq 0 \tag{7.2}$$

for any proper ensemble density operator w. If we choose for A an operator of the form:

$$A = \sum_{n,m} f(n, m)a_n a_m, \tag{7.3}$$

where $f(n, m)$ is any symmetric (antisymmetric) function of the integer variables n, m, and a_n is a single-particle boson (fermion) annihilation operator then we obtain the condition

$$E = \sum f^*(k, l)f(n, m)\,\text{tr}\,(a_l^+ a_k^+ a_n a_m w) \geq 0. \tag{7.4}$$

Using the definition of the two-particle density matrix Γ this becomes:

$$(f, \Gamma f) \geq 0. \tag{7.5}$$

Hence we in this way merely derive the fact that Γ is a nonnegative operator.

In order to derive the conditions which we made use of earlier in this paper we must choose

$$A = C + \sum f(n, m)a_n^+ a_m. \tag{7.6}$$

Then

$$\begin{aligned} A^+A &= C^*C + C^* \sum f(n, m)a_n^+ a_m \\ &+ C \sum f^*(n, m)a_m^+ a_n \\ &+ \sum f^*(k, l)f(n, m)(a_l^+ a_m \delta_{k,n} + \sigma a_l^+ a_n^+ a_k a_m) \end{aligned} \tag{7.7}$$

where $\sigma = +1$ for bosons and -1 for fermions. With this choice of A, Eq. (7.2) becomes

$$\begin{aligned} &C^*C + \sum_{n,m} [C^* f(n, m)\gamma(n \mid m) + Cf^*(n, m)\gamma^*(n \mid m)] \\ &\quad + \sum_{\substack{k,l \\ n,m}} f^*(k, l)f(n, m) \\ &\quad \times [\delta_{k,n}\gamma(l \mid m) + \sigma\Gamma(n, l \mid k, m)] \geq 0. \end{aligned} \tag{7.8}$$

Adding and subtracting a term we may put this in the form:

$$\begin{aligned} &|C + \sum f(n, m)\gamma(n \mid m)|^2 \\ &\quad + \sum_{\substack{k,l \\ n,m}} f^*(k, l)[\delta_{k,n}\gamma(l \mid m) + \sigma\Gamma(n, l \mid k, m) \\ &\quad - \gamma^*(k \mid l)\gamma(n \mid m)]f(n, m) \geq 0. \end{aligned} \tag{7.9}$$

The choice of the constant C which yields the strongest restriction is that which makes the first term vanish. For this choice of C, Eq. (7.9) tells us that

$$\begin{aligned} G(k, l \mid n, m) &= \delta_{k,n}\gamma(l \mid m) + \sigma\Gamma(n, l \mid k, m) \\ &\quad - \gamma^*(k \mid l)\gamma(n \mid m) \end{aligned} \tag{7.10}$$

is a nonnegative operator on the space of two-particle functions. This is just the condition (with the appropriate value of σ) which we have utilized in the sections on bosons and fermions. That G is a nonnegative operator means that for any function $F(k, l)$,

$$\sum_{\substack{k,l \\ m,n}} F^*(k, l)G(k, l \mid m, n)F(m, n) = (F, GF) \geq 0. \tag{7.11}$$

In order to derive the condition we have used for classical systems we must choose a function F of the form

$$F(k, l) = \delta(k - l)f(l). \tag{7.12}$$

Equation (7.11) then becomes, in the position representation:

$$\begin{aligned} \int d^3r\, d^3r' f^*(\mathbf{r})[\delta(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}') + \rho(\mathbf{r}, \mathbf{r}') \\ - \rho(\mathbf{r})\rho(\mathbf{r}')]f(\mathbf{r}') \geq 0 \end{aligned} \tag{7.13}$$

which states that the operator

$$K(\mathbf{r} \mid \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}') + \rho(\mathbf{r}, \mathbf{r}') - \rho(\mathbf{r})\rho(\mathbf{r}') \tag{7.14}$$

has no negative eigenvalues.

We now generalize the method to make it applicable to systems whose Hamiltonians do not conserve particle number. We then apply the generalized method to some simple problems. In order to prevent our equations from becoming very unwieldy, we introduce a compact notation for the variety of matrix elements we need. The notation is most easily described by a few examples. If w is any density operator, then

$$\langle k, l^+ \rangle = \text{tr}\,(a_k a_l^+ w), \tag{7.15}$$

$$\langle k^+, l, m, n^+ \rangle = \text{tr}\,(a_k^+ a_l a_m a_n^+ w), \tag{7.16}$$

$$\langle k, l \rangle = \text{tr}\,(a_k a_l w), \quad \text{etc.} \tag{7.17}$$

If we consider any operator A of the form

$$A = C + f_n a_n + g_n a_n^+, \tag{7.18}$$

where c, f_n, and g_n are scalars and we have employed the summation convention for repeated indices, then the fact that:

$$\text{tr}\,(A^+ A w) \geq 0 \tag{7.19}$$

for any density operator w yields the following inequality relating the various types of what we still call density matrix elements:

$$\begin{aligned} &C^*C + C^* f_n \langle n\rangle + C^* g_n \langle n^+\rangle + C f_n^* \langle n^+\rangle + C g_n^* \langle n\rangle \\ &\quad + f_n^* \langle n^+, m\rangle f_m + g_n^* \langle n, m\rangle f_m + f_n^* \langle n^+, m^+\rangle g_m \\ &\quad + g_n^* \langle n, m^+\rangle g_m \geq 0. \end{aligned} \tag{7.20}$$

The above inequality may be reshuffled into the form:

$$\begin{aligned} &|C + f_n\langle n\rangle + g_n\langle n^+\rangle|^2 + f_n^*(\langle n^+, m\rangle - \langle n^+\rangle\langle m\rangle) f_m \\ &\quad + g_n^*(\langle n, m\rangle - \langle n\rangle\langle m\rangle) f_m + f_n^*(\langle n^+, m^+\rangle \\ &\quad - \langle n^+\rangle\langle m^+\rangle) g_m + g_n^*(\langle n, m^+\rangle - \langle n\rangle\langle m^+\rangle) g_m \geq 0. \end{aligned} \tag{7.21}$$

The value of C which minimizes the left side of (7.21) and which therefore leads to the strongest restriction on the density matrix elements is

$$C = -f_n\langle n\rangle - g_n\langle n^+\rangle. \tag{7.22}$$

If we use this value of C and we combine f_n and g_n to form the single vector

$$F = \begin{bmatrix} f_n \\ g_n \end{bmatrix} \quad \text{and} \quad F^+ = [f_n^*, g_n^*], \tag{7.23}$$

then (7.21) becomes:

$$F^+ D F = [f_n^*, g_n^*] \begin{bmatrix} \langle n^+ : m\rangle_s & \langle n^+ : m^+\rangle_s \\ \langle n : m\rangle_s & \langle n : m^+\rangle_s \end{bmatrix} \begin{bmatrix} f_m \\ g_m \end{bmatrix} \geq 0, \tag{7.24}$$

where we have used the "subtracted" density matrix elements:

$$\langle n^+ : m\rangle_s = \langle n^+, m\rangle - \langle n^+\rangle\langle m\rangle \text{ etc.} \tag{7.25}$$

Since f_n and g_n are arbitrary, the matrix of operators,

$$D = \begin{bmatrix} \langle n^+ : m\rangle_s & \langle n^+ : m^+\rangle_s \\ \langle n : m\rangle_s & \langle n : m^+\rangle_s \end{bmatrix} \tag{7.26}$$

must be nonnegative. Let us apply this to a couple of simple variational problems.

The first problem we consider is that of a scalar field with external sources. This is characterized by the Hamiltonian

$$H = \sum_n [T_n a_n^+ a_n + \tfrac{1}{2} V_n (a_n + a_n^+)]. \tag{7.27}$$

If we assume that

$$\langle n^+\rangle = \langle n\rangle, \tag{7.28}$$

then the energy may be written

$$E = \sum (T_n \langle n^+, n\rangle + V_n\langle n\rangle). \tag{7.29}$$

If D is nonnegative, then the operators on the diagonal must both be nonnegative. This then implies that the diagonal elements of each of these operators are greater or equal to zero.

$$\langle n^+ : n\rangle_s = \langle n^+, n\rangle - \langle n^+\rangle\langle n\rangle \geq 0 \tag{7.30}$$

and

$$\langle n : n^+\rangle_s = 1 + \sigma\langle n^+, n\rangle - \langle n^+\rangle\langle n\rangle \geq 0, \tag{7.31}$$

where $\sigma = +1$ for bosons and -1 for fermions. For bosons we see

$$|\langle n\rangle| \leq (\langle n^+, n\rangle)^{\frac{1}{2}} \equiv U_n. \tag{7.32}$$

The minimum of E may then be expressed in terms of U_n as

$$E = \sum_n (T_n U_n^2 + V_n U_n). \tag{7.33}$$

Setting $\partial E/\partial U_n = 0$ we obtain

$$U_n = -\tfrac{1}{2} V_n / T_n \tag{7.34}$$

and

$$E_{\min} = \sum_n V_n^2 / 4T_n. \tag{7.35}$$

The reader may easily verify that this is the correct minimum by diagonalizing the Hamiltonian directly using the canonical transformation:

$$a_n \to a_n - \tfrac{1}{2} V_n / T_n. \tag{7.36}$$

For fermions we have two independent restrictions. If we define the quantities

$$U_n = \langle n\rangle \quad \text{and} \quad \rho_n = \langle n^+, n\rangle, \tag{7.37}$$

the restrictions are

$$U_n^2 \leq \rho_n \tag{7.38}$$

and

$$U_n^2 \leq 1 - \rho_n. \tag{7.39}$$

These imply that U_n^2 must be less than or equal to $1/2$ since ρ_n must be between zero and one. For a particular value of n, if

$$2V_n^2 \leq T_n^2, \tag{7.40}$$

the first restriction will be most effective and will

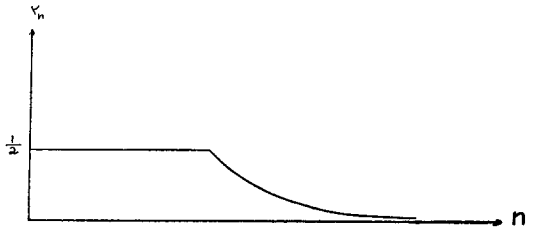


FIG. 1. Fermion momentum distribution.

yield a minimum when

$$U_n = -\tfrac{1}{2}\,|V_n|/T_n, \qquad \rho_n = U_n^2. \tag{7.41}$$

On the other hand, if

$$2V_n^2 > T_n^2, \tag{7.42}$$

the minimum energy is given by

$$U_n = -1/\sqrt{2}, \qquad \rho_n = \tfrac{1}{2}. \tag{7.43}$$

Thus the minimum total energy is:

$$E = -\sum (1/\sqrt{2})[V_n - (1/\sqrt{2})T_n] - \tfrac{1}{4}\sum V_n^2/T_n, \tag{7.44}$$

where the first sum is over those values of n for which $2V_n^2 > T_n^2$ and the second sum is over all other values of n. Hence if, as is usually the case, T_n increases monotonically with n and V_n decreases with increasing n, the single-particle density will have, in the ground-state, the general form given in Fig. 1. This is not to be confused with the "Fermi-sphere" type of distribution in which the low momentum states all have occupation equal to 1. In this case the momentum distribution in the states of small n is used as to maximize the particle number fluctuation, not the occupation of the state.

The final calculation we make, utilizing these conditions, is a variational calculation of the minimum energy for the Hamiltonian introduced by Boguliobov[18] as a model for a superfluid. For a derivation of the Hamiltonian and a physical interpretation of the results the reader is referred to Boguliobov's original article or to any of the numerous texts on the many-body problem.[19] The Hamiltonian is

$$H = \sum_k [T_k a_k^+ a_k + \tfrac{1}{2}V_k(a_k a_{-k} + a_{-k}^+ a_k^+)], \tag{7.45}$$

where the creation and annihilation operators refer to bosons. In Eq. (7.24), if we replace the subtracted density matrix elements by the corresponding density matrix elements without the subtractions, the inequality can only be strengthened. Thus the matrix

$$D = \begin{bmatrix} \langle k^+, l\rangle & \langle k^+, l^+\rangle \\ \langle k, l\rangle & \langle k, l^+\rangle \end{bmatrix} \tag{7.46}$$

must be nonnegative. A simple consequence of this is that

$$\langle k^+, k\rangle\langle l, l^+\rangle - |\langle k, l\rangle|^2 \geq 0 \text{ for any } k \text{ and } l. \tag{7.47}$$

If we choose $l = -k$ and we use the symbol $\rho(k)$ for the single-particle density $\langle k^+, k\rangle$, we obtain

$$|\langle k, -k\rangle| \leq \rho(k)(1 + \rho(-k)). \tag{7.48}$$

Now let us notice that the energy corresponding to the above Hamiltonian is given by

$$E = \sum [T_k\rho(k) + \tfrac{1}{2}V_k(\langle k, -k\rangle + \langle -k^+, k^+\rangle)]. \tag{7.49}$$

The minimum of E consistent with restriction (7.48) is

$$E = \sum \{T_k\rho(k) - V_k[\rho(k)(1 + \rho(-k))]^{\frac{1}{2}}\}. \tag{7.50}$$

Since the original Hamiltonian is symmetric with respect to the interchange of k and $-k$ we may assume that the ground-state distribution also possesses this symmetry. Thus we may assume:

$$\rho(k) = \rho(-k). \tag{7.51}$$

Doing this, and defining a function:

$$U(k) = \rho(k) + \tfrac{1}{2} \tag{7.52}$$

we express the energy as

$$E = \sum \{T_k[U(k) - \tfrac{1}{2}] - V_k[U^2(k) - \tfrac{1}{4}]^{\frac{1}{2}}\}. \tag{7.53}$$

Setting

$$\partial E/\partial U(k) = 0, \tag{7.54}$$

we obtain the equation

$$T_k - U(k)V_k/[U^2(k) - \tfrac{1}{4}]^{\frac{1}{2}} = 0. \tag{7.55}$$

Solving for $U(k)$ we find

$$U(k) = T_k/2(T_k^2 - V_k^2)^{\frac{1}{2}}. \tag{7.56}$$

This yields the energy minimum

$$E = -\tfrac{1}{2}\sum [T_k - (T_k^2 - V_k^2)^{\frac{1}{2}}], \tag{7.57}$$

which is just the result obtained by Boguliobov by his canonical transformation method.

We now return to the derivation of necessary conditions on our "generalized" density matrices. To this end let us consider an operator of the form:

[18] N. Bogoliubov, J. Phys., USSR 11, 23 (1947).
[19] D. Pines, *The Many-Body Problem* (W. A. Benjamin Press, Inc., 1961).

$$A = C + f_n a_n + g_n a_n^+ + u_{nm} a_n a_m + v_{nm} a_n^+ a_m + w_{nm} a_n^+ a_m^+. \quad (7.58)$$

(We are again using the summation convention.)

Going through a line of reasoning which is exactly analogous to that which led to Eq. (7.24), we can show that the obviously necessary condition:

$$\mathrm{tr}\,(A^+ A w) \geq 0 \quad (7.59)$$

is equivalent to the condition that

$$(f_k^*, g_k^*, u_{kl}^*, v_{kl}^*, w_{kl}^*) \begin{bmatrix} \langle k^+ : n\rangle & \langle k^+ : n^+\rangle & \langle k^+ : n, m\rangle & \langle k^+ : n^+, m\rangle & \langle k^+ : n^+, m^+\rangle \\ \langle k : n\rangle & \langle k : n^+\rangle & \langle k : n, m\rangle & \langle k : n^+, m\rangle & \langle k : n^+, m^+\rangle \\ \langle l^+, k^+ : n\rangle & \langle l^+, k^+ : n^+\rangle & \langle l^+, k^+ : n, m\rangle & \langle l^+, k^+ : n^+, m\rangle & \langle l^+, k^+ : n^+, m^+\rangle \\ \langle l^+, k : n\rangle & \langle l^+, k : n^+\rangle & \langle l^+, k : n, m\rangle & \langle l^+, k : n^+, m\rangle & \langle l^+, k : n^+, m^+\rangle \\ \langle l, k : n\rangle & \langle l, k : n^+\rangle & \langle l, k : n, m\rangle & \langle l, k : n^+, m\rangle & \langle l, k : n^+, m^+\rangle \end{bmatrix} \begin{bmatrix} f_n \\ g_n \\ u_{nm} \\ v_{nm} \\ w_{nm} \end{bmatrix} \geq 0, \quad (7.60)$$

where the elements of the array are subtracted density matrix elements which are defined to satisfy the rule that the subtracted part is split at the position of the colon. For example:

$$\langle l^+, k : n\rangle = \langle l^+, k, n\rangle - \langle l^+, k\rangle\langle n\rangle. \quad (7.61)$$

This nonnegativity condition on the above array of operators may be combined with the commutation or anticommutation relations for the creation and annihilation operators to yield relations involving the one- and two-particle density matrices. For instance, if we are dealing with a system of fermions the element in the lower right-hand corner of the array may be written in terms of Γ and γ as:

$$\langle l, k : n^+, m^+\rangle = \Gamma(n, m \mid k, l) + \delta(k, m)\gamma(n \mid l) + \delta(n, l)\gamma(m \mid k) - \delta(k, n)\gamma(m \mid l) - \delta(l, m)\gamma(n \mid k) + \delta(k, n)\delta(l, m) - \delta(l, n)\delta(k, m) - \langle l, k\rangle\langle n^+, m^+\rangle = Q(k, l \mid n, m) - \langle l, k\rangle\langle n^+, m^+\rangle,$$

where the operator Q is that which was introduced at the start of Sec. 5.

APPENDIX A

Let θ be a closed convex cone in a Hilbert space S, i.e.,

$$x_1 \in \theta \quad \text{and} \quad x_2 \in \theta$$

implies $ax_1 + bx_2 \in \theta$ whenever a and b are nonnegative. Consider a set $\tilde{\theta}$ defined as follows:

$$\tilde{\theta} = \{y : (x, y) \geq 0 \quad \text{for every} \quad x \in \theta\}.$$

Lemma: $\tilde{\theta}$ is a closed convex cone.

Proof: $(ay_1 + by_2, x) = a(y_1, x) + b(y_2, x) \geq 0$ whenever $a, b \geq 0$. Suppose $y_n \to y_\infty$. Then $(y_n, x) \to (y_\infty, x)$. But every $(y_n, x) \geq 0$. Hence $(y_\infty, x) \geq 0$ and $y_\infty \in \tilde{\theta}$.

Theorem: $\tilde{\tilde{\theta}} = \theta$.

Proof: (Part I) $\theta \subset \tilde{\tilde{\theta}}$. Suppose $x \in \theta$, then for every $y \in \tilde{\theta}$, $(x, y) \geq 0$. Hence $\theta \subset \tilde{\tilde{\theta}}$.

(Part II) $\tilde{\tilde{\theta}} \subset \theta$. Assume $\tilde{\tilde{\theta}}$ is not contained in θ. Then we may find at least one element, a, such that: $a \notin \theta$ and: $(a, y) \geq 0$ for every $y \in \tilde{\theta}$. Let $r =$ g. l. b. $|a - x|$ for all $x \in \theta$. Since θ is closed, $r > 0$.

Consider a sequence, $r_n > r$ such that $r_n \to r$. Construct the sequence of spheres of radius r_n about a. $S_n = \{u: |u - a| \leq r_n\}$. S_n is a closed convex set. Let I_n be the intersection of S_n and θ. Then I_n is a closed convex set. Consider two sequences: $v_n \in I_n$ and $v_n' \in I_n$. Then $r \leq |v_n - a| \leq r_n$. Hence $|v_n - a| \to r$. Similarly, $|v_n' - a| \to r$. We construct the midpoint of v_n and v_n':

$$\bar{v}_n = \tfrac{1}{2}v_n + \tfrac{1}{2}v_n' \in I_n.$$

Then

$$(a - \bar{v}_n, a - \bar{v}_n) = (a, a) - (a, \bar{v}_n) - (\bar{v}_n, a) + (\bar{v}_n, \bar{v}_n)$$
$$= (a, a) - \tfrac{1}{2}(a, v_n) - \tfrac{1}{2}(a, v_n') - \tfrac{1}{2}(v_n, a) - \tfrac{1}{2}(v_n', a) + \tfrac{1}{4}(v_n, v_n) + \tfrac{1}{4}(v_n, v_n') + \tfrac{1}{4}(v_n', v_n) + \tfrac{1}{4}(v_n', v_n').$$

Hence

$$(a - \bar{v}_n, a - \bar{v}_n) = \tfrac{1}{2}(a - v_n, a - v_n) - \tfrac{1}{2}(a - v_n', a - v_n') = -\tfrac{1}{4}(v_n - v_n', v_n - v_n').$$

But

$$(a - \bar{v}_n, a - \bar{v}_n) \to (a - v_n, a - v_n) \to (a - v_n', a - v_n') \to r^2.$$

Therefore,

$$|v_n - v_n'| \to 0.$$

Hence the intersection I_n converges to a point. Since the convex cone θ is closed, $I_n \to x_0$ where $x_0 \in \theta$.

We prove the following two statements which then contradict our hypothesis:

(A) $x_0 - a \in \tilde{\theta}$ (B) $(x_0 - a, a) < 0$.

Proof of (A): Let $x \in \theta$. Then $x_0 + \beta x \in \theta$ for $\beta \geq 0$. Hence $(x_0 + \beta x - a, x_0 + \beta x - a) > (x_0 - a, x_0 - a) + (x_0 - a, x) + \frac{1}{2}\beta(x, x) > 0$. Letting $\beta \to 0$, we obtain

$$(x_0 - a, x) \geq 0 \quad \text{for all} \quad x \in \theta.$$

Proof of (B):

$$\frac{d}{d\beta}\,[(1 + \beta)x_0 - a], (1 + \beta, x_0 - a)\Big|_{\beta=0} = 0.$$

Therefore $(x_0 - a, x_0) = 0$. But $-(x_0 - a, x_0 - a) = (x_0 - a, a) < 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Effect of Small Irregularities on Electromagnetic Scattering from an Interface of Arbitrary Shape*

K. M. MITZNER
Northrop Ventura, Newbury Park, California
(Received 8 May 1964; final manuscript received 28 July 1964)

A perturbation analysis is used to determine the effect of small irregularities on electromagnetic scattering from an interface between two media. The interface irregularities are replaced by approximate equivalent surface currents, and the field in space can then be found using the dyadic Green's function of the unperturbed problem. The approach is valid when the irregularity has small slope and amplitude small compared to the wavelengths and local radii of curvature. To facilitate applications, the theory of dyadic Green's functions is developed, and the necessary functions are given for some important geometries.

1. INTRODUCTION

THE problem of electromagnetic scattering from an interface between two linear, homogeneous, isotropic media can be solved directly only for fairly simple geometrics. On the other hand, it is well known that the scattered field can be affected significantly by relatively small geometric irregularities. Thus it would be desirable to have a means of calculating the field perturbation caused by a small geometric perturbation in terms of the solution for the underlying interface S_0. We give here a method valid when the irregularities have slope small compared to unity and amplitude small compared to the wavelengths of interest and compared to the local radii of curvature of S_0. Appropriate problems arise in such areas as radio wave propagation over the ocean, radar reflection from balloons, and light scattering by imperfectly polished lenses and mirrors.

The method is based on the perturbation technique first developed by Lord Rayleigh[1] to treat reflection of a scalar wave from an irregular plane wall. In this technique, the irregularity is characterized by a small displacement parameter ϵ, and the field is calculated as a power series in ϵ, the constant term being the unperturbed field. Usually only one or two additional terms of the series are actually calculated. In many treatments, including Lord Rayleigh's, the small parameter ϵ is not expressed explicitly.

Rayleigh's technique has already been applied successfully to scattering of an electromagnetic wave at an irregular plane interface. Bass and Bocharov[2] have solved this problem for an arbitrary wave incident on a perfectly conducting interface. Rice[3] has

* This paper is based on a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the California Institute of Technology, Pasadena, California.

[1] Lord Rayleigh, *The Theory of Sound* (Dover Publications, Inc., New York, 1945), Vol. II, p. 89.
[2] F. G. Bass and V. G. Bocharov, Radiotekhnika i Elektronika 3, 180 (1958). [English transl.: Radio Eng. Electron. 3, 251 (1958).]
[3] S. O. Rice, Commun. Pure Appl. Math. 4, 351 (1951).

Therefore,

$$|v_n - v_n'| \to 0.$$

Hence the intersection I_n converges to a point. Since the convex cone θ is closed, $I_n \to x_0$ where $x_0 \in \theta$.

We prove the following two statements which then contradict our hypothesis:

(A) $x_0 - a \in \tilde{\theta}$ (B) $(x_0 - a, a) < 0$.

Proof of (A): Let $x \in \theta$. Then $x_0 + \beta x \in \theta$ for $\beta \geq 0$. Hence $(x_0 + \beta x - a, x_0 + \beta x - a) > (x_0 - a, x_0 - a) + (x_0 - a, x) + \frac{1}{2}\beta(x, x) > 0$. Letting $\beta \to 0$, we obtain

$$(x_0 - a, x) \geq 0 \quad \text{for all} \quad x \in \theta.$$

Proof of (B):

$$\frac{d}{d\beta}\,[(1 + \beta)x_0 - a], (1 + \beta, x_0 - a)\Big|_{\beta=0} = 0.$$

Therefore $(x_0 - a, x_0) = 0$. But $-(x_0 - a, x_0 - a) = (x_0 - a, a) < 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Effect of Small Irregularities on Electromagnetic Scattering from an Interface of Arbitrary Shape*

K. M. MITZNER
Northrop Ventura, Newbury Park, California
(Received 8 May 1964; final manuscript received 28 July 1964)

A perturbation analysis is used to determine the effect of small irregularities on electromagnetic scattering from an interface between two media. The interface irregularities are replaced by approximate equivalent surface currents, and the field in space can then be found using the dyadic Green's function of the unperturbed problem. The approach is valid when the irregularity has small slope and amplitude small compared to the wavelengths and local radii of curvature. To facilitate applications, the theory of dyadic Green's functions is developed, and the necessary functions are given for some important geometries.

1. INTRODUCTION

THE problem of electromagnetic scattering from an interface between two linear, homogeneous, isotropic media can be solved directly only for fairly simple geometrics. On the other hand, it is well known that the scattered field can be affected significantly by relatively small geometric irregularities. Thus it would be desirable to have a means of calculating the field perturbation caused by a small geometric perturbation in terms of the solution for the underlying interface S_0. We give here a method valid when the irregularities have slope small compared to unity and amplitude small compared to the wavelengths of interest and compared to the local radii of curvature of S_0. Appropriate problems arise in such areas as radio wave propagation over the ocean, radar reflection from balloons, and light scattering by imperfectly polished lenses and mirrors.

The method is based on the perturbation technique first developed by Lord Rayleigh[1] to treat reflection of a scalar wave from an irregular plane wall. In this technique, the irregularity is characterized by a small displacement parameter ϵ, and the field is calculated as a power series in ϵ, the constant term being the unperturbed field. Usually only one or two additional terms of the series are actually calculated. In many treatments, including Lord Rayleigh's, the small parameter ϵ is not expressed explicitly.

Rayleigh's technique has already been applied successfully to scattering of an electromagnetic wave at an irregular plane interface. Bass and Bocharov[2] have solved this problem for an arbitrary wave incident on a perfectly conducting interface. Rice[3] has

* This paper is based on a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the California Institute of Technology, Pasadena, California.

[1] Lord Rayleigh, *The Theory of Sound* (Dover Publications, Inc., New York, 1945), Vol. II, p. 89.
[2] F. G. Bass and V. G. Bocharov, Radiotekhnika i Elektronika 3, 180 (1958). [English transl.: Radio Eng. Electron. 3, 251 (1958).]
[3] S. O. Rice, Commun. Pure Appl. Math. 4, 351 (1951).

solved it for a plane wave incident on an interface between a dielectric and an arbitrary medium. In both cases, the perturbed field near the interface is expanded in a double series in ϵ and in the coordinate normal to S_0. Then this expansion is inserted into the tangential boundary conditions at the perturbed interface, and the resulting equations are solved to find an intermediate result from which the perturbation field can be calculated. In the method of Bass and Bocharov, the intermediate result is the effective tangential field at S_0; the perturbation field in space is found using Kirchhoff's formulas. In Rice's method, the intermediate results are the coefficients of a Fourier series expansion of the effective field at S_0; the field in space is found by multiplying these coefficients by the appropriate mode functions and summing.

Our extension for problems in which S_0 is not necessarily plane proceeds along similar lines. The perturbed field is expanded in a double series in ϵ and the normal coordinate, the expansion is inserted into the tangential boundary conditions, and the resulting equations are solved. The intermediate results thus obtained are effective electric and magnetic surface current sources on S_0. The perturbation field in space is then calculated using the dyadic Green's functions (henceforth abbreviated d.G.f.'s) of the unperturbed problem. This treatment is similar to that used by Kur'yanov[4] to calculate the scattering of an acoustic (scalar) wave from a curved surface with small irregularities, but the present development is both more general and more detailed than Kur'yanov's.

The perturbation method is shown to work in principle if the irregularity is small enough so that the unperturbed field and perturbation fields have unique nonsingular mathematical continuations in the volume between S_0 and the perturbed interface. The practical difficulty of calculating higher-order perturbation fields further constrains the admissible irregularities, and we obtain the requirements of small slope and small amplitude already stated.

Section 2 contains background material on dyadics and on electromagnetic theory. Section 3 treats the theory of d.G.f.'s, including both material which is necessary to our development and additional material which will prove useful in applications; much of this material has not previously appeared in the literature or has appeared in incorrect form. Expressions for d.G.f.'s and related functions for some important geometries are given in Appendix 1.

[4] B. F. Kur'yanov, Akust. Zh. 8, 325 (1962). [English transl.: Soviet Phys.—Acoust. 8, 252 (1963)].

Section 4 includes the formal derivation of general expressions to second order in ϵ for the effective surface currents. These expressions are the core of our method. The conditions under which the method is applicable are discussed in Sec. 4 and Appendix 2.

The application of the method is illustrated in Sec. 5, where we calculate the first-order perturbation field for plane wave scattering from a perfectly conducting cylinder with sinusoidal surface irregularities. The results show that small irregularities of long wavelength cause a relatively large change in the scattered field, the effect decreasing as the irregularity wavelength decreases.

Harmonic time dependence $e^{-i\omega t}$ is to be understood everywhere.

2. BACKGROUND MATERIAL

Dyadics

The theory of dyadics has been treated adequately by Gibbs[5] and by Morse and Feshbach,[6] but a few additional comments are necessary here.

The transpose of a dyadic Q will be written Q^{T} and the unit dyadic will be designated by I.

The operation $Q \times \nabla$ will be defined by

$$Q \times \nabla \equiv -(\nabla \times Q^{\mathrm{T}})^{\mathrm{T}}$$
$$= \partial Q/\partial x \times \mathbf{e}_x + \partial Q/\partial y \times \mathbf{e}_y + \partial Q/\partial z \times \mathbf{e}_z. \quad (2.1)$$

This definition is consistent with the standard practice of treating the operator ∇ as a vector, for it is analogous to

$$Q \times \mathbf{A} = -(\mathbf{A} \times Q^{\mathrm{T}})^{\mathrm{T}}. \quad (2.2)$$

The expansion of a dyadic in component form,

$$Q = \sum_{i,j=1}^{3} \mathbf{e}_i Q_{ij} \mathbf{e}'_j, \quad (2.3)$$

where the $\mathbf{e}_i$ and $\mathbf{e}'_j$ are two sets of independent unit vectors, is valid for curvilinear as well as Cartesian coordinates. However, it must be remembered that in curvilinear coordinates a differential operator applied to Q operates on both the Q_{ij} and the unit vectors.

Basic Equations of Electromagnetic Theory

In this paper we are interested in problems involving two linear homogeneous isotropic media M_1 and M_2 separated by an interface S. When the order

[5] E. B. Wilson, *Vector Analysis (Founded upon the Lectures of J. Willard Gibbs)* (Dover Publications, Inc., New York, 1960), especially Chaps. V and VII.

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, Secs. 1.3–1.6.

of the media is not important, we shall designate one by M_a and the other by M_p. Medium M_a will be characterized by the material parameters μ_a and k_a, and the volume filled by M_a will be called V_a. Maxwell's equations in mks units are then, for $\mathbf{r}$ in V_a,

$$\begin{aligned} \nabla \times \mathbf{E}(\mathbf{r}) - i\omega\mu_a \mathbf{H}(\mathbf{r}) &= -\mathbf{J}_m(\mathbf{r}), \\ \nabla \times \mathbf{H}(\mathbf{r}) - (k_a^2/i\omega\mu_a)\mathbf{E}(\mathbf{r}) &= \mathbf{J}_e(\mathbf{r}), \end{aligned} \tag{2.4}$$

where $\mathbf{J}_e$ is an electric current distribution and $\mathbf{J}_m$ is a magnetic current distribution. The magnetic current has no physical existence, but effective magnetic currents appear frequently in the mathematics of electromagnetic theory.

The two parts of Eq. (2.4) can be combined to give the second-order forms

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E} - k_a^2 \mathbf{E} &= i\omega\mu_a \mathbf{J}_e - \nabla \times \mathbf{J}_m, \\ \nabla \times \nabla \times \mathbf{H} - k_a^2 \mathbf{H} &= (-k_a^2/i\omega\mu_a)\mathbf{J}_m + \nabla \times \mathbf{J}_e. \end{aligned} \tag{2.5}$$

In a source-free region

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{H} = 0, \tag{2.6}$$

and the operator $(\nabla \times \nabla \times)$ can be replaced by $(-\nabla^2)$.

If neither medium is a perfect electric or magnetic conductor, then the boundary conditions at S are

$$\Delta(\mathbf{n} \times \mathbf{H}) = \mathbf{K}_e, \qquad \Delta(\mathbf{n} \times \mathbf{E}) = -\mathbf{K}_m. \tag{2.7}$$

Here

$$\Delta F = F_2 - F_1 \tag{2.8}$$

is the jump in the function F across the interface, $\mathbf{n}$ is the unit normal from V_1 to V_2, and $\mathbf{K}_e$ and $\mathbf{K}_m$ are, respectively, electric and magnetic surface current source distributions on the interface. When one medium, say M_1, is a perfect electric conductor (k_1^2/μ_1 infinite), then electric surface current *sources* induce equal and opposite surface currents and thus have no net effect. The field in V_1 is zero, the boundary condition for $\mathbf{E}$ becomes

$$\mathbf{n} \times \mathbf{E}_2 = -\mathbf{K}_m, \tag{2.9}$$

and the electric surface current is then uniquely determined. Similarly, when M_1 is a perfect magnetic conductor (μ_1 infinite), then magnetic surface current sources have no effect, the field in V_1 is again zero, the boundary condition for $\mathbf{H}$ is

$$\mathbf{n} \times \mathbf{H}_2 = \mathbf{K}_e, \tag{2.10}$$

and the magnetic surface current is uniquely determined.

It should be noted that the fields are unchanged (except infinitesimally close to the interface) if a surface current source is shifted an infinitesimal distance into one medium; by making such a shift we can always eliminate $\mathbf{K}_e$ and $\mathbf{K}_m$ from the boundary conditions. In some situations this is the most convenient viewpoint, whereas in others it is desirable to retain the sources on the interface.

3. THE DYADIC GREEN'S FUNCTION

We now consider the theory of the d.G.f. for problems involving two media. This work is an extension of that of Schwinger and Levine[7,8] for scattering from a perfect conductor. However, in order to obtain simple boundary conditions, we have modified slightly the definition of the d.G.f. given by these authors; thus the $\Gamma^{(1)}$ of Ref. 8 is $(1/i\omega\mu)\,\Gamma_e$ in our notation, and $\Gamma^{(2)}$ is $(-i\omega\mu/k^2)\,\Gamma_m$. A previous extension to two media by Tai[9] is incorrect, apparently because it is based on an erroneous analysis by Morse and Feshbach.[10]

Let the vector $\boldsymbol{\Gamma}_{e,j}(\mathbf{r};\mathbf{r}')$ denote the electric field at $\mathbf{r}$ due to a unit impulse electric current at $\mathbf{r}'$ directed along $\mathbf{e}_j'$. The electric d.G.f., $\Gamma_e(\mathbf{r};\mathbf{r}')$, will be defined by

$$\Gamma_e(\mathbf{r};\mathbf{r}') = \sum_j \boldsymbol{\Gamma}_{e,j}(\mathbf{r};\mathbf{r}')\mathbf{e}_j'. \tag{3.1}$$

The electric field due to an arbitrary electric current source distribution $\mathbf{J}_e(\mathbf{r}')$ in volume V is then given by

$$\mathbf{E}(\mathbf{r}) = \int_V dV' \Gamma_e(\mathbf{r};\mathbf{r}') \cdot \mathbf{J}_e(\mathbf{r}'). \tag{3.2}$$

By applying Eq. (2.5) to the $\boldsymbol{\Gamma}_{e,j}$, we obtain

$$\nabla \times \nabla \times \Gamma_e(\mathbf{r};\mathbf{r}') - k^2 \Gamma_e(\mathbf{r};\mathbf{r}') = i\omega\mu I\, \delta(\mathbf{r} - \mathbf{r}'), \tag{3.3}$$

where k and μ are given their value at $\mathbf{r}$. Applying Eqs. (2.7), (2.9), (2.10) to the $\boldsymbol{\Gamma}_{e,j}$, we find the boundary conditions for $\mathbf{r}$ on the interface S to be

$$\mathbf{n}(\mathbf{r}) \times \Gamma_{e,2}(\mathbf{r};\mathbf{r}') = 0,$$

M_1 a perfect electric conductor; (3.4)

$$\mathbf{n}(\mathbf{r}) \times [\nabla \times \Gamma_{e,2}(\mathbf{r};\mathbf{r}')] = 0,$$

M_1 a perfect magnetic conductor; (3.5)

$$\Delta[\mathbf{n}(\mathbf{r}) \times \Gamma_e(\mathbf{r};\mathbf{r}')] = 0,$$

$$\Delta\{\mathbf{n}(\mathbf{r}) \times [(1/i\omega\mu)\nabla \times \Gamma_e(\mathbf{r};\mathbf{r}')]\} = 0,$$

no perfect conductors. (3.6)

[7] J. Schwinger, MIT Radiation Laboratory Report No. 205 (1943) (unpublished).

[8] H. Levine and J. Schwinger, Commun. Pure Appl. Math., 3, 355 (1950).

[9] C. T. Tai, Stanford Research Institute Technical Report 46 (1954) (unpublished).

[10] Morse and Feshbach, Ref. 6, Vol. II, Chap. XIII. The reciprocity relation given on p. 1770 is incorrect [cf., our Eq. (3.27)] and much of the subsequent development depends on this relation.

Here it is assumed (without loss of generality) that M_2 is never a perfect conductor.

In deriving Eqs. (3.3)–(3.6), the source point $\mathbf{r}'$ is required to lie in one medium or the other but not exactly astride S. This assumption is necessary because $\Gamma_e(\mathbf{r};\mathbf{r}')$ is discontinuous in $\mathbf{r}'$ at S. However, the dyadic

$$\Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}') = -[\Gamma_e(\mathbf{r};\mathbf{r}')\times\mathbf{n}(\mathbf{r}')]\times\mathbf{n}(\mathbf{r}'), \tag{3.7}$$

defined for $\mathbf{r}'$ on S, which represents the response to current sources parallel to the boundary, is unambiguously defined. This dyadic plays an important role in the perturbation theory, where all the sources of interest are surface currents.

An equation and boundary conditions for $\Gamma_e^{\parallel}$ can be obtained using Eqs. (2.5), (2.7), (2.9), (2.10). We find

$$\nabla\times\nabla\times\Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}') - k^2\Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}') = 0, \quad \mathbf{r} \text{ not in } V'; \tag{3.8}$$

$$\Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}') \equiv 0, \quad M_1 \text{ a perfect electric conductor;} \tag{3.9}$$

$$\mathbf{n}(\mathbf{r})\times[(1/i\omega\mu_2)\nabla\times\Gamma_{e,2}^{\parallel}(\mathbf{r};\mathbf{r}')] = (\mathbf{e}_{\xi'}\mathbf{e}_{\xi'}+\mathbf{e}_{\eta'}\mathbf{e}_{\eta'})(h_{\xi'}h_{\eta'})^{-1}\delta(\xi-\xi')\delta(\eta-\eta'), \quad M_1 \text{ a perfect magnetic conductor;} \tag{3.10}$$

$$\Delta[\mathbf{n}(\mathbf{r})\times\Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}')] = 0,$$
$$\Delta\{\mathbf{n}(\mathbf{r})\times[(1/i\omega\mu)\nabla\times\Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}')]\} = (\mathbf{e}_{\xi'}\mathbf{e}_{\xi'}+\mathbf{e}_{\eta'}\mathbf{e}_{\eta'})(h_{\xi'}h_{\eta'})^{-1}\delta(\xi-\xi')\delta(\eta-\eta'), \quad \text{no perfect conductors.} \tag{3.11}$$

Here V' is a small sphere around $\mathbf{r}'$, and (ξ, η, ζ) and (ξ', η', ζ') are the coordinates of $\mathbf{r}$ and $\mathbf{r}'$, respectively, in a right-handed orthogonal system with metrics $(h_{\xi'}, h_{\eta'}, h_{\zeta'})$ and with

$$\mathbf{e}_{\zeta'} = \mathbf{n} \tag{3.12}$$

on the interface.

Now let us consider the magnetic d.G.f. $\Gamma_m(\mathbf{r};\mathbf{r}')$. Its definition is similar to that of Γ_e, but with magnetic fields and currents replacing electric fields and currents. The equations analogous to Eqs. (3.2–11) are

$$H(\mathbf{r}) = \int_V dV'\ \Gamma_m(\mathbf{r};\mathbf{r}')\cdot\mathbf{J}_m(\mathbf{r}'); \tag{3.13}$$

$$\nabla\times\nabla\times\Gamma_m(\mathbf{r};\mathbf{r}') - k^2\Gamma_m(\mathbf{r};\mathbf{r}') = (-k^2/i\omega\mu)I\delta(\mathbf{r}-\mathbf{r}'); \tag{3.14}$$

$$\mathbf{n}(\mathbf{r})\times\Gamma_{m,2}(\mathbf{r};\mathbf{r}') = 0, \quad M_1 \text{ a perfect magnetic conductor;} \tag{3.15}$$

$$\mathbf{n}(\mathbf{r})\times[\nabla\times\Gamma_{m,2}(\mathbf{r};\mathbf{r}')] = 0, \quad M_1 \text{ a perfect electric conductor;} \tag{3.16}$$

$$\Delta[\mathbf{n}(\mathbf{r})\times\Gamma_m(\mathbf{r};\mathbf{r}')] = 0,$$
$$\Delta\{\mathbf{n}(\mathbf{r})\times[(i\omega\mu/k^2)\nabla\times\Gamma_m(\mathbf{r};\mathbf{r}')]\} = 0, \quad \text{no perfect conductors;} \tag{3.17}$$

$$\Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}') = -[\Gamma_m(\mathbf{r};\mathbf{r}')\times\mathbf{n}(\mathbf{r}')]\times\mathbf{n}(\mathbf{r}') \tag{3.18}$$

for $\mathbf{r}'$ on S;

$$\nabla\times\nabla\times\Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}') - k^2\Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}') = 0, \quad \mathbf{r} \text{ not in } V'; \tag{3.19}$$

$$\Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}') \equiv 0, \quad M_1 \text{ a perfect magnetic conductor;} \tag{3.20}$$

$$\mathbf{n}(\mathbf{r})\times[(i\omega\mu_2/k_2^2)\nabla\times\Gamma_{m,2}^{\parallel}(\mathbf{r};\mathbf{r}')] = -(\mathbf{e}_{\xi'}\mathbf{e}_{\xi'}+\mathbf{e}_{\eta'}\mathbf{e}_{\eta'})(h_{\xi'}h_{\eta'})^{-1}\delta(\xi-\xi')\delta(\eta-\eta'), \quad M_1 \text{ a perfect electric conductor;} \tag{3.21}$$

$$\Delta[\mathbf{n}(\mathbf{r})\times\Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}')] = 0,$$
$$\Delta\{\mathbf{n}(\mathbf{r})\times[(i\omega\mu/k^2)\nabla\times\Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}')]\} = -(\mathbf{e}_{\xi'}\mathbf{e}_{\xi'}+\mathbf{e}_{\eta'}\mathbf{e}_{\eta'})(h_{\xi'}h_{\eta'})^{-1}\delta(\xi-\xi')\delta(\eta-\eta'), \quad \text{no perfect conductors.} \tag{3.22}$$

In a general problem, both electric and magnetic current sources are present, so that the total field is expressed in terms of the sources and d.G.f.'s by[11]

$$\mathbf{E}(\mathbf{r}) = \int_V dV'\ \Gamma_e(\mathbf{r};\mathbf{r}')\cdot\mathbf{J}_e(\mathbf{r}') + (i\omega\mu/k^2)\nabla\times\int_V dV'\ \Gamma_m(\mathbf{r};\mathbf{r}')\cdot\mathbf{J}_m(\mathbf{r}'), \tag{3.23}$$

$$\mathbf{H}(\mathbf{r}) = \int_V dV'\ \Gamma_m(\mathbf{r};\mathbf{r}')\cdot\mathbf{J}_m(\mathbf{r}') + (1/i\omega\mu)\nabla\times\int_V dV'\ \Gamma_e(\mathbf{r};\mathbf{r}')\cdot\mathbf{J}_e(\mathbf{r}'). \tag{3.24}$$

If the only sources are surface currents on S, then these equations become

$$\mathbf{E}(\mathbf{r}) = \int_S dS'\ \Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}')\cdot\mathbf{K}_e(\mathbf{r}') + (i\omega\mu/k^2)\nabla\times\int_S dS'\ \Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}')\cdot\mathbf{K}_m(\mathbf{r}'), \tag{3.25}$$

$$\mathbf{H}(\mathbf{r}) = \int_S dS'\ \Gamma_m^{\parallel}(\mathbf{r};\mathbf{r}')\cdot\mathbf{K}_m(\mathbf{r}') + (1/i\omega\mu)\nabla\times\int_S dS'\ \Gamma_e^{\parallel}(\mathbf{r};\mathbf{r}')\cdot\mathbf{K}_e(\mathbf{r}'); \tag{3.26}$$

[11] We are assuming here that there are no source-free solutions and that a radiation condition is satisfied.

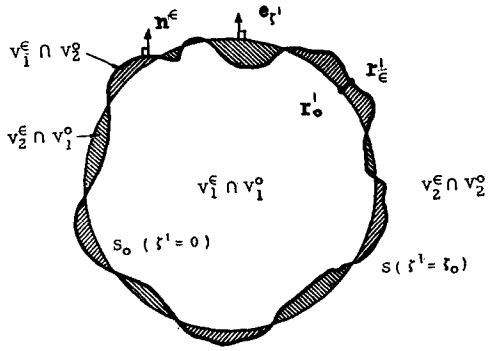


FIG. 1. The irregular interface.

these are the forms which are of most use in the perturbation theory.

Reciprocity relations analogous to Eqs. (3.15) and (3.16) of Ref. 8 can be obtained by applying the vector Green's theorem[12] in each medium and cancelling the surface integrals on the interface against each other. The results are

$$\Gamma(\mathbf{r};\mathbf{r}') = \Gamma^T(\mathbf{r}';\mathbf{r}), \tag{3.27}$$

$$(1/i\omega\mu)\nabla \times \Gamma_e(\mathbf{r};\mathbf{r}') = [i\omega\mu'/(k')^2]\Gamma_m(\mathbf{r};\mathbf{r}') \times \nabla', \tag{3.28}$$

$$(i\omega\mu/k^2)\nabla \times \Gamma_m(\mathbf{r};\mathbf{r}') = (1/i\omega\mu')\Gamma_e(\mathbf{r};\mathbf{r}') \times \nabla'. \tag{3.29}$$

Expressions for Γ, Γ^{II}, and $\nabla \times \Gamma^{\text{II}}$ for some important geometries are given in Appendix 1.

4. THE EFFECTIVE SURFACE CURRENTS

The field scattered from an irregular interface S can be considered as the sum of the field scattered from the underlying interface S_0 plus a perturbation field. When the irregularity is of sufficiently small scale, equivalent surface currents on S_0 can be found which produce the same perturbation field to some approximation. We shall now develop expressions for these effective currents accurate to second order in a small parameter ϵ characterizing the irregularity. These expressions can be used in Eqs. (3.25–26) to find the perturbation field in space.

The situation of interest is shown in Fig. 1 (with the irregularity exaggerated). Here V_q^ϵ is the volume to the side of S occupied by medium M_q and V_q^0 is the volume to the side of S_0 occupied by M_q in the unperturbed case. The points common to two volumes are denoted using the standard symbol $\cap$ for an intersection. A right-handed orthogonal coordinate system (ξ', η', ζ') is established such that S_0 is the coordinate surface $\zeta' = 0$ and $\zeta' > 0$ in V_2^0; this system need not be defined everywhere in space, but S must be within the volume in which it is defined.

The equation of S is written

$$Q = \zeta' - \zeta_0(\xi', \eta') = \zeta' - \zeta_0(\mathbf{r}_0') = \zeta' - w_0(\mathbf{r}_0')\epsilon = 0. \tag{4.1}$$

Here w_0 is to be considered as defined in a volume but independent of ζ'; it satisfies the inequalities

$$\begin{aligned} \max\{h_{\zeta'}(\mathbf{r}')\,|w_0|\} &\leq 1, \\ \max\{h_{\zeta'}(\mathbf{r}')\,|\nabla_T' w_0|\} &\leq B, \end{aligned} \tag{4.2}$$

where B is a given constant, ∇_T' is the two-dimensional gradient tangent to S_0, and the maximum is taken over the volume between S_0 and S. A point $\mathbf{r}_0'$ on S_0 is mapped uniquely into a point $\mathbf{r}_\epsilon'$ on S by

$$\mathbf{r}_\epsilon' = \mathbf{r}_0' + \int_0^{\zeta_0} d\zeta'\, h_{\zeta'}\mathbf{e}_{\zeta'}. \tag{4.3}$$

Now let $\mathbf{r}'$ be a point removed from $\mathbf{r}_0'$ by a curvilinear distance of order ϵ along a ζ' curve. Let $\mathbf{F}_q(\mathbf{r}')$ be some field vector (either an $\mathbf{E}$ or an $\mathbf{H}$) satisfying Maxwell's equations for medium M_q. Then $\mathbf{F}_q$ can be expanded formally in ζ' as

$$\mathbf{F}_q(\mathbf{r}') = \mathbf{F}_q(\mathbf{r}_0') + \zeta' \frac{\partial}{\partial\zeta'}\mathbf{F}_q(\mathbf{r}_0') + \tfrac{1}{2}(\zeta')^2\left(\frac{\partial}{\partial\zeta'}\right)^2\mathbf{F}_q(\mathbf{r}_0') + O(\epsilon^3). \tag{4.4}$$

The perturbed field $\mathbf{F}_q^\epsilon$ can be expanded formally in ϵ as

$$\mathbf{F}_q^\epsilon(\mathbf{r}') = \mathbf{F}_q^0(\mathbf{r}') + \delta\mathbf{F}_q(\mathbf{r}') + \delta^2\mathbf{F}_q(\mathbf{r}') + O(\epsilon^3). \tag{4.5}$$

Here $\mathbf{F}_q^0$ is the unperturbed field (incident plus scattered) and $\delta^n\mathbf{F}_q$ is the perturbation field of order ϵ^n; the $\delta^n\mathbf{F}_q$ are required to be solutions of Maxwell's equations for medium M_q. If now the three fields on the right-hand side in Eq. (4.5) have expansions of the form (4.4), we obtain the double series expansion

$$\begin{aligned} \mathbf{F}_q^\epsilon(\mathbf{r}') = \mathbf{F}_q^0(\mathbf{r}_0') &+ \left[\zeta' \frac{\partial}{\partial\zeta'}\mathbf{F}_q^0(\mathbf{r}_0') + \delta\mathbf{F}_q(\mathbf{r}_0')\right] \\ &+ \left[\tfrac{1}{2}(\zeta')^2\left(\frac{\partial}{\partial\zeta'}\right)^2\mathbf{F}_q^0(\mathbf{r}_0') + \zeta' \frac{\partial}{\partial\zeta'}\delta\mathbf{F}_q(\mathbf{r}_0') + \delta^2\mathbf{F}_q(\mathbf{r}_0')\right] + O(\epsilon^3). \end{aligned} \tag{4.6}$$

The boundary conditions on both fields at S are of form

$$\Delta[\mathbf{n}^\epsilon(\mathbf{r}_\epsilon') \times \mathbf{F}^\epsilon(\mathbf{r}_\epsilon')] = 0, \tag{4.7}$$

[12] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Sec. 4.14.

where the jump operator Δ is defined as in Eq. (2.8) and $\mathbf{n}^\epsilon$ is the unit normal to S directed from V_1^ϵ to V_2^ϵ. Condition (4.7) can also be written as

$$\Delta[\mathbf{X}(\mathbf{r}_\epsilon') \times \mathbf{F}^\epsilon(\mathbf{r}_\epsilon')] = 0, \tag{4.8}$$

where the vector

$$\mathbf{X}(\mathbf{r}_\epsilon') = (h_{\xi'} |\nabla' Q| \mathbf{n}^\epsilon)]_{\mathbf{r}'=\mathbf{r}_\epsilon'} = (h_{\xi'} \nabla' Q)]_{\mathbf{r}'=\mathbf{r}_\epsilon'} \tag{4.9}$$

has the same direction as $\mathbf{n}^\epsilon$ everywhere on S. This vector can be expressed in terms of quantities measured at $\mathbf{r}_0'$ as[13]

$$\mathbf{X}(\mathbf{r}_\epsilon') = \left\{ \mathbf{e}_{\xi'} - \nabla_T'(h_{\xi'}\zeta_0) - \frac{1}{2}\frac{\partial}{\partial \zeta'} \nabla_T'(h_{\xi'}\zeta_0^2) \right\}_{\mathbf{r}'=\mathbf{r}_0'} + O(\epsilon^3). \tag{4.10}$$

Setting Eqs. (4.6) and (4.10) into Eq. (4.8) and equating terms of the same order gives expressions for the jump in the perturbation fields at S_0. Converting these jumps to surface currents by use of Eq. (2.7), we obtain the effective surface currents

$$\delta\mathbf{K}_m = \mathbf{e}_{\xi'} \times \left[(\Delta E_{\xi'}^0) \nabla_T'(h_{\xi'}\zeta_0) + \zeta_0 \Delta \frac{\partial}{\partial \zeta'} \mathbf{E}^0 \right], \tag{4.11}$$

$$\delta^2\mathbf{K}_m = \mathbf{e}_{\xi'} \times \left[\tfrac{1}{2}\zeta_0^2 \Delta \left(\frac{\partial}{\partial \zeta'}\right)^2 \mathbf{E}^0 + \tfrac{1}{2}(\Delta E_{\xi'}^0) \frac{\partial}{\partial \zeta'} \nabla_T'(h_{\xi'}\zeta_0^2) + \left(\Delta \frac{\partial}{\partial \zeta'} E_{\xi'}^0\right) \zeta_0 \nabla_T'(h_{\xi'}\zeta_0) + (\Delta \delta E_{\xi'}) \nabla_T'(h_{\xi'}\zeta_0) + \zeta_0 \Delta \frac{\partial}{\partial \zeta'} \delta\mathbf{E} \right], \tag{4.12}$$

$$\delta\mathbf{K}_e = -\mathbf{e}_{\xi'} \times \left[(\Delta H_{\xi'}^0) \nabla_T'(h_{\xi'}\zeta_0) + \zeta_0 \Delta \frac{\partial}{\partial \zeta'} \mathbf{H}^0 \right], \tag{4.13}$$

$$\delta^2\mathbf{K}_e = -\mathbf{e}_{\xi'} \times \left[\tfrac{1}{2}\zeta_0^2 \Delta \left(\frac{\partial}{\partial \zeta'}\right)^2 \mathbf{H}^0 + \tfrac{1}{2}(\Delta H_{\xi'}^0) \frac{\partial}{\partial \zeta'} \nabla_T'(h_{\xi'}\zeta_0^2) + \left(\Delta \frac{\partial}{\partial \zeta'} H_{\xi'}^0\right) \zeta_0 \nabla_T'(h_{\xi'}\zeta_0) + (\Delta \delta H_{\xi'}) \nabla_T'(h_{\xi'}\zeta_0) + \zeta_0 \Delta \frac{\partial}{\partial \zeta'} \delta\mathbf{H} \right], \tag{4.14}$$

where the argument $\mathbf{r}_0'$ is understood on both sides of the equations.

This completes the formal development, but now we must investigate the validity of the results. A technical difficulty arises in connection with the quantity $\mathbf{F}_\epsilon(\mathbf{r}')$ of Eq. (4.4), which we have implicitly assumed to be defined for $\mathbf{r}'$ in V_p^0 even though it clearly has no physical existence there. Our assumption is conditionally justified in Appendix 2, where we show that, for $|\zeta'|$ small enough (compared to the local radii of curvature), the physical field has a well-defined and unique mathematical continuation.

[13] The derivation uses Eq. (1.3.6) of Ref. 6.

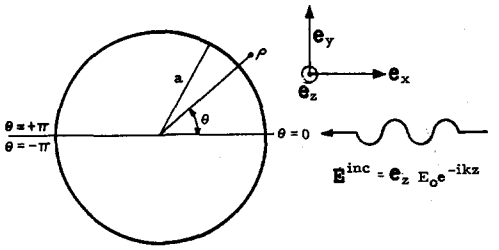


FIG. 2. Geometry of the unperturbed problem.

Additional limitations on the applicability of our results are apparent from the development. First of all, it is necessary that no significant sources of the unperturbed field lie in the volumes $V_1^\epsilon \cap V_2^0$ and $V_2^\epsilon \cap V_1^0$ through which the interface is perturbed. Also, since the difference in the unperturbed fields at S and S_0 must be small, it is necessary that the irregularity be small compared to the wavelengths in both media[14] and small compared to the local radii of curvature of S_0.[15] Lastly, the slope between corresponding points on S and S_0 must be small in order that the rotation and distortion of field lines at the boundary not be too great.

5. EXAMPLE. SCATTERING FROM A CYLINDER WITH SINUSOIDAL IRREGULARITIES

To illustrate the material developed above, we shall now calculate $\delta\mathbf{E}$ for the scattering of a plane wave from a perfectly conducting cylinder with sinusoidal irregularities.

The geometry of the unperturbed problem is shown in Fig. 2. A plane wave

$$\mathbf{E}^{inc} = \mathbf{e}_z E_0 e^{-ikz} = \mathbf{e}_z E_0 \sum_{n=-\infty}^{+\infty} J_n(k\rho) e^{in(\theta - \pi/2)} \tag{5.1}$$

is incident on the cylinder of radius a. The scattered field is

$$\mathbf{E}^{scat} = -\mathbf{e}_z E_0 \sum_{n=-\infty}^{+\infty} [H_n^{(1)}(ka)]^{-1} \times H_n^{(1)}(k\rho) J_n(ka) e^{in(\theta - \pi/2)}. \tag{5.2}$$

In the far zone this can be rewritten as

[14] If, however, one medium—say M_1—is quite lossy, then the wavelength in M_1 is of no consequence in determining whether the perturbation theory can be used in M_2.

[15] This restriction appears to be stronger than the one imposed by continuation considerations.

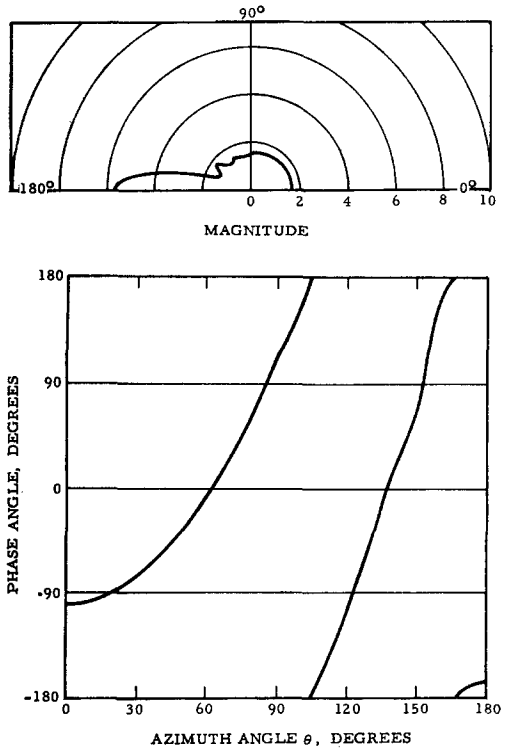


FIG. 3. Magnitude and phase of the unperturbed field pattern $M(\theta, 6)$.

$$\mathbf{E}^{\text{scat}} = \mathbf{e}_z E_0(k\rho)^{-\frac{1}{2}} e^{i(k\rho - \pi/4)} M(\theta, ka), \tag{5.3}$$

where M is the field pattern

$$M(\theta, ka) = -\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \sum_{n=-\infty}^{+\infty} [H_n^{(1)}(ka)]^{-1} J_n(ka) e^{in(\theta - \pi)}, \tag{5.4}$$

an even function of θ. In Fig. 3, the magnitude and and phase of M are plotted for $ka = 6$; the phase curve is actually continuous, but it has been telescoped into a 360° range.

Now let us perturb the interface to

$$Q = (\rho - a) - \rho_0 = (\rho - a) - b \cos(p\theta + \psi) = 0, \tag{5.5}$$

with b small compared to both the wavelength and the radius and p of necessity an integer.

An equation for $\delta\mathbf{E}$ can be obtained using Eqs. (3.25) and (4.11) and Part C of Appendix 1. Taking advantage of the facts that the cylinder is a perfect conductor and that we have essentially a two-dimensional scalar problem, we find

$$\delta\mathbf{E} = \mathbf{e}_z(i\omega\mu a/k^2) \int_{-\pi}^{\pi} d\theta [\mathbf{e}_{\theta'} \cdot \delta\mathbf{K}_m(\mathbf{r}_0')] G(\mathbf{r}; \mathbf{r}_0'), \tag{5.6}$$

with

$$\mathbf{e}_{\theta'} \cdot \delta\mathbf{K}_m = -\rho_0 \, \partial E_z^0/\partial\rho' = \frac{2ib}{\pi a} E_0 \cos(p\theta' + \psi) \times \sum_{n=-\infty}^{+\infty} [H_n^{(1)}(ka)]^{-1} e^{in(\theta' - \pi/2)} \tag{5.7}$$

and

$$G(\mathbf{r}; \mathbf{r}_0') = \int_{-\infty}^{\infty} dz' \, \mathbf{e}_z \cdot \nabla \times \Gamma_m^{\natural}(\mathbf{r}; \mathbf{r}_0') \cdot \mathbf{e}_{\theta'} = \frac{k^2}{2\pi i\omega\mu a} \sum_{n=-\infty}^{+\infty} [H_n^{(1)}(ka)]^{-1} H_n^{(1)}(k\rho) e^{in(\theta - \theta')}. \tag{5.8}$$

The integration in Eq. (5.6) gives, after some rearrangement,

$$\delta\mathbf{E} = \mathbf{e}_z(b/a) E_0 [P_e \cos\psi + P_o \sin\psi], \tag{5.9}$$

with

$$\left.\begin{matrix} P_e \\ P_o \end{matrix}\right\} = \frac{2i}{\pi} \sum_{n=-\infty}^{+\infty} [H_n^{(1)}(ka) H_{n+p}^{(1)}(ka)]^{-1} \times H_n^{(1)}(k\rho) e^{-i(n+p)\pi/2} \begin{cases} \cos n\theta \\ \sin n\theta \end{cases}. \tag{5.10}$$

In the case $\psi = 0$, we find in the far zone

$$\delta\mathbf{E} = \mathbf{e}_z E_0 (k\rho)^{-\frac{1}{2}} e^{i(k\rho - \pi/4)} [(b/a) M^{\epsilon}(\theta, ka; p)], \tag{5.11}$$

where $(b/a)M^{\epsilon}$ is the field pattern analogous to M

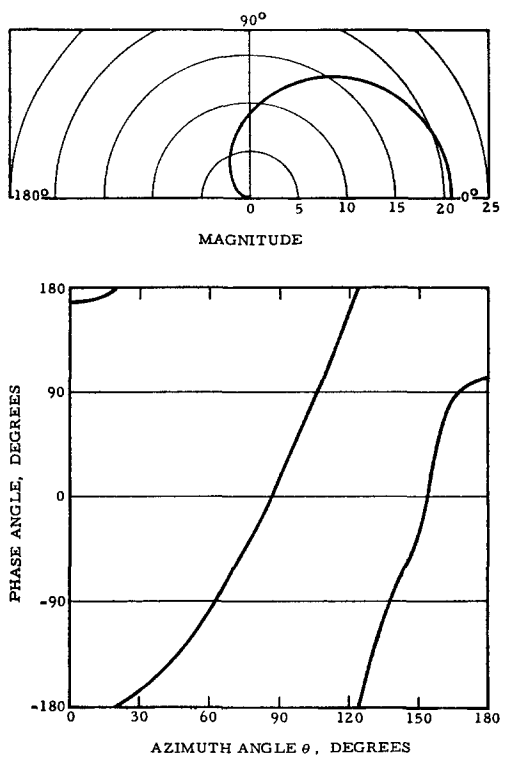


FIG. 4. Magnitude and phase of the perturbation field pattern $M^{\epsilon}(\theta, 6; 1)$.

of Eq. (5.4) and

$$M^e(\theta, ka; p) = i\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \times \sum_{n=-\infty}^{+\infty} [H_n^{(1)}(ka)H_{n+p}^{(1)}(ka)]^{-1} e^{-i(n+\frac{1}{2}p)\pi} \cos n\theta. \quad (5.12)$$

The series for M^e can also be expressed in the more rapidly convergent forms

$$M^e(\theta, ka; 2s+1) = 2\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \cos[(s+\tfrac{1}{2})\theta] \times \sum_{n=0}^{\infty} [A(s,n)]^{-1} \cos[(n+\tfrac{1}{2})\theta], \quad (5.13)$$

with

$$A(s,n) = (-1)^s H_{s+n+1}^{(1)}(ka)H_{s-n}^{(1)}(ka), \quad n \le s, \quad (5.14)$$
$$= (-1)^n H_{s+n+1}^{(1)}(ka)H_{n-s}^{(1)}(ka), \quad n \ge s+1;$$

and

$$M^e(\theta, ka; 2s) = (-1)^s\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \cos s\theta\left\{[H_s^{(1)}(ka)]^{-2} + 2\sum_{n=1}^{\infty} [B(s,n)]^{-1} \cos n\theta\right\}, \quad (5.15)$$

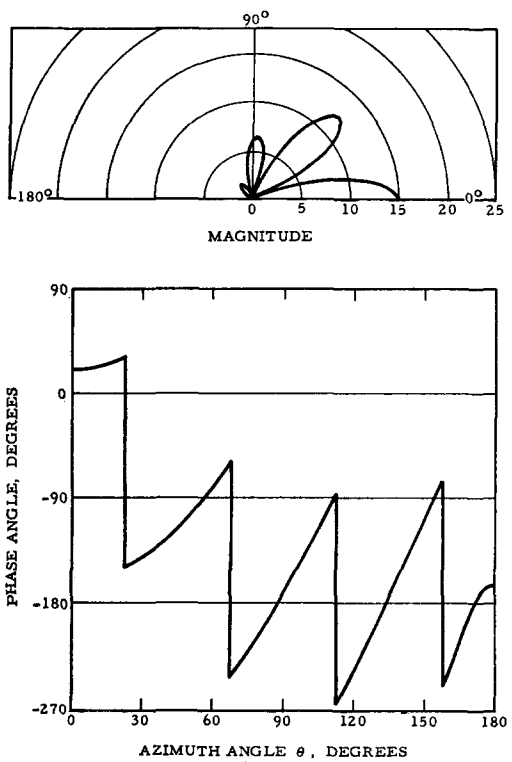


FIG. 5. Magnitude and phase of the perturbation field pattern $M^e(\theta, 6; 8)$.

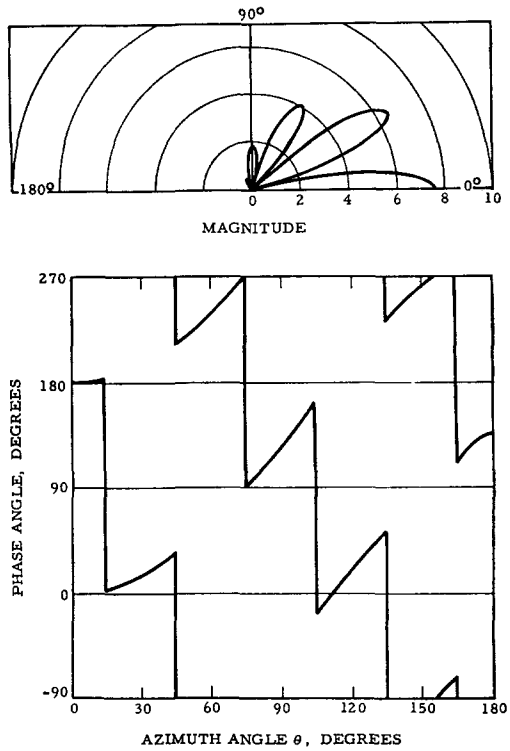


FIG. 6. Magnitude and phase of the perturbation field pattern $M^e(\theta, 6; 12)$.

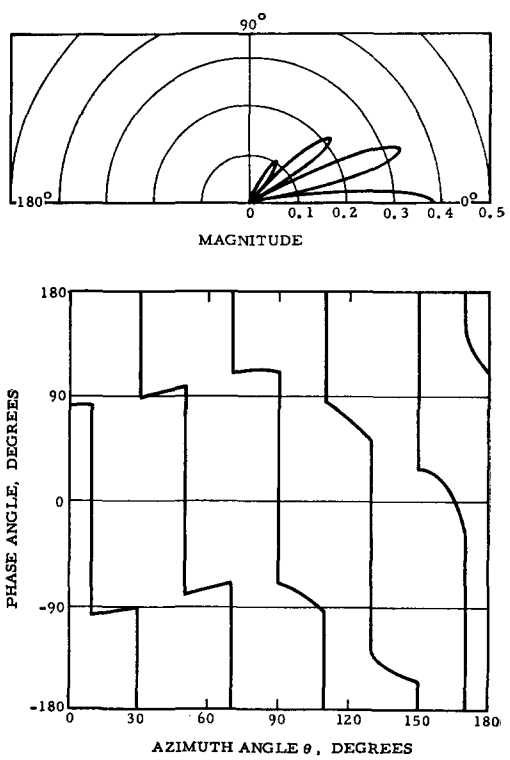


FIG. 7. Magnitude and phase of the perturbation field pattern $M^e(\theta, 6; 18)$.

with

$$B(s,n) = H^{(1)}_{s+n}(ka)H^{(1)}_{s-n}(ka), \quad n \leq s,$$
$$= (-1)^{n-s}H^{(1)}_{s+n}(ka)H^{(1)}_{n-s}(ka), \quad n \geq s+1. \quad (5.16)$$

The rearranged series show clearly that the perturbation field has p evenly spaced nulls and that there is a phase discontinuity of 180° at each null except the one at $\pm 180°$ for p odd.

In Figs. 4–7, the magnitude and phase of M' are plotted for $ka = 6$ and $p = 1, 8, 12, 18$. In all cases, the magnitude is greatest for backscattering ($\theta = 0°$); furthermore, for all but the largest p considered, the magnitude of M' is greater than that of M at $\theta = 0°$. Thus, if p is small enough, a small surface perturbation produces a relatively large change in the backscattered field.

6. CONCLUDING REMARKS

We have presented here an approximate method for determining the effect of irregularities of small slope and amplitude on the scattering from an interface of arbitrary underlying shape. The analysis has been carried out to second order in the maximum irregularity amplitude ϵ. An extension to higher-order perturbation terms is straightforward but tedious.

Problems involving somewhat larger irregularities could be treated by a modification of our method in which the effective surface currents are determined by solving integral equations rather than by expanding in ϵ. In such an approach, the calculation of the perturbation fields from the effective surface currents would still be carried out using the $\Gamma^{\parallel}$'s of the unperturbed problem.

Indeed, the Γ's and $\Gamma^{\parallel}$'s are useful in a wide range of scattering problems. This consideration, as well as the need in applications of our method for an understanding of the d.G.f and for information about specific d.G.f.'s, has led us to emphasize the theory of the d.G.f. in our development.

Some further details of the material of this paper have been given elsewhere.[16] They include a verification of the agreement of the results in Refs. 2 and 3 with the present theory and a discussion of statistically irregular interfaces.

APPENDIX 1. EXPRESSIONS FOR Γ AND RELATED FUNCTIONS FOR SOME IMPORTANT GEOMETRIES

We compile here some of the more important d.G.f.'s and related quantities. To avoid unnecessary repetition, we note that an expression for Γ_m can always be obtained by replacing $i\omega\mu$ by $(-k^2/i\omega\mu)$ in an expression for Γ_e. Where the d.G.f. is of complicated form, we shall give $\Gamma^{\parallel}$ and $\nabla \times \Gamma^{\parallel}$ rather than the complete d.G.f.; these functions suffice for calculation of the perturbation fields.

The d.G.f.'s for an unbounded region and for a half-space with perfectly conducting boundary have been taken from Ref. 8. The $\Gamma^{\parallel}_e$ and $\nabla \times \Gamma^{\parallel}_e$ for the general plane, cylindrical, and spherical boundaries have been found using known mode expansions of the field[17] and matching the discontinuity in tangential $\mathbf{H}$ at the interface; the straightforward but tedious details will be omitted. Results are presented in order of increasing geometric complexity.

A. Unbounded Medium

Here

$$\Gamma_e(\mathbf{r};\mathbf{r}') = \Gamma^0_e(\mathbf{r};\mathbf{r}') = i\omega\mu[I - (1/k^2)\nabla\nabla']G_f(\mathbf{r};\mathbf{r}'), \quad (A1.1)$$

where G_f is the scalar Green's function

$$G_f(\mathbf{r};\mathbf{r}') = \frac{1}{4\pi\,|\mathbf{r}-\mathbf{r}'|}\exp(ik\,|\mathbf{r}-\mathbf{r}'|). \quad (A1.2)$$

B. Half-Space

If the half-space $z > 0$ is bounded by a perfect electric conductor, then

$$\Gamma_e(\mathbf{r};\mathbf{r}') = \Gamma^0_e(\mathbf{r};\mathbf{r}') - \Gamma^0_e(\mathbf{r};\mathbf{r}' - 2\mathbf{e}_z\cdot\mathbf{r}'\mathbf{e}_z)\cdot(I - 2\mathbf{e}_z\mathbf{e}_z), \quad (A1.3)$$

$$\Gamma_m(\mathbf{r};\mathbf{r}') = \Gamma^0_m(\mathbf{r};\mathbf{r}') + \Gamma^0_m(\mathbf{r};\mathbf{r}' - 2\mathbf{e}_z\cdot\mathbf{r}'\mathbf{e}_z)\cdot(I - 2\mathbf{e}_z\mathbf{e}_z). \quad (A1.4)$$

In accordance with Eq. (3.9),

$$\Gamma^{\parallel}_e(\mathbf{r};\mathbf{r}') \equiv 0; \quad (A1.5)$$

we also find

$$\Gamma^{\parallel}_m(\mathbf{r};\mathbf{r}') = 2\Gamma^0_m(\mathbf{r};\mathbf{r}')\cdot(\mathbf{e}_x\mathbf{e}_x + \mathbf{e}_y\mathbf{e}_y), \quad z > 0. \quad (A1.6)$$

The image method used to obtain these equations can be used to find in closed form the d.G.f. for any wedge of angle π/n with a perfectly conducting boundary.

When the half-space $z < 0$ is filled with medium M_1 and the half-space $z > 0$ is filled with medium M_2, then for $\mathbf{r}$ in medium M_q we have

[16] K. M. Mitzner, California Institute of Technology Antenna Laboratory Report No. 30 (1964), Part III (unpublished).

[17] See, for example, Stratton, Ref. 12, Secs. 7.2, 7.11, and 9.29. The expansions to be given below for $\Gamma^{\parallel}$ are not quite correct at $\mathbf{r} = \mathbf{r}'$, but must there be augmented by the irrotational part of the δ-function source.

$$\Gamma^{\parallel}_{e,q}(\mathbf{r};\mathbf{r}') = L_q T_q, \quad \nabla \times \Gamma^{\parallel}_{e,q}(\mathbf{r};\mathbf{r}') = L_q W_q. \tag{A1.7}$$

Here L_q is the operator

$$L_q = \frac{\omega\mu_1\mu_2}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} du\,dv \times \exp\{iu(x-x') + iv(y-y') + ih_q z\} \times \frac{[\cdots]}{(\mu_2 h_1 - \mu_1 h_2)(\mu_2 h_2 k_1^2 - \mu_1 h_1 k_2^2)}, \tag{A1.8}$$

and

$$\begin{aligned} T_q = {} & \mathbf{e}_x\mathbf{e}_x[\mu_2 h_2(k_1^2 - u^2) - \mu_1 h_1(k_2^2 - u^2)] \\ & + \mathbf{e}_x\mathbf{e}_y uw(\mu_1 h_1 - \mu_2 h_2) \\ & + \mathbf{e}_y\mathbf{e}_x uw(\mu_1 h_1 - \mu_2 h_2) \\ & + \mathbf{e}_y\mathbf{e}_y[\mu_2 h_2(k_1^2 - v^2) - \mu_1 h_1(k_2^2 - v^2)] \\ & - \mathbf{e}_x\mathbf{e}_x uh_1 h_2(\mu_2 h_1 - \mu_1 h_2)/h_q \\ & - \mathbf{e}_x\mathbf{e}_y vh_1 h_2(\mu_2 h_1 - \mu_1 h_2)/h_q, \end{aligned} \tag{A1.9}$$

$$\begin{aligned} W_q = {} & \mathbf{e}_x\mathbf{e}_x iw\mu_q(k_2^2 - k_1^2) \\ & + \mathbf{e}_x\mathbf{e}_y[k_q^2 h_1 h_2(\mu_2 h_1 - \mu_1 h_2) + \mu_q h_q u^2(k_2^2 - k_1^2)]/(ih_q) \\ & - \mathbf{e}_y\mathbf{e}_x[k_q^2 h_1 h_2(\mu_2 h_1 - \mu_1 h_2) + \mu_q h_q v^2(k_2^2 - k_1^2)]/(ih_q) \\ & - \mathbf{e}_y\mathbf{e}_y iw\mu_q(k_2^2 - k_1^2) + \mathbf{e}_x\mathbf{e}_x iv(\mu_1 h_1 k_2^2 - \mu_2 h_2 k_1^2) \\ & - \mathbf{e}_x\mathbf{e}_y iu(\mu_1 h_1 k_2^2 - \mu_2 h_2 k_1^2), \end{aligned} \tag{A1.10}$$

with

$$\begin{aligned} h_1 &= -[k_1^2 - (u^2 + v^2)]^{\frac{1}{2}}, \\ h_2 &= +[k_2^2 - (u^2 + v^2)]^{\frac{1}{2}}. \end{aligned} \tag{A1.11}$$

C. Circular Cylinder

When the interior of the circular cylinder $\rho = a$ is filled with medium M_1 and the exterior is of medium M_2, then for $\mathbf{r}$ in medium M_q we have

$$\begin{aligned} \Gamma^{\parallel}_{e,q}(\mathbf{r};\mathbf{r}') &= L[\mathbf{s}_q(\rho)\ \mathbf{c}_q + k_q^{-1}\mathbf{t}_q(\rho)\mathbf{d}_q], \\ \nabla \times \Gamma^{\parallel}_{e,q}(\mathbf{r};\mathbf{r}') &= L[k_q\mathbf{s}_q(\rho)\mathbf{d}_q + \mathbf{t}_q(\rho)\mathbf{c}_q]. \end{aligned} \tag{A1.12}$$

Here L is the operator

$$\begin{aligned} L = {} & \frac{i\omega k_1 k_2 \mu_1\mu_2}{4\pi^2}\sum_{n=-\infty}^{+\infty}\int_{-\infty}^{+\infty} dh \\ & \times \exp\{in(\theta - \theta') + ih(z - z')\}[v_1^2 v_2^2 J_n(v_1)H_n^{(1)}(v_2)]^{-2} \\ & \times \left\{n^2h^2\left(\frac{1}{v_1^2} - \frac{1}{v_2^2}\right)^2 - \left[\frac{\mu_1}{v_1}\frac{J_n'(v_1)}{J_n(v_1)} - \frac{\mu_2}{v_2}\frac{H_n^{(1)\prime}(v_2)}{H_n^{(1)}(v_2)}\right]\right. \\ & \left.\times\left[\frac{k_1^2}{\mu_1 v_1}\frac{J_n'(v_1)}{J_n(v_1)} - \frac{k_2^2}{\mu_2 v_2}\frac{H_n^{(1)\prime}(v_2)}{H_n^{(1)}(v_2)}\right]\right\}^{-1}[\cdots]; \end{aligned} \tag{A1.13}$$

and

$$\mathbf{s}_q(\rho) = (in/\rho)Z_n^q(\lambda_q\rho)\mathbf{e}_\rho - \frac{\partial}{\partial\rho}Z_n^q(\lambda_q\rho)\mathbf{e}_\theta; \tag{A1.14}$$

$$\begin{aligned} \mathbf{t}_q(\rho) = {} & ih\frac{\partial}{\partial\rho}Z_n^q(\lambda_q\rho)\mathbf{e}_\rho \\ & - \frac{nh}{\rho}Z_n^q(\lambda_q\rho)\mathbf{e}_\theta + \lambda_q^2 Z_n^q(\lambda_q\rho)\mathbf{e}_z; \end{aligned} \tag{A1.15}$$

$$\begin{aligned} \mathbf{c}_1 = {} & \left\{\frac{n^2h^2(v_1^2 - v_2^2)}{k_1k_2\mu_2}J_n(v_1)[H_n^{(1)}(v_2)]^2\right. \\ & - \frac{k_2v_1^2v_2^2}{k_1\mu_2}J_n(v_1)[H_n^{(1)\prime}(v_2)]^2 \\ & \left. + \frac{k_1v_1v_2^3}{k_2\mu_1}J_n'(v_1)H_n^{(1)}(v_2)H_n^{(1)\prime}(v_2)\right\}a\mathbf{e}_\theta, \\ & + \frac{nhv_2^2(v_2^2 - v_1^2)}{k_1k_2\mu_2}J_n(v_1)[H_n^{(1)}(v_2)]^2\mathbf{e}_z; \\ \mathbf{c}_2 = {} & \left\{\frac{n^2h^2(v_1^2 - v_2^2)}{k_1k_2\mu_1}[J_n(v_1)]^2H_n^{(1)}(v_2)\right. \\ & + \frac{k_1v_1^2v_2^2}{k_2\mu_1}[J_n'(v_1)]^2H_n^{(1)}(v_2) \\ & \left. - \frac{k_2v_1^3v_2}{k_1\mu_2}J_n(v_1)J_n'(v_1)H_n^{(1)\prime}(v_2)\right\}a\mathbf{e}_\theta, \\ & + \frac{nhv_1^2(v_2^2 - v_1^2)}{k_1k_2\mu_1}[J_n(v_1)]^2H_n^{(1)}(v_2)\mathbf{e}_z; \end{aligned} \tag{A1.16}$$

$$\mathbf{d}_1 = (v_2^2/k_2)H_n^{(1)}(v_2)\mathbf{d}_0, \quad \mathbf{d}_2 = (v_1^2/k_1)J_n(v_1)\mathbf{d}_0; \tag{A1.17}$$

$$\begin{aligned} \mathbf{d}_0 = {} & \left[\frac{v_1}{\mu_2}J_n'(v_1)H_n^{(1)}(v_2) - \frac{v_2}{\mu_1}J_n(v_1)H_n^{(1)\prime}(v_2)\right]anh\mathbf{e}_\theta, \\ & + \left[\frac{v_1}{\mu_1}J_n(v_1)H_n^{(1)\prime}(v_2) - \frac{v_2}{\mu_2}J_n'(v_1)H_n^{(1)}(v_2)\right]v_1v_2\mathbf{e}_z; \end{aligned}$$

$$v_q = a\lambda_q = +a(k_q^2 - h^2)^{\frac{1}{2}}; \tag{A1.19}$$

$$\begin{aligned} Z_n^1 &= J_n \text{ (Bessel function)}; \\ Z_n^2 &= H_n^{(1)} \text{ (Hankel function)}; \end{aligned} \tag{A1.20}$$

$$Z'(v) = dZ(v)/dv. \tag{A1.21}$$

D. Sphere

When the interiod of the sphere $r = a$ is filled with medium M_1 and the exterior is of medium M_2, then for $\mathbf{r}$ in medium M_q we have

$$\begin{aligned} \Gamma^{\parallel}_{e,q}(\mathbf{r};\mathbf{r}') = {} & L[c_0^{-1}\mathbf{m}_{q,l,m}(\mathbf{r})\mathbf{m}_{p,l,-m}(\mathbf{r}') \\ & + d_0^{-1}\mathbf{n}_{q,l,m}(\mathbf{r})\hat{\mathbf{n}}_{p,l,-m}(\mathbf{r}')], \end{aligned} \tag{A1.22}$$

$$\begin{aligned} \nabla \times \Gamma^{\parallel}_{e,q}(\mathbf{r};\mathbf{r}') = {} & L[c_0^{-1}\mathbf{n}_{q,l,m}(\mathbf{r})\mathbf{m}_{p,l,-m}(\mathbf{r}') \\ & + d_0^{-1}k_q^2\mathbf{m}_{q,l,m}(\mathbf{r})\hat{\mathbf{n}}_{p,l,-m}(\mathbf{r}')]. \end{aligned} \tag{A1.23}$$

Here L is the operator

$$L = \frac{i\omega\mu_1\mu_2}{a}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{(-1)^m}{l(l+1)}[\cdots]; \tag{A1.24}$$

and

$$\begin{aligned} \mathbf{m}_{q,l,m}(\mathbf{r}) = {} & z_l^q(\rho)\left[\mathbf{e}_\theta\frac{im}{\sin\theta}Y_{l,m}(\theta,\phi)\right. \\ & \left. - \mathbf{e}_\phi\frac{d}{d\theta}Y_{l,m}(\theta,\phi)\right], \\ \hat{\mathbf{n}}_{q,l,m}(\mathbf{r}) = {} & \frac{1}{r}[\rho z_l^q(\rho)]'\left[\mathbf{e}_\theta\frac{d}{d\theta}Y_{l,m}(\theta,\phi)\right. \\ & \left. + \mathbf{e}_\phi\frac{im}{\sin\theta}Y_{l,m}(\theta,\phi)\right], \end{aligned}$$

$$\mathbf{n}_{q,l,m}(\mathbf{r}) = \mathbf{e}_r \frac{l(l+1)}{r} z_l^q(\rho) Y_{l,m}(\theta, \phi) + \hat{\mathbf{n}}_{q,l,m}(\mathbf{r}); \tag{A1.25}$$

$$c_0 = \mu_2 h_l^{(1)}(\rho_2)[\rho_1 j_l(\rho_1)]' - \mu_1 j_l(\rho_1)[\rho_2 h_l^{(1)}(\rho_2)]',$$

$$d_0 = \mu_1 k_2^2 h_l^{(1)}(\rho_2)[\rho_1 j_l(\rho_1)]' - \mu_2 k_1^2 j_l(\rho_1)[\rho_2 h_l^{(1)}(\rho_2)]'; \tag{A1.26}$$

$$\rho = k_q r, \qquad \rho_q = k_q a,$$

$$[\rho z_l(\rho)]' = \frac{\partial}{\partial \rho}[\rho z_l(\rho)]. \tag{A1.27}$$

The radial functions are

$$z_l^1 = j_l \text{ (spherical Bessel function)},$$

$$z_l^2 = h_l^{(1)} \text{ (spherical Hankel function)}, \tag{A1.28}$$

and the $Y_{l,m}$ are the spherical harmonics of angular momentum theory,[18] which can be expressed in terms of the more common associated Legendre functions by

$$Y_{l,m}(\theta, \phi) = (-1)^m \left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right]^{\frac{1}{2}} P_l^m(\cos\theta) e^{im\phi}. \tag{A1.29}$$

APPENDIX 2. CONTINUATION OF FIELDS

If the formal development in Sec. 4 is to be valid, then it is necessary that everywhere on S the field vectors $\mathbf{F}_q^0$, $\delta\mathbf{F}_q$, and $\delta^2\mathbf{F}_q$ satisfy equations of form (4.4) with remainder of order ϵ^3. It is therefore necessary that these fields, which exist physically only in V_q^0, have unique mathematical continuations through $V_q^\epsilon \cap V_p^0$ to S. Furthermore, since the continuations are to be expressed in Taylor series, they cannot have any singularities in $V_q^\epsilon \cap V_p^0$.

The existence of field continuations is readily seen in the problem of reflection of a plane wave from a plane interface and in any problem which can be solved by images. Another example is the conducting half-plane problem, in which the field can be continued across the plane onto another sheet of the Riemann surface, but no continuation is possible across the knife edge (at which the field is singular).

In most cases it is difficult to exhibit the continuation, and thus it is desirable to have general rules which aid us in determining when it exists. To this end, let us consider a continuous deformation of S_0 accomplished by fixing w_0 in Eq. (4.1) and increasing ϵ from zero to ϵ^*, where ϵ^* bounds the set of ϵ for which $\mathbf{F}_q$ has a nonsingular continuation in $V_q^\epsilon \cap V_p^0$. For $\epsilon < \epsilon^*$, it must be possible to establish $\mathbf{F}_q$ by the combination of a volume source distribution independent of ϵ (needed only for F_q^0) and a surface current distribution on S; for ϵ just greater than ϵ^*, it is necessarily impossible to establish $\mathbf{F}_q$ in this manner. Thus the critical condition determining ϵ^* is the necessity of introducing a new volume source into $V_q^\epsilon \cap V_p^0$. Since the source cannot suddenly appear in the interior of the region, the critical condition is characterized by the field being singular at some point on S.

We immediately see that any $\mathbf{F}_q$ which is not singular on S_0 can be continued uniquely some distance into V_p^0, the continuation being the field produced in $V_q^\epsilon \cap V_p^0$ by sources on S which produce $\mathbf{F}_q$ in $V_q^\epsilon \cap V_q^0$. On the other hand, no continuation at all is possible at a knife edge or other geometrical feature at which the field is singular.

To proceed further, we must ask why it eventually becomes necessary to introduce a source into $V_q^\epsilon \cap V_p^0$. The answer is that the spatial variation of the field somewhere on S_0 is so rapid that it cannot be maintained by sources further removed. Thus we are led to the following important generalization:

> The more rapid the spatial variation of the field near a point on S_0, the smaller the depth is to which the field can be continued near that point.

An important corollary of this principle is that the depth of continuation will be small where the radii of curvature are small. Another corollary is that the depth of continuation will usually decrease as a source distribution is brought closer to S_0.

Furthermore, $\delta\mathbf{F}_q$ will usually represent finer details of the field than will $\mathbf{F}_q^0$, and thus $\delta\mathbf{F}_q$ will tend to have the more rapid spatial variation. Likewise, $\delta^2\mathbf{F}_q$ will tend to vary more rapidly than $\delta\mathbf{F}_q$, and so on for higher order terms. Thus, as the order of perturbation increases, the depth of continuation tends to decrease. This suggests that the perturbation technique may give an asymptotic approximation to the field rather than a convergent one.

ACKNOWLEDGMENTS

I am indebted to Professor C. H. Papas for his guidance and advice in the course of this research. Thanks are also due J. Lo, who assisted in the machine computation of numerical results. Lastly, I wish to express my gratitude to the Hughes Aircraft Company for the generous support I received as a Howard Hughes Fellow.

[18] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1960), 2nd ed., Sec. 2.5.

Comparison of Two Methods for Lower Bounds to Eigenvalues*

WOLFGANG BÖRSCH-SUPAN
Institut für Angewandte Mathematik der Universität, Mainz, Germany
(Received 7 August 1964)

It is shown that two procedures given by Bazley and Fox for the calculation of lower bounds to eigenvalues of self-adjoint operators are essentially equivalent.

RECENTLY Bazley and Fox gave several methods for estimating the eigenvalues of self-adjoint operators from below. Among these procedures there is the method of "truncation of the base operator"[1] and the method of "double projection."[2,3] Closer inspection shows that in both procedures almost the same information is used for construction of the lower bounds. The only difference is that in the method of truncation the $(l+1)$th eigenvalue λ_{l+1}^0 of the base operator shows up, whereas, instead of this, in the method of double projection a certain positive parameter γ is used which may be chosen freely. It was shown[2] that the lower bound for the νth eigenvalue of the problem is monotonically increasing in γ until it reaches the value $\lambda_{l+1}^0 - \gamma$, which leads to the conclusion that the best value of γ for estimating the νth eigenvalue is $\gamma = \lambda_{l+1}^0 - \lambda$, where λ is the νth eigenvalue of the comparison operator in the double projection method.

If we now modify this method by introducing the best value of γ, though this is not known in advance, then exactly the same information is used as in the method of truncation. Now it is natural to ask which of the methods makes better use of that information, i.e., which one leads to closer bounds. It will be shown in this note that both methods lead to exactly the same bounds. In the method of double projection, γ must be chosen from other information and is, in general, not the best value; thus the method of truncation turns out to be superior, as long as computational questions are not taken into consideration. However, observe that the double projection leads to a kth-order linear matrix problem, while the truncation in general leads to a $(k+l)$th-order linear problem.

We are going now to prove the above-mentioned equivalence. Notations are taken as far as possible from the papers referred to above.[1-3]

In the method of double projection one has to determine the eigenvalue λ and the eigenfunction u in the span $\{u_1^0, \cdots, u_l^0\}$ from the equation

$$[A^0 - \gamma + (A'P^k + \gamma)Q^l]u = \lambda u, \qquad (1)$$

where P^k is the A'-orthogonal projection[4] on the span $\{p_1, \cdots, p_k\}$ and Q^l is the $(A'P^k + \gamma)$-orthogonal projection on the span $\{q_1, \cdots, q_l\}$ with q_i defined[5] by $(A'P^k + \gamma)q_i = u_i^0$ for $i = 1, \cdots, l$; i.e., for any function v, $Q^l v$ is the (uniquely determined) element of the span $\{q_1, \cdots, q_l\}$ with the property that $v - Q^l v$ is orthogonal to $u_1^0, \cdots, u_l^0$.

On the other hand, in the method of truncation, the eigenvalue λ and the eigenfunction w in the span $\{u_1^0, \cdots, u_l^0, A'p_1, \cdots, A'p_k\}$ are defined to be solutions of

$$[A^0T^l + \lambda_{l+1}^0(I - T^l) + A'P^k]w = \lambda w, \qquad (2)$$

where I is the identity and T^l the orthogonal projection on the span $\{u_1^0, \cdots, u_l^0\}$; i.e., for any function v, $T^l v$ is the (uniquely determined) element of the span $\{u_1^0, \cdots, u_l^0\}$ such that $v - T^l v$ is orthogonal to $u_1^0, \cdots, u_l^0$.

Let now u and λ form a solution of (1) for a positive value of γ which satisfies $\gamma = \lambda_{l+1}^0 - \lambda$. This condition can be fulfilled because of the monotonicity properties of the operator in Eq. (1) as shown by Bazley and Fox.[2] Define $w = Q^l u$. Then Eq. (1) becomes

$$(A^0 - \lambda_{l+1}^0)u + (A'P^k + \lambda_{l+1}^0 - \lambda)w = 0, \qquad (3)$$

where w, being in the span of the q_i, by definition of the q_i, is also in the span of the u_i^0 and the $A'p_i$. Now, u is in the span $\{u_1^0, \cdots, u_l^0\}$ and, by definition of Q^l, $u - w$ is orthogonal to $u_1^0, \cdots, u_l^0$. Hence, by definition of T^l, $u = T^l w$, and from Eq. (3) follows

* This work was sponsored in part by the NATO Research Grants Programme under the number SA 5-2-05(170)/278 (64)HH.

[1] N. W. Bazley and D. W. Fox, J. Res. Nat. Bur. Std. **65B**, 105(1961).

[2] N. W. Bazley and D. W. Fox, J. Math. Phys. **3**, 469(1962).

[3] It is assumed that the reader is familiar with the contents of both papers (Refs. 1 and 2).

[4] For the sake of simplicity, we assume that A' is positive definite.

[5] Since $\gamma > 0$, $A'P^k + \gamma$ is positive definite, and q^i can be defined this way.

Eq. (2). On the other hand let w and λ form a solution of Eq. (2) with $\lambda < \lambda^0_{l+1}$. Define $u = T^l w$ and $\gamma = \lambda^0_{l+1} - \lambda$. Then Eq. (3) holds, and, according to this, $(A'P^k + \gamma)w$ is spanned by $u^0_1, \cdots, u^0_l$, and, therefore, w by $q_1, \cdots, q_l$. Since, by definition of T^l, $w - u$ is orthogonal to $u^0_1, \cdots, u^0_l$, it follows from the definition of Q^l that $w = Q^l u$, and, as u is in the span of the u^0_i, (1) follows from Eq. (3).

By these arguments it is shown that both methods provide the same eigenvalues smaller than λ^0_{l+1}, and that their respective eigenfunctions obey the relations $u = T^l w$ and $w = Q^l u$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Generalized Perturbation Expansion for the Klein–Gordon Equation with a Small Nonlinearity*

DAVID MONTGOMERY†
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
(Received 5 May 1964)

A previously given method for deriving secularity-free perturbation expansions for the Klein–Gordon equation with a "small" nonlinear term is generalized to include situations in which the lowest-order solution is not restricted to be a monochromatic wave.

I. INTRODUCTION

IT is the purpose of this paper to present a general method for developing well-behaved perturbation expansions in ϵ/λ^2 for the Klein–Gordon equation with a small nonlinear term:

$$\left(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2} + \lambda^2\right) f(x, t) = \epsilon F\left(f, \frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}\right), \qquad (1a)$$

where F is some arbitrary function of f, $\partial f/\partial t$, and $\partial f/\partial x$. Both f and F are real. For $\epsilon = 0$, the solution to (1a) is easily given as a Fourier series or integral, which is a superposition of waves of frequency ω and wavenumber κ related by the dispersion relation $\omega^2 = c^2\kappa^2 + \lambda^2$. However, a straightforward attempt to expand in powers of ϵ about the $\epsilon = 0$ solution leads to secular (i.e., t or x proportional) terms in the corrections to f, making more refined methods necessary. The physical reason for this phenomenon is that the equations which result from the perturbation theory are equivalent to the zeroth order equation, with the difference that there is an effective inhomogeneous driving term provided by the F. The various $\epsilon = 0$ wavenumbers and frequencies can be combined by the F to yield wavenumbers and frequencies which are in resonance with the zeroth-order normal modes. An expansion which avoids this difficulty for Eq. (1a) has recently been given,[1] adapting the Krylov–Bogoliubov–Mitropolskii techniques of nonlinear mechanics.[2,3] However, the treatment of Ref. 1 (which was designed for a specific plasma problem) suffers from one rather severe limitation on the zeroth-order solution: only a monochromatic wave led to manageable equations in the higher orders.

Here, this limitation is removed, and the treatment of Ref. 1 is generalized to include all situations in which the zeroth order is expressible as a Fourier series which is summed over a discrete spectrum of frequencies. In practice, we can avoid questions associated with the convergence of the Fourier series by limiting ourselves to boundary conditions and F's which lead to only a finite number of terms in f, to any given order in ϵ. The restriction to a discrete spectrum does not eliminate any interesting phenomena; the continuous spectrum case actually appears *less* pathological than the discrete one.

* This work was supported in part by the National Aeronautics and Space Administration, under Contract Ns G 220–62.

† Present address: Instituut voor Theoretische Fysica der Rijksuniversiteit, Utrecht, The Netherlands.

[1] D. Montgomery and D. A. Tidman, Phys. Fluids 7, 242 (1964). A more complete bibliography is given in this reference.

[2] An elegant approach to the wave equation using similar methods has been given by M. D. Kruskal and N. J. Zabusky, J. Math. Phys. 5, 231 (1964). The perturbation series for the wave equation, however, possesses much more pathological behavior than in the case of the Klein–Gordon equation.

[3] N. Bogolyubov and Y. A. Mitropolskii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon and Breach Science Publishers, New York, 1961; translated from the Russian).

Eq. (2). On the other hand let w and λ form a solution of Eq. (2) with $\lambda < \lambda^0_{l+1}$. Define $u = T^l w$ and $\gamma = \lambda^0_{l+1} - \lambda$. Then Eq. (3) holds, and, according to this, $(A'P^k + \gamma)w$ is spanned by $u^0_1, \cdots, u^0_l$, and, therefore, w by $q_1, \cdots, q_l$. Since, by definition of T^l, $w - u$ is orthogonal to $u^0_1, \cdots, u^0_l$, it follows from the definition of Q^l that $w = Q^l u$, and, as u is in the span of the u^0_i, (1) follows from Eq. (3).

By these arguments it is shown that both methods provide the same eigenvalues smaller than λ^0_{l+1}, and that their respective eigenfunctions obey the relations $u = T^l w$ and $w = Q^l u$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Generalized Perturbation Expansion for the Klein–Gordon Equation with a Small Nonlinearity*

DAVID MONTGOMERY†
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
(Received 5 May 1964)

A previously given method for deriving secularity-free perturbation expansions for the Klein–Gordon equation with a "small" nonlinear term is generalized to include situations in which the lowest-order solution is not restricted to be a monochromatic wave.

I. INTRODUCTION

IT is the purpose of this paper to present a general method for developing well-behaved perturbation expansions in ϵ/λ^2 for the Klein–Gordon equation with a small nonlinear term:

$$\left(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2} + \lambda^2\right) f(x, t) = \epsilon F\left(f, \frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}\right), \qquad (1a)$$

where F is some arbitrary function of f, $\partial f/\partial t$, and $\partial f/\partial x$. Both f and F are real. For $\epsilon = 0$, the solution to (1a) is easily given as a Fourier series or integral, which is a superposition of waves of frequency ω and wavenumber κ related by the dispersion relation $\omega^2 = c^2\kappa^2 + \lambda^2$. However, a straightforward attempt to expand in powers of ϵ about the $\epsilon = 0$ solution leads to secular (i.e., t or x proportional) terms in the corrections to f, making more refined methods necessary. The physical reason for this phenomenon is that the equations which result from the perturbation theory are equivalent to the zeroth order equation, with the difference that there is an effective inhomogeneous driving term provided by the F. The various $\epsilon = 0$ wavenumbers and frequencies can be combined by the F to yield wavenumbers and frequencies which are in resonance with the zeroth-order normal modes. An expansion which avoids this difficulty for Eq. (1a) has recently been given,[1] adapting the Krylov–Bogoliubov–Mitropolskii techniques of nonlinear mechanics.[2,3] However, the treatment of Ref. 1 (which was designed for a specific plasma problem) suffers from one rather severe limitation on the zeroth-order solution: only a monochromatic wave led to manageable equations in the higher orders.

Here, this limitation is removed, and the treatment of Ref. 1 is generalized to include all situations in which the zeroth order is expressible as a Fourier series which is summed over a discrete spectrum of frequencies. In practice, we can avoid questions associated with the convergence of the Fourier series by limiting ourselves to boundary conditions and F's which lead to only a finite number of terms in f, to any given order in ϵ. The restriction to a discrete spectrum does not eliminate any interesting phenomena; the continuous spectrum case actually appears *less* pathological than the discrete one.

* This work was supported in part by the National Aeronautics and Space Administration, under Contract Ns G 220–62.

† Present address: Instituut voor Theoretische Fysica der Rijksuniversiteit, Utrecht, The Netherlands.

[1] D. Montgomery and D. A. Tidman, Phys. Fluids 7, 242 (1964). A more complete bibliography is given in this reference.

[2] An elegant approach to the wave equation using similar methods has been given by M. D. Kruskal and N. J. Zabusky, J. Math. Phys. 5, 231 (1964). The perturbation series for the wave equation, however, possesses much more pathological behavior than in the case of the Klein–Gordon equation.

[3] N. Bogolyubov and Y. A. Mitropolskii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon and Breach Science Publishers, New York, 1961; translated from the Russian).

Many of the difficulties, given the idea of Ref. 1, are largely matters of bookkeeping. The only problem that is really essential is one of closure among the frequencies and wavenumbers of the $\epsilon = 0$ solution. By this we mean closure *only* among frequencies and wavenumbers which satisfy the $\epsilon = 0$ dispersion relation, closure among the other frequencies and wavenumbers being unimportant at any given finite order in ϵ. Whether such closure occurs can, unhappily, only be determined for each case individually.

The solution to (1a) is determined completely, as is well known, by specifying f and its normal derivative along any line which cuts all the characteristics once[4] [the characteristics of (1a) are the lines $\xi \equiv \frac{1}{2}\lambda(t + x/c) = \text{const.}$, $\eta \equiv \frac{1}{2}\lambda(t - x/c) = \text{const.}$]. Two such lines are the x and t axes, for instance, and to these two cases we confine our attention, asking for the solution to (1a) either above the x axis, or to the right of the t axis. The formalism can be specialized to cover both situations.

The function f and its normal derivative will be assumed periodic over the boundary. Since the method of images can often be used to replace a bounded problem by an equivalent unbounded one which is periodic,[5] these conditions include, for example, the case in which f and $\partial f/\partial t$ are given for $0 < x < L$ at $t = 0$, with the added requirement that $f = 0$ at $x = 0$ and $x = L$ for all t.

It turns out to be easier to work in the characteristic coordinates rather than x and t. Therefore, in Sec. II, the method is developed entirely in these coordinates. In Sec. III, the role of boundary conditions is discussed, and two simple examples are treated in Sec. IV.

In the interests of simplicity, we go only to $O(\epsilon)$ in the main body of this paper, although an example is carried to $O(\epsilon^2)$ in the Appendix. Also for simplicity, we confine ourselves to very simple forms for F, for the F's which occur in problems of genuine physical interest seem always to generate so much algebra that an understanding of the method becomes unnecessarily difficult.

II. THE METHOD

The substitutions

$$\xi = \tfrac{1}{2}\lambda(t + x/c), \quad \eta = \tfrac{1}{2}\lambda(t - x/c), \quad \varepsilon = \epsilon/\lambda^2,$$

$$F(f, \partial f/\partial t, \partial f/\partial x) = \mathfrak{F}(f, \partial f/\partial \xi, \partial f/\partial \eta),$$

reduce (1a) to the form

$$\left(\frac{\partial^2}{\partial \xi\, \partial \eta} + 1\right) f = \varepsilon \mathfrak{F}\left(f, \frac{\partial f}{\partial \xi}, \frac{\partial f}{\partial \eta}\right). \tag{1b}$$

We seek a solution to (1b) of the form

$$f = \sum_{K,L} a(K, L) e^{i\psi(K,L)} + \varepsilon \mu_1(a, \psi) + \varepsilon^2 \mu_2(a, \psi) + \cdots, \tag{2}$$

where a and ψ stand symbolically for all the amplitudes $a(K, L)$ and phases $\psi(K, L)$.

We must take some pains to identify the labels K, L; to do this requires several steps. Throughout, we must bear in mind that the zeroth-order part of (2) must coincide with the $\varepsilon = 0$ solution of Eqs. (1).

(i) We introduce an enumerable sequence of two-component *basic vectors* (k_i, l_i), the components of which satisfy

$$k_i l_i = 1, \quad \text{all} \quad j.$$

Referring to the $\epsilon = 0$ form of (1a), we find that the numbers $\lambda(k_i + l_i)/2$ are the allowed frequencies, and $\lambda(k_i - l_i)/2c$ are the allowed wavenumbers, for the $\epsilon = 0$ problem. These are determined by boundary or initial conditions in a way that need not concern us yet.

(ii) The notation (K, L) means the *derived vector*

$$(K, L) = n_1(k_1, l_1) + n_2(k_2, l_2) + \cdots + n_N(k_N, l_N),$$

where $n_1, n_2, \cdots, n_N$ are any collection of integers, positive, negative, or zero. Clearly, the basic vectors are also derived vectors. This sum is *formal*, in the sense of (iii) below.

(iii) If

$$(K, L) = \sum_i n_i(k_i, l_i)$$

and

$$(K', L') = \sum_i n_i'(k_i, l_i),$$

then (K, L) is regarded as the same vector as (K', L') if and only if

$$n_i = n_i', \quad \text{all} \quad j.$$

(iv) With each basic vector (k_i, l_i) there is associated a *basic phase* $\psi(k_i, l_i)$.

(v) For the derived vector (K, L), the phase $\psi(K, L)$ is defined by

$$\psi(K, L) = \psi\left(\sum_i n_i k_i, \sum_i n_i l_i\right) = \sum_i n_i \psi(k_i, l_i).$$

(vi) For the basic phases $\psi(k_i, l_i)$ there is assumed

[4] See, e.g., P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, Chap. 6.

[5] A quite similar recipe to that given on p. 686 of Ref. 4 for the wave equation can readily be given for the Klein–Gordon equation.

to be an expansion of the form

$$\begin{aligned} \partial\psi(k_i, l_i)/\partial\xi &= k_i + \varepsilon C_{k_i l_i}(a) + \cdots, \\ \partial\psi(k_i, l_i)/\partial\eta &= l_i + \varepsilon D_{k_i l_i}(a) + \cdots, \end{aligned} \tag{3}$$

where the coefficients C_{k_i,l_i}, D_{k_i,l_i} depend only upon the amplitudes, but remain otherwise unspecified, as yet. The "$\cdots$" means higher powers of ε.

These six statements determine the sort of expansion which is being sought for the ψ's. Note that (v) and (vi) require that for the derived phases,

$$\begin{aligned} &\partial\psi(K, L)/\partial\xi \\ &\quad = \sum_i n_i k_i + \varepsilon \sum_i n_i C_{k_i l_i}(a) + \cdots, \\ &\partial\psi(K, L)/\partial\eta \\ &\quad = \sum_i n_i l_i + \varepsilon \sum_i n_i D_{k_i l_i}(a) + \cdots. \end{aligned} \tag{4}$$

From Eqs. (4), we may see the necessity for Condition (iii) above. It will turn out that the $C_{k_i l_i}(a)$ and $D_{k_i l_i}(a)$ are determined by the form of F, the zeroth order solution, and the boundary conditions. From these, the $O(\varepsilon)$ parts of $\partial\psi(K, L)/\partial\xi$ and $\partial\psi(K, L)/\partial\eta$ follow completely from Eq. (4). Even if the zeroth-order parts of the right-hand sides of Eqs. (4) happen to add to the same value for two different values of K and L, there is no guarantee that the equality will persevere through $O(\varepsilon)$.

To finish specifying the sort of solution to (2) that we are seeking, we assume for the amplitudes an expansion of the form

$$\begin{aligned} \partial a(K, L)/\partial\xi &= \varepsilon A_{KL}(a) + \cdots, \\ \partial a(K, L)/\partial\eta &= \varepsilon B_{KL}(a) + \cdots. \end{aligned} \tag{5}$$

By the symbols k and l, we shall mean the numerical value of $n_1k_1 + n_2k_2 + \cdots + n_Nk_N$ and $n_1l_1 + n_2l_2 + \cdots + n_Nl_N$. Thus, (K, L) is a basic vector if and only if $kl = 1$, but different basic vectors may have the same values of k and l [see (iii) above].

It is no loss of generality to require that $a(K, L)$ and $\psi(K, L)$ be real, and that

$$\begin{aligned} a(K, L) &= a(-K, -L), \\ A_{KL}(a) &= A_{-K,-L}(a), \\ B_{KL}(a) &= B_{-K,-L}(a), \\ C_{k_i l_i}(a) &= -C_{-k_i,-l_i}(a), \\ D_{k_i l_i}(a) &= -D_{-k_i,-l_i}(a), \\ \psi(K, L) &= -\psi(-K, -L). \end{aligned} \tag{6}$$

For the solution to be consistent for $\varepsilon = 0$, we require that

$$a(K, L) = O(\varepsilon), \qquad kl \neq 1, \tag{7}$$

so that if (2) is written as

$$f = f^{(0)} + \varepsilon f^{(1)} + \varepsilon^2 f^{(2)} + \cdots, \tag{8}$$

then $f^{(0)}$ may be unambiguously written as

$$f^{(0)} = \sum_{kl=1;K,L} a(K, L)e^{i\psi(K,L)}.$$

With this set of notations, and for any $\mathfrak{F}$ which is expandable about zero in f, $\partial f/\partial\xi$, and $\partial f/\partial\eta$, we may always write

$$\mathfrak{F}(f^{(0)}, \partial f^{(0)}/\partial\xi, \partial f^{(0)}/\partial\eta) = \sum_{K,L} F_{KL}(a)e^{i\psi(K,L)}, \tag{9}$$

where the $F_{KL}(a)$ are *known* functions of the amplitudes which satisfy, for $\mathfrak{F}$ real,

$$F_{KL}(a) = F^*_{-K,-L}(a). \tag{10}$$

In practice, $\mathfrak{F}$ will be some polynomial if f, $\partial f/\partial\xi$, and $\partial f/\partial\eta$, and the boundary conditions will be such that the summation in (9) will contain only a finite number of terms.

Using Eqs. (2) through (9) in Eq. (1b), we get, upon noting that the zeroth-order coefficient vanishes identically, and equating coefficients of ε:

$$\begin{aligned} &(\partial^2/\partial\xi\,\partial\eta + 1)\mu_1(a, \psi) \\ &\quad + \sum_{kl\neq 1;K,L} (-kl + 1)a(K, L)e^{i\psi(K,L)} \\ &\quad + \sum_{k_i,l_i} e^{i\psi(k_i,l_i)}\{i(k_iB_{k_il_i} + l_iA_{k_il_i}) \\ &\quad - a(k_i, l_i)(k_iD_{k_il_i} + l_iC_{k_il_i})\} \\ &\quad = \sum_{K,L} F_{KL}(a)e^{i\psi(K,L)}. \end{aligned} \tag{11}$$

If we define

$$v_1 = \mu_1 + \sum_{kl\neq 1;K,L} a(K, L)e^{i\psi(K,L)},$$

then (11) becomes a differential equation for v_1:

$$\begin{aligned} &(\partial^2/\partial\xi\,\partial\eta + 1)v_1 = \sum_{K,L} F_{KL}(a)e^{i\psi(K,L)} \\ &\quad - \sum_{k_i,l_i} e^{i\psi(k_i,l_i)}\{i(k_iB_{k_il_i} + l_iA_{k_il_i}) \\ &\quad - a(k_i, l_i)(k_iD_{k_il_i} + l_iC_{k_il_i})\}. \end{aligned} \tag{12}$$

Equation (12) has a secularity-free solution for v_1 of the form

$$v_1 = \sum_{K,L} v^{(1)}(K, L)e^{i\psi(K,L)}$$

if and only if

$$\begin{aligned} &i(k_iB_{k_il_i} + l_iA_{k_il_i}) \\ &\quad - a(k_i, l_i)(k_iD_{k_il_i} + l_iC_{k_il_i}) = F_{k_il_i}(a) \end{aligned} \tag{13}$$

for all (k_i, l_i). Note that nowhere have we as yet been required to commit ourselves as to the values of $A_{k_il_i}$, $B_{k_il_i}$, $C_{k_il_i}$, $D_{k_il_i}$; we are free to *choose* them to satisfy (13). Only in this way can we avoid ξ or η proportional terms in v_1, i.e., "secularity."

Recalling (7), it is clear that the only $a(K, L)$ which contribute to the right-hand side of (13) are the $a(k_i, l_i)$. Equating real and imaginary parts of the left and right-hand sides of (13),

$$k_iB_{k_il_i}(a) + l_iA_{k_il_i}(a) = \text{Im}\,\{F_{k_il_i}(a)\},$$

$$a(k_i, l_i)(l_iC_{k_il_i}(a) + k_iD_{k_il_i}(a)) = -\text{Re}\,\{F_{k_il_i}(a)\}. \quad (14)$$

One consequence of (14) and (5) is worth noting right away, in connection with the closure problem: we can never have, simultaneously, $a(k_i, l_i) = 0$ and $F_{k_il_i}(a) \neq 0$ for any j. Thus all the various $\varepsilon = 0$ normal modes of the system, if coupled by the $\mathfrak{F}$, must be excited to $O(1)$. This leaves us with two possibilities: the subset of the basic vectors for which

$$F_{k_il_i}(a) \neq 0, \qquad a(k_il_i) \neq 0$$

is either finite of infinite. Naturally, the former case has many more calculational possibilities. The latter case is calculable when two circumstances happen to exist:

$$\text{Im}\,\{F_{k_il_i}(a)\} = 0, \quad \text{all} \quad j;$$

$$a(k_i, l_i) \neq 0, \quad \text{all} \quad j \quad \text{for which} \quad F_{k_il_i}(a) \neq 0.$$

It is regrettable that more satisfying general statements about when it is possible to achieve closure among the (k_i, l_i) cannot be made. This is due to the very general possibilities for $\mathfrak{F}$; to go much farther, we must specialize $\mathfrak{F}$, which we do in Sec. IV.

We close this section with a proof that the situation in Ref. 1, with only one monochromatic wave in zeroth order, always leads trivially to closure within the set of basic vectors. There, we had

$$f^{(0)} = a(k_0, l_0)e^{i\psi(k_0, l_0)} + a(-k_0, -l_0)e^{i\psi(-k_0, -l_0)}$$

initially, where $k_0l_0 = 1$, and all the other $a(k_i, l_i) = 0$. We will show that the vectors $\pm(k_0, l_0)$ cannot be combined by any form of $\mathfrak{F}$ to lead to a different basic vector. It is no problem, of course, that closure is not achieved among the derived vectors. The act of solving for $v^{(1)}(K, L)$ for (K, L) *outside* the set of basic vectors is a straightforward algebraic operation.

The proof is as follows. From

$$\mathfrak{F}(f^{(0)}, \partial f^{(0)}/\partial\xi, \partial f^{(0)}/\partial\eta),$$

we get terms of the type $F_{k_il_i}(a) \neq 0$ only if

$$k_i = (n_1 + n_2 + \cdots)k_0, \qquad l_i = (n_1 + n_2 + \cdots)l_0.$$

However, if (k_i, l_i) is to be a basic vector, we must have $k_il_i = 1$, or since $k_0l_0 = 1$,

$$(n_1 + n_2 + \cdots)^2 = 1$$

or

$$(k_i, l_i) = \pm(k_0, l_0),$$

so that we can never be led outside the set (k_0, l_0), $(-k_0, -l_0)$ by any form of $\mathfrak{F}$.

Note added in proof: As in Ref. 1, we choose the boundary condition on v_1 so that the homogeneous solution to Eq. (12) is zero. This is in contrast to the convention used on the examples of Sec. IV of this paper.]

III. THE ROLE OF BOUNDARY CONDITIONS

From Eqs. (3)–(5), it is clear that $A_{k_il_i}(a)$ and $B_{k_il_i}(a)$ are not completely independent, nor are $C_{k_il_i}(a)$ and $D_{k_il_i}(a)$. Such connection as exists between them is largely determined by boundary conditions, which we now discuss.

In Sec. I, we limited ourselves to the case where f and its normal derivative are given as periodic along the x or t axis. It is clear from the form of Eq. (1) that periodicity, once given, must be preserved throughout the rest of the region of interest in the xt plane in both cases. Inspection of our solution also makes it clear that it is the interval of periodicity which fixes the spectra of the basic vectors.

Consider first the case where the boundary is the x axis, and the interval of periodicity is L. Recall that $\partial/\partial x \sim (\partial/\partial\xi - \partial/\partial\eta)$, and that our phase variables are fixed along the x axis independently of the amplitudes, and amplitude-dependent effects. From these statements, it follows that, to $O(1)$:

$$\lambda(k_i - l_i)/2c = 2\pi j/L,$$

$$k_il_i = 1, \qquad j = 0, \pm1, \pm2, \cdots, \quad (15a)$$

and to $O(\varepsilon)$:

$$C_{k_il_i}(a) = D_{k_il_i}(a),$$
$$A_{k_il_i}(a) = B_{k_il_i}(a). \quad (15b)$$

The second of Eqs. (15b) does not follow uniquely from the boundary condition, but is the most natural way to satisfy it. It is easy to show that this choice does not cost us any generality.

Similarly, if the periodicity is in t with period T (with the whole half-plane in x as the region of interest) we have:

$$\lambda(k_i + l_i)/2 = 2\pi j/T,$$

$$k_i l_i = 1,$$
$$C_{k_i l_i}(a) = -D_{k_i l_i}(a),$$
$$A_{k_i l_i}(a) = -B_{k_i l_i}(a),$$
$$j = \pm j_0, \quad \pm(j_0 + 1), \quad \pm(j_0 + 2), \cdots,$$
$$j_0 = \text{Integral part of } \{|\lambda T/2\pi| + 1\}. \tag{16}$$

In both cases, $\lambda(k_i + l_i)/2$ is the frequency of the jth zeroth-order mode, and $\lambda(k_i - l_i)/2c$ is the wavenumber. In the first case, we call $\lambda \varepsilon C_{k_i l_i}$ the jth "frequency shift," and in the second, $\lambda \varepsilon C_{k_i l_i}/c$ is the jth "wavenumber shift."

In both cases the program is, in words, the following. Determine the $a(k_i, l_i)$ and $\psi(k_i, l_i)$ over the boundary from the boundary conditions. Compute the $F_{k_i l_i}(a)$ there, using the $a(k_i, l_i)$ and $\psi(k_i, l_i)$. From Eqs. (5) and the first of Eqs. (14), we then have a system of differential equations for the $a(k_i, l_i)$—i.e., for the amplitudes whose associated wave vectors lie in the basic set—which may or may not have a simple solution. Assuming that the system is solvable for the $a(k_i, l_i)$, we can then solve the second of Eqs. (14) for the frequency or wavenumber shifts, as the case may be. We then solve Eqs. (3) and (4) for the $\psi(K, L)$, for (K, L) outside as well as inside the set of basic vectors. Finally, we determine v_1 from Eq. (12).

The solution for v_1 involves some arbitrariness, which is most handily resolved by putting all of f into the zeroth order over the boundary—i.e., picking v_1 and its normal derivative zero over the boundary. This is, of course, by no means the only way to split up the disturbance among the various orders in the expansion. This program is, of course, too hard to carry out for all $\mathfrak{F}$'s and $f^{(0)}$'s.

IV. TWO TRACTABLE CASES

A. A Case in Which $\mathfrak{F} = \mathfrak{F}(f)$ Only.

In this case, Im $\{F_{k_i l_i}(a)\} = 0$, all j, so we may pick $A_{k_i l_i} = B_{k_i l_i} = 0$ for all j, and $a(k_i, l_i) =$ const., all j. This leads at once to:

$$\psi(k_i, l_i) = (k_i + \varepsilon C_{k_i l_i})\xi + (l_i + \varepsilon D_{k_i l_i})\eta + \varphi(k_i, l_i), \tag{17}$$

all j, where $\varphi(k_i, l_i)$ is constant.

For definiteness, let us give f and its normal derivative over the x axis in the form

$$f(x, 0) = \sum_n c_n e^{2\pi n i x/L}, \qquad c_n = c^*_{-n},$$
$$\partial f(x, 0)/\partial t = \sum_n d_n e^{2\pi n i x/L}, \qquad d_n = d^*_{-n}.$$

The function f given here must be matched up with the $t = 0$ value of

$$f^{(0)} = \sum_{k_j, l_j} a(k_j, l_j) e^{i\psi(k_j, l_j)}, = \sum_{k_j, l_j} a(k_j, l_j) e^{i(k_j - l_j)\lambda x/2c + i\varphi(k_j, l_j)}.$$

In this "Fourier series," the coefficient of

$$e^{i(k_n - 1/k_n)\lambda x/2c}$$

is

$$a(k_n, 1/k_n) e^{i\varphi(k_n, 1/k_n)} + a(-k_n, -1/k_n) e^{i\varphi(-1/k_n, -k_n)}$$

so that

$$c_n = a(k_n, 1/k_n) e^{i\varphi(k_n, 1/k_n)} + a(-1/k_n, -k_n) e^{i\varphi(-1/k_n, -k_n)}. \tag{18}$$

The derivative is

$$\partial f^{(0)}/\partial t = \sum_{k_j, l_j} \frac{i\lambda}{2}(k_i + l_i) \times a(k_i, l_i) e^{i\varphi(k_j, l_j) + i\lambda(k_j - l_j)x/2c}$$

so that

$$d_n = \tfrac{1}{2} i(k_n + 1/k_n)\lambda [a(k_n, 1/k_n) e^{i\varphi(k_n, 1/k_n)} - a(-1/k_n, -k_n) e^{i\varphi(-1/k_n, -k_n)}]. \tag{19}$$

Solving (18) and (19),

$$a(k_n, 1/k_n) = a(k_n, l_n) = \frac{1}{2}\left[c_n + \frac{2d_n}{i\lambda(k_n + 1/k_n)}\right] e^{-i\varphi(k_n, 1/k_n)}, \tag{20}$$

which determines $a(k_i, l_i)$ and $\varphi(k_i, l_i)$ for all j. This therefore determines $\psi(k_i, l_i)$.

We assume throughout the rest of this subsection, therefore, that we know the a's and φ's.

The particular situation we now wish to consider is that of two zeroth-order traveling waves in the presence of a nonlinear term $\mathfrak{F} = -\nu f^3$, $\nu =$ const. The equation is then formally that of a stretched string imbedded in a nonlinear elastic medium, if a physical example is desired. We determine, in particular, the frequency shifts for the traveling waves 1 and 2, with

$$f^{(0)} = a_1 \cos \psi_1 + a_2 \cos \psi_2, \tag{21}$$

where for $\varepsilon = 0$,

$$\psi_1 = k_1\xi + l_1\eta + \varphi_1, \quad \psi_2 = k_2\xi + l_2\eta + \varphi_2,$$
$$a(k_1, l_1) = a(-k_1, -l_1) = \tfrac{1}{2}a_1, \qquad a(k_2, l_2) = a(-k_2, -l_2) = \tfrac{1}{2}a_2. \tag{22}$$

Closure among the set of basic vectors is guaranteed by picking $k_1 l_2 + k_2 l_1$ as any irrational number, or $(k_1/k_2) + (k_2/k_1)$ irrational, which is equivalent; although weaker conditions will suffice.

Computing $\varepsilon\mathfrak{F}(f^{(0)})$, it is readily seen that the nonvanishing $F_{KL}(a)$'s are:

(K, L)	$[-8F_{KL}(a)/\nu]$
$3(k_1, l_1)$	a_1^3
(k_1, l_1)	$3a_1^3 + 6a_1a_2^2$
$3(k_2, l_2)$	a_2^3
(k_2, l_2)	$3a_2^3 + 6a_1^2a_2$
$(k_1, l_1) + 2(k_2, l_2)$	$3a_1a_2^2$
$(k_1, l_1) - 2(k_2, l_2)$	$3a_1a_2^2$
$2(k_1, l_1) + (k_2, l_2)$	$3a_1^2a_2$
$2(k_1, l_1) - (k_2, l_2)$	$3a_1^2a_2$,

with $F_{KL}(a) = F_{-K,-L}(a)$, for all (K, L). Of the various (K, L), only (k_1, l_1) and (k_2, l_2) are basic vectors.

Equations (14) and (15) now give

$$\tfrac{1}{2}a_1(k_1 + 1/k_1)C_{k_1l_1}(a) = (\nu/8)(3a_1^3 + 6a_1a_2^2),$$
$$\tfrac{1}{2}a_2(k_2 + 1/k_2)C_{k_2l_2}(a) = (\nu/8)(3a_2^3 + 6a_1^2a_2),$$

which can be solved for the frequency shifts of waves 1 and 2:

$$\Delta\omega_1 = \lambda\varepsilon C_{k_1l_1} = \left(\frac{3\lambda\varepsilon\nu}{4}\right)\frac{a_1^2 + 2a_2^2}{k_1 + 1/k_1},$$
$$\Delta\omega_2 = \lambda\varepsilon C_{k_2l_2} = \left(\frac{3\lambda\varepsilon\nu}{4}\right)\frac{a_2^2 + 2a_1^2}{k_2 + 1/k_2}. \tag{23}$$

The computations for $a(K, L)$ for v_1 and its normal derivative equal to zero over the boundary are straightforward, and will not be written out. This example is carried to $O(\varepsilon^2)$ in the Appendix.

B. A Dissipative Case

We consider again the traveling waves of Eqs. (21) and (22), but now in the presence of a frictional dissipation, $F = -\sigma(\partial f/\partial t)^3$. (The case of linear friction, $F \sim -\partial f/\partial t$, is trivial.) Thus

$$\mathfrak{F}(f^{(0)}) = -(\sigma\lambda^3/8)(\partial f^{(0)}/\partial\xi + \partial f^{(0)}/\partial\eta)^3,$$
$$= (\sigma\lambda^3/8)[(k_1 + l_1)a_1 \sin\psi_1 + (k_2 + l_2)a_2 \sin\psi_2]^3. \tag{24}$$

The $F_{KL}(a)$ are again readily computed, and all have zero real parts, so that the shifts all vanish. The nonvanishing $F_{KL}(a)$ are:

(K, L)	$-64iF_{KL}(a)/\sigma\lambda^3$
$3(k_1, l_1)$	α_1^3
(k_1, l_1)	$-(3\alpha_1^3 + 6\alpha_1\alpha_2^2)$
$3(k_2, l_2)$	α_2^3
(k_2, l_2)	$-(3\alpha_2^3 + 6\alpha_1^2\alpha_2)$
$(k_1, l_1) + 2(k_2, l_2)$	$3\alpha_1\alpha_2^2$
$(k_1, l_1) - 2(k_2, l_2)$	$3\alpha_1\alpha_2^2$
$2(k_1, l_1) + (k_2, l_2)$	$3\alpha_1^2\alpha_2$
$2(k_1, l_1) - (k_2, l_2)$	$-3\alpha_1^2\alpha_2$,

where $\alpha_1 \equiv a_1(k_1 + l_1)$, $\alpha_2 \equiv a_2(k_2 + l_2)$, and $F_{-K,-L}(a) = -F_{KL}(a)$.

From (14), we find

$$(k_1 + l_1)A_{k_1l_1}(a) = -\tfrac{1}{64}\sigma\lambda^3[3\alpha_1^3 + 6\alpha_1\alpha_2^2],$$
$$(k_2 + l_2)A_{k_2l_2}(a) = -\tfrac{1}{64}\sigma\lambda^3[3\alpha_2^3 + 6\alpha_2\alpha_1^2]. \tag{25}$$

Taking into account Eqs. (15), the equations for the time development of the amplitudes can be written in normalized form as:

$$\partial\beta_1/\partial\tau = 1 + 2(\beta_1/\beta_2),$$
$$\partial\beta_2/\partial\tau = 1 + 2(\beta_2/\beta_1), \tag{26}$$

where $\beta_i = 1/\alpha_i^2 = (k_i + l_i)^{-2}a_i^{-2}$, for $i = 1$ or 2, and $\tau = (3\sigma\lambda^4\varepsilon/16)t$. Since β_1 and β_2 are ≥ 0, Eqs. (26) show that both $\partial\beta_1/\partial\tau$ and $\partial\beta_2/\partial\tau$ are always ≥ 0, or that the absolute values of the amplitudes are monotonically decreasing functions of time. In the special case $\beta_1(0) = \beta_2(0)$, we may solve Eq. (26) in terms of elementary functions, and get:

$$\beta_i(\tau) = 3\tau + \beta_i(0) \tag{27}$$

for $i = 1$ or 2. Since the $\psi(k_i, l_i)$ retain their simple $k_i\xi + l_i\eta +$ const. form for all x, t, the rest of the solution is straightforward.

ACKNOWLEDGMENTS

Valuable discussions with Dr. D. A. Tidman, Dr. J. K. Hale, and Dr. A. Stokes are gratefully acknowledged. The referee has also made some constructive criticisms.

APPENDIX: EXTENSION OF THE EXAMPLE OF SEC. IVA TO $O(\varepsilon^2)$

Here, we calculate the second-order effects associated with the two traveling waves of Sec. IVA. The principal obstacle to extending the calculation to higher order in epsilon is that the number of different phases which are excited grows very rapidly with ascending powers of ε. Thus, for the example of Sec. IVA, whereas there are only two linearly independent phases in zeroth order, there are 18 (or a total of 36 new terms) present in $O(\varepsilon)$; 6 of these come from the F_{KL} terms that appear in the table between Eqs. (22) and (23), and 12 more come about from the requirement that v_1 and its normal derivative, $\partial v_1/\partial t$, shall vanish on the line $t = 0$ (or the line $\xi = -\eta$). The total number of phases in $O(\varepsilon^2)$, if we were to write them out, would in general number $18 \times 5 + 2 \times (18 \times 5) = 270$.

The only qualitative modification in the calculation that results from going to higher order in ε is the presence of more and more points on the real k_1/k_2 axis to be avoided in the initial conditions. Associated with each of these points is a neighborhood of width of $O(\varepsilon)$ which must be avoided, if the perturbation corrections are to be appropriately small. This comes about from the requirement of closure among the basic phases, in the following way. Recall that in Sec. IVA, Eqs. (21) and (22) could represent an acceptable zeroth-order solution only if k_1 and k_2 were such that $\psi(k_1, l_1)$ and $\psi(k_2, l_2)$ when combined by the $\mathfrak{F}$, gave only new phases which were neither of them. Our demand that $k_1/k_2 + k_2/k_1$ be irrational was an unnecessarily strong condition which would guarantee this; a weaker condition which would have sufficed is that

$$(n_1k_1 + n_2k_2)(n_1l_1 + n_2l_2) \neq 1,$$

with (n_1, n_2) chosen from among the pairs $\pm(3, 0)$, $\pm(0, 3)$, $\pm(1, 2)$, $\pm(1, -2)$, $\pm(2, 1)$, $\pm(2, -1)$. This condition leads to a finite discrete set of points on the real k_1/k_2 axis which must be avoided if closure is to be achieved—that is, if Eqs. (21), (22) are to represent the lowest order in a perturbation-theoretic solution in which the "perturbations" agree to stay small. More such values of k_1/k_2 to be avoided are added at each order in epsilon, but they remain a finite discrete set of values. What happens when k_1/k_2 lies within a distance of $O(\varepsilon)$ of one of these values is beyond the scope of this paper. At any given order in perturbation theory, this means that there are a finite discrete set of pieces, of length of $O(\varepsilon)$, of the k_1/k_2 axis which we cannot treat with the method given here.

We now compute the $O(\varepsilon^2)$ corrections to the frequency shifts of Eqs. (23).

First we must modify the formalism given in the main body of the paper to take account of $O(\varepsilon^2)$ effects. Since we are working with an $\mathfrak{F} = \mathfrak{F}(f)$ only, we shall anticipate the fact that the amplitudes $a(K, L)$ are all constants, and not expand them in ε. [An analysis in which $\partial a(K, L)/\partial\xi$ and $\partial a(K, L)/\partial\eta$ had expansions in ε would yield the conclusion that the expansion coefficients were all zero, at the end.] For the phases $\psi(k_i, l_i)$, we seek expansions

$$\begin{aligned}\partial\psi(k_i, l_i)/\partial\xi &= k_i + \varepsilon C^{(1)}_{k_il_i}(a) + \varepsilon^2 C^{(2)}_{k_il_i}(a) + \cdots,\\ \partial\psi(k_i, l_i)/\partial\eta &= l_i + \varepsilon D^{(1)}_{k_il_i}(a) + \varepsilon^2 D^{(2)}_{k_il_i}(a) + \cdots,\end{aligned} \tag{A1}$$

and make use of the result that, for the initial value problem, we may set $C^{(n)}_{k_il_i} = D^{(n)}_{k_il_i}$, $n \geq 1$.

The solution of Eq. (1b) is most usefully written as

$$f = \sum_{k_i, l_i} a(k_i, l_i)e^{i\psi(k_i, l_i)} + \varepsilon v_1(a, \psi) + \varepsilon^2 v_2(a, \psi) + \cdots, \tag{A2}$$

where now, only the basic phases are formally present in the zeroth-order solution [we *deduced* this, between Eqs. (10) and (12), as a consequence of the requirement that Eq. (2) solve Eq. (1b) for $\varepsilon = 0$].

The $O(\varepsilon)$ solution is found, as in the main body of the paper; the $O(\varepsilon^2)$ equation becomes, using the above simplifications:

$$\begin{aligned}(\partial^2/\partial\xi\,\partial\eta + 1)v_2 &= \sum_{K,L} C^{(1)}_{KL}v^{(1)}(K, L)(l + k)e^{i\psi(K,L)}\\ &+ \sum_{k_i, l_i} a(k_i, l_i)[C^{(1)}_{k_il_i}]^2 e^{i\psi(k_i, l_i)}\\ &+ \sum_{k_i, l_i} a(k_i, l_i)C^{(2)}_{k_il_i}[k_i + l_i]e^{i\psi(k_i, l_i)}\\ &- 3\nu[\sum_{k_i, l_i} a(k_i, l_i)e^{i\psi(k_i, l_i)}]^2 \sum_{K,L} v^{(1)}(K, L)e^{i\psi(K,L)},\end{aligned} \tag{A3}$$

where the relation $C^{(1)}_{KL} \equiv \sum_i n_iC^{(1)}_{k_il_i}$ follows from Eq. (4) and Condition (v) of Sec. II. The last term in Eq. (A3) is the $O(\varepsilon^2)$ term from the expansion of $-\varepsilon\nu(f)^3$, and to know it, we must know the coefficients $v^{(1)}(K, L)$.

The coefficients $v^{(1)}(K, L)$ fall into two classes: those which come from the $F_{KL}(a)$, or inhomogeneous terms in Eq. (12), and those which come from the requirement that the total v_1 and $\partial v_1/\partial t$ shall vanish on the line $\xi = -\eta$. These latter terms necessarily contain only basic phases, since they come from solutions of the homogeneous equation for v_1, which results from striking out the right-hand side of Eq. (12). We represent the situation by writing

$$\begin{aligned}v_1(a, \psi) &= \sum_{K,L} v^{(1)}(K, L)e^{i\psi(K,L)},\\ &= \sum_{\substack{K,L\\ kl\neq 1}} \left(\frac{F_{KL}(a)}{-kl + 1}\right)e^{i\psi(K,L)}\\ &\quad + \sum_{k_j, l_j} h(k_i, l_i)e^{i\psi(k_j, l_j)}.\end{aligned} \tag{A4}$$

The first term of the right-hand side of Eq. (A4) can be read off the table given after Eq. (22), and the second term is determined by the boundary condition that $v_1(a, \psi)$ and $\partial v_1(a, \psi)/\partial t$ shall vanish at $t = 0$. For each nonvanishing F_{KL}, $kl \neq 1$, we define

$$\begin{aligned}k^{\pm}_i(K, L) &= [(k - l) \pm ((k - l)^2 + 4)^{\frac{1}{2}}]/2,\\ l^{\pm}_i(K, L) &= 1/k^{\pm}_i(K, L).\end{aligned} \tag{A5}$$

Then for each (K, L) which is present in the first sum on the right-hand side of Eq. (A4), we have nonvanishing $h(k_i, l_i)$ of the form

$$h(k_i^+(K, L), l_i^+(K, L)) = \left[\frac{k + l - k_i^-(K, L) - l_i^-(K, L)}{k_i^-(K, L) + l_i^-(K, L) - k_i^+(K, L) - l_i^+(K, L)}\right] \times \left[\frac{F_{KL}(a)}{-kl + 1}\right],$$

$$h(k_i^-(K, L), l_i^-(K, L)) = \left[\frac{k + l - k_i^+(K, L) - l_i^+(K, L)}{k_i^+(K, L) + l_i^+(K, L) - k_i^-(K, L) - l_i^-(K, L)}\right] \times \left[\frac{F_{KL}(a)}{-kl + 1}\right]. \tag{A6}$$

Purely for simplicity, we have set the constants $\varphi(k_i, l_i)$ of Eq. (17) equal to zero.

The total $O(\varepsilon)$ part of f is, therefore,

$$\begin{aligned}\varepsilon v_1(a, \psi) &= \varepsilon \sum_{K,L} v^{(1)}(K, L) e^{i\psi(K,L)} \\ &= \varepsilon \sum_{\substack{K,L \\ kl \neq 1}} \left\{ \frac{F_{KL}(a)}{-kl + 1} e^{i\psi(K,L)} \right. \\ &\quad + h(k_i^+(K, L), l_i^+(K, L)) e^{i\psi(k_i^+(K,L), l_i^+(K,L))} \\ &\quad \left. + h(k_i^-(K, L), l_i^-(K, L)) e^{i\psi(k_i^-(K,L), l_i^-(K,L))} \right\}.\end{aligned} \tag{A7}$$

The expression (A7) satisfies both Eq. (12) and the boundary conditions for v_1. The $\sum_{K,L}$ runs over $\pm$ all those values represented in the table of F_{KL}'s after Eq. (22), except for $(K, L) = \pm(k_1, l_1)$ and $\pm(k_2, l_2)$.

Computation of the last term of Eq. (A3) is now straightforward, but runs to terms which number in the hundreds. All of them, however, have the form of coefficients (which depend only on the a's, k's, and ν), multiplying exponentials of phases which have the form

$$n_1\psi(k_1, l_1) + n_2\psi(k_2, l_2) + [\text{any phase which appears in Eq. (A7)}],$$

with (n_1, n_2) chosen from among the pairs $\pm(2, 0)$, $\pm(0, 2)$, $\pm(1, 1)$, $\pm(1, -1)$, $(0, 0)$. This set of phases always contains some phases which are just $\psi(k_1, l_1)$ and $\psi(k_2, l_2)$, and always those *basic* phases which are present in Eq. (A7). The closure condition—that no *new* basic phases, beyond these, be present— is an algebraic condition which implies that a finite number of ratios k_1/k_2 must be avoided in the initial conditions.

The conditions which determine the $O(\varepsilon^2)$ frequency shifts for waves 1 and 2, and the $O(\varepsilon)$ shifts for the new basic phases which are present in Eq. (A7), are given exactly as before, by the condition that Eq. (A3) have a secularity-free solution for v_2. We limit ourselves to the situation in which closure is achieved in the above sense, and to calculating the $O(\varepsilon^2)$ frequency shifts for waves 1 and 2. These are, in general, nonvanishing, and are given by

$$\Delta\omega_1^{(2)} = \lambda\varepsilon^2 C_{k_1 l_1}^{(2)},$$
$$\Delta\omega_2^{(2)} = \lambda\varepsilon^2 C_{k_2 l_2}^{(2)},$$

with the C's determined by the no-secularity requirement as:

$$C_{k_i l_i}^{(2)} = -\frac{a\nu^2}{16}\frac{(a_i^2 + 2a_j^2)^2}{(k_i + l_i)^3} + \frac{3\nu^2}{128}\frac{a_i^4}{k_i + l_i} + \frac{9\nu^2}{32}\frac{(a_i^4 + 2a_i^2a_j^2)}{(k_i + l_i)[1 - \frac{1}{4}(k_il_j + k_jl_i)^2]}, \tag{A8}$$

where $i, j = 1, 2$ or $2, 1$.

In closing, we feel that there is one point in Ref. 1 which requires clarification. We refer, in particular, to Sec. III of that paper. There, the first-order solution satisfies a different initial condition that is physically somewhat obscure, and does not correspond to the present situation where the first-order solution and its normal derivative are chosen to vanish initially. Only the part of the first-order solution which comes directly from the nonlinear term (i.e., the inhomogeneous part) is present. It will be seen, however, that the simpler and more natural-appearing boundary condition of the present paper does lead to much more algebra.

Note added in proof: Professor N. G. van Kampen and Professor Th. W. Ruijgrok have kindly pointed out that a particular *exact* solution to the equation treated in Sec. IVA has been given by D. F. Kurdgelaidze [Soviet Phys.—JETP **5**, 941 (1959)], and provides an interesting check on the accuracy of the present method. The solution is of the monochromatic, plane-wave type, periodic in both x and t, but not sinusoidal, being a Jacobian elliptic function. This function possesses a Fourier-like asymptotic expansion in powers of the amplitude, and can be compared with the results obtained here by fixing the wavenumber and setting one of the two amplitudes a_1, a_2 equal to zero. For the first three orders in the amplitude, which are all that are given here, the agreement is exact.

One word of caution is in order. If Kurdgelaidze's solution is expanded, to reproduce it by the present method requires a different initial condition on the perturbation than the one used here. Here, all the "fundamentals" are collected in the zeroth order initially; in the cited paper, they are split up among the various orders in ε in a special way. Account must be taken of this different split-up (it is a purely arbitrary choice) if the correct expression for the second-order frequency shift is to be obtained.

S Theorem and Construction of the Invariants of the Semisimple Compact Lie Algebras

B. GRUBER*
Dublin Institute for Advanced Studies, Ireland
AND
L. O'RAIFEARTAIGH†
Dublin Institute for Advanced Studies, Ireland
and Institute of Mathematical Sciences, Madras, India
(Received 21 May, 1964)

An infinitesimal proof of the S theorem, which states that the invariants of a compact semisimple Lie algebra are symmetric with respect to the discrete Weyl group of the algebra, is given. The complete set of invariants of the various compact semisimple Lie algebras found by Racah are rederived in a somewhat different and explicit way, the S theorem being used to establish their completeness.

1. INTRODUCTION

THE primary purpose of this paper is to give an infinitesimal proof of the S theorem (Sec. 4), which states that any "invariant" of a compact semisimple Lie algebra is symmetric with respect to the reflexions which generate the discrete Weyl group of the algebra. An important application of this theorem is its use in establishing the completeness of the sets of invariants found by Racah[1] for the various compact semisimple Lie algebras. A secondary purpose of the paper is to rederive the invariants found by Racah, in a somewhat different and more explicit way. In particular, we show that the type of invariants considered by Racah arise as a natural further generalization of the generalized Casimir operators,[2] and we give explicitly an elimination procedure for selecting from the general class of invariants considered by Racah, the l independent ones, where l is the rank of the algebra. We also exhibit the l-independent invariants, and their leading terms with respect to the Cartan basis, in a simple explicit form. Throughout, we shall confine our attention to the four main classes of compact semisimple Lie algebras A_l, B_l, C_l, and D_l, and to G_2. The data[3] for the four other exceptional groups, F_4, E_6, E_7, and E_8 can be found in Refs. 1 and 3.

2. GENERAL FORM OF INVARIANTS

Let

$$[X_\lambda, X_\mu] = c^\nu_{\lambda\mu} X_\nu, \qquad \lambda, \mu, \nu = 1, \cdots, r, \tag{2.1}$$

be any compact semisimple Lie algebra of order r ($g_{\lambda\mu} = c^\tau_{\lambda\sigma} c^\sigma_{\mu\tau}$ negative definite). The generalized Casimir operators

$$C_n = c^{\sigma_1}_{\lambda\sigma_2} c^{\sigma_2}_{\mu\sigma_3} c^{\sigma_3}_{\nu\sigma_4} \cdots c^{\sigma_n}_{\rho\sigma_1} X^\lambda X^\mu X^\nu \cdots X^\rho, \tag{2.2}$$

are *invariants* of the algebra, that is to say, they satisfy the relations

$$[C_n, X_\lambda] = 0, \tag{2.3}$$

for all X_λ, $\lambda = 1, \cdots, r$, but it is known that, in general, they are not the only independent invariants. To obtain all the independent invariants, the C_n must be generalized, and the natural generalization which suggests itself is the following: One introduces the *adjoint* representation of the algebra

$$X^{(A)}_\lambda = c^{\sigma_1}_{\lambda\sigma_2}, \tag{2.4}$$

where σ_1 and σ_2 are to be regarded as matrix indices. One sees then that an alternative form for the C_n is

$$C_n = \mathrm{Tr}(X^{(A)}_\mu \cdots X^{(A)}_\tau) X^\mu \cdots X^\tau, \tag{2.5}$$

and the generalization which suggests itself is to replace the $X^{(A)}_\mu$ in (2.5) by the $\hat{X}_\mu$ of *any* representation. If one does this, one obtains the quantities

$$I_n = \mathrm{Tr}(\hat{X}_\mu \cdots \hat{X}_\tau) X^\mu \cdots X^\tau, \tag{2.6}$$

which are easily seen to be invariants, because if we let

$$U = 1 + \epsilon_\lambda X_\lambda, \tag{2.7}$$

where the ϵ_λ are arbitrary small numbers, we have

$$U X_\lambda U^{-1} = a^\sigma_\lambda X_\sigma, \tag{2.8}$$

where

$$a^\sigma_\lambda = \delta^\sigma_\lambda + \epsilon_\tau c^\sigma_{\tau\lambda}, \tag{2.9}$$

* Present address: Istituto di Fisica Teorica, Università di Napoli, Naples, Italy.
† Present address: Department of Physics, Syracuse University, Syracuse, New York.
[1] G. Racah, Rend. Lincei 8, 108 (1950).
[2] G. Racah, Princeton Lecture Notes, CERN 61-8 (1961) (unpublished).
[3] M. Umezawa, Nucl. Phys. (to be published).

and so

$$\begin{aligned} UI_nU^{-1} &= \mathrm{Tr}(\hat{X}_\mu \cdots \hat{X}_\tau)UX^\mu \cdots X^\tau U^{-1}, \\ &= \mathrm{Tr}(\hat{X}_\mu \cdots \hat{X}_\tau)a^\mu_\sigma \cdots a^\tau_\lambda X^\sigma \cdots X^\lambda, \\ &= \mathrm{Tr}(U\hat{X}_\sigma \cdots \hat{X}_\lambda U^{-1})X^\sigma \cdots X^\lambda, \\ &= I_n. \end{aligned} \tag{2.10}$$

These invariants were originally found by Racah[2] in a somewhat different form. Our task will be to pick out from all the possible "Racah-type" invariants I_n, the independent ones, and to show that these form a complete set of invariants. As a preliminary, we shall express the I_n in a somewhat simpler form, as follows: from (2.6), we have,

$$I_n = \mathrm{Tr}(\hat{X}_\lambda \cdots \hat{X}_\tau \times X^\lambda \cdots X^\tau), \tag{2.11}$$

where $\times$ is the direct product symbol throughout this paper and the trace is to be taken with respect to the $\hat{X}$ space only. Hence,

$$\begin{aligned} I_n &= \mathrm{Tr}(\hat{X}_\lambda \times X^\lambda) \cdots (\hat{X}_\tau \times X^\tau), \\ &= \mathrm{Tr}A^n, \end{aligned} \tag{2.12}$$

where,

$$A = (\hat{X}_\lambda \times X^\lambda). \tag{2.13}$$

Here the direct product means that A is a matrix of the same dimension, d, as the representation $\hat{X}_\lambda$, and the X_λ (or linear combinations thereof) are the elements of this matrix. For example, if $\hat{X}_\lambda$ are the infinitesimal generators of the three-dimensional representation of the three-dimensional rotation group and X_1, X_2, and X_3 are the conventional base elements of the corresponding Lie algebra,

$$A = \begin{pmatrix} 0 & X_3 & -X_2 \\ -X_3 & 0 & X_1 \\ X_2 & -X_1 & 0 \end{pmatrix}. \tag{2.14}$$

From (2.12) and the fact that every matrix A satisfies its own characteristic equation, we see that for any representation X_λ, the invariants of order $n > d$ can not be independent of those of order $n \leq d$. We see also that an equivalent set of invariants are the coefficients of ρ^r, $r = 0, 1, \cdots, d-1$, in the expansion of

$$|A - \rho|. \tag{2.15}$$

This is the form in which the invariants were originally given by Racah.

3. SELECTION OF THE INDEPENDENT INVARIANTS

In this section we should like to give an elimination procedure for selecting out the independent Racah-type invariants.

We begin by recalling how the representations of the compact Lie algebras constructed. First, the reducible representations are fully reducible, and so are direct sums of the irreducible representations. Secondly, the irreducible representations are finite dimensional and are obtained by extracting the highest-dimensional representations from the various direct products of the l-fundamental representations. Finally, the fundamental representations are constructed as follows: Letting Γ_1 be the self-representation (classical group), the quantities Γ_r = completely antisymmetric part of the representation $\Gamma_1 \times \Gamma_1 \times \cdots \times \Gamma_1$, (for r terms) are constructed for $r = 1, \cdots, l$. It then turns out that

For A_l: the Γ_r are just the l fundamental representations.[2]

B_l: the Γ_r are fundamental for $r = 1, \cdots, l-1$. There exists independently a fundamental spinor representation Δ. Γ_l is irreducible but not fundamental.[2,4]

C_l: the Γ_r are reducible, on account of the symplectic condition which leaves a certain skew form invariant. But the leading representation in each Γ_r is a fundamental representation, and the remainder consists of a direct sum of Γ_s, $s < r$.

D_l: the Γ_r are fundamental for $r = 1, \cdots, l-2$. There exist independently two fundamental spinor representations, Δ^+ and Δ^-. Γ_{l-1} is irreducible but not fundamental. $\Gamma_l = \Gamma_l^+ \dotplus \Gamma_l^-$, where Γ_l^+ and Γ_l^- are of the same dimension, are irreducible, but are not fundamental.[4]

G_2: $\Gamma_2 = \Gamma_{(14)} \dotplus \Gamma_1$, where $\Gamma_{(14)}$ is fundamental.[5]

In this way we see that all the representations are "built up" out of Γ_1 and the spinor representations. This makes it very plausible that the invariants formed with these representations will be "built up" out of the invariants formed with Γ_1 and Δ, $\Delta^\pm$. In fact, it can be proved fairly easily that this is the case [using the form (2.12) for the I_n]. However, we prefer to omit the proof and to use only the plausibility argument here. The methods used below in the case of the spinor representations should help to make the lines along which a proof can be constructed fairly clear.

Our next step is to show that (with one exception) we can dispense with the spinor representa-

[4] H. Boerner, *Representations of Groups* (North-Holland Publishing Company, Amsterdam, Holland, 1963), Chap. VII, Sec. 14; Chap. VIII, Sec. 3–5.

[5] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. 34, 1 (1962).

tions also. The argument here is more mathematical. We consider first the case of B_l, and form the direct product[6]

$$\Delta \times \Delta = 1 + \Gamma_1 + \Gamma_2 + \cdots + \Gamma_{l-1} + \Gamma_l, \quad (3.1)$$

from which it follows immediately that any invariant constructed with $\Delta \times \Delta$ will be a function of the invariants constructed with the Γ_r, and hence a function of the invariants constructed with Γ_1. On the other hand, if $A(\Delta)$ is the A-matrix of (2.13) for the representation $\Delta = 1 + \epsilon_\alpha X_\alpha(\Delta)$, and $A(\Delta \times \Delta)$ the A matrix for the representation

$$\Delta \times \Delta = 1 + \epsilon_\alpha[1 \times X_\alpha(\Delta) + X_\alpha(\Delta) \times 1] + \cdots, \quad (3.1a)$$

we have

$$A(\Delta \times \Delta) = A(\Delta) \times 1 + 1 \times A(\Delta), \quad (3.2)$$

whence,

$$\begin{aligned} I_n(\Delta \times \Delta) &= \mathrm{Tr}[A(\Delta) \times 1 + 1 \times A(\Delta)]^n, \\ &= \sum_{r=0}^{n} \binom{n}{r} \mathrm{Tr}A^r(\Delta) \times A(\Delta)^{n-r}, \\ &= \sum_{r=0}^{n} \binom{n}{r} \mathrm{Tr}A^r(\Delta)\mathrm{Tr}A(\Delta)^{n-r}, \\ &= 2dI_n(\Delta) + \text{lower-order invariants}, \end{aligned} \quad (3.3)$$

where d is the dimension of Δ. Hence, by induction in n, we see that the $I_n(\Delta)$ are functions of the $I_n(\Delta \times \Delta))$, which, we have already seen, are functions of the $I_n(\Gamma_1)$. Thus we can dispense with Δ.

For D_l, the situation is not so simple. If we take the direct product[6]

$$\Delta^+ \times \Delta^- = \Gamma_1 + \Gamma_3 + \cdots + \Gamma_{l-3} + \Gamma_{l+1}, \quad (3.4)$$

we can deduce in the same way as above that $I_n(\Delta^+) + I_n(\Delta^-)$ is a function of the $I_n(\Gamma_1)$, so that at most one of the two representations Δ^+ or Δ^- can give rise to independent invariants; but if we take the product

$$\begin{aligned} \Delta^+ \times \Delta^+ &= 1 + \Gamma_2 + \cdots + \Gamma_{l-2} + \Gamma_l^+ \quad (l \text{ even}), \\ &= \Gamma_1 + \Gamma_3 + \cdots + \Gamma_{l-2} + \Gamma_l^- \quad (l \text{ odd}) \end{aligned} \quad (3.5)$$

(or the similar product for $\Delta^- \times \Delta^-$), the proof breaks down because of the appearance of the representations Γ_l^+ and Γ_l^-. On account of the way in which Γ_l breaks up into $\Gamma_l^+ + \Gamma_l^-$ it can not be shown that the invariants constructed with Γ_l^+ and Γ_l^- separately are functions of the $I_n(\Gamma_1)$.

[6] R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).

By a refinement of the argument, however, we can show that while we cannot dispense with Δ^+ (or alternatively Δ^-) completely, we shall need it to construct only one single independent invariant. To show this we consider not $\Delta^+ \times \Delta^+$, but only its antisymmetric (AS) part. We have[6]

$$\begin{aligned} (\Delta^+ \times \Delta^+)_{AS} &= \Gamma_2 + \Gamma_6 + \cdots + \Gamma_{l-2}, \quad l = 0 \pmod 4, \\ &= \Gamma_3 + \Gamma_7 + \cdots + \Gamma_{l-2}, \quad l = 1 \pmod 4, \\ &= 1 + \Gamma_4 + \cdots + \Gamma_{l-2}, \quad l = 2 \pmod 4, \\ &= \Gamma_1 + \Gamma_5 + \cdots + \Gamma_{l-2}, \quad l = 3 \pmod 4, \end{aligned} \quad (3.6)$$

so that $I_n((\Delta^+ \times \Delta^+)_{AS})$ is a function of the $I_n(\Gamma_r)$, and so of the $I_n(\Gamma_1)$ only. In analogy to (3.3) we have

$$I_n((\Delta^+ \times \Delta^+)_{AS}) = \sum_{r=0}^{n} \binom{n}{r} \mathrm{Tr}[A(\Delta^+)^r \times A(\Delta^+)^{n-r}]_{AS}; \quad (3.7)$$

but because

$$\mathrm{Tr}(B \times C)_{AS} = \tfrac{1}{2}\mathrm{Tr}B\mathrm{Tr}C - \tfrac{1}{2}\mathrm{Tr}BC, \quad (3.8)$$

we obtain from the next step

$$\begin{aligned} I_n((\Delta^+ \times \Delta^+)_{AS}) &= \tfrac{1}{2} \sum_{r=0}^{n} \binom{n}{r} [\mathrm{Tr}A(\Delta^+)^r \mathrm{Tr}A(\Delta^+)^{n-r} - \mathrm{Tr}A(\Delta^+)^n], \\ &= \tfrac{1}{2}(2d^+ - 2^n) I_n(\Delta^+) + \text{lower-order invariants}. \end{aligned} \quad (3.9)$$

Thus, unless $2d^+ = 2^n$, the argument goes through as before, and $I_n(\Delta^+)$ is a function of the $I_n(\Gamma_1)$. The exceptional case $2d^+ = 2^n$ occurs, because d^+, the dimension of the representation Δ^+, is just 2^{l-1}. Thus the argument breaks down for $n = l$, i.e., for $I_l(\Delta^+)$.

We see, therefore, that with the exception of the single invariant $I_l(\Delta^+)$ for D_l, all of the invariants can be constructed from the self-representation. The only question remaining is that of the *orders* of the invariants which we should construct. But, as we have seen in Sec. 2, if d is the dimension of the self-representation, we can confine ourselves to the invariants of order $n \leq d$. Furthermore, on account of the orthogonality condition $A = -\tilde{A}$ for B_l and D_l, and the symplectic condition $A = -JA\tilde{J}^{-1}$, where J is skew, for C_l, the odd invariants drop out for these algebras. Similarly for G_2, which is a subalgebra of B_3. Finally, for A_l, the unimodular condition means that $I_1 = 0$.

Thus we are left with

$$\begin{aligned}
&A_l(d = l+1) : I_2, I_3, \cdots, I_{l+1},\\
&B_l(d = 2l+1) : I_2, I_4, \cdots, I_{2l},\\
&C_l(d = 2l) : I_2, I_4, \cdots, I_{2l},\\
&D_l(d = 2l) : I_2, I_4, \cdots, I_{2l}, \text{ and } I_l(\Delta^+),\\
&G_2(d = 7) : I_2, I_4, I_6,
\end{aligned} \tag{3.10}$$

where it is understood that the I_r are constructed with the self-representation. In Sec. 5, it is shown that these sets of invariants form complete sets (in fact, for D_l and G_2 we can omit I_{2l} and I_4, respectively), but before going on to this proof it is necessary to establish the S theorem, which is done in the next section.

We conclude this section by giving simple explicit forms of the above invariants. Since the algebras are semisimple and compact, we can choose the basis so that $g_{\lambda\mu} = \delta_{\lambda\mu}$. Then, from (2.13), using a double-index instead of a single-index notation for the X's, we have

$$A = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & & & \\ \vdots & & & \\ X_{n1} & & & X_{nn} \end{pmatrix}, \tag{3.11}$$

where we have the conditions

$$\begin{aligned}
\sum X_{ii} &= 0, \quad \text{for} \quad A_l,\\
X_{ij} + X_{ji} &= 0, \quad \text{for} \quad B_l \text{ and } D_l,\\
X_{ij}J_{jk} + J_{ij}X_{kj} &= 0, \quad \text{for} \quad C_l,
\end{aligned} \tag{3.12}$$

where J is $\mathbb{1} \times \tau$, where $\mathbb{1}$ is the unit matrix in l dimensions and $\tau = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Hence for the I_r in (3.10), we have

$$I_r = X_{ij}X_{jk}X_{ks} \cdots X_{tq}X_{qi}, \tag{3.13}$$

with the conditions (3.12).

If one uses the Cartan canonical basis H_i, E_α, $E_{-\alpha}$, the invariants become polynomials in these quantities. For reasons which will become clear in the next section, one is interested in obtaining the leading terms in the H_i in these polynomials. As these are obtained by setting the $E_{\pm\alpha}$ equal to zero, they can be obtained directly from the A of (3.11) by letting $A \to A_0 = (A)_{E\pm\alpha=0}$. One obtains in this way,

$$A_0 = \begin{pmatrix} H_1 & & & \\ & H_2 & & 0 \\ & & \ddots & \\ 0 & & & H_{l+1} \end{pmatrix}_{\sum_{r=1}^{l+1} H_r = 0},$$

$$\begin{pmatrix} 0 & H_1 & & & & & \\ -H_1 & 0 & & & & & \\ & & 0 & H_2 & & & \\ & & -H_2 & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & H_l \\ & & & & & -H_l & 0 \\ & & & & & & & 0 \end{pmatrix},$$

$$\begin{pmatrix} H_1 & & & & & \\ & -H_1 & & & & \\ & & H_2 & & & \\ & & & -H_2 & & \\ & & & & \ddots & \\ & & & & & H_l \\ & & & & & & -H_l \end{pmatrix},$$

$$\begin{pmatrix} 0 & H_1 & & & & \\ -H_1 & 0 & & & & \\ & & 0 & H_2 & & \\ & & -H_2 & 0 & & \\ & & & & \ddots & \\ & & & & 0 & H_l \\ & & & & -H_l & 0 \end{pmatrix}, \tag{3.14}$$

for A_l, B_l, C_l, and D_l, since for these algebras

$$\begin{aligned}
H_r &= X_{r,r}, \quad H_r = X_{2r-1,2r},\\
H_r &= X_{2r-1,2r-1}(= -X_{2r,2r}),
\end{aligned}$$

and $H_r = X_{2r-1,2r}$, respectively. Thus for A_l,

$$I_1 = 0, \qquad I_r = \sum_{i=1}^{l+1} (H_i)^r + \cdots,$$

$$r = 2, 3, \cdots, l+1, \tag{3.15}$$

and for the others,

$$I_{2r} = \sum_{i=1}^{l} (H_i)^{2r} + \cdots, \quad r = 1, 2, \cdots, l. \tag{3.16}$$

It remains to obtain the leading terms for $I_l(\Delta^\pm)$. The representations $\Delta^\pm$ are obtained as follows[4]:

One constructs

$$\rho_i = \tau_3 \times \cdots \times \tau_3 \times \tau_1 \times 1 \times \cdots \times 1, \quad i = 1, 2, \cdots, l-1,$$
$$\rho_l = \tau_3 \times \cdots \times \tau_3 \times \tau_3 \times \tau_3 \times \cdots \times \tau_3, \tag{3.17}$$
$$\rho_i = \tau_3 \times \cdots \times \tau_3 \times \tau_2 \times 1 \times \cdots \times 1, \quad i = l+1, \cdots 2l-1,$$

where the τ_i are the Pauli matrices, and each direct product contains $l-1$ factors, and sets

$$X_\alpha^\pm = \{\cdots \pm \rho_\mu \cdots [\rho_\mu, \rho_\nu] \cdots\}, \quad \mu, \nu = 1, \cdots, 2l-1, \quad \alpha = 1, \cdots, \frac{2l(2l-1)}{2}. \tag{3.18}$$

A convenient Cartan algebra is

$$H_1 = \rho_l, \qquad H_r = -i[\rho_r, \rho_{l+r}] = 1 \times \cdots \times 1 \times \tau_3 \times 1 \times \cdots \times 1. \tag{3.19}$$

Instead of $I_l(\Delta^+)$ or $I_l(\Delta^-)$ it is convenient to use the invariant $I_l(\Delta^+) - I_l(\Delta^-)$, which is equivalent to $I_l(\Delta^+)$ or $I_l(\Delta^-)$ from (3.4), and which has the property of changing sign under the transformation $\Delta^+ \rightleftarrows \Delta^-$. From (3.18) this transformation is effected by $\rho_\mu \to -\rho_\mu$, which for the H_i in particular is $H_1 \to -H_1$ and $H_i \to H_i$, $i \neq 1$. Thus the leading terms in $I_l(\Delta^+) - I_l(\Delta^-)$ change sign under the transformation $H_1 \to -H_1$ and $H_i \to H_i$, $i \neq 1$. On the other hand, since no H_i is preferred, it follows that the leading terms change sign under the transformation $H_j \to -H_j$, $H_i \to H_i$, $i \neq j$, for any given j. Thus the leading terms in $I_l(\Delta^+) - I_l(\Delta^-)$ must be odd in each of the H_i. Hence they must contain each H_i at least once. But since there are l H_i and $I_l(\Delta^+) - I_l(\Delta^-)$ is of order l it follows that they can contain each H_i only once. Thus,

$$I_l(\Delta^+) - I_l(\Delta^-) = \text{const}\, H_1 H_2 \cdots H_l + \cdots. \tag{3.20}$$

Finally, we consider the group $\mathcal{G}_2$. For this group if we use the self-representation of H_1 and H_2 given in Ref. 5 [Fig. 3(a)], we see that A_0 is a 7-by-7 diagonal matrix with diagonal elements 0, $\pm H_1/2\sqrt{3}$, and $\pm H_2/4 \pm H_1/4\sqrt{3}$. Thus

$$I_2 = \mathrm{Tr}A_0^2 = \tfrac{1}{4}(H_1^2 + H_2^2) + \cdots,$$
$$I_4 = \mathrm{Tr}A_0^4 = \frac{1}{4^3}(H_1^2 + H_2^2)^2 + \cdots, \tag{3.21}$$
$$I_6 = \mathrm{Tr}A_0^6 = \frac{1}{4^5}(H_1^2 + H_2^2)^3 + \frac{2}{4^5} H_1^2(H_2^2 - \tfrac{1}{3}H_1^2)^2 + \cdots.$$

4. STATEMENT AND PROOF OF S THEOREM

Statement. Let $F(X_\lambda)$ be any invariant and $|j\rangle$ the unit vector belonging to any highest weight j, and let

$$\langle j|\, F(X_\lambda)\, |j\rangle = \phi_0(j). \tag{4.1}$$

Let Σ denote the sum of the positive roots, and S the Weyl group generated by the reflexions in the planes orthogonal to the roots. Then, if

$$\phi_0'(j + \tfrac{1}{2}\Sigma) = \phi_0(j), \tag{4.2}$$
$$\phi_0'(j + \tfrac{1}{2}\Sigma) = \phi_0'(S(j + \tfrac{1}{2}\Sigma)). \tag{4.3}$$

In other words, ϕ_0' is invariant under the Weyl group.

Proof. Before presenting the proof proper, we find it convenient to introduce first the concept of *primitive* roots, and to list some of their properties.

Definition. The primitive roots of a compact semi-simple Lie algebra of rank l are a set of positive roots $r, s, \cdots, t$, such that every root α can be expressed in the form

$$\alpha = k_r r + k_s s + \cdots + k_t t, \tag{4.4}$$

where the ks are either all positive or zero (for the positive roots) or all negative or zero (for the negative roots). The proof of the existence of such a set of l roots has been given by Dynkin.[7] (For the Lie algebras of low rank, the existence of such a set of roots can be seen by inspection of the root diagrams).

Properties. (a) E_α, the E corresponding to any positive root α, can be expressed as polynomial in the E_r, the Es corresponding to the primitive roots $r, s, \cdots, t$.

Proof. The proof follows by applying the relation

$$\alpha = \beta + \gamma \to [E_\beta, E_\gamma] = N_{\beta\gamma} E_\alpha, \quad N_{\beta\gamma} \neq 0, \tag{4.5}$$

repeatedly to (4.4).

(b) $$[E_r, E_{-s}] = 0, \qquad r \neq s. \tag{4.6}$$

Proof. Otherwise $\alpha = r - s$ would be a root, in contradiction to (4.4).

(c) $$[E_r, E_{-r}^\nu] = E_{-r}^{\nu-1} g_\nu(H), \tag{4.7}$$

where

$$g_\nu(H) = \nu r \circ H - \frac{\nu(\nu-1)}{2} r \circ r. \tag{4.8}$$

Proof. The proof follows easily by induction from

$$[E_r, E_{-r}] = r \circ H, \tag{4.9}$$

[7] E. B. Dynkin, Am. Math. Soc. Trans. 9, 105 (1950).

and holds, in fact, for all roots α, though we shall need it only for the primitive roots.

(d) If Σ_r is the sum of the positive roots excluding r, and S_r is the reflection in the plane orthogonal to r,

$$S_r \Sigma_r = \Sigma_r. \tag{4.10}$$

Proof. Let α be a positive root. If $\alpha = k_r r$, $\alpha = r$. If $\alpha \neq k_r r$, at least one of the other k's (k_q say) in (4.4) is nonzero. But (4.4) implies that if $k_q > 0$

$$S_r\alpha = \alpha - \frac{2(\alpha\cdot r)}{(r\cdot r)} r = \left[k_r - \frac{2(\alpha\cdot r)}{(r\cdot r)}\right] r + \cdots + k_q q + \cdots \tag{4.11}$$

is positive. Thus S_r transforms the set of all positive roots excluding r, into itself. Since S_r is also nonsingular ($S_r^2 = 1$), Σ_r is left invariant under S_r as required.

(e) The Weyl group S is generated by the reflexions in the planes orthogonal to the primitive roots alone.

Proof. For every positive α in (4.4), there exists at least one r, s, $\cdots$ (q_1, say) such that $(\alpha\cdot q_1) > 0$ since otherwise

$$\alpha\cdot\alpha = k_r(\alpha\cdot r) + k_s(\alpha\cdot s) + \cdots < 0.$$

Since for such a q_1

$$S_{q_1}\alpha = \alpha - 2\frac{(\alpha\cdot q_1)}{(q_1\cdot q_1)} q_1 < \alpha, \tag{4.12}$$

it is easy to see that by a finite number (ν, say) of such reflections, α can be transformed into a primitive root (q, say), i.e.,

$$S_{q_\nu} S_{q_{\nu-1}} \cdots S_{q_2} S_{q_1}\alpha = q. \tag{4.13}$$

But then, by a standard theorem on reflexions

$$S_\alpha = S_{q_1} \cdots S_{q_\nu}(S_q) S_{q_\nu} \cdots S_{q_1}, \tag{4.14}$$

and the reflexion S_α is generated by primitive reflexions, as required.

We proceed now to the proof of the S theorem. Let $F(X_\lambda)$ be any polynomial. If we change from the basis X_λ to the Cartan basis H_i, E_α, $E_{-\alpha}$, $F(X_\lambda)$ becomes a polynomial in the latter base elements, i.e.,

$$F(X_\lambda) = \Psi(H_i, E_\alpha, E_{-\alpha}). \tag{4.15}$$

Using property (a) above, this polynomial can be reduced to a polynomial in the H_i, E_r, and E_{-r}, i.e.,

$$F(X_\lambda) = \Psi(H_i, E_\alpha, E_{-\alpha}) = \Phi(H_i, E_r, E_{-r}). \tag{4.16}$$

By using property (b) above, (4.9), and

$$[H_i, E_{\pm r}] = \pm r_i E_{\pm r}, \tag{4.17}$$

it is easy to see that we can rearrange the terms in $\Phi(H_i, E_r, E_{-r})$ so that the E_{-r} are to the left, and the E_r to the right in each term, i.e., we can assume without loss of generality that $\Phi(H_i, E_r, E_{-r})$ is of the form

$$\Phi(H_i, E_r, E_{-r}) = \sum E_{-r}E_{-s} \cdots E_{-t}\phi(H)E_{r'} \cdots E_{t'}, \tag{4.18}$$

where $\phi(H)$ is a polynomial in the H_i, and is, of course, different in each term in the sum. Note that the procedure of moving the E_{-r} to the left and the E_r to the right does not reintroduce nonprimitive roots if we leave the order of the positive roots among themselves and the negative roots among themselves unchanged.

For any highest weight j,

$$E_r |j\rangle = E_s |j\rangle = \cdots = E_t |j\rangle = 0, \tag{4.19}$$

we construct the unit vectors $|j\rangle$, $|j - r\rangle$, $\cdots$, $|j - Nr\rangle$, belonging to the complete "string" of weights j, $j - r$, $\cdots$, $j - Nr$, $N = 2(r\cdot j)/(r\cdot r)$, formed with any primitive root r.

If we now define (since the algebra is compact we can choose our basis so that $E_r^+ = E_{-r}$);

$$c_\nu = \langle j - \nu r| \Phi(H_i, E_r, E_{-r}) |j - \nu r\rangle = \frac{\langle j| E_r'\Phi(H_i, E_r, E_{-r})E_{-r}' |j\rangle}{\langle j| E_r'E_{-r}' |j\rangle}, \quad \nu = 0, 1, \cdots N, \tag{4.21}$$

it is easy to see, from property (b) above and (4.19), that the only terms in $\Phi(H_i, E_r, E_{-r})$ which will contribute to c_ν are of the form

$$E_{-r}^\mu \phi_\mu(H) E_r^\mu, \qquad \mu \leq \nu. \tag{4.22}$$

Hence,

$$c_\nu = \sum_{\mu=0}^{\nu} c(\nu\mu j)\phi_\mu(j - \nu r + \mu r), \tag{4.23}$$

where

$$c(\nu\mu j) = \frac{\langle j| E_r'E_{-r}^\mu E_r^\mu E_{-r}' |j\rangle}{\langle j| E_r'E_{-r}^\nu |j\rangle}. \tag{4.24}$$

Using the property (c) above, it is easy to verify that the coefficients $c(\nu\mu j)$ satisfy the recurrence relation

$$c(\nu + 1\mu j) = c(\nu\mu j) + c(\nu\mu - 1j)g_\mu(j - \nu r + \mu r - r), \quad \nu = 0, 1, \cdots N - 1, \tag{4.25}$$

and that

$$c(\nu\nu j) \neq 0, \qquad \nu = 0, 1, \cdots N. \tag{4.26}$$

So far, we have not assumed that $F(X_\lambda) = \Phi(H_i, E_r, E_{-r})$ is an invariant. If it is an invariant then we have the condition

$$c_\nu = c_{\nu+1}, \qquad \nu = 0, 1, \cdots, N-1. \tag{4.27}$$

Our method of procedure will be to use this condition to obtain a relation between the ϕ_μ and $\phi_{\mu+1}$ of (4.22) and (4.23), and to deduce from this relation the required property (4.3) of ϕ_0. From (4.23) we have

$$\begin{aligned} c_{\nu+1} &= \sum_{\mu=0}^{\nu+1} c(\nu+1\mu j)\phi_\mu(j-\nu r+\mu r-r), \\ &= \phi_0(j-\nu r-r) \\ &\quad + \sum_{\mu=1}^{\nu} c(\nu+1\mu j)\phi_\mu(j-\nu r+\mu r-r) \\ &\quad + c(\nu+1\nu+1j)\phi_{\nu+1}(j), \\ &= \phi_0(j-\nu r-r) + \sum_{\mu=1}^{\nu} c(\nu\mu j)\phi_\mu(j-\nu r+\mu r-r) \\ &\quad + \sum_{\mu=1}^{\nu} c(\nu\mu-1j)g_\mu(j-\nu r+\mu r-r) \\ &\quad \times \phi_\mu(j-\nu r+\mu r-r) + c(\nu+1\nu+1j)\phi_{\nu+1}(j), \\ &= \sum_{\mu=0}^{\nu} c(\nu\mu j)\phi_\mu(j-\nu r+\mu r-r) \\ &\quad + \sum_{\mu=0}^{\nu} c(\nu\mu j)g_{\mu+1}(j-\nu r+\mu r)\phi_{\mu+1}(j-\nu r+\mu r). \end{aligned} \tag{4.28}$$

Hence, from (4.27)

$$\begin{aligned} &\sum_{\mu=0}^{\nu} c(\nu\mu j)[\phi_\mu(j-\nu r+\mu r) - \phi_\mu(j-\nu r+\mu r-r) \\ &\quad - g_{\mu+1}(j-\nu r+\mu r)\phi_{\mu+1}(j-\nu r+\mu r)] = 0, \\ &\qquad\qquad \nu \leq N-1. \end{aligned} \tag{4.29}$$

Suppose now that for $\mu = 1, \cdots, M$, the ϕ_μ satisfy the relations

$$\phi_\mu(\lambda) - \phi_\mu(\lambda - r) = g_{\mu+1}(\lambda)\phi_{\mu+1}(\lambda), \tag{4.30}$$

where λ is any l-vector. Then from (4.29), with $\nu = M+1$, and using (4.26), we have

$$\phi_{M+1}(j) - \phi_{M+1}(j-r) = g_{M+2}(j)\phi_{M+2}(j), \tag{4.31}$$

where j is any j such that $2(r\cdot j)/(r\cdot r) = N \geq \nu + 1 = M+2$. Note that the ϕs and gs depend on r but not on j, so that (4.29) is valid for all such j. But the number of such j's is denumerably infinite, and they span the whole l space. Hence, since the ϕ's are polynomials, we have

$$\phi_{M+1}(\lambda) - \phi_{M+1}(\lambda - r) = g_{M+2}(\lambda)\phi_{M+2}(\lambda), \tag{4.32}$$

for any λ. Thus (4.30) holds for $\mu = M+1$. Since it holds for $\mu = 0$ [by setting $\nu = 0$ in (4.29)] it holds for all μ. This is the required relation between ϕ_μ and $\phi_{\mu+1}$.

The next step is to substitute $S_r\lambda + \mu r$ for λ in (4.30). We obtain

$$\begin{aligned} &\phi_\mu(S_r\lambda + \mu r) - \phi_\mu(S_r\lambda + \mu r - r) \\ &\qquad = -g_{\mu+1}(\lambda)\phi_{\mu+1}(S_r\lambda + \mu r), \end{aligned} \tag{4.33}$$

because, as is easily verified from (4.8),

$$g_{\mu+1}(\lambda) = -g_{\mu+1}(S_r\lambda + \mu r). \tag{4.34}$$

Adding (4.33) and (4.30), we obtain

$$\psi_\mu(\lambda) - \psi_\mu(\lambda - r) = g_{\mu+1}(\lambda)\psi_{\mu+1}(\lambda), \tag{4.35}$$

which is the same equation as (4.30), but for the quantities

$$\psi_\mu(\lambda) = \phi_\mu(\lambda) - \phi_\mu(S_r\lambda + \mu r - r), \tag{4.36}$$

instead of the $\phi_\mu(\lambda)$.

From (4.35), it is clear that if $\psi_{\mu+1}(\lambda) = 0$, $\psi_\mu(\lambda)$ is periodic in the r direction. But the only periodic polynomial is a constant. Hence, from (4.35),

$$\psi_{\mu+1}(\lambda) = 0 \rightarrow \psi_\mu(\lambda) = \psi_\mu(\lambda + kr), \tag{4.37}$$

where k is any real number.

But then,

$$\psi_\mu(\lambda) = \psi_\mu(\lambda_\perp + \tfrac{1}{2}\mu r),$$

where $\lambda_\perp$ is that part of λ which is orthogonal to r,

$$\begin{aligned} &= \phi_\mu\left(\lambda_\perp + \frac{\mu r}{2}\right) - \phi_\mu\left(S_r\left(\lambda_\perp + \frac{\mu r}{2}\right) + \mu r - r\right), \\ &= \phi_\mu\left(\lambda_\perp + \frac{\mu r}{2}\right) - \phi_\mu\left(\lambda_\perp + \frac{\mu r}{2} - r\right), \\ &= g_{\mu+1}\left(\lambda_\perp + \frac{\mu r}{2}\right)\phi_{\mu+1}\left(\lambda_\perp + \frac{\mu r}{2}\right), \quad \text{from (4.30)} \\ &= 0, \qquad\qquad \text{from (4.34).} \end{aligned} \tag{4.38}$$

Thus

$$\psi_{\mu+1}(\lambda) = 0 \rightarrow \psi_\mu(\lambda) = 0. \tag{4.39}$$

On the other hand, it is clear from (4.22) that, for an invariant $F(X_\lambda)$ of given degree (n, say) $\phi_\mu(\lambda)$ is a constant for $\mu =$ some $\mu_0 \leq n/2$. Hence $\psi_\mu(\lambda) = 0$ for $\mu = \mu_0$, and so $\psi_\mu(\lambda) = 0$ all μ. In

particular,

$$\psi_0(\lambda) = \phi_0(\lambda) - \phi_0(S_r\lambda - r) = 0. \tag{4.40}$$

It remains only to express this result in terms of ϕ_0', i.e.,

$$\phi_0'(\lambda + \tfrac{1}{2}\Sigma) = \phi_0'(S_r\lambda - r + \tfrac{1}{2}\Sigma) \tag{4.41}$$

and then we have

$$\phi_0'\Big(\lambda + \tfrac{1}{2}\Sigma\Big) = \phi_0'\Big(S_r\lambda - \frac{r}{2} + \tfrac{1}{2}\Sigma_r\Big),$$

$$= \phi_0'\Big(S_r\lambda + S_r\frac{r}{2} + S_r\tfrac{1}{2}\Sigma_r\Big),$$

using property (d) above,

$$= \phi_0'(S_r(\lambda + \tfrac{1}{2}\Sigma)),$$

$$= \phi_0'(S(\lambda + \tfrac{1}{2}\Sigma)), \tag{4.42}$$

using property (e). Q.E.D.

5. PROOF OF COMPLETENESS

In this section we shall use the S theorem to prove the completeness of the set of invariants listed in Sec. 3. The method of proof is due to Racah[1,2]

Let $I(X_\lambda)$ be any invariant. In any irreducible representation, we have, from Sec. 4,

$$I(X_\lambda) = \phi_0(j), \tag{5.1}$$

where j is the highest weight. If we can now find a set of I's, $I^{(k)}$, say, such that, conversely, for each irreducible representation

$$j_i = j_i(I^{(k)}), \tag{5.2}$$

where the j_i, $i = 1, \cdots, l$, are the components of j, then the set $I^{(k)}$ will be a complete set, since any other invariant can then be expressed in terms of the $I^{(k)}$ using (5.1) and (5.2). Our task is to show that the invariants listed in Sec. 3 constitute such a set.

To show this, we note first of all that (if we drop I_{2l} for D_l and I_4 for G_2) we have listed in Sec. 3 exactly l invariants for each Lie algebra. Thus for each Lie algebra we have exactly l equations

$$I^{(k)} = \phi_0^{(k)}(j), \tag{5.3}$$

which we can regard as l equations for the l unknowns j_i. One can see that these l equations are algebraically independent by noting that the leading terms in the $\phi_0^{(k)}(j)$ are just the leading terms in the $I^{(k)}$, with $H_i \to j_i$, and the latter, which are exhibited explicitly in Sec. 3, are algebraically independent. (Note that this is not true for I_{2l} of D_l and I_4 of G_2.) This means that the l equations (5.3) will have only discrete solutions, the number of different solutions, p, being less than or equal to the product of the highest powers of the $\phi_0^{(k)}(j)$. What we have to show is that among the p solutions of (5.3), for a given set of values $I^{(k)}$, there can occur at most one highest weight j.

It is to show this that we use the S theorem. The S theorem tells us that if j is a solution of (5.3), so is

$$j_s = Sj + S\tfrac{1}{2}\Sigma - \tfrac{1}{2}\Sigma, \tag{5.4}$$

where S is any element of the Weyl group. Thus the solutions of (5.3) occur in *sets* $\{\cdots j_s \cdots\}$. On the other hand, by calculating Σ explicitly from the standard form of the roots for each Lie algebra one easily sees that

$$S\Sigma < \Sigma, \qquad S \neq 1. \tag{5.5}$$

[For example, for A_l the roots are $e_i - e_k$, $i, k = 1, \cdots, l+1$, $\sum = (l, l-2, \cdots, -l)$ and S is the group of permutations of the components of an $l+1$ vector ($s = (l+1)!$).] Hence, if j is a highest weight ($Sj \leq j$), we have (a)

$$j_s = j_{s'} \to (j + \tfrac{1}{2}\Sigma) = S^{-1}S'(j + \tfrac{1}{2}\Sigma) \to S = S', \tag{5.6}$$

and (b)

$$j_s = Sj + S\tfrac{1}{2}\Sigma - \tfrac{1}{2}\Sigma < j + \tfrac{1}{2}\Sigma - S^{-1}\tfrac{1}{2}\Sigma$$

$$= S^{-1}j_s, \qquad S \neq 1, \tag{5.7}$$

which means that, if the set $\{\cdots j_s \cdots\}$ contains a highest weight j, (a) it contains s distinct elements j_s; (b) only one of these (j) can be a highest weight.

This result holds for any set of l algebraically independent invariants $I^{(k)}$. What distinguishes the l invariants of Sec. 3 is that these allow only *one* set $\{\cdots j_s \cdots\}$ for each set of values $I^{(k)}$, and hence allow only one highest weight (at most), as required. To see this, one simply notes that p, the maximum number of distinct solutions permitted, is equal to the product of the orders of the invariants used, and that for the l invariants of Sec. 3, this product is, in the case of each Lie algebra, just equal to the order s of the corresponding Weyl group. Q.E.D.

In conclusion, we note that if for D_l we had used I_{2l} instead of $I_l(\Delta^+)$ we would have obtained $p = 2s$ and thus two sets of solutions $\{\cdots j_s \cdots\}$ of (5.3), and two possible highest weights j, for some values of the $I^{(k)}$. This corresponds to the fact that the

I_{2r}, $r = 1, \cdots, l$, do not distinguish between the representations Δ^+ and Δ^-.

ACKNOWLEDGMENTS

The authors would like to thank Dr. J. T. Lewis, who brought this problem to their notice, and to thank him and Dr. N. Straumann and Dr. F. Kamber for many illuminating discussions concerning it.

One of them (O'R) would like to thank the Director and Board of Governors of the Madras Institute of Mathematical Sciences for their invitation to visit that Institute, during which visit some of the above work was carried out, and to thank the Director and Governing Board of the School of Theoretical Physics, Dublin Institute for Advanced Studies, for granting him leave of absence for that visit.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Addendum: One-Speed Neutron Transport in Two Adjacent Half-Spaces†

M. R. MENDELSON*
AND
G. C. SUMMERFIELD
Department of Nuclear Engineering, The University of Michigan
Ann Arbor, Michigan
(Received 7 July, 1964)
[J. Math. Phys. 5, 668 (1964)]

The interface current for the problem of two half-spaces with a constant source in one half-space is obtained in closed form.

THE interface current is

$$j(0) = \int_{-1}^{1} \mu\psi(0, \mu)\, d\mu,$$

or

$$j(0) = -\lim_{z\to\infty} z \int_{-1}^{1} d\mu \frac{\mu\psi(0, \mu)}{\mu - z}.$$

Using[1] Eq. (V-9) for $\psi(0, \mu)$, Eq. (III-13) and (III-14), and

$$z^2\chi(-z) + 1 = \frac{c_1}{2(1 - c_1)} \int_{-1}^{0} \frac{\mu^3 X_2(\mu)\, d\mu}{X_1(\mu)(\nu_{01}^2 - \mu^2)(\mu - z)} + \frac{c_2}{2(1 - c_2)} \int_{0}^{1} \frac{\mu^3 X_1(-\mu)\, d\mu}{X_2(-\mu)(\nu_{02}^2 - \mu^2)(\mu - z)},$$

we obtain

$$j(0) = \frac{2s(\nu_{02} - \nu_{01})(1 - c_2)}{c_1 - c_2}.$$

† Work supported in part by the U. S. Atomic Energy Commission and in part by the Marquardt Corporation.
* Rackham predoctoral fellow, now at the Knolls Atomic Power Laboratory, Schenectady, New York.

[1] I. Kuscer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. (N. Y.) (To be published).

I_{2r}, $r = 1, \cdots, l$, do not distinguish between the representations Δ^+ and Δ^-.

ACKNOWLEDGMENTS

The authors would like to thank Dr. J. T. Lewis, who brought this problem to their notice, and to thank him and Dr. N. Straumann and Dr. F. Kamber for many illuminating discussions concerning it.

One of them (O'R) would like to thank the Director and Board of Governors of the Madras Institute of Mathematical Sciences for their invitation to visit that Institute, during which visit some of the above work was carried out, and to thank the Director and Governing Board of the School of Theoretical Physics, Dublin Institute for Advanced Studies, for granting him leave of absence for that visit.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 12 DECEMBER 1964

Addendum: One-Speed Neutron Transport in Two Adjacent Half-Spaces†

M. R. MENDELSON*
AND
G. C. SUMMERFIELD
Department of Nuclear Engineering, The University of Michigan
Ann Arbor, Michigan
(Received 7 July, 1964)
[J. Math. Phys. 5, 668 (1964)]

The interface current for the problem of two half-spaces with a constant source in one half-space is obtained in closed form.

THE interface current is

$$j(0) = \int_{-1}^{1} \mu\psi(0, \mu)\, d\mu,$$

or

$$j(0) = -\lim_{z\to\infty} z \int_{-1}^{1} d\mu \frac{\mu\psi(0, \mu)}{\mu - z}.$$

Using[1] Eq. (V-9) for $\psi(0, \mu)$, Eq. (III-13) and (III-14), and

$$z^2\chi(-z) + 1 = \frac{c_1}{2(1 - c_1)} \int_{-1}^{0} \frac{\mu^3 X_2(\mu)\, d\mu}{X_1(\mu)(\nu_{01}^2 - \mu^2)(\mu - z)} + \frac{c_2}{2(1 - c_2)} \int_{0}^{1} \frac{\mu^3 X_1(-\mu)\, d\mu}{X_2(-\mu)(\nu_{02}^2 - \mu^2)(\mu - z)},$$

we obtain

$$j(0) = \frac{2s(\nu_{02} - \nu_{01})(1 - c_2)}{c_1 - c_2}.$$

† Work supported in part by the U. S. Atomic Energy Commission and in part by the Marquardt Corporation.
* Rackham predoctoral fellow, now at the Knolls Atomic Power Laboratory, Schenectady, New York.

[1] I. Kuscer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. (N. Y.) (To be published).