JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 7, NUMBER 2 FEBRUARY 1966

Structure of Space and the Formalism of Relativistic Quantum Theory. III

YRJÖ AHMAVAARA
Institute of Mathematical Physics, University of Turku, Turku, Finland
(Received 1 June 1965)

The elastic scattering of two scalar particles having equal masses is considered in the formalism of relativistic quantum theory over a Galois field $GF(q)$. The scattering function σ determining the cross section is introduced. It is determined by the geometrical relations of Euclidicity, to be imposed on observable 4-momenta in a finite geometry. Thus the requirement of Euclidicity of observable 4-momenta can be considered as the counterpart, in a finite geometry, of the requirement of the analyticity of the invariant amplitude used in conventional S-matrix theory for the determination of the cross section.

I. MAXIMAL ORDER IN A GALOIS FIELD

LET q be a prime, and a be an integer. Let us denote

$$a_q = \{n; n = \text{integer}, \quad n \equiv a(\text{mod } q)\},$$

$$GF(q) = \{0_q, 1_q, 2_q, \cdots, (q-1)_q\}.$$

By the definitions

$$a_q + b_q = (a+b)_q \quad \text{and} \quad a_q b_q = (ab)_q,$$

$GF(q)$ becomes a field of numbers known as a "Galois field".[1]

Let us from now on restrict ourselves to such Galois fields for which the fundamental prime q is of the particular form

$$q = 8nq_1q_2 \cdots q_k - 1, \tag{1}$$

where n is a positive integer, and the $q_1, q_2, \cdots, q_k$ are the k first primes. Let us denote

$$E[a] = \{a_q, (a+1)_q, \cdots, (a+q_{k+1}-1)_q\} \subset GF(q).$$

By the definition

$$a_q > b_q \quad \text{if and only if} \quad a_q - b_q = \rho^{2h},$$

where ρ is a primitive number[1] of $GF(q)$ and h is an integer depending on $a_q - b_q$, every set $E[a]$ of $GF(q)$ becomes transitively ordered.[2] In other words, for every two elements $(a+j)_q \in E[a]$ and $(a+l)_q \in E[a]$ one has

$$(a+j)_q > (a+l)_q \quad \text{if and only if} \quad j > l.$$

If, in particular,

$$(q_{k+1})_q = \rho^{2h+1}, \tag{2}$$

where h is an integer, then the sequences $E[a]$ are the longest sequences of successive integers of the Galois field that can be transitively ordered. Let us call such $E[a]$ the "Euclidean chains"[4] of $GF(q)$. Their "length" can be defined to be equal to $E[a] = q_{k+1}$. A Galois field $GF(q)$ obeying the conditions (1) and (2) will be called a "maximally ordered Galois field," there being a "maximal order of the length q_{k+1}" in this field.

The importance of maximally ordered Galois fields is due to the fact that the Euclidean chains of such a field can be isomorphically mapped to q_{k+1} points

[1] See, e.g., L. Dickson, *Linear Groups* (Dover Publications, Inc., New York, 1958).

[2] P. Kustaanheimo, Soc. Sci. Fenn. 15, No. 19 (1950). See also Part I of this series of papers.[3]

[3] Part I: Y. Ahmavaara, J. Math. Phys. 6, 87 (1965). Part II: Y. Ahmavaara, J. Math. Phys. 6, 220 (1965).

[4] In accordance with F. Levi, Zentr. Math. 39, 156 (1951); a review of Kustaanheimo, Ref. 2.

of the real axis. Let R be the field of the real numbers, and let us denote

$$R[a] = \{a\epsilon, (a+1)\epsilon, \cdots, (a+q_{k+1}-1)\epsilon\} \subset R,$$

where a is an integer, and ϵ is a positive real number. Then the function φ from $E[a]$ to $R[a]$ defined by

$$\varphi(b_a) = b\epsilon$$

is an isomorphism. This isomorphism holds both for the addition, for the multiplication, and for the order of the respective elements of $E[a]$ and $R[a]$:

$$\begin{cases} \varphi(a_a + b_a) = \varphi((a+b)_a) = (a+b)\epsilon, \\ \varphi(a_a b_a) = \varphi((ab)_a) = ab\epsilon, \\ a_a > b_a \Leftrightarrow \varphi(a_a) = a\epsilon > \varphi(b_a) = b\epsilon \Leftrightarrow a > b. \end{cases}$$

II. THE MOMENTUM SPACE OVER A MAXIMALLY ORDERED GALOIS FIELD

Let $EG(4, q)$ be a four-dimensional linear space over a maximally ordered Galois field $GF(q)$. Every element p of $EG(4, q)$ is a vector having four components, each of the components being a Galois number. Let us denote these components as follows:

$$p = (p_0, p_1, p_2, p_3) \in EG(4, q),$$
$$p_\alpha \in GF(q) \forall \qquad \alpha = 0, 1, 2, 3.$$

Consider $p \in EG(4, q)$, such that

$$\begin{aligned} p_1^2 + p_2^2 + p_3^2 &\geq 0_a, \\ p_0^2 - p_1^2 - p_2^2 - p_3^2 &\geq 0_a. \end{aligned} \tag{3}$$

For such p there exist the Galois numbers κ and μ such that

$$\begin{aligned} p_1^2 + p_2^2 + p_3^2 &= \kappa^2, \\ p_0^2 - p_1^2 - p_2^2 - p_3^2 &= \mu^2. \end{aligned} \tag{4}$$

Let us define a subset E of $EG(4, q)$ by $p \in E$ if and only if the relations (4) and the formulas

$$\begin{aligned} &p_1 \in E[-\tfrac{1}{2}(q_{k+1}-1)], \quad p_2 \in E[-\tfrac{1}{2}(q_{k+1}-1)], \\ &p_3 \in E[-\tfrac{1}{2}(q_{k+1}-1)], \quad p_0 \in E[0], \\ &\kappa \in E[0], \qquad\qquad \mu \in E[0] \end{aligned} \tag{5}$$

hold good. Denoting

$$p_1 = a_a, \quad p_2 = b_a, \quad p_3 = c_a, \quad p_0 = d_a,$$
$$\kappa = k_a, \quad \mu = m_a,$$

the relations (4) and (5) read

$$a^2 + b^2 + c^2 = k^2, \qquad d^2 - k^2 = m^2, \tag{4'}$$

$$a, b, c = 0, \pm 1, \pm 2, \cdots, \pm\tfrac{1}{2}(q_{k+1}-1), \tag{5'}$$
$$d, k, m = 0, 1, 2, \cdots, q_{k+1} - 1,$$

respectively.

Let $\Sigma(m)$ be the number of the solutions (a, b, c, d, k, m) of the Diophantine equations (4') by the values (5') for a fixed value of m. Let $\Sigma(m, d)$ be the number of such solutions for fixed values of m and d.

Performing the mapping φ from $E[-\frac{1}{2}(q_{k+1}-1)]$ to $R[-\frac{1}{2}(q_{k+1}-1)]$ and from $E[0]$ to $R[0]$ the elements $p \in E$ of $EG(4, q)$ are mapped to the elements r of the real four-dimensional linear space R^4, such that

$$r_1 = a\epsilon, \quad r_2 = b\epsilon, \quad r_3 = c\epsilon, \quad r_0 = d\epsilon,$$
$$r_1^2 + r_2^2 + r_3^2 = k^2\epsilon^2, \qquad r_0^2 - r_1^2 - r_2^2 - r_3^2 = m^2\epsilon^2.$$

The elements r form a finite lattice L of points in R^4. The mapping Φ from E to the lattice L so induced is an isomorphism both with respect to all the rational operations on the coordinates of the points, and with respect to the order of the points on straight lines. In view of this isomorphism E may be called a particular "Euclidean lattice" of points of the finite space $EG(4, q)$.

Physical Interpretation

The vectors $p \in EG(4, q)$ obeying the relativistic energy–momentum relations (4) represent the 4-momenta in a finite geometry. In particular, the Euclidean lattice $E \subset EG(4, q)$ is the "physical domain" of momentum, that is, the vectors $p \in E$ represent the observable 4-momenta. If a physical system is in a quantum theoretical state characterized by a momentum $p \in E$, the measurement of 4-momentum yields the result $\Phi(p) = r$ (in a suitably chosen unit ϵ of energy). The function $\Sigma(m)$ thus gives the number of the observable states of momentum for a fixed value of mass, and the function $\Sigma(m, d)$ gives the number of the observable states of momentum for fixed values of mass and energy. Accordingly, $\Sigma(m)$ can be considered as the mass spectrum, and $\Sigma(m, d)$ as the energy spectrum of a particle with mass m, in a finite geometry. All these interpretations were introduced in Part I of this series of papers.[3]

III. THE SCATTERING FUNCTION

Consider four elements p, n, p', and n' of $E \subset EG(4, q)$, such that

$$\begin{aligned} p_0^2 - p_1^2 - p_2^2 - p_3^2 &= n_0^2 - n_1^2 - n_2^2 - n_3^2 = m_a^2, \\ p_0'^2 - p_1'^2 - p_2'^2 - p_3'^2 &= n_0'^2 - n_1'^2 - n_2'^2 - n_3'^2 = m_a^2, \\ p_1^2 + p_2^2 + p_3^2 &= n_1^2 + n_2^2 + n_3^2 = k_a^2, \\ p_1'^2 + p_2'^2 + p_3'^2 &= n_1'^2 + n_2'^2 + n_3'^2 = k_a^2, \end{aligned} \tag{6}$$

$$p_1 + n_1 = p_2 + n_2 = p_3 + n_3 = 0_q,$$
$$p_1' + n_1' = p_2' + n_2' = p_3' + n_3' = 0_q,$$
$$p_0 + n_0 = p_0' + n_0' = E_q \in E[0],$$
$$(p_1 - p_1')^2 + (p_2 - p_2')^2 + (p_3 - p_3')^2 = \Delta_q^2,$$
$$\Delta_q \in E[0].$$

Denoting

$$\begin{aligned} p_1 &= a_q, & p_2 &= b_q, & p_3 &= c_q, & p_0 &= d_q, \\ p_1' &= a_q', & p_2' &= b_q', & p_3' &= c_q', & p_0' &= d_q', \\ n_1 &= A_q, & n_2 &= B_q, & n_3 &= C_q, & n_0 &= D_q, \\ n_1' &= A_q', & n_2' &= B_q', & n_3' &= C_q', & n_0' &= D_q', \end{aligned} \tag{7}$$

the conditions (6) can be written in the form of the Diophantine equations

$$\begin{aligned} a^2 + b^2 + c^2 &= A^2 + B^2 + C^2 = k^2, \\ a'^2 + b'^2 + c'^2 &= A'^2 + B'^2 + C'^2 = k^2, \\ d^2 &= D^2 = d'^2 = D'^2 = m^2 + k^2, \\ a + A &= b + B = c + C = 0, \\ a' + A' &= b' + B' = c' + C' = 0, \\ 4(k^2 + m^2) &= E^2, \\ (a - a')^2 + (b - b')^2 + (c - c')^2 &= \Delta^2, \end{aligned} \tag{8}$$

for the integers

$$\begin{aligned} &a, b, c, A, B, C, a', b', c', A', B', C' \\ &\qquad = 0, \pm 1, \cdots, \pm(q_{k+1} - 1)/2, \\ &d, D, d', D', k, m, E, \Delta = 0, 1, 2, \cdots, (q_{k+1} - 1). \end{aligned} \tag{9}$$

Let $\sigma(m, E, \Delta)$ be the number of the solutions $(a, b, c, d, A, B, C, D, a', b', c', d', A', B', C', D', k, m, E, \Delta)$ of the Diophantine equations (8) by the values (9) for fixed m, E, and Δ.

Physical Interpretation

Consider an elastic scattering process $p + n \rightarrow p' + n'$, where p, n, p', and n' are the Galois-number 4-vectors representing (by the isomorphism Φ) the 4-momenta of the particles in question. Let all these particles be scalar particles with identical masses represented by the Galois number m_q. The conservation of momentum requires that

$$p + n = p' + n', \tag{10}$$

and the three quadratic invariants of the process are represented by

$$\begin{aligned} s &= (p_0 + n_0)^2 - (p_1 + n_1)^2 - (p_2 + n_2)^2 - (p_3 + n_3)^2, \\ t &= (p_0 - p_0')^2 - (p_1 - p_1')^2 - (p_2 - p_2')^2 - (p_3 - p_3')^2, \\ u &= (p_0 - n_0')^2 - (p_1 - n_1')^2 - (p_2 - n_2')^2 - (p_3 - n_3')^2. \end{aligned} \tag{11}$$

These are connected with one another by the relation

$$s + t + u = 4m_q^2.$$

In the center-of-mass system of the incoming particles p and n, one has

$$p_1 + n_1 = p_2 + n_2 = p_3 + n_3 = 0_q, \qquad p_0 = n_0,$$

and, in view of the conservation of momentum (10), the further relations

$$p_1' + n_1' = p_2' + n_2' = p_3' + n_3' = 0_q,$$
$$p_0' = n_0' = p_0 = n_0.$$

The variables k_q, E_q, and Δ_q defined by (6) thus represent the center-of-mass (c.m.) 3-momentum, the c.m. total energy, and the c.m. transfer of momentum of the process, respectively. If θ is the scattering angle between the particles p and p' in the c.m. system, one has

$$\begin{aligned} s &= E_q^2 = 4k_q^2 + 4m_q^2, \\ t &= -\Delta_q^2 = -2k_q^2(1 - \cos\theta), \\ u &= -2k_q^2(1 + \cos\theta). \end{aligned}$$

The value of the function σ for fixed m, E, and Δ evidently gives the number of open channels of observable momentum for the process $p + n \rightarrow p' + n'$ for fixed values of the mass, the scattering energy, and the transfer of momentum. This function can be considered as the theoretical prediction for the cross section [apart from a factor of proportionality, cf. (13′)] in a finite geometry.

IV. AN OUTLINE OF THE MODEL COMPUTATIONS

The spectral function $\Sigma(m)$ and the scattering function $\sigma(m, E, \Delta)$ together determine completely the existing scalar particles (the masses) and their elastic scattering processes in the present formalism of relativistic quantum theory over a Galois field.

For every prime q_{k+1} there is a particular spectral function $\Sigma(m)$ and a particular scattering function $\sigma(m, E, \Delta)$. After the choice of a particular value of q_{k+1} one first has to compute the function $\Sigma(m)$. Hereafter one can compute the scattering function $\sigma(m, E\Delta)$ for the particular values of m for which there is a peak in the function $\Sigma(m)$. These m-values represent the common (scalar) particles of the finite

world in question. It would be interesting to see how the functions $\Sigma(m)$ and $\sigma(m, E, \Delta)$ change when q_{k+1} is changed and, in particular, how these functions behave for large q_{k+1} (this should correspond to a realistic physical situation).

After the generalization of the scattering function σ to the case of nonelastic scattering of scalar particles having different masses (this can be done in principle in a similar way that was used here for elastic scattering of equal-mass particles), one can also compute these processes of scattering in the chosen model.

In order to consider the interactions of nonscalar particles in the finite model recourse must be taken to the general group formalism of interactions described in Part II of this series of papers.[3] One can, for instance, consider the inhomogeneous Lorentz group (the Poincaré group) over a field $GF(q)$ with a maximal order of the length q_{k+1} [it follows from a theorem of Dirichlet that for any chosen value of q_{k+1} there is an infinite sequence of the primes q satisfying the formula (1)]. If this group is chosen to be the relativity group of space–time, one has to construct the irreducible unitary representations U of this group in terms of ordinary complex-number matrices. Each of these representations is characterized by two indices M and S (see Part I[3]): U_{MS}. To be specific, pick out three particular mass peaks m_1, m_2, and m_3 of the model function $\Sigma(m)$, and consider the irreducible unitary representations corresponding to the values $M_1 = -(m_1)_q^2$, $M_2 = -(m_2)_q^2$, and $M_3 = -(m_3)_q^2$. Choose also some values S_1, S_2, and S_3 of the index S. Find the reduction

$$U_{M_1S_1} \otimes U^*_{M_2S_2} \sim U^*_{M_3S_3} \oplus \cdots .$$

The reduction coefficients g (cf. Part II[3]) can now be substituted to the S-matrix elements of any desired graph describing the mutual interactions of the particles (M_1S_1), (M_2S_2), and (M_3S_3). If

$$\langle |S| \rangle = \langle p_1s_1, p_2s_2, \cdots |S| p_1's_1', p_2's_2', \cdots \rangle \qquad (12)$$

is an S-matrix element calculated with the help of the g-coefficients, one must at the next stage compute the sum

$$\sum_{\text{channels}} |\langle |S| \rangle|^2 \qquad (13)$$

over all the channels of the process discussed. The number of these channels of course depends on what variables are observed in the process. The sums of the form (13) for fixed values of the observed variables then give the final theoretical prediction of interaction in question.

V. COMPARISONS OF DIFFERENT FORMALISMS

The Scattering Function and the Group Formalism of Interactions

For the scattering $p + n \to p' + n'$ of scalar particles the S-matrix elements (12) reduce to a form

$$\langle |S| \rangle = \langle p, n \, |S| \, p', n' \rangle .$$

On the other hand, the only information contained in the reduction coefficients g in the case of scalar particles is the conservation of momentum: $p + n = p' + n'$. Accordingly, the S-matrix element describing the scattering of scalar particles reduces further to the form

$$\langle p, n \, |S| \, p', n' \rangle = \langle p, n \mid p', n' \rangle + A \langle p + n \mid p' + n' \rangle, \qquad (12')$$

which can also be written as

$$\langle p, n | \, S - 1 \, | p', n' \rangle = A \langle p + n \mid p' + n' \rangle. \qquad (12'')$$

Here A is a (complex) constant, and $\langle p + n | p' + n' \rangle$ is equal to the Kronecker function $\delta_{p+n,p'+n'}$. Thus the sum (13) over the channels of the process becomes simply the sum over the channels of observable momentum for which the total momentum is conserved. Leaving aside the trivial case in which $p = p'$ and $n = n'$, one has then for particles having identical masses the result

$$\sum_{\text{channels}} |\langle p, n \, |S| \, p', n' \rangle|^2 = |A|^2 \times \times \sum_{\text{channels}} \langle p + n \mid p' + n' \rangle = |A|^2 \, \sigma(m, E, \Delta). \qquad (13')$$

Thus the description of the scattering of scalar particles by the scattering function is in accordance with the general group formalism of interaction developed in Part II of this series of papers.[3]

The Finite Formalism and the Conventional S-Matrix Theory

If the momentum space $EG(4, q)$ over the Galois field $GF(q)$ is replaced by the momentum space R^4 over the real numbers, the spectral function $\Sigma(m)$, as well as the scattering function $\sigma(m, E, \Delta)$ become trivial density functions. Accordingly, there is no (nontrivial) geometrical mass spectrum in the momentum space over the real numbers, and no (nontrivial) geometrical scattering function. Thus the finite formalism of relativistic quantum theory developed in this series of papers does not give nontrivial results in the ordinary geometry. The nontrivial results of the finite formalism are due to the application of a finite geometry.

In the conventional S-matrix theory, associated with a momentum space over the real numbers, one assumes that the invariant amplitude A of the formulas (12′), (12″), and (13′) is not a constant but a function of the mutually independent variables s and t: $A = A(s, t)$. The requirement of the analyticity of the function $A(s, t)$ when extended to complex variables s and t, is then used for the determination of the cross sections.

In the finite formalism of relativistic quantum theory, where the general group formalism of Part II is applied, the amplitude A is a constant. Thus the amplitude is now trivial, while the density functions $\Sigma(m)$ and $\sigma(m, E, \Delta)$ are not. The scattering of scalar particles is now determined by the nontrivial scattering function $\sigma(m, E, \Delta)$. Thus the requirement of the Euclidicity of the observable 4-momenta, on the basis of which the function $\sigma(m, e, \Delta)$ was constructed, can be considered as the counterpart, in a finite geometry, of the requirement of the analyticity of the invariant amplitude, used for the construction of cross sections in the ordinary S-matrix theory applying the ordinary geometry.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 2 FEBRUARY 1966

Structure of Space and the Formalism of Relativistic Quantum Theory. IV

YRJÖ AHMAVAARA
Institute of Mathematical Physics, University of Turku, Turku, Finland
(Received 20 July 1965)

The decay functions determining the lifetimes and the stability of particles in finite models of relativistic quantum theory are considered.

I. THE DECAY FUNCTION: DECAY INTO TWO PARTICLES

THIS is the last one in a series of papers considering the formalism of relativistic quantum theory over a finite geometry. One of the possibilities of finite geometry is the purely group theoretical ("unitary symmetry") character of the formalism, as was emphasized in Parts I and II before.[1] Another consequence of finite geometry is the existence of the conditions of Euclidicity to be imposed on the observable 4-momenta. These conditions determine three important functions of momentum, viz.

(1) *the spectral function* Σ determining the mass spectrum in finite geometry, and introduced in Part I,

(2) *the scattering function* σ determining the momentum dependence of the scattering of particles, and introduced for the elastic scattering of scalar particles in Part III,[2] and

(3) *the decay function* δ determining the lifetimes and the stability of particles in a finite geometry. This last function will be considered in the present paper.

The Euclidean lattice of the momentum space (cf. Ref. 2) could be chosen in several ways. Applying the same choice that has been made in the previous papers of this series, and using again the center-of-mass system of the process under consideration, one obtains the following decay function for a decay into two particles, in the simplest possible case.

Consider the decay of a scalar particle with rest mass M into two scalar particles having the rest masses m_1 and m_2, respectively. Let the 4-momenta of these particles be denoted as P, p_1, and p_2, respectively, so that the decay process in question can be denoted as $P \rightarrow p_1 + p_2$.

In order that the masses and the 4-momenta are observable the masses, the energies, and the absolute values of the 3-momenta in question must belong to the Euclidean chain $E[0]$ of the underlying Galois field (for a closer study see Ref. 2), and the components of the 3-momenta must belong to the chain $E[-\frac{1}{2}(q_{k+1} - 1)]$. Here q_{k+1} is the prime giving the length of the maximal order in the underlying Galois field (see Ref. 2).

[1] Y. Ahmavaara, J. Math. Phys. **6**, 87 (1965) (Part I); *ibid.* 223 (1965) (Part II).
[2] Y. Ahmavaara, J. Math. Phys. **7**, 197 (1966) (previous paper).

In the conventional S-matrix theory, associated with a momentum space over the real numbers, one assumes that the invariant amplitude A of the formulas (12′), (12″), and (13′) is not a constant but a function of the mutually independent variables s and t: $A = A(s, t)$. The requirement of the analyticity of the function $A(s, t)$ when extended to complex variables s and t, is then used for the determination of the cross sections.

In the finite formalism of relativistic quantum theory, where the general group formalism of Part II is applied, the amplitude A is a constant. Thus the amplitude is now trivial, while the density functions $\Sigma(m)$ and $\sigma(m, E, \Delta)$ are not. The scattering of scalar particles is now determined by the nontrivial scattering function $\sigma(m, E, \Delta)$. Thus the requirement of the Euclidicity of the observable 4-momenta, on the basis of which the function $\sigma(m, e, \Delta)$ was constructed, can be considered as the counterpart, in a finite geometry, of the requirement of the analyticity of the invariant amplitude, used for the construction of cross sections in the ordinary S-matrix theory applying the ordinary geometry.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 2 FEBRUARY 1966

Structure of Space and the Formalism of Relativistic Quantum Theory. IV

YRJÖ AHMAVAARA
Institute of Mathematical Physics, University of Turku, Turku, Finland
(Received 20 July 1965)

The decay functions determining the lifetimes and the stability of particles in finite models of relativistic quantum theory are considered.

I. THE DECAY FUNCTION: DECAY INTO TWO PARTICLES

THIS is the last one in a series of papers considering the formalism of relativistic quantum theory over a finite geometry. One of the possibilities of finite geometry is the purely group theoretical ("unitary symmetry") character of the formalism, as was emphasized in Parts I and II before.[1] Another consequence of finite geometry is the existence of the conditions of Euclidicity to be imposed on the observable 4-momenta. These conditions determine three important functions of momentum, viz.

(1) *the spectral function* Σ determining the mass spectrum in finite geometry, and introduced in Part I,

(2) *the scattering function* σ determining the momentum dependence of the scattering of particles, and introduced for the elastic scattering of scalar particles in Part III,[2] and

(3) *the decay function* δ determining the lifetimes and the stability of particles in a finite geometry. This last function will be considered in the present paper.

The Euclidean lattice of the momentum space (cf. Ref. 2) could be chosen in several ways. Applying the same choice that has been made in the previous papers of this series, and using again the center-of-mass system of the process under consideration, one obtains the following decay function for a decay into two particles, in the simplest possible case.

Consider the decay of a scalar particle with rest mass M into two scalar particles having the rest masses m_1 and m_2, respectively. Let the 4-momenta of these particles be denoted as P, p_1, and p_2, respectively, so that the decay process in question can be denoted as $P \rightarrow p_1 + p_2$.

In order that the masses and the 4-momenta are observable the masses, the energies, and the absolute values of the 3-momenta in question must belong to the Euclidean chain $E[0]$ of the underlying Galois field (for a closer study see Ref. 2), and the components of the 3-momenta must belong to the chain $E[-\frac{1}{2}(q_{k+1} - 1)]$. Here q_{k+1} is the prime giving the length of the maximal order in the underlying Galois field (see Ref. 2).

[1] Y. Ahmavaara, J. Math. Phys. **6**, 87 (1965) (Part I); *ibid.* 223 (1965) (Part II).
[2] Y. Ahmavaara, J. Math. Phys. **7**, 197 (1966) (previous paper).

To simplify the notation one can assume that the momentum variables in question are equal to the integers associated with the Euclidean chains in question. Accordingly, the masses M, m_1, and m_2, the energies P_0, p_{10}, and p_{20}, and the absolute values of the 3-momenta, K, k_1, and k_2, are elements of the sequence

$$0, 1, 2, \cdots, q_{k+1} - 1, \tag{1}$$

and the components of the 3-momenta, P_j, p_{1j}, and p_{2j}, for $j = 1, 2, 3$, are elements of the sequence

$$-\tfrac{1}{2}(q_{k+1} - 1), \cdots, -2, -1, 0, 1, 2, \cdots, \tfrac{1}{2}(q_{k+1} - 1). \tag{2}$$

The conservation of momentum thus reads:

$$\begin{gathered} 0 = p_{1j} + p_{2j}, \qquad j = 1, 2, 3, \\ (p_{11}^2 + p_{12}^2 + p_{13}^2)^{\frac{1}{2}} = (p_{21}^2 + p_{22}^2 + p_{32}^2)^{\frac{1}{2}} = k, \\ M = p_{10} + p_{20}. \end{gathered} \tag{3}$$

The relativistic mass relations give the further equations

$$m_1^2 = p_{10}^2 - k^2, \qquad m_2^2 = p_{20}^2 - k^2. \tag{4}$$

The problem to be considered is the following: Given three integers M, m_1, and m_2 from the sequence (1), such that, say, $m_1 \geq m_2$, find the number $\delta(M, m_1, m_2)$ of the solutions of (3) and (4) by the allowed integer values (1) or (2) of the respective variables. This number δ then gives the number of the open channels of momentum for the process $P \rightarrow p_1 + p_2$.

One can readily solve the equations for k, p_{10}, and p_{20}:

$$\begin{aligned} p_{10} &= (M^2 + m_1^2 - m_2^2)/2M, \\ p_{20} &= M - p_{10}, \\ k &= (p_{10}^2 - m_1^2)^{\frac{1}{2}}. \end{aligned} \tag{5}$$

In order that p_{10} belong to the sequence (1) it is thus necessary that

$$(M^2 + m_1^2 - m_2^2)/2M = \text{integer} \tag{6}$$

from the sequence (1). Accordingly, the integer $M^2 + m_1^2 - m_2^2$ must be divisible by $2M$.

In order that p_{20} belongs to the sequence (1) one obviously must have

$$M - p_{10} = M/2 - (m_1^2 - m_2^2)/2M \geq 0. \tag{7}$$

This is equivalent to

$$M^2 \geq m_1^2 - m_2^2. \tag{7'}$$

In order that k belongs to the sequence (1) it is necessary that $p_{10} \geq m_1$ or, what is the same,

$$M/2 + (m_1^2 - m_2^2)/2M \geq m_1. \tag{8}$$

In view of the assumed relation $m_1 \geq m_2$ this condition is equivalent to

$$M \geq m_1 + m_2. \tag{8'}$$

Evidently, (7') is a consequence of (8').

Furthermore one must have the condition that

$$p_{10}^2 - m_1^2 = \text{square of an integer} \tag{9}$$

from the sequence (1).

The decay function $\delta(M, m_1, m_2)$ can now be constructed in the following way: Given the numbers M, m_1, and m_2, such that $m_1 \geq m_2$,

(1) Check whether $M \geq m_1 + m_2$. If not, $\delta(M, m_1, m_2) = 0$.

(2) Check whether $(M^2 + m_1^2 - m_2^2)/2M$ is an integer from the sequence (1). If not, $\delta(M, m_1, m_2) = 0$. If so, put

$$(M^2 + m_1^2 - m_2^2)/2M = \nu. \tag{10}$$

(3) Check whether $(\nu^2 - m_1^2)^{\frac{1}{2}}$ is an integer from the sequence (1). If not, $\delta(M, m_1, m_2) = 0$. If so, put

$$(\nu^2 - m_1^2)^{\frac{1}{2}} = k. \tag{11}$$

(4) Find the number of the solutions (k_1, k_2, k_3) of

$$k^2 = k_1^2 + k_2^2 + k_3^2 \tag{12}$$

by the values k_1, k_2, and k_3 taken from the sequence (2). This number gives the value of $\delta(M, m_1, m_2)$ in the case that $\delta(M, m_1, m_2) \neq 0$.

Connection to the S-Matrix Formalism

Evidently, the decay function should have a similar kind of connection to the S-matrix formalism as the scattering function σ was shown to have.[2] If one considers all the channels through which a scalar particle having the mass M can decay into two scalar particles having the masses m_1 and m_2, and sums over the probabilities of the decays through the different channels, one should obtain

$$\sum_{\text{channels}} |\langle M| S |m_1, m_2\rangle|^2 = |B|^2 \delta(M, m_1, m_2). \tag{13}$$

Here B is a complex constant [cf. Ref. 2, Eq. (22')]. Thus the decay function $\delta(M, m_1, m_2)$ should represent, apart from a constant factor, the sum of all the Feynman graphs beginning with an incoming scalar particle M, and ending with two outgoing scalar particles m_1 and m_2.

II. THE DECAY FUNCTION: DECAY INTO THREE PARTICLES

Consider now a decay $P \to p_1 + p_2 + p_3$ of a scalar particle having the mass M into three scalar particles having the masses m_1, m_2, and m_3. In the center-of-mass system the conservation of momentum reads

$$\begin{aligned} 0 &= p_{1j} + p_{2j} + p_{3j}, \qquad j = 1, 2, 3, \\ M &= p_{10} + p_{20} + p_{30}, \end{aligned} \tag{14}$$

and the relativistic mass relations give

$$\begin{aligned} m_1^2 &= p_{10}^2 - q_1^2, \qquad q_1^2 = p_{11}^2 + p_{12}^2 + p_{13}^2, \\ m_2^2 &= p_{20}^2 - q_2^2, \qquad q_2^2 = p_{21}^2 + p_{22}^2 + p_{23}^2, \\ m_3^2 &= p_{30}^2 - q_3^2, \qquad q_3^2 = p_{31}^2 + p_{32}^2 + p_{33}^2. \end{aligned} \tag{15}$$

Here q_1, q_2, and q_3 are the absolute values of the three 3-momenta in question.

The decay function $\delta(M, m_1, m_2, m_3)$ can now be constructed, evidently, in the following way:

(1) Find the solutions

$$\begin{aligned} &(m_1, q_1, p_{10}, p_{11}, p_{12}, p_{13}), \\ &(m_2, q_2, p_{20}, p_{21}, p_{22}, p_{23}), \\ &(m_3, q_3, p_{30}, p_{31}, p_{32}, p_{33}) \end{aligned} \tag{16}$$

of the equations (15) by the integers of the respective sequences (1) or (2), the integers m_1, m_2, m_3, q_1, q_2, q_3, p_{10}, p_{20}, and p_{30} being to be taken from the sequence (1), and the remaining integers from the sequence (2).

(2) Compute, for every combination (p_{10}, p_{20}, p_{30}), the value

$$M = p_{10} + p_{20} + p_{30}. \tag{17}$$

(3) Find, for every value of M, the number of the solutions (16) which obey the further condition that

$$p_{1j} + p_{2j} + p_{3j} = 0, \qquad j = 1, 2, 3. \tag{18}$$

This number gives the value of $\delta(M, m_1, m_2, m_3)$.

Of course, the function $\delta(M, m_1, m_2, m_3)$ should represent, apart from a constant factor, the sum of the Feynman graphs having an incoming scalar particle M, and three outgoing scalar particles m_1, m_2, and m_3.

III. THE STABILITY OF PARTICLES IN A FINITE MODEL

Models of Scalar Particles

Choose the prime q_{k+1} and thus the length of the Euclidean chains. Compute the spectral function $\Sigma(m)$. Choose the peaks of this function. Compute the decay functions $\delta(M, m_1, m_2)$ and $\delta(M, m_1, m_2, m_3)$ for the peaks M of the spectral function. The maximum δ-value for a fixed M then gives a number inversely proportional to the lifetime of the particle having the mass M. If this maximum is zero, the particle in question is stable.

Models of Nonscalar Particles

In a realistic finite model of relativistic quantum theory the particles are associated with the irreducible representations of the Dieudonné group (see Part I) over a Galois field $GF(q)$, such that $q = 8nq_1q_2 \cdots q_k - 1$. Here n is an integer, and the $q_1, q_2, \cdots, q_k$ are the k first primes preceding the prime q_{k+1} which gives the length of the Euclidean chains. These irreducible representations and, accordingly, the particles of a finite world model, seem to be characterized by three labels representing mass M, spin S, and charge Q (cf. Part I). The labels S and Q may be functions of mass but these functions are not known. Also the representations themselves are unknown. In the present situation one could think of the following kind of models of nonscalar particles.

Choose the prime q_{k+1}. Compute again the spectral function $\Sigma(m)$. Choose the peaks M of this function, and compute the decay functions $\delta(M, m_1, m_2)$ and $\delta(M, m_1, m_2, m_3)$ for these peaks M. Let the values of spin S and charge Q be distributed in some way over the mass peaks M. Impose the requirement of the conservation of spin and charge on the decays of particles. Let the particles with zero mass $M = 0$ have the charge zero. Then the smallest mass peak $M_0 > 0$ can always be made stable by associating with it the elementary charge. The smallest mass peak M_0 then represents the electron of the model.

The nucleon must be represented in the model by the smallest mass peak $M_1 > M_0$ for which the decay functions are zero. Then the stability of the nucleon is not due to the conservation of spin and charge but to the decay functions characteristic of finite geometry. This way the conservation of the "barionic number" could be given a geometric interpretation in terms of the finite geometry of the momentum space.

The unstable particles of the model are associated with the mass peaks M with $\max \delta \neq 0$ which can decay violating not the conservation of spin and charge. The only quantum numbers characterizing the particles should be the geometric quantum

numbers M, S, and Q (maybe a combination including parity). The isotopic spin quantum numbers and all the other approximative quantum numbers should be given a geometric interpretation in terms of the approximative properties of the decay and scattering functions characteristic of the finite geometry of the momentum space.

IV. NOTE OF DISCUSSION

All the three types of functions, the spectral function Σ, the decay function δ, and the scattering function σ, give the number of the solutions of some set of Diophantine equations. The integer values allowed for the variables in these equations are associated with the Euclidean chains of the underlying Galois field $GF(q)$. Of course the number of the solutions—and thus the functions Σ, δ, and σ—depends on the choice of these Euclidean chains. In the present series of articles one has chosen the masses, the energies, and the absolute values of the 3-momenta to be taken from the Euclidean chain $E[0]$, and the components of 3-momenta to be taken from the Euclidean chain $E[-\frac{1}{2}(q_{k+1} - 1)]$. Other choices could possibly have been made, giving other functions Σ, δ, and σ.

There is another possibility of choice which also may affect the values of the δ- and the σ-functions, viz., the choice of the system of reference in which the particle process in question is considered. In the present series of articles one has always chosen the center-of-mass system. When computing these functions in finite models of relativistic quantum theory one should experiment also by making other choices of the system of reference.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7. NUMBER 2 FEBRUARY 1966

Applications of the Gel'fand–Moshinsky Bases in Unitary Symmetry and its Breaking

J. A. CASTILHO ALCARÁS, V. C. AGUILERA NAVARRO, AND P. LEAL FERREIRA
Instituto de Física Teórica, São Paulo, Brazil
(Received 20 May 1965)

The Gel'fand–Moshinsky bases for the unitary irreducible representations of the SU_3 group are applied to the unitary scheme for the classification of elementary particles and resonances. A unified method is given for the derivation of the matrix elements of the octet tensor operators in a way which makes transparent its application for other irreducible tensors. The results are given in a remarkably simple form. The systematic use of the basis provides a methodological alternative to the current tensorial methods.

1. INTRODUCTION

THE present work is concerned with the applications of the bases for the unitary irreducible representation (IR) of the SU_3 group to the current scheme[1] for the classification of elementary particles and resonances and its symmetry breaking.[2] The bases here considered for the U_3 group correspond to those derived by Moshinsky[3] in his recent treatment of the nuclear many-body problem.[4] In Sec. 2 we give a brief account of Moshinsky's theory from which easily follows the corresponding basis for SU_3.

The use of the bases presents a number of methodological advantages, the first of which is that they correspond to irreducible representations. On the other hand, Moshinsky's realization of the generators of the group in terms of creation and destruction Fermion operators gives,[5] when applied to a given IR, the structure of the corresponding configuration in terms of the quarks or aces.[6]

From the practical viewpoint the use of the bases provides a direct and clear-cut method for the computation of matrix elements (ME) of irreducible unitary tensors between states of any irreducible representations. In Sec. 3, after defining the irreducible unitary tensor operators of a given rank we give in some detail the method of calculation of the tensor operators associated with the octet representation of SU_3, between states of any IR's, thus unifying the previous works of Okubo[2] and Lurié and MacFarlane.[7] It is convenient to introduce the regular tensor operator[8] in order to simplify the algebraic calculations. The ME are obtained in terms of "reduced matrix elements" which depend only on the IR's considered and on the tensor itself. In spite of the fact that these "reduced matrix elements" are not the reduced matrix elements in the Wigner–Eckart sense[9] they appear in a natural way in our algebraic treatment. In the case of the ME between states of the same IR the "reduced matrix elements" are linear combinations of the Wigner–Eckart ones and for states of different IR's they are simply proportional to them.

The use of Gel'fand's labels for the states, instead of current labeling[7] in terms of λ, μ, I, ν, and Y makes the expressions for the ME much simpler. Relations for going from Gel' fand's labels to $\lambda\mu I\nu Y$ are given in the text.

As it is explained in Sec. 3, the ME of the regular tensor operator between states of different IR's are given in terms of the ME of the singlet component of this tensor, which are given in Table I. It is remarkable that all these ME factorize a function t that depends only on the states envolved and not on the particular tensor components.

The ME between states of the same IR are given in Table II.

The present systematic use of the basis, its direct significance and application provides a method-

[1] M. Gell-Mann, Phys. Rev. 125, 1067 (1962); Y. Ne'eman, Nucl. Phys. 26, 222 (1961).
[2] S. Okubo, Progr. Theor. Phys. (Kyoto) 27, 949(1962).
[3] M. Moshinsky, *Physics of Many Particle Systems*, edited by E. Meeron (Gordon and Breach Science Publishers, Inc., New York, 1964).
[4] This basis was fully discussed by G. E. Baird and L. C. Biedenharn in a series of papers: Phys. Letters 3, 69 (1962); J. Math. Phys. 4, 436 and 1449 (1963); J. Math. Phys. 5, 1723 and 1730 (1964). These authors refer to it as the Gel'fand basis, after the name of whom first considered it: I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR 71, 825 (1958). However in the following we shall stick to Moshinsky's work, since he first gave an explicit construction of the basis.
[5] For a realization of the generators in terms of boson operators see M. Moshinsky, J. Math. Phys. 4, 1128 (1963) and Rev. Mod. Phys. 34, 813 (1962).
[6] G. Zweig, CERN report (unpublished).
[7] D. Lurié and A. J. MacFarlane, J. Math. Phys. 5, 565 (1964). In comparing our results with those of the quoted authors, one misprint (a minus sign) in the 7th equation of their Table III was found.
[8] J. Ginibre, J. Math. Phys. 4, 720 (1963).
[9] J. J. de Swart, Rev. Mod. Phys. 35, 916 (1963).

TABLE I. Matrix elements of Z_3^3 for $(\bar{h}) \neq (h)$.

$(\bar{h})$	$\langle(\bar{h}), h_1'h_2', h_1''\mid Z_3^3 \mid (h), h_1'h_2', h_1''\rangle\langle\bar{h}\mid\mid Z \mid\mid h\rangle^{-1}$
$h_1 + 1\ h_2 - 1\ h_3$	$\left[\frac{(h_1 - h_2' + 2)(h_1 - h_1' + 1)(h_1' - h_2 + 1)(h_2 - h_2')}{(h_1 - h_3 + 2)(h_1 - h_2 + 1)(h_1 - h_2 + 2)(h_2 - h_3)}\right]^{\frac{1}{2}}$
$h_1\ h_2 + 1\ h_3 - 1$	$\left[\frac{(h_2' - h_3 + 1)(h_1' - h_3 + 2)(h_2 - h_2' + 1)(h_1' - h_2)}{(h_2 - h_3 + 1)(h_2 - h_3 + 2)(h_1 - h_3 + 2)(h_1 - h_2)}\right]^{\frac{1}{2}}$
$h_1 + 1\ h_2\ h_3 - 1$	$\left[\frac{(h_1 - h_2' + 2)(h_1' - h_3 + 2)(h_2' - h_3 + 1)(h_1 - h_1' + 1)}{(h_2 - h_3 + 1)(h_1 - h_2 + 1)(h_1 - h_3 + 2)(h_1 - h_3 + 3)}\right]^{\frac{1}{2}}$
$h_1\ h_2 - 1\ h_3 + 1$	$\left[\frac{(h_2 - h_2')(h_1' - h_2 + 1)(h_2' - h_3)(h_1' - h_3 + 1)}{(h_1 - h_2 + 1)(h_2 - h_3)(h_2 - h_3 - 1)(h_1 - h_3 + 1)}\right]^{\frac{1}{2}}$
$h_1 - 1\ h_2\ h_3 + 1$	$\left[\frac{(h_1 - h_1')(h_1 - h_2' + 1)(h_2' - h_3)(h_1' - h_3 + 1)}{(h_2 - h_3)(h_1 - h_3 + 1)(h_1 - h_3)(h_1 - h_2)}\right]^{\frac{1}{2}}$
$h_1 - 1\ h_2 + 1\ h_3$	$\left[\frac{(h_1 - h_1')(h_1 - h_2' + 1)(h_1' - h_2)(h_2 - h_2' + 1)}{(h_2 - h_3 + 1)(h_1 - h_3 + 1)(h_1 - h_2)(h_1 - h_2 - 1)}\right]^{\frac{1}{2}}$

TABLE II. Matrix elements of Z_α^β for $(\bar{h}) = (h)$. A, B, C are defined by (3.24).

$\alpha\beta$	$\bar{h}_1'\ \bar{h}_2'\ \bar{h}_1''$	$\langle(h), \bar{h}_1'\bar{h}_2', \bar{h}_1''\mid Z_\alpha^\beta \mid (h), h_1'h_2', h_1''\rangle$
33	$h_1'\ h_2'\ h_1''$	$H_3 + \frac{(h_2 - h_2')(h_1' - h_2 + 1)(H_2 - H_3)}{(h_1 - h_2 + 1)(h_2 - h_3)} + \frac{(h_1 - h_1')(h_1 - h_2' + 1)[(h_1 - h_2 + 1)(H_1 - H_3) + H_3 - H_2]}{(h_1 - h_2)(h_1 - h_2 + 1)(h_1 - h_3 + 1)}$
31	$h_1' - 1\ h_2'\ h_1'' - 1$	$[B + (h_1 + h_2 + h_3 - h_2' - 1)C]\langle(h), h_1' - 1\ h_2', h_1'' - 1\mid \mathcal{C}_3^1 \mid (h), h_1'h_2', h_1''\rangle$
	$h_1'\ h_2' - 1\ h_1'' - 1$	$[B + (h_1 + h_2 + h_3 - h_1' - 2)C]\langle(h), h_1'h_2' - 1, h_1'' - 1\mid \mathcal{C}_3^1 \mid (h), h_1'h_2', h_1''\rangle$
13	$h_1' + 1\ h_2'\ h_1'' + 1$	$[B + (h_1 + h_2 + h_3 - h_2' - 1)C]\langle(h), h_1'h_2', h_1''\mid \mathcal{C}_3^1 \mid (h), h_1' + 1\ h_2', h_1'' + 1\rangle$
	$h_1'\ h_2' + 1\ h_1'' + 1$	$[B + (h_1 + h_2 + h_3 - h_1' - 2)C]\langle(h), h_1'h_2', h_1''\mid \mathcal{C}_3^1 \mid (h), h_1'h_2' + 1, h_1'' + 1\rangle$
32	$h_1' - 1\ h_2'\ h_1''$	$[B + (h_1 + h_2 + h_3 - h_2' - 1)C]\langle(h), h_1' - 1\ h_2', h_1''\mid \mathcal{C}_3^2 \mid (h), h_1'h_2', h_1''\rangle$
	$h_1'\ h_2' - 1\ h_1''$	$[B + (h_1 + h_2 + h_3 - h_1' - 2)C]\langle(h), h_1'h_2' - 1, h_1''\mid \mathcal{C}_3^2 \mid (h), h_1'h_2', h_1''\rangle$
23	$h_1' + 1\ h_2'\ h_1''$	$[B + (h_1 + h_2 + h_3 - h_2' - 1)C]\langle(h), h_1'h_2', h_1''\mid \mathcal{C}_3^2 \mid (h), h_1' + 1\ h_2', h_1''\rangle$
	$h_1'\ h_2' + 1\ h_1''$	$[B + (h_1 + h_2 + h_3 - h_1' - 2)C]\langle(h), h_1'h_2', h_1''\mid \mathcal{C}_3^2 \mid (h), h_1'h_2' + 1, h_1''\rangle$
	$h_1' + 1\ h_2' - 1\ h_1'' - 1$	$C\langle(h), h_1' + 1\ h_2' - 1, h_1'' - 1\mid \mathcal{C}_3^1 \mid (h), h_1' + 1\ h_2', h_1''\rangle \times \langle(h), h_1'h_2', h_1''\mid \mathcal{C}_3^2 \mid (h), h_1' + 1\ h_2', h_1''\rangle$
21	$h_1' - 1\ h_2' + 1\ h_1'' - 1$	$C\langle(h), h_1' - 1\ h_2' + 1, h_1'' - 1\mid \mathcal{C}_3^1 \mid (h), h_1'h_2' + 1, h_1''\rangle \times \langle(h), h_1'h_2', h_1''\mid \mathcal{C}_3^2 \mid (h), h_1'h_2' + 1, h_1''\rangle$

TABLE II. *Continued.*

$\alpha\beta$	$\bar{h}_1' \bar{h}_2' \bar{h}_1''$	$\langle (h), \bar{h}_1'\bar{h}_2', \bar{h}_1''\vert\, Z_\alpha^\beta \,\vert (h), h_1'h_2', h_1''\rangle$
	$h_1' \; h_2' \; h_1'' - 1$	$[(h_1' - h_1'' + 1)(h_1'' - h_2')]^{\frac{1}{2}}$ $\times \left\{B + C\left[h_1' + h_2' - 1 + \frac{(h_1 - h_1')(h_1' - h_2 + 1)(h_1' - h_3 + 2)}{(h_1' - h_2' + 1)(h_1' - h_2' + 2)} - \frac{(h_2' - h_3 + 1)(h_2 - h_2')(h_1 - h_2' + 1)}{(h_1' - h_2')(h_1' - h_2' + 1)}\right]\right\}$
	$h_1' + 1 \; h_2' - 1 \; h_1'' + 1$	$C\langle (h), h_1' + 1 \; h_2' - 1, h_1'' + 1\vert \; \mathcal{C}_3^2 \,\vert (h), h_1' + 1 \; h_2', h_1'' + 1\rangle$ $\times \langle (h), h_1'h_2', h_1''\vert \; \mathcal{C}_3^1 \,\vert (h), h_1' + 1 \; h_2', h_1'' + 1\rangle$
12	$h_1' - 1 \; h_2' + 1 \; h_1'' + 1$	$C\langle (h), h_1' - 1 \; h_2' + 1, h_1'' + 1\vert \; \mathcal{C}_3^2 \,\vert (h), h_1'h_2' + 1, h_1'' + 1\rangle$ $\times \langle (h), h_1'h_2', h_1''\vert \; \mathcal{C}_3^1 \,\vert (h), h_1'h_2' + 1, h_1'' + 1\rangle$
	$h_1' \; h_2' \; h_1'' + 1$	$[(h_1' - h_1'')(h_1'' - h_2' + 1)]^{\frac{1}{2}}$ $\times \left\{B + C\left[h_1' + h_2' - 1 + \frac{(h_1 - h_1')(h_1' - h_2 + 1)(h_1' - h_3 + 2)}{(h_1' - h_2' + 1)(h_1' - h_2' + 2)} - \frac{(h_2' - h_3 + 1)(h_2 - h_2')(h_1 - h_2' + 1)}{(h_1' - h_2')(h_1' - h_2' + 1)}\right]\right\}$
	$h_1' + 1 \; h_2' - 1 \; h_1''$	$C\langle (h), h_1' + 1 \; h_2' - 1, h_1''\vert \; \mathcal{C}_3^1 \,\vert (h), h_1' + 1 \; h_2', h_1'' + 1\rangle$ $\times \langle (h), h_1'h_2', h_1''\vert \; \mathcal{C}_3^1 \,\vert (h), h_1' + 1 \; h_2', h_1'' + 1\rangle$
11	$h_1' - 1 \; h_2' + 1 \; h_1''$	$C\langle (h), h_1' - 1 \; h_2' + 1, h_1''\vert \; \mathcal{C}_3^1 \,\vert (h), h_1'h_2' + 1, h_1'' + 1\rangle$ $\times \langle (h), h_1'h_2', h_1''\vert \; \mathcal{C}_3^1 \,\vert (h), h_1'h_2' + 1, h_1'' + 1\rangle$
	$h_1' \; h_2' \; h_1''$	$A + Bh_1'' + C\left[(h_1'')^2 + (h_1' - h_1'')(h_1'' - h_2' + 1) + \frac{(h_1 - h_1')(h_1' - h_2 + 1)(h_1' - h_3 + 2)(h_1'' - h_2' + 1)}{(h_1' - h_2' + 1)(h_1' - h_2' + 2)} + \frac{(h_2' - h_3 + 1)(h_2 - h_2')(h_1 - h_2' + 1)(h_1' - h_1'')}{(h_1' - h_2')(h_1' - h_2' + 1)}\right]$
	$h_1' + 1 \; h_2' - 1 \; h_1''$	$C\langle (h), h_1' + 1 \; h_2' - 1, h_1''\vert \; \mathcal{C}_3^2 \,\vert (h), h_1' + 1 \; h_2', h_1''\rangle$ $\times \langle (h), h_1'h_2', h_1''\vert \; \mathcal{C}_3^2 \,\vert (h), h_1' + 1 \; h_2', h_1''\rangle$
22	$h_1' - 1 \; h_2' + 1 \; h_1''$	$C\langle (h), h_1' - 1 \; h_2' + 1, h_1''\vert \; \mathcal{C}_3^2 \,\vert (h), h_1'h_2' + 1, h_1''\rangle$ $\times \langle (h), h_1'h_2', h_1''\vert \; \mathcal{C}_3^2 \,\vert (h), h_1'h_2' + 1, h_1''\rangle$
	$h_1' \; h_2' \; h_1''$	$A + B(h_1' + h_2' - h_1'') + C\left[(h_1'' - h_2')(h_1' - h_1'' + 1) + (h_1' + h_2' - h_1'')^2 + \frac{(h_1 - h_1')(h_1' - h_2 + 1)(h_1' - h_3 + 2)(h_1' - h_1'' + 1)}{(h_1' - h_2' + 1)(h_1' - h_2' + 2)} + \frac{(h_2' - h_3 + 1)(h_2 - h_2')(h_1 - h_2' + 1)(h_1'' - h_2')}{(h_1' - h_2')(h_1' - h_2' + 1)}\right].$

ological alternative to the current tensor methods.[10] As far as the ME of irreducible tensor operators are concerned the method here presented for the octet operators may be applied, *mutatis mutandis*, to the tensor operators of a given rank, associated with a given IR of the SU_3 group.

2. THE BASES FOR THE SU_3 GROUP

A

The bases for the unitary IR of U_3 labeled by $[h_1h_2h_3]$ are given by[3,11]

$$|h_1h_2h_3,\ h_1'h_2',\ h_1''\rangle = B^{h_1'h_2'}_{h_1''}B^{h_1h_2h_3}_{h_1'h_2'}(\mathcal{C}_2^1)^{h_1'-h_1''}(\mathbf{L}_3^1)^{h_1-h_1'} \times (\mathcal{C}_3^2)^{h_2-h_2'}\mathbf{P}^{[h_1h_2h_3]}(b^\dagger_{\mu s})\,|0\rangle \qquad (2.1)$$

where $\mathcal{C}_\mu^{\mu'}$ are the infinitesimal generators of U_3 which may be expressed in terms of fermion creation and destruction operators $b^\dagger_{\mu s}$, $b^{\mu' s}$ as

$$\mathcal{C}_\mu^{\mu'} = \sum_s b^\dagger_{\mu s}b^{\mu' s} \qquad (\mu, \mu' = 1, 2, 3). \qquad (2.2)$$

From the anticommutation rules for the fermion operators it follows that

$$[\mathcal{C}^{\mu_1'}_{\mu_1}, \mathcal{C}^{\mu_2'}_{\mu_2}] = \mathcal{C}^{\mu_2'}_{\mu_1}\,\delta^{\mu_1'}_{\mu_2} - \mathcal{C}^{\mu_1'}_{\mu_2}\,\delta^{\mu_2'}_{\mu_1}. \qquad (2.3)$$

The index μ is the one affected by the unitary transformations of the group and the index s is here considered as a particle index ($s = 1, 2, \cdots, n$).

In (2.1), $\mathbf{P}^{[h_1h_2h_3]}(b^\dagger_{\mu s})|0\rangle = |h_1h_2h_3,\ h_1h_2,\ h_1\rangle$ is the highest-weight polynomial in the $b^\dagger_{\mu s}$ and is associated with a Young tableau of three rows, as described in Ref. 3. Further,

$$\mathbf{L}_3^1 = \mathcal{C}_3^1(\mathcal{C}_1^1 - \mathcal{C}_2^2 + 1) + \mathcal{C}_2^1\mathcal{C}_3^2, \qquad (2.4)$$

which, like $\mathcal{C}_2^1$ and $\mathcal{C}_3^2$ is a lowering weight operator. The normalization constants are found to be

$$B^{h_1'h_2'}_{h_1''} = \left[\frac{(h_1'' - h_2')!}{(h_1' - h_2')!\,(h_1' - h_1'')!}\right]^{\frac{1}{2}}, \qquad (2.5)$$

$$B^{h_1h_2h_3}_{h_1'h_2'} = \left[\frac{(h_1' - h_2' + 1)!\,(h_2' - h_3)!\,(h_1' - h_3 + 1)!\,(h_1' - h_2)!}{(h_1 - h_2' + 1)!\,(h_2 - h_3)!\,(h_1 - h_3 + 1)!\,(h_1 - h_2)!\,(h_2 - h_2')!\,(h_1 - h_1')!}\right]^{\frac{1}{2}}. \qquad (2.6)$$

The nonnegative integers h_1', h_2', h_1'' satisfy the branching rules

$$\begin{gathered} h_1 \geq h_1' \geq h_2 \geq h_2' \geq h_3 \geq 0, \\ h_1' \geq h_1'' \geq h_2'. \end{gathered} \qquad (2.7)$$

The states (2.1) of a given IR are defined up to an over-all phase factor.[12]

The ME of the generators of U_3 may be calculated using (2.1) and (2.3).[3] For future reference, we give here the results of the nonvanishing ones:

$$\begin{aligned}
\langle(h),\ h_1'h_2',\ h_1''|\ \mathcal{C}_1^1\ |(h),\ h_1'h_2',\ h_1'\rangle &= h_1'', \\
\langle(h),\ h_1'h_2',\ h_1''|\ \mathcal{C}_2^2\ |(h),\ h_1'h_2',\ h_1'\rangle &= h_1' + h_2' - h_1'', \\
\langle(h),\ h_1'h_2',\ h_1''|\ \mathcal{C}_3^3\ |(h),\ h_1'h_2',\ h_1'\rangle &= h_1 + h_2 + h_3 - h_1' - h_2', \\
\langle(h),\ h_1'h_2',\ h_1'' - 1|\ \mathcal{C}_2^1\ |(h),\ h_1'h_2',\ h_1'\rangle &= [(h_1' - h_1'' + 1)(h_1'' - h_2')]^{\frac{1}{2}}, \\
\langle(h),\ h_1' - 1\ h_2',\ h_1''|\ \mathcal{C}_3^2\ |(h),\ h_1'h_2',\ h_1'\rangle &= \left[\frac{(h_1 - h_1' + 1)(h_1' - h_2)(h_1' - h_3 + 1)(h_1' - h_1'')}{(h_1' - h_2')(h_1' - h_2' + 1)}\right]^{\frac{1}{2}}, \\
\langle(h),\ h_1'h_2' - 1,\ h_1''|\ \mathcal{C}_3^2\ |(h),\ h_1'h_2',\ h_1'\rangle &= \left[\frac{(h_2' - h_3)(h_2 - h_2' + 1)(h_1 - h_2' + 2)(h_1'' - h_2' + 1)}{(h_1' - h_2' + 1)(h_1' - h_2' + 2)}\right]^{\frac{1}{2}}, \\
\langle(h),\ h_1' - 1\ h_2',\ h_1'' - 1|\ \mathcal{C}_3^1\ |(h),\ h_1'h_2',\ h_1'\rangle &= \left[\frac{(h_1 - h_1' + 1)(h_1' - h_2)(h_1' - h_3 + 1)(h_1'' - h_2')}{(h_1' - h_2')(h_1' - h_2' + 1)}\right]^{\frac{1}{2}}, \\
\langle(h),\ h_1'h_2' - 1,\ h_1'' - 1|\ \mathcal{C}_3^1\ |(h),\ h_1'h_2',\ h_1'\rangle &= -\left[\frac{(h_2' - h_3)(h_2 - h_2' + 1)(h_1 - h_2' + 2)(h_1' - h_1'' + 1)}{(h_1' - h_2' + 1)(h_1' - h_2' + 2)}\right]^{\frac{1}{2}},
\end{aligned} \qquad (2.8)$$

$$(h) \equiv h_1h_2h_3.$$

[10] N. Mukunda and L. K. Pandit, J. Math. Phys. **6**, 746 (1965). See this article for other references to the literature on this field.

[11] J. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).

[12] However, later on we shall adopt a more convenient phase convention, in connection with the definition of the regular operators.

The ME of $\mathcal{C}_1^2$, $\mathcal{C}_2^3$, and $\mathcal{C}_1^3$ are not given explicitly as they can be obtained from (2.8) since $(\mathcal{C}_\mu^\nu)^\dagger = \mathcal{C}_\nu^\mu$.

Notice that the ME of $\mathcal{C}_\mu^{\mu'}$ between states of different IR's vanish.

B

To get the corresponding basis for the SU_3 group, we note that the operator $b_{1s}^\dagger b_{2s}^\dagger b_{3s}^\dagger$ is an invariant under SU_3 and corresponds to a complete column of a Young tableau of U_3. In fact, making use of the relation $b_{\alpha s}^\dagger b_{\beta s}^\dagger b_{\gamma s}^\dagger = \epsilon_{\alpha\beta\gamma} b_{1s}^\dagger b_{2s}^\dagger b_{3s}^\dagger$, we have $(b_{1s}^\dagger b_{2s}^\dagger b_{3s}^\dagger)' = \det U \cdot b_{1s}^\dagger b_{2s}^\dagger b_{3s}^\dagger$, and one sees at once that for unimodular unitary transformations the mentioned operator is an invariant. Therefore, h_3 complete columns of the Young tableau associated with $\mathbf{P}^{[h_1 h_2 h_3]}(b_{\mu s}^\dagger)\ |0\rangle$ is an invariant under SU_3 which can be factorized in (2.1) since it commutes with the lowering weight operators. Then a basis for the IR of SU_3 is obtained suppressing the block of h_3 complete columns and reenumerating the index s (due to the irrelevant shift of h_3 units in s) and we have

$$|f_1 f_2, f_1' f_2', f_1''\rangle \equiv |h_1 - h_3\ h_2 - h_3\ 0, h_1' - h_3\ h_2' - h_3, h_1'' - h_3\rangle. \tag{2.9}$$

For the f's defined by (2.9) similar branching rules (2.7) hold.

By counting the states (2.9) it follows that the dimension of a given IR $[f_1 f_2]$ of SU_3 is

$$\begin{aligned}\dim [f_1 f_2] &= \sum_{f_3'=0}^{f_3} \sum_{f_1'=f_3}^{f_1} (f_1' - f_2' + 1) \\ &= \tfrac{1}{2}(f_1 - f_2 + 1)(f_1 + 2)(f_2 + 1).\end{aligned} \tag{2.10}$$

C

Contact with physics is made by identifying, among the generators of SU_3 $\overline{\mathcal{C}}_\mu^{\mu'} = \mathcal{C}_\mu^{\mu'} - \frac{1}{3}\delta_\mu^{\mu'} \operatorname{Tr} \mathcal{C}$, physical quantities like the hypercharge and isospin.[2] These are defined by

$$\begin{aligned} I_1 &= \tfrac{1}{2}(\mathcal{C}_1^2 + \mathcal{C}_2^1), \\ I_2 &= -\tfrac{1}{2}i(\mathcal{C}_1^2 - \mathcal{C}_2^1), \\ I_3 &= \tfrac{1}{2}(\mathcal{C}_1^1 - \mathcal{C}_2^2), \\ Y &= -[\mathcal{C}_3^3 - \tfrac{1}{3}(\mathcal{C}_1^1 + \mathcal{C}_2^2 + \mathcal{C}_3^3)]. \end{aligned} \tag{2.11}$$

It follows from (2.8) that

$$\begin{aligned} I^2\ |f_1' f_2', f_1''\rangle &= \tfrac{1}{4}(f_1' - f_2')(f_1' - f_2' + 2)\ |f_1' f_2', f_1''\rangle, \\ I_3\ |f_1' f_2', f_1''\rangle &= \tfrac{1}{2}(2f_1'' - f_1' - f_2')\ |f_1' f_2', f_1''\rangle, \\ Y\ |f_1' f_2', f_1''\rangle &= [(f_1' + f_2') - \tfrac{2}{3}(f_1 + f_2)]\ |f_1' f_2', f_1''\rangle. \end{aligned} \tag{2.12}$$

Here $|f_1' f_2', f_1''\rangle \equiv |f_1 f_2, f_1' f_2', f_1''\rangle$.

From the first of Eqs. (2.12) one sees at once that $I = \frac{1}{2}(f_1' - f_2')$. Further, assuming the Gell-Mann–Nishijima relation $Q = I_Z + \frac{1}{2}Y$, we see that the eigenvalues of Q for a given IR $[f_1 f_2]$ of SU_3 are given by

$$Q = f_1'' - \tfrac{1}{3}(f_1 + f_2). \tag{2.13}$$

Then, integral charges appear for IR such that $f_1 + f_2 \equiv 0 \pmod 3$, otherwise one gets fractional charges.

The full basis for the representations [10] and [11] are easily seen to be

$$\begin{aligned} |10, 1\rangle &= b_{11}^\dagger\ |0\rangle & |11, 1\rangle &= b_{11}^\dagger b_{21}^\dagger\ |0\rangle \\ |10, 0\rangle &= b_{21}^\dagger\ |0\rangle \quad \text{and} & |10, 1\rangle &= b_{11}^\dagger b_{31}^\dagger\ |0\rangle \\ |00, 0\rangle &= b_{31}^\dagger\ |0\rangle & |10, 0\rangle &= b_{21}^\dagger b_{31}^\dagger\ |0\rangle. \end{aligned}$$

The states of [10] are associated with a triplet of "fundamental particles" of fractional charges while those of [11] may be associated with the corresponding antiparticles.

In fact, let us consider the operator

$$c^{\dagger\alpha 1} = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma} b_{\beta 1}^\dagger b_{\gamma 1}^\dagger \tag{2.14}$$

which transforms under a unitary transformation U as

$$(c^{\dagger\alpha 1})' = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma} U_\beta^{\beta'} U_\gamma^{\gamma'} b_{\beta' 1}^\dagger b_{\gamma' 1}^\dagger.$$

As $\epsilon^{\alpha\beta\gamma} = \delta_{\alpha'}^\alpha \epsilon^{\alpha'\beta\gamma} = U_{\alpha'}^\eta U_\eta^{\dagger\alpha} \epsilon^{\alpha'\beta\gamma}$, we have

$$\begin{aligned}(c^{\dagger\alpha 1})' &= \tfrac{1}{2} U_\eta^{\dagger\alpha} \epsilon^{\alpha'\beta\gamma} U_{\alpha'}^\eta U_\beta^{\beta'} U_\gamma^{\gamma'} b_{\beta' 1}^\dagger b_{\gamma' 1}^\dagger \\ &= (\det U) U_\eta^{\dagger\alpha} c^{\dagger\eta 1}.\end{aligned}$$

Therefore, for unimodular transformations, the $c^{\dagger\alpha 1}$ transforms contragrediently with respect to the $b_{\mu s}^\dagger$. In terms of the $c^{\dagger\alpha 1}$ the states of [11] are given by

$$\begin{aligned} |10, 0\rangle &= c^{\dagger 11}\ |0\rangle, \\ |10, 1\rangle &= -c^{\dagger 21}\ |0\rangle, \\ |11, 1\rangle &= c^{\dagger 31}\ |0\rangle, \end{aligned} \tag{2.15}$$

and we may interpret $c^{\dagger\alpha 1}$ as a creation operator for the "fundamental antiparticles."

For the representation [21] we have

$$\begin{aligned} |21, 2\rangle &= b_{11}^\dagger b_{12}^\dagger b_{21}^\dagger\ |0\rangle, \\ |21, 1\rangle &= -b_{11}^\dagger b_{21}^\dagger b_{22}^\dagger\ |0\rangle, \\ |20, 2\rangle &= b_{11}^\dagger b_{12}^\dagger b_{31}^\dagger\ |0\rangle, \\ |20, 1\rangle &= 2^{-\frac{1}{2}}(b_{11}^\dagger b_{22}^\dagger b_{31}^\dagger - b_{12}^\dagger b_{21}^\dagger b_{31}^\dagger)\ |0\rangle, \\ |20, 0\rangle &= b_{21}^\dagger b_{22}^\dagger b_{31}^\dagger\ |0\rangle, \\ |11, 1\rangle &= 6^{-\frac{1}{2}}(b_{12}^\dagger b_{21}^\dagger b_{31}^\dagger - 2b_{11}^\dagger b_{21}^\dagger b_{32}^\dagger + b_{11}^\dagger b_{22}^\dagger b_{31}^\dagger)\ |0\rangle, \\ |10, 1\rangle &= -b_{11}^\dagger b_{31}^\dagger b_{32}^\dagger\ |0\rangle, \\ |10, 0\rangle &= -b_{21}^\dagger b_{31}^\dagger b_{32}^\dagger\ |0\rangle. \end{aligned} \tag{2.16}$$

According to the eightfold way[1] both mesons and baryons are included in this representation of SU_3. This is realized in (2.16), the baryons being described by products of three creation operators associated with the particles of the IR [10]. Equations (2.16) conduct to the same pictorial representation of Zweig[6] of a triangular structure for the baryons. On the other hand we can rewrite (2.16) as products of the form $c^{\dagger\alpha\beta}b^{\dagger}_{\beta s'}$, using (2.14). For example,

$$|11, 1\rangle = 6^{-\frac{1}{2}}(c^{\dagger 11}b^{\dagger}_{12} - 2c^{\dagger 31}b^{\dagger}_{32} + c^{\dagger 21}b^{\dagger}_{22})\,|0\rangle$$

describes a composite structure of a "fundamental particle" and a "antiparticle" associated with a meson. This conducts to the dumbell structure of mesons of Zweig. Similar considerations are valid for other representations.

3. THE MATRIX ELEMENTS OF THE REGULAR OPERATORS

A

We now introduce the concept of irreducible tensor operators of the SU_3 group in a similar way to that usually done for the R_3 group. An irreducible tensor operator of rank $[f] \equiv [f_1 f_2]$ is a set of dim $[f]$ operators $T^{[f]}_{f'_1 f'_2 f'_3 f'_4}$ that transform under unitary unimodular transformations U like the basic elements of the representation $[f]$ namely

$$T'^{[f]}_{f'} \equiv UT^{[f]}_{f'}U^{\dagger} = \sum_{\bar{f}'} D^{[f]}_{\bar{f}'f'}(U)T^{[f]}_{\bar{f}'}, \qquad f' \equiv (f'_1 f'_2, f''_1). \tag{3.1}$$

For infinitesimal unitary unimodular transformations we have

$$U = 1 - \epsilon^{\alpha}_{\beta}\bar{\mathcal{C}}^{\beta}_{\alpha}, \tag{3.2}$$

$$D^{[f]}_{\bar{f}'f'}(U) = \langle (f), \bar{f}'|\, 1 - \epsilon^{\nu}_{\mu}\bar{\mathcal{C}}^{\mu}_{\nu}\,|(f), f'\rangle, \tag{3.3}$$

where

$$(\epsilon^{\alpha}_{\beta})^{*} = -\epsilon^{\beta}_{\alpha}; \qquad (\bar{\mathcal{C}}^{\mu}_{\nu})^{\dagger} = \bar{\mathcal{C}}^{\nu}_{\mu}.$$

Substituting (3.2) and (3.3) into (3.1) we have in the first order in $\epsilon^{\alpha}_{\beta}$

$$[\bar{\mathcal{C}}^{\beta}_{\alpha}, T^{[f]}_{f'}] = \sum_{\bar{f}'} \langle (f), \bar{f}'|\, \bar{\mathcal{C}}^{\beta}_{\alpha}\,|(f), f'\rangle T^{[f]}_{\bar{f}'}. \tag{3.4}$$

In the following we are interested in the calculation of the ME of irreducible tensor operators associated with the octet representation [21] of SU_3. In this case, (3.4) may be put in a more convenient form with the introduction of the so called regular operators Z^{β}_{α}.[8] These operators transform under SU_3 like the generators $\bar{\mathcal{C}}^{\beta}_{\alpha}$ and are related to the components of the tensor operators $T^{[21]}_{f_1' f_2' f_3' f_4''}$ by

$$\begin{aligned}
T^{[21]}_{100} &= -Z^1_3, & T^{[21]}_{201} &= 2^{-\frac{1}{2}}(Z^2_2 - Z^1_1),\\
T^{[21]}_{101} &= Z^2_3, & T^{[21]}_{202} &= Z^2_1,\\
T^{[21]}_{111} &= (\tfrac{3}{2})^{\frac{1}{2}}Z^3_3, & T^{[21]}_{211} &= Z^3_2,\\
T^{[21]}_{200} &= -Z^1_2, & T^{[21]}_{212} &= Z^3_1.
\end{aligned} \tag{3.5}$$

From (3.4) we get

$$[\bar{\mathcal{C}}^{\beta}_{\alpha}, Z^{\epsilon}_{\gamma}] = Z^{\epsilon}_{\alpha}\delta^{\beta}_{\gamma} - Z^{\beta}_{\gamma}\delta^{\epsilon}_{\alpha} \tag{3.6}$$

with

$$\mathrm{Tr}\, Z = 0 \tag{3.7}$$

provided we take the phases of the states $|21, f'_1 f'_2, f''_1\rangle$ according to the following convention: the states $|21, 21, 2\rangle$, $|21, 21, 1\rangle$, $|21, 11, 1\rangle$ have all the same phase, opposite to that of the remaining states of the octet (Biedenharn's phase convention[4]).

The use of the simpler commutation relation (3.6) introduces considerable simplification for the calculation of the matrix elements. The ME of Z^{β}_{α} are in following calculated without the restriction (3.7), therefore our results hold for the U_3 group. The condition (3.7) and the correspondences

$$\begin{aligned}
h_i - h_3 &= f_i\\
h'_i - h_3 &= f'_i \qquad (i = 1, 2)\\
h''_1 - h_3 &= f''_1
\end{aligned} \tag{3.8}$$

will be used when we go to the SU_3 case.

B

We now outline the method to obtain the ME of Z^{β}_{α} between states of any IR's of the U_3 group, making use of (3.6), (2.8), (2.7), and (2.1).

First of all, we look for the nonvanishing ME of Z^{β}_{α}. From $[\bar{\mathcal{C}}^{\alpha}_{\alpha}, Z^3_3] = 0$ $(\alpha = 1, 2)$ and $[\bar{\mathcal{C}}^1_2\bar{\mathcal{C}}^2_1, Z^3_3] = 0$ we see that for the ME of Z^3_3, namely

$$\langle (\bar{h}), \bar{h}'_1\bar{h}'_2, \bar{h}''_1\,|Z^3_3|\,(h), h'_1h'_2, h''_1\rangle,$$

the following relations hold

$$\begin{aligned}
&\bar{h}''_1 - h''_1 = \tfrac{1}{3}(\bar{h}_1 + \bar{h}_2 + \bar{h}_3 - h_1 - h_2 - h_3),\\
&\bar{h}'_1 + \bar{h}'_2 - \bar{h}''_1 = h'_1 + h'_2 - h''_1\\
&\qquad\qquad + \tfrac{1}{3}(\bar{h}_1 + \bar{h}_2 + \bar{h}_3 - h_1 - h_2 - h_3),\\
&(\bar{h}'_1 - \bar{h}''_1)(\bar{h}''_1 - \bar{h}'_2 + 1)\\
&\qquad\qquad = (h'_1 - h''_1)(h''_1 - h'_2 + 1).
\end{aligned}$$

Putting

$$\bar{h}''_1 - h''_1 = n, \qquad \bar{h}'_1 - h'_1 = m, \qquad \bar{h}'_2 - h'_2 = p,$$

we have $n = m = p$. Then, the only nonvanishing ME of Z^3_3 is

$$\langle (\bar{h}), h'_1 + n\; h'_2 + n, h''_1 + n|\; Z^3_3\;|(h), h'_1h'_2, h''_1\rangle \tag{3.9}$$

with

$$\bar{h}_1 + \bar{h}_2 + \bar{h}_3 = h_1 + h_2 + h_3 + 3n. \qquad (3.10)$$

By (2.7), n is an integer equal or greater than $-h_3$. From now on, we restrict ourselves to the case $n = 0$. As far as ME are concerned, this restriction implies that (3.6) holds also for $\mathcal{C}_\alpha^\beta$. Further, for ME between states of SU_3, (3.8) show that the value of n is irrelevant, since they deal with differences of h's. For $n \neq 0$ see Footnote 13.

By similar considerations we find out that the nonvanishing ME of the remaining tensor components Z_α^β are those listed in formulas (3.21).

One sees at once that maintaining fixed (h) and $(\bar{h})$ we have 21 nonvanishing ME. Since, as will be shown later, we have 7 possibilities for $(\bar{h})$, the total number of nonvanishing ME is 147.

Next, we calculate the ME of Z_3^3, since later on it will be seen that the sole knowledge of them allows us the determination of the ME of all the remaining components Z_α^β.

Making use of (2.1), (2.4) and $[Z_3^3, (\mathcal{C}_3^2)^n] = n(\mathcal{C}_3^2)^{n-1}Z_3^2$, $n = 1, 2, \cdots$, we have

$$\begin{aligned}
&\langle(\bar{h}), h_1'h_2', h_1''|\ Z_3^3\ |(h), h_1'h_2', h_1''\rangle \\
&= B^{h_1h_2h_3}_{h_1'h_2'}[\langle(\bar{h}), h_1'h_2', h_1'|\ (\mathbf{L}_3^1)^{h_1-h_1'} \\
&\quad\times (\mathcal{C}_3^2)^{h_2-h_2'}Z_3^3\ |(h), h_1h_2, h_1\rangle + (h_1 - h_1') \\
&\quad\times (h_1 - h_2' + 1)\langle(\bar{h}), h_1'h_2', h_1'|\ (\mathbf{L}_3^1)^{h_1-h_1'-1} \\
&\quad\times (\mathcal{C}_3^2)^{h_2-h_2'}Z_3^1\ |(h), h_1h_2, h_1\rangle + (h_2 - h_2') \\
&\quad\times \langle(\bar{h}), h_1'h_2', h_1'|\ (\mathbf{L}_3^1)^{h_1-h_1'} \\
&\quad\times (\mathcal{C}_3^2)^{h_2-h_2'-1}Z_3^2\ |(h), h_1h_2, h_1\rangle]. \qquad (3.11)
\end{aligned}$$

On the other hand by (3.9) we get

$$\begin{aligned}
Z_3^3\ |(h), h_1h_2, h_1\rangle &= \sum_{(\bar{h})} |(\bar{h}), h_1h_2, h_1\rangle\langle(\bar{h}), h_1h_2, h_1| \\
&\quad\times Z_3^3\ |(h), h_1h_2, h_1\rangle, \qquad (3.12)
\end{aligned}$$

where according to (2.7), the $(\bar{h})$ is subject to

$$\bar{h}_1 \geq h_1 \geq \bar{h}_2 \geq h_2 \geq \bar{h}_3. \qquad (3.13)$$

Putting

$$\bar{h}_1 - h_1 = n_1 \geq 0, \qquad \bar{h}_2 - h_2 = n_2 \geq 0,$$

we get

$$\begin{aligned}
Z_3^3\ |(h), h_1h_2, h_1\rangle &= \sum_{\bar{h}_1\bar{h}_2} [B^{\bar{h}_1\bar{h}_2\bar{h}_3}_{h_1h_2}\ |\bar{h}_1h_2\bar{h}_3, h_1h_2, h_1\rangle \\
&\quad\times \langle\bar{h}_1h_2\bar{h}_3, \bar{h}_1h_2, \bar{h}_1|\ (\mathbf{L}_1^3)^{n_1}Z_3^3\ |(h), h_1h_2, h_1\rangle \\
&\quad+ \sum_{n_2\geq 1} n_2 B^{\bar{h}_1\bar{h}_2\bar{h}_3}_{h_1h_2}\ |(\bar{h}), h_1h_2, h_1\rangle\langle(\bar{h}), \bar{h}_1\bar{h}_2, \bar{h}_1| \\
&\quad\times (\mathbf{L}_1^3)^{n_1}(\mathcal{C}_2^3)^{n_2-1}Z_2^3\ |(h), h_1h_2, h_1\rangle] \qquad (3.14) \\
&\qquad \mathbf{L}_1^3 = (\mathbf{L}_3^1)^\dagger,
\end{aligned}$$

where use was made of $[(\mathcal{C}_2^3)^m, Z_3^3] = mZ_2^3(\mathcal{C}_2^3)^{m-1}$, $m = 1, 2, \cdots$.

As $\mathcal{C}_2^3\ |(h), h_1h_2, h_1\rangle = 0$ we have only two possibilities for n_2; these are: $n_2 = 0$ and $n_2 = 1$.

(i) For $n_2 = 0$, i.e., $\bar{h}_2 = h_2$,

$$\begin{aligned}
Z_3^3\ |(h), h_1h_2, h_1\rangle &= \sum_{\bar{h}_3} [B^{h_1h_2\bar{h}_3}_{h_1h_2}\ |h_1h_2\bar{h}_3, h_1h_2, h_1\rangle \\
&\quad\times \langle h_1h_2\bar{h}_3, h_1h_2, h_1|\ Z_3^3\ |(h), h_1h_2, h_1\rangle \\
&\quad+ (h_1 - h_2 + 2) \sum_{n_1\geq 1} B^{\bar{h}_1h_2\bar{h}_3}_{h_1h_2}\ |\bar{h}_1h_2\bar{h}_3, h_1h_2, h_1\rangle \\
&\quad\times \langle\bar{h}_1h_2\bar{h}_3, \bar{h}_1h_2, n_1|\ (\mathbf{L}_1^3)^{n_1-1}Z_1^3\ |(h), h_1h_2, h_1\rangle]. \qquad (3.15)
\end{aligned}$$

Since $\mathbf{L}_1^3Z_1^3\ |(h), h_1h_2, h_1\rangle = 0$, we have two possibilities for n_1: $n_1 = 0$ and $n_1 = 1$. When $n_1 = 0$, i.e., $\bar{h}_1 = h_1$, $\bar{h}_3 = h_3$, (3.14) reduces to an identity. When $n_1 = 1$, i.e., $\bar{h}_1 = h_1 + 1$, $\bar{h}_3 = h_3 - 1$,

$$\begin{aligned}
Z_3^3\ |(h), h_1h_2, h_1\rangle &= (h_1 - h_2 + 2) \\
&\quad\times B^{h_1+1\,h_2h_3-1}_{h_1h_2}\ |h_1 + 1\ h_2h_3 - 1, h_1h_2, h_1\rangle \\
&\quad\times \langle h_1 + 1\ h_2h_3 - 1, h_1 + 1\ h_2, h_1 + 1| \\
&\quad\times Z_1^3\ |(h), h_1h_2, h_1\rangle. \qquad (3.16)
\end{aligned}$$

(ii) For $n_2 = 1$, i.e., $\bar{h}_2 = h_2 + 1$, we see by an analogous calculation that only one possibility for n_1 arises, viz. $n_1 = 0$, i.e., $\bar{h}_1 = h_1$, $\bar{h}_3 = h_3 - 1$,

$$\begin{aligned}
&Z_3^3\ |(h), h_1h_2, h_1\rangle \\
&= B^{h_1h_2+1\,h_3-1}_{h_1h_2}\ |h_1h_2 + 1\ h_3 - 1, h_1h_2, h_1\rangle \\
&\quad\times \langle h_1h_2 + 1\ h_3 - 1, h_1h_2 + 1, h_1| \\
&\quad\times Z_2^3\ |(h), h_1h_2, h_1\rangle. \qquad (3.17)
\end{aligned}$$

The possibilities contained in relations (3.16) and (3.17) give for (3.12) the final result

$$\begin{aligned}
&Z_3^3\ |(h), h_1h_2, h_1\rangle \\
&= |(h), h_1h_2, h_1\rangle\langle(h), h_1h_2, h_1|\ Z_3^3\ |(h), h_1h_2, h_1\rangle \\
&\quad+ [(h_1 - h_2 + 1)(h_1 - h_3 + 3)]^{-1} \\
&\quad\times \mathbf{L}_3^1\ |h_1 + 1\ h_2h_3 - 1, h_1 + 1\ h_2, h_1 + 1\rangle \\
&\quad\times \langle h_1 + 1\ h_2h_3 - 1, h_1 + 1\ h_2, h_1 + 1| \\
&\quad\times Z_1^3\ |(h), h_1h_2, h_1\rangle + (h_2 - h_3 + 2)^{-1} \\
&\quad\times \mathcal{C}_3^2\ |h_1h_2 + 1\ h_3 - 1, h_1h_2 + 1, h_1\rangle \\
&\quad\times \langle h_1h_2 + 1\ h_3 - 1, h_1h_2 + 1, h_1| \\
&\quad\times Z_2^3\ |(h), h_1h_2, h_1\rangle. \qquad (3.18)
\end{aligned}$$

By similar calculations one can easily evaluate $Z_3^1\ |(h), h_1h_2, h_1\rangle$ and $Z_3^2\ |(h), h_1h_2, h_1\rangle$. These results give for (3.11) seven different ME for Z_3^3 which are

shown in the Tables I and II. The possibilities for $(\bar{h})$ contained in the tables, for the SU_3 case, are related to the IR's envolved in the decomposition of the Kronecker product $(f_1 f_2) \otimes (21)$.

Finally, we notice that the ME of Z_α^β are obtainable from those of Z_3^3 by judicious choice of the commutation relations (3.6). For example, from $[\mathcal{C}_1^3, Z_3^3] = Z_1^3$ we have

$$\begin{aligned}\langle(\bar{h}), \bar{h}_1'\bar{h}_2', \bar{h}_1''|\ Z_1^3\ |(h), h_1'h_2', h_1''\rangle &= \langle(\bar{h}), \bar{h}_1'\bar{h}_2', \bar{h}_1''|\ \mathcal{C}_1^3\ |(\bar{h}), h_1'h_2', h_1''\rangle \\ &\times \langle(\bar{h}), h_1'h_2', h_1''|\ Z_3^3\ |(h), h_1'h_2', h_1''\rangle \\ &- \langle(\bar{h}), \bar{h}_1'\bar{h}_2', \bar{h}_1''|\ Z_3^3\ |(h), \bar{h}_1'\bar{h}_2', \bar{h}_1''\rangle \\ &\times \langle(h), \bar{h}_1'\bar{h}_2', \bar{h}_1''|\ \mathcal{C}_1^3\ |(h), h_1'h_2', h_1''\rangle \qquad (3.19)\end{aligned}$$

and therefore, using (2.8) and Table I we get the ME of Z_1^3.

In the case $(\bar{h}) \neq (h)$ each component Z_α^β conducts to a "reduced matrix element"

$$\langle(\bar{h}), \bar{h}_1\bar{h}_2, \bar{h}_1|\ Z_\alpha^{\beta\prime}\ |(h), h_1h_2, h_1\rangle$$

which can be now written as a single one,

$$\langle\bar{h}||\ Z\ ||h\rangle = \sum_{\alpha\beta} \langle(\bar{h}), \bar{h}_1\bar{h}_2, \bar{h}_1|\ Z_\alpha^\beta\ |(h), h_1h_2, h_1\rangle. \qquad (3.20)$$

Notice that $\langle\bar{h}||\ Z||h\rangle$ is actually a reduced matrix element since the bras and kets occurring in (3.20) are highest-weight polynomials.

The results for $(\bar{h}) \neq (h)$ may be cast in a remarkably simple form by introducing a function $t(\bar{h}_1\bar{h}_2\bar{h}_3h_1'h_2')$. We have

$$\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1'' - 1|\ Z_3^1\ |(h), h_1'h_2', h_1''\rangle = t(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_1'\bar{h}_2')\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\ Z_3^3\ |(h), \bar{h}_1'\bar{h}_2', h_1''\rangle\langle(h), \bar{h}_1'\bar{h}_2', h_1'' - 1|\ \mathcal{C}_3^1\ |(h), h_1'h_2', h_1''\rangle, \qquad (3.21a)$$

$$\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1'' + 1|\ Z_1^3\ |(h), h_1'h_2', h_1''\rangle = t(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_1'\bar{h}_2')\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\ Z_3^3\ |(h), \bar{h}_1'\bar{h}_2', h_1''\rangle\langle(h), h_1'h_2', h_1''|\ \mathcal{C}_3^1\ |(h), \bar{h}_1'\bar{h}_2', h_1'' + 1\rangle, \qquad (3.21b)$$

$$\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\ Z_3^2\ |(h), h_1'h_2', h_1''\rangle = t(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_1'\bar{h}_2')\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\ Z_3^3\ |(h), \bar{h}_1'\bar{h}_2', h_1''\rangle\langle(h), \bar{h}_1'\bar{h}_2', h_1''|\ \mathcal{C}_3^2\ |(h), h_1'h_2', h_1''\rangle, \qquad (3.21c)$$

$$\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\ Z_2^3\ |(h), h_1'h_2', h_1''\rangle = t(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_1'\bar{h}_2')\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\ Z_3^3\ |(h), \bar{h}_1'\bar{h}_2', h_1''\rangle\langle(h), h_1'h_2', h_1''|\ \mathcal{C}_3^2\ |(h), \bar{h}_1'\bar{h}_2', h_1''\rangle, \qquad (3.21d)$$

$$\begin{aligned}\langle(\bar{h}), h_1' + 1\ h_2' - 1, h_1'' - 1|\ Z_2^1\ |(h), h_1'h_2', h_1''\rangle &= t(\bar{h}_1\bar{h}_2\bar{h}_3h_1' + 1\ h_2' - 1)\langle(\bar{h}), h_1' + 1\ h_2' - 1, h_1''|\ Z_3^3\ |(h), h_1' + 1\ h_2' - 1, h_1''\rangle \\ &\times \langle(h), h_1'h_2' - 1, h_1'' - 1|\ \mathcal{C}_3^1\ |(h), h_1'h_2', h_1''\rangle\langle(h), h_1'h_2' - 1, h_1'' - 1|\ \mathcal{C}_3^2\ |(h), h_1' + 1\ h_2' - 1, h_1'' - 1\rangle, \qquad (3.21e)\end{aligned}$$

$$\begin{aligned}\langle(\bar{h}), h_1' - 1\ h_2' + 1, h_1'' - 1|\ Z_2^1\ |(h), h_1'h_2', h_1''\rangle &= t(\bar{h}_1\bar{h}_2\bar{h}_3h_1' - 1\ h_2' + 1)\langle(\bar{h}), h_1' - 1\ h_2' + 1, h_1''|\ Z_3^3\ |(h), h_1' - 1\ h_2' + 1, h_1''\rangle \\ &\times \langle(h), h_1' - 1\ h_2', h_1'' - 1|\ \mathcal{C}_3^1\ |(h), h_1'h_2', h_1''\rangle\langle(h), h_1' - 1\ h_2', h_1'' - 1|\ \mathcal{C}_3^2\ |(h), h_1' - 1\ h_2' + 1, h_1'' - 1\rangle, \qquad (3.21f)\end{aligned}$$

$$\begin{aligned}\langle(\bar{h}), h_1'h_2', h_1'' - 1|\ Z_2^1\ |(h), h_1'h_2', h_1''\rangle &= -[(h_1' - h_2')(h_1' - h_2' + 2)t(\bar{h}_1\bar{h}_2\bar{h}_3h_1'h_2')]^{-1}[(h_1'' - h_2')(h_1' - h_1'' + 1)]^{\frac{1}{2}}\langle(\bar{h}), h_1'h_2', h_1''|\ Z_3^3\ |(h), h_1'h_2', h_1''\rangle, \qquad (3.21g)\end{aligned}$$

$$\begin{aligned}\langle(\bar{h}), h_1' + 1\ h_2' - 1, h_1'' + 1|\ Z_1^2\ |(h), h_1'h_2', h_1''\rangle &= t(\bar{h}_1\bar{h}_2\bar{h}_3h_1' + 1\ h_2' - 1)\langle(\bar{h}), h_1' + 1\ h_2' - 1, h_1''|\ Z_3^3\ |(h), h_1' + 1\ h_2' - 1, h_1''\rangle \\ &\times \langle(h), h_1'h_2' - 1, h_1''|\ \mathcal{C}_3^2\ |(h), h_1'h_2', h_1''\rangle\langle(h), h_1'h_2' - 1, h_1''|\ \mathcal{C}_3^1\ |(h), h_1' + 1\ h_2' - 1, h_1'' + 1\rangle, \qquad (3.21h)\end{aligned}$$

$$\begin{aligned}\langle(\bar{h}), h_1' - 1\ h_2' + 1, h_1'' + 1|\ Z_1^2\ |(h), h_1'h_2', h_1''\rangle &= t(\bar{h}_1\bar{h}_2\bar{h}_3h_1' - 1\ h_2' + 1)\langle(\bar{h}), h_1' - 1\ h_2' + 1, h_1''|\ Z_3^3\ |(h), h_1' - 1\ h_2' + 1, h_1''\rangle \\ &\times \langle(h), h_1' - 1\ h_2', h_1''|\ \mathcal{C}_3^2\ |(h), h_1'h_2', h_1''\rangle\langle(h), h_1' - 1\ h_2', h_1''|\ \mathcal{C}_3^1\ |(h), h_1' - 1\ h_2' + 1, h_1'' + 1\rangle, \qquad (3.21i)\end{aligned}$$

$$\langle(\bar{h}), h_1'h_2', h_1''+1|\, Z_1^2\,|(h), h_1'h_2', h_1''\rangle$$
$$= -[(h_1'-h_2')(h_1'-h_2'+2)t(\bar{h}_1\bar{h}_2\bar{h}_3h_1'h_2')]^{-1}[(h_1'-h_1'')(h_1''-h_2'+1)]^{\frac{1}{2}}\langle(\bar{h}), h_1'h_2', h_1''|\, Z_3^3\,|(h), h_1'h_2', h_1''\rangle, \tag{3.21j}$$

$$\langle(\bar{h}), h_1'+1\ h_2'-1, h_1''|\, Z_2^2\,|(h), h_1'h_2', h_1''\rangle$$
$$= -\langle(\bar{h}), h_1'+1\ h_2'-1, h_1''|\, Z_1^1\,|(h), h_1'h_2', h_1''\rangle$$
$$= t(\bar{h}_1\bar{h}_2\bar{h}_3h_1'+1\ h_2'-1)\langle(\bar{h}), h_1'+1\ h_2'-1, h_1''|\, Z_3^3\,|(h), h_1'+1\ h_2'-1, h_1''\rangle$$
$$\times\langle(h), h_1'+1\ h_2'-1, h_1''|\, \mathcal{C}_3^2\,|(h), h_1'+1\ h_2', h_1''\rangle\langle(h), h_1'h_2', h_1''|\, \mathcal{C}_3^2\,|(h), h_1'+1\ h_2', h_1''\rangle, \tag{3.21k}$$

$$\langle(\bar{h}), h_1'-1\ h_2'+1, h_1''|\, Z_2^2\,|(h), h_1'h_2', h_1''\rangle$$
$$= -\langle(\bar{h}), h_1'-1\ h_2'+1, h_1''|\, Z_1^1\,|(h), h_1'h_2', h_1''\rangle$$
$$= t(\bar{h}_1\bar{h}_2\bar{h}_3h_1'-1\ h_2'+1)\langle(\bar{h}), h_1'-1\ h_2'+1, h_1''|\, Z_3^3\,|(h), h_1'-1\ h_2'+1, h_1''\rangle$$
$$\times\langle(h), h_1'-1\ h_2'+1, h_1''|\, \mathcal{C}_3^2\,|(h), h_1'h_2'+1, h_1''\rangle\langle(h), h_1'h_2', h_1''|\, \mathcal{C}_3^2\,|(h), h_1'h_2'+1, h_1''\rangle, \tag{3.21l}$$

$$\langle(\bar{h}), h_1'h_2', h_1''|\, Z_2^2\,|(h), h_1'h_2', h_1''\rangle$$
$$= -\langle(\bar{h}), \bar{h}_1'\bar{h}_2', h_1''|\, Z_3^3\,|(h), h_1'h_2', h_1''\rangle - \langle(\bar{h}), h_1'h_2', h_1''|\, Z_1^1\,|(h), h_1'h_2', h_1''\rangle$$
$$= -\tfrac{1}{2}\langle(\bar{h}), h_1'h_2', h_1''|\, Z_3^3\,|(h), h_1'h_2', h_1''\rangle\{1+[(h_1'-h_2')(h_1'-h_2'+2)]^{-1}(h_1'+h_2'-2h_1'')t(\bar{h}_1\bar{h}_2\bar{h}_3h_1'h_2')\}, \tag{3.21m}$$

where $t(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_1'\bar{h}_2')$ is defined as follows.

Let us consider the matrix element $\langle(\bar{h}), \bar{h}_1'\bar{h}_2', \bar{h}_1''|\, Z_\alpha^\beta\,|(h), h_1'h_2', h_1''\rangle$. The indices i, j, k, p, q are such that in the bra we have

$$\bar{h}_i = h_i + 1; \qquad \bar{h}_j = h_j - 1; \qquad \bar{h}_k = h_k;$$
$$\bar{h}_p' = h_p' + 1; \qquad \bar{h}_q' = h_q' - 1,$$

p or q vanishing when there is no $\bar{h}'$ correspondingly. For example, in the bra $\langle h_1 - 1\ h_2h_3 + 1, h_1' - 1\ h_2', h_1'' - 1|$ we have $i = 3$, $j = 1$, $k = 2$, $p = 0$, $q = 1$. Then $t = t_1t_2t_3$ with

$$t_1 = \begin{cases} 1 & \text{if } p = 0 \\ h_i - h_p' + p - i & \text{if } p \neq 0, \end{cases}$$

$$t_2 = \begin{cases} 1 & \text{if } q = 0 \\ h_j - h_q' + q - j + 1 & \text{if } q \neq 0, \end{cases}$$

$$t_3 = \begin{cases} 1 & \text{if } p \neq 0 \text{ or } q \neq 0 \\ h_1' + h_2' - 2h_k + 2k - 4 & \text{if } p, q \text{ both} \\ & \text{equal to } 0. \end{cases}$$

The case $(\bar{h}) = (h)$ is shown in Table II, where the results are written in terms of the "reduced matrix elements"

$$H_i(h) = \langle(h), h_1h_2, h_1|\, Z_i^i\,|(h), h_1h_2, h_1\rangle. \tag{3.22}$$

This separation was made in view of the different structure of the formulas for $(\bar{h}) \neq (h)$ and $(\bar{h}) = (h)$.

Formulae (3.21a) to (3.21m) are valid for the U_3 case and $n = 0$.[13] To get the ME for $T^{[21]}_{f_1'f_2'f_1''}$ we simply go to the SU_3 case through the relations (3.8) and use the definition (3.5).

The case $(\bar{h}) = (h)$ corresponds to that considered by Okubo and find direct application when the breaking interaction (transforming like a definite component of Z_α^β) is taken in first order of perturbation theory, thereby neglecting the representation mixing.

The formulas contained in Table II may be written in the form of Okubo's theorem[2]

$$\langle(h), \bar{h}_1'\bar{h}_2', \bar{h}_1''\,|Z_\alpha^\beta|\,(h), h_1'h_2', h_1''\rangle$$
$$= \langle(h), \bar{h}_1'\bar{h}_2', \bar{h}_1''|\, A\ \delta_\alpha^\beta$$
$$+ B\mathcal{C}_\alpha^\beta + C\mathcal{C}_\lambda^\beta\mathcal{C}_\alpha^\lambda\,|(h), h_1'h_2', h_1''\rangle \tag{3.23}$$

where

$$A = H_1 - h_1B - h_1^2C,$$
$$B = \frac{(h_1+h_2-1)(H_2-H_3)}{(h_1-h_3+1)(h_2-h_3)} - \frac{(h_2+h_3-2)(H_1-H_2)}{(h_1-h_3+1)(h_1-h_2)}, \tag{3.24}$$
$$C = \frac{1}{h_1-h_3+1}\left[\frac{H_1-H_2}{h_1-h_2} - \frac{H_2-H_3}{h_2-h_3}\right],$$

where the H_i's are defined by (3.22).

[13] The case $n \geq -h_3$ referred to in (3.10) only introduces a shift of n unities in the h's of the bra appearing in the "reduced matrix element" (3.20) and (3.22).

C

We wish now to connect our results with the Wigner–Eckart theorem for SU_3 as discussed by de Swart.[9] The Wigner–Eckart theorem in his notation is

$$(\phi_{\nu_3}^{(\mu_3)}, T_{\nu_3}^{(\mu_3)}\phi_{\nu_1}^{(\mu_1)}) = \sum_{\gamma} \begin{pmatrix} \mu_1 & \mu_2 & \mu_3\gamma \\ \nu_1 & \nu_2 & \nu_3 \end{pmatrix} (\mu_3 || T^{(\mu_3)} ||\mu_1)_\gamma \tag{3.25}$$

where the index γ runs over the number of times the representation $\{\mu_3\}$ is contained in the product $\{\mu_1\} \otimes \{\mu_2\}$.

The function

$$\begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \\ \nu_1 & \nu_2 & \nu_3 \end{pmatrix}$$

is the Clebsch–Gordan coefficient of SU_3 and $(\mu_3|| T^{(\mu_3)} ||\mu_1)$ is the reduced matrix element for which the following result holds

$$(\mu_3|| T^{(\mu_3)} ||\mu_1) = \frac{1}{N_3} \sum_{\lambda_1\lambda_2\lambda_3} \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix} (\phi_{\lambda_3}^{(\mu_3)}, T_{\lambda_3}^{(\mu_3)}\phi_{\lambda_1}^{(\mu_1)}), \tag{3.26}$$

where N_3 is the dimension of $\{\mu_2\}$.

The relation between our notation and his one is

$$\begin{aligned} (\mu) &= (\lambda\mu) = (f_1 - f_2, f_2), \\ (\nu) &= (I_3 Y) \\ &= (\tfrac{1}{2}(f_1' - f_2'), \tfrac{1}{2}(2f_1'' - f_1' - f_2'), f_1' + f_2' - \tfrac{2}{3}(f_1 + f_2)), \\ \phi_\nu^{(\mu)} &\to \epsilon\, |(f), f_1'f_2', f_1''\rangle, \end{aligned} \tag{3.27}$$

where ϵ is a phase factor that accounts for the relative phases between states of the same representation according to the phase convention[4]

(i) states with the same $f_1' + f_2'$ and $f_1' - f_2'$ have the same phase;

(ii) if for a given state $|(f), f_1'f_2', f_1''\rangle$ the branching rules allow the existence of $|(f), f_1' + 1\, f_2', f_1'' + 1\rangle$ and $|(f), f_1'f_2' + 1, f_1'' + 1\rangle$, we attach to $|(f), f_1' + 1f_2', f_1'' + 1\rangle$ the same phase of $|(f), f_1' f_2', f_1''\rangle$ and to $|(f), f_1'f_2' + 1, f_1'' + 1\rangle$ the opposite one.

This phase convention is in agreement with that used by McNamee and Chilton[14] in the computation of his tables of Clebsch–Gordan coefficients of SU_3 and for $(f) = [21]$ is the same as that used in the derivation of Eq. (3.6).

Taking into account this correspondence of notations we may put our results in the form of the Wigner–Eckart theorem. It may be shown that when $(\bar{f}) = (f)$ our "reduced matrix elements" are linear combinations of those of the Wigner–Eckart theorem. When $(\bar{f}) \neq (f)$, Eqs. (3.25) and (3.26) give the result

$$(\bar{f}|| T^{[21]} ||f)^2 = \tfrac{1}{8} \sum_{f'\bar{f}'\bar{\bar{f}}'} \langle (\bar{f}), \bar{f}' |\, T_{\bar{\bar{f}}'}^{[21]}\, |(f), f'\rangle^2 \tag{3.28}$$

and one sees that in this case our "reduced matrix elements" are proportional to the Wigner–Eckart ones.

As a consequence of the above comparison, we may derive analytical formulas for the Clebsch–Gordan coefficients of SU_3 involved in the given ME.

D

Finally we wish to point out that the same method may be applied for the calculation of the ME of tensor operators associated with an IR of SU_3 starting from the definition (3.4).

The fact that for a given IR of SU_3, there exists a single state $|(f), f_2f_2, f_2\rangle$ [a result which follows directly from the branching rules (2.7) for the f's] may be of considerable help since the associated tensor component $T_{f_3f_3,f_3}^{[f]}$ plays a similar role as the $T_{111}^{[21]} = (\frac{3}{2})^{\frac{1}{2}} Z_3^3$ in the present context. Therefore, the ME of the remaining components can be obtained from that of the $T_{f_3f_3,f_3}^{[f]}$.

ACKNOWLEDGMENTS

We are grateful to Professor A. H. Zimerman for his continued interest and encouragment. Two of us (J.A.C.A. and V.C.A.N.) wish to thank the "Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP" for the grant of fellowships under which the present work was performed.

[14] P. McNamee, S. J. and F. Chilton, Rev. Mod. Phys. **36**, 1005 (1964).

Faddeev-Type Equations for the Four-Body Problem*

VICTOR A. ALESSANDRINI†
Lawrence Radiation Laboratory, University of California, Berkeley, California
(Received 16 June 1965)

The technique used by Faddeev to obtain connected equations for the nonrelativistic three-body T matrix is generalized for four particles. It is shown that the four-body equations are completely determined by the solutions of all the possible two-body subsystems, as is the case in the three-body problem. This approach can be extended to more complicated multiparticle systems.

1. INTRODUCTION

THE study of nonrelativistic scattering processes that involve more than two particles has recently received considerable attention.[1-4] When the particles interact only by pairs, and there are no multiparticle forces, the problem cannot be approached by means of the Lippmann–Schwinger equation.[5] The main reason for this is that the Lippmann–Schwinger kernel $\sum_{i<j} V_{ij}(E - H_0)^{-1}$ is the sum of disconnected parts in each of which $(N - 2)$ particles are not interacting. In momentum space, this yields $(N - 2)$ delta functions in addition to the over-all delta function representing conservation of momentum. Consequently, the kernel is unbounded and the equation is strongly singular. This difficulty cannot be removed by iterating the equations; any iterated kernel will still contain disconnected parts. The only possibility of obtaining equations that may be solvable by one of the standard methods is to apply one of the usual tricks for handling singular integral equations. It consists of solving in some way the singular part of the kernel in a closed form, in such a way that the remaining equation is nonsingular. In the case we are considering, it amounts to recasting the Lippmann–Schwinger equation into a connected form, by previously solving some pieces of the kernel in an explicit way.

This problem was solved for the general N-body problem by Weinberg.[3] We refer to his paper for a very lucid discussion of the difficulties associated with the multiparticle scattering problem. Huntziker[6] has given a general proof of the compactness of the Weinberg kernel, providing certain assumptions are made about the potentials. In the four-body problem, for example, the Weinberg equations require a knowledge of the solutions of all the possible two- and three-body problems involved, as well as of the potentials V_{ij}.

In the three-body problem, another possible solution was proposed previously by Faddeev.[2] In place of having only one equation for the three-body T matrix, he proposed a set of three coupled integral equations. But the counterpart of this slight complication is that the Faddeev equations do not depend upon the original potentials. The inhomogeneous term and the kernel of the Faddeev equations are completely determined by the off-the-energy-shell two-body amplitudes. This property of the Faddeev equations has been used by Lovelace[2] to propose a practical theory for three-particle processes, in which experimental information about the two-particle subsystems is used to determine partially the off-shell two-body amplitudes.

The purpose of this paper is to generalize the Faddeev approach to the four-body problem; that is, to get connected equations in which the two-body potentials do not appear explicitly. It is possible to go on and get similar equations for more than four particles, but we will not do so explicitly in this paper because the four-body problem is sufficiently complicated to illustrate the general technique. In Sec. II we review briefly the derivation of the three-body Faddeev equations. In Sec. III the four-body problem is formulated and some preliminary results are derived. In Sec. III, the four-body equations are derived; and finally their properties and possible importance are discussed in Sec. V.

* This work was done under the auspices of the U. S. Atomic Energy Commission.

† Permanent Address: Universidad Nacional de La Plata, Argentina. Fellow of the Consejo Nacional de Investigaciones Cientificas y Technicas of Argentina. This institution neither approves nor assumes any liabilities for the information contained in publications by its Fellows.

[1] L. D. Faddeev, Zh. Eksperim. i Teor. Fiz. **39**, 1459 (1960) [English transl.: Soviet Phys.—JETP **12**, 1014 (1961)]; Dokl. Akad. Nauk **138**, 565 (1961) and **145**, 301 (1962) [English transl.: Soviet Phys.—Doklady **6**, 384 (1961) and **1**, 600 (1962).

[2] C. Lovelace, in *Strong Interaction Physics*, edited by R. G. Moorhouse (Oliver and Boyd, London, 1964); also C. Lovelace, Phys. Rev. **135**, B1225 (1964).

[3] S. Weinberg, Phys. Rev. **133**, B232 (1964).

[4] L. Rosenberg, Phys. Rev. **135**, B715 (1964).

[5] B. A. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).

[6] W. Huntziker, Phys. Rev. **135**, B800 (1964).

II. THE THREE-BODY PROBLEM

Consider the Hamiltonian

$$H = H_0 + V, \tag{2.1}$$

where

$$H_0 = \sum_{i=1}^{3} \frac{\mathbf{p}_i^2}{2m_i} \text{ and } V = V_{12} + V_{13} + V_{23}. \tag{2.2}$$

When the resolvent operators of H_0 and H,

$$G_0(z) = (z - H_0)^{-1}; \qquad G(z) = (z - H)^{-1}, \tag{2.3}$$

are introduced, the three-body T matrix is defined by

$$G(z) = G_0(z) + G_0(z)\hat{T}(z)G_0(z). \tag{2.4}$$

Using the resolvent identity

$$G(z) = G_0(z) + G_0(z)VG(z), \tag{2.5}$$

one obtains the Lippmann–Schwinger equation,[6]

$$\hat{T}(z) = V + VG(z)V = V + VG_0(z)\hat{T}(z). \tag{2.6}$$

Faddeev[1] defines the following operators

$$\hat{T}_{ij}(z) = V_{ij} + V_{ij}G(z)V. \tag{2.7}$$

Clearly, the three-body T matrix is given by the sum

$$\hat{T}(z) = \hat{T}_{12}(z) + \hat{T}_{13}(z) + \hat{T}_{23}(z). \tag{2.8}$$

The Faddeev equations are coupled integral equations for the $\hat{T}_{ij}(z)$. In order to obtain them, let us consider the resolvent of the Hamiltonian $H_{ij} = H_0 + V_{ij}$,

$$G_{ij}(z) = [z - H_{ij}]^{-1}. \tag{2.9}$$

The two-body T matrix for particles i and j in the three-body Hilbert space—i.e., with particle k as a spectator particle—is defined by

$$t_{ij}(z) = V_{ij} + V_{ij}G_{ij}(z)V_{ij}, \tag{2.10}$$

and satisfies the Lippmann–Schwinger equation,

$$t_{ij}(z) = V_{ij} + V_{ij}G_0(z)t_{ij}(z). \tag{2.11}$$

It is trivially related to the solutions of the two-body problem, $\hat{t}_{ij}(z)$, by

$$\langle \mathbf{p}_i\mathbf{p}_j\mathbf{p}_k|\ t_{ij}(z)\ |\mathbf{p}_i'\mathbf{p}_j'\mathbf{p}_k'\rangle = \delta(\mathbf{p}_k - \mathbf{p}_k')\langle \mathbf{p}_i\mathbf{p}_j|\ \hat{t}_{ij}(z - \mathbf{p}_k^2/2m_k)\ |\mathbf{p}_i'\mathbf{p}_j'\rangle. \tag{2.12}$$

The identity

$$G(z) = G_{ij}(z) + G_{ij}(z)[V_{ik} + V_{jk}]G(z); \qquad i \neq j \neq k \tag{2.13}$$

can easily be shown, and by inserting (2.13) into (2.7) we get

$$\begin{aligned}\hat{T}_{ij}(z) &= V_{ij} + V_{ij}G_{ij}(z)V \\ &\quad + V_{ij}G_{ij}(z)[V_{ik} + V_{jk}]G(z)V \\ &= V_{ij} + V_{ij}G_{ij}(z)V_{ij} \\ &\quad + V_{ij}G_{ij}(z)[V_{ik} + V_{jk}] \\ &\quad + V_{ij}G_{ij}(z)[V_{ik} + V_{jk}]G(z)V.\end{aligned}$$

By using (2.11), and also the Lippmann–Schwinger equation for $t_{ij}(z)$ in the form $G_{ij}(z)V_{ij} = G_0(z)t_{ij}(z)$, one obtains

$$\begin{aligned}\hat{T}_{ij}(z) &= t_{ij}(z) + t_{ij}(z)G_0(z) \\ &\quad \times [V_{ik} + V_{ik}G(z)V + V_{jk} + V_{jk}G(z)V].\end{aligned}$$

Finally, using the definitions (2.7), this equation becomes

$$\hat{T}_{ij}(z) = t_{ij}(z) + t_{ij}(z)G_0(z)[\hat{T}_{ik}(z) + \hat{T}_{jk}(z)] \quad \text{for } i, j, k = 1, 2, 3 \text{ and } i \neq j \neq k. \tag{2.14}$$

These are the Faddeev equations. Because of the fact that $T_{ij}(z)$ is not coupled to itself, the first iterated kernel is connected. Assuming that the potential satisfies

$$|v_{ij}(\mathbf{q} - \mathbf{q}')| = C[1 + (q - q')]^{-1-\epsilon_0}; \ \epsilon_0 > 0, \tag{2.15}$$

Faddeev proved that the first iterated kernel is compact, except when z is on the real positive axis.[1] It is also possible to prove[7] that the fifth iterated kernel is compact for any value of z.

III. THE FOUR-BODY PROBLEM

In this section, we consider a Hamiltonian of the form $H = H_0 + V$, where

$$H_0 = \sum_{i=1}^{4} \frac{\mathbf{p}_i^2}{2m_i}, \tag{3.1}$$

$$V = \sum_{i<j} V_{ij}, \qquad \text{for } i, j = 1, 2, 3, 4. \tag{3.2}$$

Here again we define

$$G_0(z) = (z - H_0)^{-1}, \qquad G(z) = (z - H)^{-1}. \tag{3.3}$$

The four-body amplitude $\mathfrak{I}(z)$ is defined by the relations

$$G(z) = G_0(z) + G_0(z)\,\mathfrak{I}(z)G_0(z) \tag{3.4}$$

or

$$\mathfrak{I}(z) = V + VG(z)V. \tag{3.5}$$

We introduce next six operators, in analogy with (2.7),

$$\mathfrak{I}_{ij}(z) = V_{ij} + V_{ij}G(z)V, \quad \text{for } i < j; \qquad i, j = 1, 2, 3, 4. \tag{3.6}$$

[7] L. D. Faddeev, Stoklov Mathematical Institute Report No. 69, 1963 (unpublished), translated by J. B. Sykes.

The four-body $\mathfrak{I}(z)$ operator is then given by the sum

$$\mathfrak{I}(z) = \sum_{i<j} \mathfrak{I}_{ij}(z). \tag{3.7}$$

Our aim is to get a set of coupled integral equations for the $\mathfrak{I}_{ij}(z)$, such that they are connected, and do not contain the potentials. This will be done in the next section. Here, for the sake of clarity, we want to make a few comments about the notation we will use in the rest of the paper. If we use the indices i, j, k, l it will be understood that their range of values is from 1 to 4. When we use the subindices ij, ijk, or $ijkl$ in an operator, it will also be understood that $i<j$, $i<j<k$, and $i<j<k<l$, respectively. The two-body amplitudes of particles i and j in the four-body Hilbert space will be denoted by $t_{ij}(z)$; the three-body amplitude of particles ijk in the four-body Hilbert space will be denoted by $T^{(l)}(z)$, where the upper index indicates the spectator particles. We will use $\mathfrak{I}(z)$ for the four-body amplitudes.

The matrix elements of $t_{ij}(z)$ and $T^{(l)}(z)$ can be written in terms of the matrix elements of the operators defined in the previous section, in the following way[2]:

$$\langle \mathbf{p}_i\mathbf{p}_j\mathbf{p}_k\mathbf{p}_l |\, t_{ij}(z)\, |\mathbf{p}_i'\mathbf{p}_j'\mathbf{p}_k'\mathbf{p}_l'\rangle = \delta(\mathbf{p}_k-\mathbf{p}_k')\delta(\mathbf{p}_l-\mathbf{p}_l')\langle \mathbf{p}_i\mathbf{p}_j |\, \hat{t}_{ij}(z-\omega_k-\omega_l)\, |\mathbf{p}_i'\mathbf{p}_j'\rangle, \tag{3.8}$$

$$\langle \mathbf{p}_i\mathbf{p}_j\mathbf{p}_k\mathbf{p}_l |\, T^{(l)}(z)\, |\mathbf{p}_i'\mathbf{p}_j'\mathbf{p}_k'\mathbf{p}_l'\rangle = \delta(\mathbf{p}_l-\mathbf{p}_l')\langle \mathbf{p}_i\mathbf{p}_j\mathbf{p}_k |\, \hat{T}(z-\omega_l)\, |\mathbf{p}_i'\mathbf{p}_j'\mathbf{p}_k'\rangle, \tag{3.9}$$

where $\omega_i = \mathbf{p}_i^2/2m_i$.

Let us define the operators

$$H_{ij} = H_0 + V_{ij}, \tag{3.10a}$$

$$H_{ijk} = H_0 + V_{ij} + V_{ik} + V_{jk}, \tag{3.11a}$$

$$H_{ij,kl} = H_0 + V_{ij} + V_{kl}, \tag{3.12a}$$

and their resolvents,

$$G_{ij}(z) = (z - H_{ij})^{-1}, \tag{3.10b}$$

$$G_{ijk}(z) = (z - H_{ijk})^{-1}, \tag{3.11b}$$

$$G_{ij,kl}(z) = (z - H_{ij,kl})^{-1}. \tag{3.12b}$$

We will need to use several properties of the two- and three-body amplitudes. The two-body amplitudes are given by

$$t_{ij}(z) = V_{ij} + V_{ij}G_{ij}(z)V_{ij}, \tag{3.13}$$

and the Lippmann–Schwinger equations read

$$V_{ij}G_{ij}(z) = t_{ij}(z)G_0(z). \tag{3.14}$$

The three-body amplitudes $T^{(l)}(z)$ are defined by

$$T^{(l)}(z) = (V_{ij} + V_{ik} + V_{jk}) + (V_{ij} + V_{ik} + V_{jk})G_{ijk}(z) \times (V_{ij} + V_{ik} + V_{jk}), \tag{3.15}$$

and the Faddeev operators (2.7) in the four-body Hilbert space read

$$T_{ij}^{(l)}(z) = V_{ij} + V_{ij}G_{ijk}(z)(V_{ij} + V_{ik} + V_{jk}). \tag{3.16}$$

Their matrix elements are trivially related to the matrix elements of the operators $\hat{T}_{ij}(z)$ studied in the preceding section; the relation is given by Eq. (3.9) by writing $T_{ij}^{(l)}(z)$ and $\hat{T}_{ij}(z-\omega_l)$ in place of $T^l(z)$ and $\hat{T}(z-\omega_l)$, respectively.

The Faddeev equations for $T_{ij}^{(l)}(z)$ are

$$T_{ij}^{(l)}(z) = t_{ij}(z) + t_{ij}(z)G_0(z)[T_{ik}^{(l)}(z) + T_{jk}^{(l)}(z)]. \tag{3.17}$$

Before going on to derive the four-body equations, it is convenient to consider in some detail the Green's function, $G_{ij,kl}(z)$. To calculate it is to solve a four-body problem in which the only nonvanishing potentials are V_{ij} and V_{kl}. We shall show that this problem can be solved in a closed form, in terms of the two-body amplitudes $\hat{t}_{ij}(z)$ and $\hat{T}_{kl}(z)$ only.

Let us call $\mathfrak{a}_{ij,kl}(z)$ the four-body amplitude associated with the Hamiltonian $H_{ij,kl}$. Obviously, we have

$$\mathfrak{a}_{ij,kl}(z) = (V_{ij} + V_{kl}) + (V_{ij} + V_{kl})G_{ij,kl}(z)(V_{ij} + V_{kl}). \tag{3.18}$$

Again following Faddeev's idea, we introduce the operators

$$\mathfrak{a}_{ij}(z) = V_{ij} + V_{ij}G_{ij,kl}(z)(V_{ij} + V_{kl}), \tag{3.19}$$

$$\mathfrak{a}_{kl}(z) = V_{kl} + V_{kl}G_{ij,kl}(z)(V_{ij} + V_{kl}), \tag{3.20}$$

$$\mathfrak{a}_{ij,kl}(z) = \mathfrak{a}_{ij}(z) + \mathfrak{a}_{kl}(z). \tag{3.21}$$

Using the identity

$$G_{ij,kl}(z) = G_{ij}(z) + G_{ij}(z)V_{kl}G_{ij,kl}(z), \tag{3.22}$$

one can very simply obtain for $\mathfrak{a}_{ij}(z)$ and $\mathfrak{a}_{kl}(z)$ the equations

$$\begin{aligned} \mathfrak{a}_{ij}(z) &= t_{ij}(z) + t_{ij}(z)G_0(z)\mathfrak{a}_{kl}(z), \\ \mathfrak{a}_{kl}(z) &= t_{kl}(z) + t_{kl}(z)G_0(z)\mathfrak{a}_{ij}(z). \end{aligned} \tag{3.23}$$

Although these equations will help us in simplifying the algebra in the next section, it is not necessary to solve them to calculate e.g., $\mathfrak{a}_{ij}(z)$. Remember that the Hamiltonian $H_{ij,kl}$ is

$$\begin{aligned} H_{ij,kl} &= H_0 + V_{ij} + V_{kl} \\ &= h_0^{(ij)} + V_{ij} + h_0^{(kl)} + V_{kl} \\ &= h_{ij} + h_{kl}, \end{aligned} \tag{3.24}$$

where

$$h_0^{(ij)} = \frac{\mathbf{p}_i^2}{2m_i} + \frac{\mathbf{p}_j^2}{2m_j} = \omega_i + \omega_j, \quad \text{for } \omega = \frac{\mathbf{p}^2}{2m}, \tag{3.25}$$

$$h_0^{(kl)} = \omega_k + \omega_l.$$

Therefore, $G_{ij,kl}(z)$ is the resolvent of the sum of the two-body Hamiltonians h_{ij} and h_{kl}. These two operators commute, because they act upon different spaces. Therefore, we know that if $g_{ij}(z) = (z - h_{ij})^{-1}$, and $g_{kl}(z) = (z - h_{kl})^{-1}$, the resolvent of $h_{ij} + h_{kl}$ is given by[3,8]

$$G_{ij,kl}(z) = \frac{1}{2\pi_i} \int_c g_{ij}(z') g_{kl}(z - z')\, dz', \tag{3.26}$$

where the contour of integration encircles the spectrum of $g_{ij}(z')$ in a counterclockwise way (or the spectrum of $g_k(z - z')$ in a clockwise way). The reader should bear in mind that $g_{ij}(z)$ and $g_{kl}(z)$ are the two-body Green's functions in different two-body Hilbert spaces. Therefore, the matrix element of the right-hand side is trivial,

$$\langle \mathbf{p}_i\mathbf{p}_j\mathbf{p}_k\mathbf{p}_l |\, G_{ij,kl}(z)\, | \mathbf{p}_i'\mathbf{p}_j'\mathbf{p}_k'\mathbf{p}_l' \rangle = \frac{1}{2\pi i} \int_c \langle \mathbf{p}_i\mathbf{p}_j |\, g_{ij}(z')\, | \mathbf{p}_i'\mathbf{p}_j' \rangle \times \langle \mathbf{p}_k\mathbf{p}_l |\, g_{kl}(z - z')\, | \mathbf{p}_k'\mathbf{p}_l' \rangle\, dz'. \tag{3.27}$$

Using Eq. (3.26) for $G_{ij,kl}(z)$ one can obtain for the operator $\mathfrak{A}_{ij}(z)$ the formula

$$\mathfrak{A}_{ij}(z) = t_{ij}(z) + \frac{1}{2\pi i} \int_c \hat{t}_{ij}(z') g_0^{(kl)}(z - z') \hat{t}_{kl}(z - z') g_0^{(kl)}(z - z')\, dz' + \frac{1}{2\pi i} \int_c \hat{t}_{ij}(z') g_0^{(ij)}(z') g_0^{(kl)}(z - z') \hat{t}_{kl}(z - z')\, dz'. \tag{3.28}$$

A proof of this formula is given in the Appendix. The matrix elements of $\mathfrak{A}_{ij}(z)$ are given by

$$\begin{aligned}
&\langle \mathbf{p}_i\mathbf{p}_j\mathbf{p}_k\mathbf{p}_l |\, \mathfrak{A}_{ij}(z)\, | \mathbf{p}_i'\mathbf{p}_j'\mathbf{p}_k'\mathbf{p}_l' \rangle \\
&= \langle \mathbf{p}_i\mathbf{p}_j |\, \hat{t}_{ij}(z - \omega_k - \omega_l)\, | \mathbf{p}_i'\mathbf{p}_j' \rangle \delta(\mathbf{p}_k - \mathbf{p}_k') \delta(\mathbf{p}_l - \mathbf{p}_l') \\
&\quad + \frac{1}{2\pi i} \int_c dz' \langle \mathbf{p}_i\mathbf{p}_j |\, \hat{t}_{ij}(z')\, | \mathbf{p}_i'\mathbf{p}_j' \rangle \\
&\quad \times [(z - z') - (\omega_k + \omega_l)]^{-1} \langle \mathbf{p}_k\mathbf{p}_l |\, \hat{t}_{kl}(z - z')\, | \mathbf{p}_k'\mathbf{p}_l' \rangle \\
&\quad \times [(z - z') - (\omega_k' + \omega_l')]^{-1} \\
&\quad + \frac{1}{2\pi i} \int_c dz' \langle \mathbf{p}_i\mathbf{p}_j |\, \hat{t}_{ij}(z')\, | \mathbf{p}_i'\mathbf{p}_j' \rangle \\
&\quad \times [z' - (\omega_i' + \omega_j')]^{-1} [(z - z') - (\omega_k + \omega_l)]^{-1} \\
&\quad \times \langle \mathbf{p}_k\mathbf{p}_l |\, \hat{t}_{kl}(z - z')\, | \mathbf{p}_k'\mathbf{p}_l' \rangle,
\end{aligned} \tag{3.29}$$

[8] L. Bianchi and L. Favella, Nuovo Cimento **34**, 1823 (1964).

and in the case in which sample pole approximations are used for the two-body amplitudes, the integrals can easily be evaluated. A similar formula can be written for $\mathfrak{A}_{kl}(z)$.

IV. THE FOUR-BODY EQUATIONS

In the previous section we have introduced the operators $\mathfrak{I}_{ij}(z)$; our intention here is to derive the system of coupled integral equations satisfied by them. Let us consider, for example, the operator $\mathfrak{I}_{12}(z)$, defined by

$$\mathfrak{I}_{12}(z) = V_{12} + V_{12} G(z) V. \tag{4.1}$$

When we derived the Faddeev equations for $\hat{T}_{ij}$ we used in the definition (2.7) of this operator an identity between $G(z)$ and $G_{ij}(z)$. We could use in (4.1) a similar resolvent identity connecting $G(z)$ with $G_{12}(z)$, but the resulting equations would not be connected. If we are to obtain connected equations, we must use in (4.1) an identity connecting $G(z)$ with $G_{12}(z)$ and all the other Green's functions containing the subindices 12, namely: $G_{123}(z)$, $G_{124}(z)$ and $G_{12,34}(z)$.

The following identities can be easily shown:

$$G(z) = G_{12}(z) + G_{12}(z) \times [V_{13} + V_{14} + V_{23} + V_{34} + V_{34}] G(z), \tag{4.2}$$

$$G(z) = G_{123}(z) + G_{123}(z)[V_{14} + V_{24} + V_{34}] G(z), \tag{4.3}$$

$$G(z) = G_{124}(z) + G_{124}(z)[V_{13} + V_{23} + V_{34}] G(z), \tag{4.4}$$

$$G(z) = G_{12,34}(z) + G_{12,34}(z) \times [V_{13} + V_{14} + V_{23} + V_{34}] G(z). \tag{4.5}$$

Next we rewrite (4.2) as

$$G(z) = G_{12}(z) + G_{12}(z)[V_{13} + V_{23}] G(z) + G_{12}(z)[V_{14} + V_{24}] G(z) + G_{12}(z) V_{34} G(z), \tag{4.6}$$

and insert the identities (4.3), (4.4), and (4.5) in place of the $G(z)$ which are multipled (in the operator sense) by $[V_{13} + V_{23}]$, $[V_{14} + V_{24}]$, and V_{34}, respectively. In this way, we obtain

$$\begin{aligned}
G(z) &= G_{12}(z) + G_{12}(z)[V_{13} + V_{23}] G_{123}(z) \\
&\quad + G_{12}(z)[V_{14} + V_{24}] G_{124}(z) + G_{12}(z) V_{34} G_{12,34}(z) \\
&\quad + G_{12}(z)[V_{13} + V_{23}] G_{123}(z)[V_{14} + V_{24} + V_{34}] G(z) \\
&\quad + G_{12}(z)[V_{14} + V_{34}] G_{124}(z)[V_{13} + V_{23} + V_{34}] G(z) \\
&\quad + G_{12}(z) V_{34} G_{12,34}(z)[V_{13} + V_{14} + V_{23} + V_{24}] G(z).
\end{aligned} \tag{4.7}$$

This is the resolvent identity which we next insert in (4.1), to get an equation for $\mathfrak{I}_{12}(z)$, Using

(3.14), (3.16), (3.17), (3.18), and (3.6), we find

$$\begin{aligned}\mathfrak{I}_{12}(z) &= t_{12}(z) + [T_{12}^{(4)}(z) - t_{12}(z)] + [T_{12}^{(3)}(z) - t_{12}(z)] \\ &\quad + [\mathfrak{A}_{12}(z) - t_{12}(z)] \\ &\quad + t_{12}(z)G_0(z)[V_{13} + V_{23}]G_{123}(z) \\ &\quad \times [\mathfrak{I}_{14}(z) + \mathfrak{I}_{24}(z) + \mathfrak{I}_{34}(z)] \\ &\quad + t_{12}(z)G_0(z)[V_{14} + V_{24}]G_{124}(z) \\ &\quad \times [\mathfrak{I}_{13}(z) + \mathfrak{I}_{23}(z) + \mathfrak{I}_{34}(z)] \\ &\quad + t_{12}(z)G_0(z)V_{34}G_{12,34}(z) \\ &\quad \times [\mathfrak{I}_{13}(z) + \mathfrak{I}_{14}(z) + \mathfrak{I}_{23}(z) + \mathfrak{I}_{24}(z)]. \quad (4.8)\end{aligned}$$

The potentials can be completely eliminated from the equations by using the following relations which may be obtained with the help of (3.16)–(3.19):

$$[V_{ik} + V_{jk}]G_{ijk}(z) = [T_{ik}^{(l)}(z) + T_{jk}^{(l)}(z)]G_0(z), \quad (4.9)$$

$$V_{kl}G_{ij,kl}(z) = \mathfrak{A}_{ij}(z)G_0(z). \quad (4.10)$$

Therefore, using again the Faddeev equations (3.17), as well as (3.23) one has

$$\begin{aligned}&t_{12}(z)G_0(z)[V_{13} + V_{23}]G_{123}(z) \\ &\quad = [T_{12}^{(4)}(z) - t_{12}(z)]G_0(z) = T_{12}^{(4)c}(z)G_0(z), \quad (4.11)\end{aligned}$$

$$\begin{aligned}&t_{12}(z)G_0(z)[V_{14} + V_{24}]G_{124}(z) \\ &\quad = [T_{12}^{(3)}(z) - t_{12}(z)]G_0(z) = T_{12}^{(3)c}(z)G_0(z), \quad (4.12)\end{aligned}$$

$$\begin{aligned}&t_{12}G_0(z)V_{34}G_{12,34}(z) \\ &\quad = [\mathfrak{A}_{12}(z) - t_{12}(z)]G_0(z) = \mathfrak{A}_{12}^{c}(z)G_0(z). \quad (4.13)\end{aligned}$$

The final four-body equations are obtained by inserting (4.11)–(4.13) into (4.8). In general, they read

$$\begin{aligned}\mathfrak{I}_{ij}(z) &= t_{ij}(z) + T_{ij}^{(k)c}(z) + T_{ij}^{(l)c}(z) + \mathfrak{A}_{ij}^{c}(z) \\ &\quad + T_{ij}^{(k)}(z)G_0(z)[\mathfrak{I}_{ik}(z) + \mathfrak{I}_{jk}(z) + \mathfrak{I}_{lk}(z)] \\ &\quad + T_{ij}^{(l)(c)}(z)G_0(z)[\mathfrak{I}_{il}(z) + \mathfrak{I}_{jl}(z) + \mathfrak{I}_{lk}(z)] \\ &\quad + \mathfrak{A}_{ij}^{c}(z)G_0(z) \\ &\quad \times [\mathfrak{I}_{ik}(z) + \mathfrak{I}_{jk}(z) + \mathfrak{I}_{il}(z) + \mathfrak{I}_{jl}(z)]. \quad (4.14)\end{aligned}$$

The operator $T_{ij}^{(k)c}(z)$ is defined to be $[T_{ij}^{(k)}(z) - t_{ij}(z)]$. The operator $\mathfrak{A}_{ij}^{c}(z)$ is also defined to be $[\mathfrak{A}_{ij}(z) - t_{ij}(z)]$. Given the two-body scattering amplitudes $t_{ij}(z)$, one can calculate these operators by solving the three-body Faddeev equations and computing the integrals involved in our formulas (3.29).

Recalling that the four-body $\mathfrak{I}(z)$ operator is the sum of all the $\mathfrak{I}_{ij}(z)$ operators, one can check very easily that the sum of the inhomogeneous terms of the six equations yields correctly all the disconnected parts of the four-body amplitude. The first iterated kernel is connected because in (4.14) $\mathfrak{I}_{ij}(z)$ is not coupled to itself.

V. CONCLUSIONS

This approach can in principle be generalized to the N-body problem. The basic idea is to introduce $N(N-1)/2$ amplitudes $\mathfrak{I}_{ij}(z)$, in analogy with Eq. (3.6). In order to get an equation for $\mathfrak{I}_{ij}(z)$ one has to insert in its definition the resolvent identity between the full N-body Green's function $G(z)$ and all the possible disconnected N-body Green's functions that contain the potential V_{ij}. These are known from the solutions for systems with a smaller number of particles, and from generalizations of Eq. (3.26). By following this approach, we are guaranteed that the potentials V_{ij} will not appear in the final equations.

We come then to the conclusion that, in the absence of multiparticle forces, the multiparticle $T(z)$ operators are completely determined by the two-body $t_{ij}(z)$ operators, with no reference to the original potentials whatsoever. However, one must bear in mind that in order to solve the Faddeev equations for the three-body problem, or the equations we proposed for the four-body problem, it is necessary to know the matrix elements of $t_{ij}(z)$ off the energy shell. The experimental data determine them only on the energy shell, so that all we can measure is $\langle \mathbf{p}|\ \hat{t}_{ij}(z)\ |\mathbf{p}'\rangle = t_{ij}(\mathbf{p}, \mathbf{p}'; z)$ when $\mathbf{p}^2 = \mathbf{p}'^2 = z$. The only way of obtaining the off-shell extension is through the Lippmann–Schwinger equation, which requires a knowledge of the potential. Nevertheless, the Faddeev approach still has its advantages in some cases. For example, if one is dealing with singular potentials, the mathematical difficulties associated with them need to be solved only at the two-body level, since they are not directly relevant to multiparticle calculations. If the two-body scattering amplitudes appear to be dominated by poles near the physical region—i.e., bound state or resonance poles—the problem of their off-shell extension can be overcome by using phenomenological form factors. If one considers an off-shell partial-wave amplitude $t_l(p, p'; z)$ the poles will be poles in z. It is possible to prove that in the neighborhood of a pole z_p the off-shell amplitude is factorizable in its dependence upon the variables p and p'.[2] Therefore, one can write

$$t_l(p, p'; z) \simeq g_l(p)t_l(z)g_l(p'). \quad (5.1)$$

A simple form for $t_l(z)$ is just a pole term, $1/(z-z_p)$, in the case of a bound state. However, more complicated expressions for resonance poles can be used

if one wants to satisfy two-body unitarity. The functions $g_l(p)$ are the so-called form factors; in the case of bound-state poles they are given in terms of the bound-state wavefunction by $(p^2 - E_p)\psi(p)$, E_p being the binding energy. In the case of a resonance, they are not so well defined, but in any case we know their behavior at the origin ($\sim p^l$) and at infinity ($\sim p^{l-2}$) for superpositions of Yukawa potentials.[9] They also contain the left-hand cuts of the partial-wave amplitudes,[9] and merely express the fact that the bound state and resonance poles by which one is approximating the two-body amplitudes are not elementary systems but composite ones with internal structure. All these requirements can be used to construct phenomenological expressions for the form factors. The Faddeev approach is very useful in performing semiphenomenological calculations to investigate the effects of two-body resonances and bound states in multiparticle systems.

After this paper was written, we received a paper by L. Rosenberg,[10] which includes most of the conclusions presented here.

ACKNOWLEDGMENT

The author is indebted to Dr. David Judd for the kind hospitality extended to him by the Theoretical Division of the Lawrence Radiation Laboratory.

APPENDIX

In order to prove Eq. (3.29), we have to use Eq. (3.26) for $G_{ij,kl}(z)$ in the definition (3.19) of the operator $\mathcal{A}_{ij}(z)$. In so doing, we obtain

$$\mathcal{A}_{ij}(z) = V_{ij} + V_{ij}\frac{1}{2\pi i}\int_c g_{ij}(z')g_{kl}(z-z')\,dz'V_{ij} + V_{ij}\frac{1}{2\pi i}\int_c g_{ij}(z')g_{kl}(z-z')\,dz'V_{kl}. \quad \text{(A1)}$$

Using the Lippmann–Schwinger equations

$$V_{ij}g_{ij}(z') = \hat{t}_{ij}(z')g_0^{(ij)}(z'),$$

$$g_{kl}(z-z')V_{kl} = g_0^{(kl)}(z-z')\hat{t}_{kl}(z-z'),$$

and

$$V_{ij}g_{ij}(z')V_{ij} = \hat{t}_{ij}(z') - V_{ij},$$

we obtain

$$\mathcal{A}_{ij}(z) = V_{ij} + \frac{1}{2\pi i}\int_c [\hat{t}_{ij}(z') - V_{ij}]g_{kl}(z-z')\,dz' + \frac{1}{2\pi i}\int_c \hat{t}_{ij}(z')g_0^{(ij)}(z')g_0^{(kl)}(z-z')\hat{t}_{kl}(z-z')\,dz'. \quad \text{(A2)}$$

Using next the definitions of $\hat{t}_{kl}(z)$,

$$g_{kl}(z-z') = g_0^{(kl)}(z-z') + g_0^{(kl)}(z-z')\hat{t}_{kl}(z-z')g_0^{(kl)}(z-z'),$$

Equation (A2) becomes

$$\mathcal{A}_{ij}(z) = V_{ij}\left[1 - \frac{1}{2\pi i}\int_c g_{kl}(z-z')\,dz'\right] + \frac{1}{2\pi i}\int_c \hat{t}_{ij}(z')g_0^{(kl)}(z-z')\,dz' + \frac{1}{2\pi i}\int_c \hat{t}_{ij}(z')g_0^{(kl)}(z-z')\hat{t}_{kl}(z-z')g_0^{(kl)}(z-z')\,dz' + \frac{1}{2\pi i}\int_c \hat{t}_{ij}(z')g_0^{(ij)}(z')g_0^{(kl)}(z-z')\hat{t}_{kl}(z-z')\,dz'. \quad \text{(A3)}$$

By taking the contour c of integration as enclosing the singularities of $g_{kl}(z-z')$ and $g_0^{(kl)}(z-z')$ in a clockwise way, the first two terms of the right-hand side can be simplified. Recalling that

$$g_0^{(kl)}(z+i\epsilon) - g_0^{(kl)}(z-i\epsilon) = -2\pi i\,\delta[z - h_0^{(ij)}],$$

one gets

$$\frac{1}{2\pi i}\int_c \hat{t}_{ij}(z')g_0^{(kl)}(z-z')\,dz' = t_{ij}(z). \quad \text{(A4)}$$

The bracket multiplying the potential V_{ij} in Eq. (A3) can be shown to vanish because of the completeness relation for the eigenstates of the Hamiltonian $h_{kl} = h_0^{(kl)} + V_{kl}$. We know that the Green's function $g_{kl}(z)$ can be represented as

$$g_{kl}(z) = \sum_n \frac{|\psi_n\rangle\langle\psi_n|}{z+E_n} + \int_0^\infty dE\,\frac{|\psi(E)\rangle\langle\psi(E)|}{z-E}, \quad \text{(A5)}$$

where $|\psi_n\rangle$ are the discrete eigenstates of h_{kl} with binding energy $(-E_n)$, and $|\psi(E)\rangle$ are these belonging to the continuum. Therefore

$$\frac{1}{2\pi i}\int_c g_{kl}(z-z')\,dz' = -\frac{1}{2\pi i}\int_{c'} g_{kl}(\omega)\,d\omega = \sum_n |\psi_n\rangle\langle\psi_n| + \int_0^\infty dE\,|\psi(E)\rangle\langle\psi(E)|,$$

where the contour c' encloses the spectrum of $g_{kl}(\omega)$ in a clockwise way. Therefore, the completeness of the eigenstates of $h_{kl}(z)$ guarantees that

$$1 - \int_c g_{kl}(z-z')\,dz' = 0. \quad \text{(A6)}$$

Using (A4) and (A6), one can reduce Eq. (A3) to the Eq. (3.28) of the text.

[9] V. A. Alessandrini and R. L. Omnes, Phys. Rev. **139**, B167 (1965).

[10] L. Rosenberg, Phys. Rev. **140**, B217 (1965).

Transfer-Matrix Method for Gamma-Ray and Neutron Penetration*

RAPHAEL ARONSON
New York University, New York, New York†
AND
DAVID LEON YARMUSH
TRG, Incorporated, Melville, New York‡
(Received 10 May 1965)

The problem of radiation transport is formulated in terms of a transfer matrix H which is a 2 × 2 matrix of operators. H is simply related to the more intuitive transmission and reflection operators T and R. An explicit expression for H is derived in slab geometry for radiation distributions that depend on the angle with the slab normal and on energy. H for a multilayer slab is the matrix product of the transfer matrices for the individual layers. A formal expression for H for a homogeneous slab of finite thickness is found in terms of the T and R appropriate to an infinitesimally thin slab. These in turn are related to the single-scattering distribution and therefore can be computed from the microscopic cross section. For purposes of computation, finite matrix representations for the operators must be introduced corresponding to the finite vectors which approximate the distributions. Expansions of the distributions in the cosine of the angle and group representations in energy were chosen in the present work. Some numerical results are presented for gamma rays on aluminum. The extension to problems with internal sources and to nonplanar geometries is outlined.

I. INTRODUCTION

IN many physical problems one is interested in inputs to and outputs from a linear system, with the system itself treated as a black box. It has proved useful to look at such problems in terms of scattering or transfer matrices. Matrix methods have been used in the study of transmission lines, wave guides, nuclear scattering, elementary particles, crystal lattice dynamics, acoustics, and other types of problems. The purpose of this article is to formulate such an approach and to apply it to the detailed calculation of the penetration of neutrons and gamma rays in various media. Some of the ideas, however, may be useful in treating other kinds of physical systems.

The prototype of such systems can be taken as the transmission line, which can sustain waves propagating in either direction. These waves are partially reflected and partially transmitted by a circuit element; the over-all effect can be described by a scattering matrix giving the outputs in both directions in terms of the inputs. Alternatively, the element may be described by a transfer matrix giving the input and output on one side in terms of those on the other. The latter point of view is the one adopted in this paper. It leads to a convenient description of the effect of an array or cascade of elements in terms of the effects of the individual elements. Redheffer[1] has given an extensive discussion of the relation between the two views. He considers transmission lines primarily but also discusses the application of both approaches to a variety of other physical systems.

More specifically, the transfer matrix approach described in Sec. II below has proved useful not only for transmission lines[1,2] but for electromagnetic radiation in an array of dielectrics.[3] It has also been applied to the transverse vibration of a bar[4] and propagation in a bifurcated waveguide.[5] The latter systems have both propagating and evanescent modes in each direction. More closely related to the present work is that of Bobrowsky[6] and of Jones.[7] Bobrowsky[6] recognized that the transmission of neutrons and gamma rays through a succession of slabs could be found as a product of transfer matrices. Jones[7] was concerned with the transmission of polarized light through an array of optically active elements. The relation of these papers to the present work is discussed in more detail in Sec. XI.

There is also a generic relation to the invariance

* This work was supported in part by the U. S. Air Force under contracts AF 33(616)-3616 and AF 33(616)-6081 and by the U. S. Army Ballistic Research Laboratories under contract DA-30-069-AMC-96(R).
† Work done in large part while at TRG, Incorporated.
‡ A subsidiary of Control Data Corporation.

[1] R. Redheffer, J. Math. Phys. (Cambridge) 41, 1 (1962).
[2] H. E. Rowe, Bell System Tech. J. 43, 261 (1964).
[3] R. Aronson, unpublished.
[4] D. Yarmush, TRG-142-TN-64-10 (TRG, Incorporated), unpublished.
[5] J. R. Pace and R. Mittra, in *Microwave Research Institute Symposia Series* (Polytechnic Press, Brooklyn, 1964), Vol. 14, p. 177.
[6] A. R. Bobrowsky, NACA Technical Note 1712 (1948).
[7] R. C. Jones, J. Opt. Soc. Am. 46, 126 (1956) and earlier references given there.

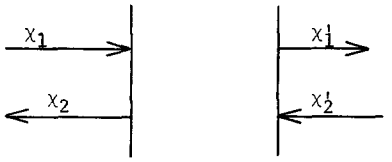

FIG. 1. Schematic single-layer geometry.

approach formulated and applied by Ambartsumian[8] and by Chandrasekhar[9] to astrophysical problems and by Bellman, Kalaba, and Wing[10] to neutron transport.

The problem of neutron and gamma ray transport is more complicated than the others mentioned above because the elements of the transfer matrix are in general integral operators in several variables.

We will be concerned with determining the distribution in direction and energy of transmitted and reflected neutron and gamma radiation in terms of the incident distribution. The transfer matrix method has been used for machine calculations on gamma rays, and so the exposition deals with gamma rays rather than neutrons when it is necessary to specialize. Extensive numerical results for transmission and reflection of gamma rays incident on aluminum slabs are given elsewhere.[11]

In Sec. II we introduce the concept of transfer matrices. In Secs. III–V expressions are given for the transfer matrices and for the transmission and reflection operators for finite slabs in terms of the differential cross sections. Section VI deals with the introduction of finite matrix representations in direction and energy. The explicit matrix algebra for computing the matrices of interest is given in Sec. VII. Section VIII discusses some aspects of the approximation problem. Numerical results for some small problems are given in Sec. IX. The method is evaluated for large scale computation in Sec. X and the general conclusions are stated in Sec. XI.

Appendices A–D refer to special aspects of the general discussion in the text. Appendix E discusses the relation to the P_n method, and application to general geometries is outlined in Appendix F.

II. GENERAL FORMULATION

Consider a slab of material, which need not be homogeneous. Let the distribution of radiation incident from the left be denoted by χ_1, and that from the right by χ_2'. These distributions are in general functions of direction, energy, position of incidence, and possibly other variables. They may represent either amplitudes or intensities, depending on the problem. Let the distribution emerging on the right be designated by χ_1' and that on the left by χ_2. The situation is shown symbolically in Fig. 1.

It is assumed that the problem is linear, that is, that the outputs are linearly related to the inputs. Assuming no fixed sources in the interior, one can express the linear property by the relations

$$\begin{aligned} \chi_1' &= \mathbf{T}\chi_1 + \mathbf{R}^*\chi_2', \\ \chi_2 &= \mathbf{R}\chi_1 + \mathbf{T}^*\chi_2'. \end{aligned} \tag{1}$$

$\mathbf{T}$ and $\mathbf{R}$ are, respectively, transmission and reflection operators for radiation incident from the left. $\mathbf{T}^*$ and $\mathbf{R}^*$ are the operators for radiation incident from the right. $\mathbf{T} = \mathbf{T}^*$ and $\mathbf{R} = \mathbf{R}^*$ only if the slab is symmetric.

Equations (1) can be solved for χ_1' and χ_2' to give in matrix form

$$\begin{bmatrix} \chi_1' \\ \chi_2' \end{bmatrix} = \mathbf{H}\begin{bmatrix} \chi_1 \\ \chi_2 \end{bmatrix} \tag{2}$$

where $\mathbf{H}$ is a 2×2 matrix of operators:

$$\mathbf{H} = \begin{bmatrix} \mathbf{T} - \mathbf{R}^*\mathbf{U}^*\mathbf{R} & \mathbf{R}^*\mathbf{U}^* \\ -\mathbf{U}^*\mathbf{R} & \mathbf{U}^* \end{bmatrix}, \tag{3}$$

with

$$\begin{aligned} \mathbf{U} &= \mathbf{T}^{-1}, \\ \mathbf{U}^* &= (\mathbf{T}^*)^{-1}. \end{aligned} \tag{4}$$

The form (2) leads immediately to a composition law for $\mathbf{H}$-matrices. Consider a two-layer configuration, such as is shown in Fig. 2.

One has

$$\begin{bmatrix} \chi_1'' \\ \chi_2'' \end{bmatrix} = \mathbf{H}_2\begin{bmatrix} \chi_1' \\ \chi_2' \end{bmatrix} = \mathbf{H}_2\mathbf{H}_1\begin{bmatrix} \chi_1 \\ \chi_2 \end{bmatrix},$$

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are, respectively, the $\mathbf{H}$-matrices for slabs 1 and 2. If the $\mathbf{H}$-matrix for the entire configuration is denoted by $\mathbf{H}$, then

$$\mathbf{H} = \mathbf{H}_2\mathbf{H}_1. \tag{5}$$

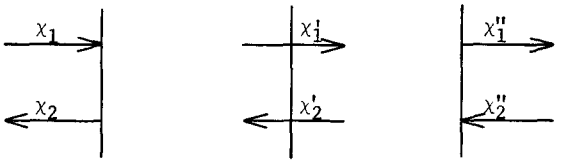

FIG. 2. Schematic two-layer geometry.

[8] V. A. Ambartsumian, Soviet Astron–AJ 19, 30 (1942); Dokl. Akad. Nauk SSSR 38, 229 (1943).
[9] S. Chandrasenkhar, *Radiative Transfer* (Dover Publications, Inc., New York, 1960).
[10] R. Bellman, R. Kalaba, and G. M. Wing, J. Math. Phys. 1, 280 (1960); G. M. Wing, *An Introduction to Transport Theory* (John Wiley & Sons, Inc., New York, 1962).
[11] D. Yarmush, J. Zell, and R. Aronson, WADC-TR-59-772 (1960).

The composition of n layers evidently gives

$$\mathbf{H} = \mathbf{H}_n \cdots \mathbf{H}_2\mathbf{H}_1. \tag{6}$$

From Eq. (5) and the explicit form (3) for **H**, one derives the results of Peebles and Plesset,[12] with **I** the unit operator:

$$\mathbf{T} = \mathbf{T}_2(\mathbf{I} - \mathbf{R}_1^*\mathbf{R}_2)^{-1}\mathbf{T}_1 = \mathbf{T}_2 \sum_{n=0}^{\infty} (\mathbf{R}_1^*\mathbf{R}_2)^n\mathbf{T}_1, \tag{7}$$

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_1 + \mathbf{T}_1^*(\mathbf{I} - \mathbf{R}_2\mathbf{R}_1^*)^{-1}\mathbf{R}_2\mathbf{T}_1 \\ &= \mathbf{R}_1 + \mathbf{T}_1^*\mathbf{R}_2(\mathbf{I} - \mathbf{R}_1^*\mathbf{R}_2)^{-1}\mathbf{T}_1 \\ &= \mathbf{R}_1 + \mathbf{T}_1^*\mathbf{R}_2 \sum_{n=0}^{\infty} (\mathbf{R}_1^*\mathbf{R}_2)^n\mathbf{T}_1. \end{aligned} \tag{8}$$

Each term of the series expansions in (7) and (8) has a physical interpretation in terms of reflections from and transmission through the two slabs. For instance, $\mathbf{T}_1^*\mathbf{R}_2(\mathbf{R}_1^*\mathbf{R}_2)^n\mathbf{T}_1$ represents transmission through slab 1, n pairs of reflections back and forth at the interface, a final reflection from slab 2, and transmission back through slab 1.

While these results have been stated for slabs for definiteness, no use was made in the derivation of any slab properties. Equations (3) and (6) hold for any configuration whose boundary can be divided into two parts. Figures 3 (a)–(c) show typical cross sections of three such configurations.

Figure 3(a) represents a region with an irregular boundary naturally divided into two parts. Figure 3(b) represents a region which has an inside and an outside boundary rather than "right" and "left." In Fig. 3(c) the boundary surface is divided arbitrarily into two parts. One part is indicated in the figure by the curve going clockwise from A to B and the other by the curve going counterclockwise from A to B. One part of the boundary is arbitrarily designated the right-hand face and the other part is then the left-hand face.[13]

If there are sources in the medium, let

Q_+ = flux, due to sources in interior, emerging from right-hand face;

Q_- = flux, due to sources in interior, emerging from left-hand face.

Then Eq. (2) becomes

$$\mathbf{\Phi}' = \mathbf{H}\mathbf{\Phi} + \mathbf{Q}, \tag{9}$$

where

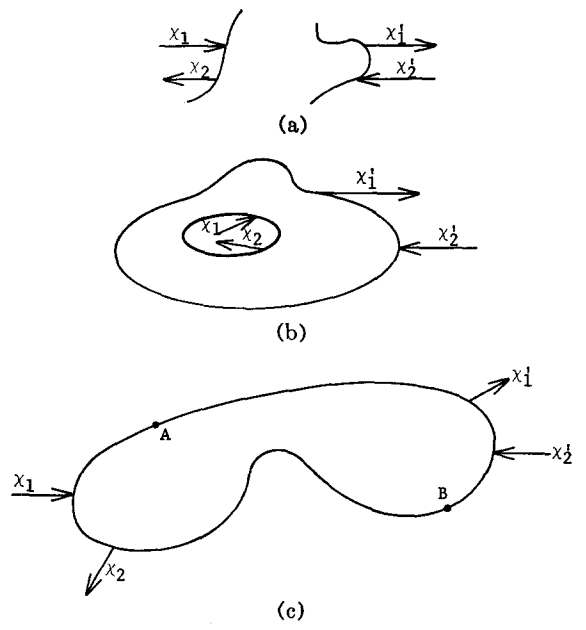

FIG. 3(a) Two-surfaced configuration; (b) hollow configuration; (c) one-surfaced configuration.

$$\mathbf{\Phi} = \begin{bmatrix}\chi_1 \\ \chi_2\end{bmatrix}, \qquad \mathbf{\Phi}' = \begin{bmatrix}\chi_1' \\ \chi_2'\end{bmatrix} \tag{10}$$

and

$$\mathbf{Q} = \begin{bmatrix}Q_+ - \mathbf{R}^*\mathbf{U}^*Q_- \\ -\mathbf{U}^*Q_-\end{bmatrix}. \tag{11}$$

One must distinguish between true sources, described by **Q**, and multiplication or gain, which is taken into account in **H**.

III. DETERMINATION OF H FOR HOMOGENEOUS SLABS

Consider radiation incident uniformly over the surface of a slab. If the slab is composed of homogeneous layers, then the **H**-matrix for the composite is found by applying the relation (6). If the properties of the slab material depend continuously on the depth into the slab, the **H**-matrix can be found to any desired degree of accuracy by breaking the slab up into laminas thin enough to be regarded as homogeneous. Thus it will be sufficient to determine **H** for homogeneous layers.

For a homogeneous slab, $\mathbf{T} = \mathbf{T}^*$ and $\mathbf{R} = \mathbf{R}^*$. To first order in the thickness t,

$$\mathbf{T}(t) = \mathbf{I} - \alpha t + \cdots, \tag{12a}$$

$$\mathbf{R}(t) = \beta t + \cdots, \tag{12b}$$

where **I** is the unit operator. α and β are operators related to the differential cross sections. The nega-

[12] G. H. Peebles and M. S. Plesset, Phys. Rev. **81**, 430 (1951).

[13] Precautions are necessary if the surfaces are concave or re-entrant. See the discussion in Appendix F.

tive sign is chosen for the first-order term in equation (12a) so that the elements of α will be nonnegative for pure absorption.

Since we have assumed that the slab is homogeneous, $\mathbf{H}$ is a function of the thickness only. The composition law (6) gives in this case

$$\mathbf{H}(t_1 + t_2) = \mathbf{H}(t_1)\mathbf{H}(t_2) = \mathbf{H}(t_2)\mathbf{H}(t_1).$$

Since

$$\mathbf{H} = \mathbf{I} \quad \text{for} \quad t = 0,$$

$\mathbf{H}$ must have the form

$$\mathbf{H}(t) = \exp - \mathbf{W}t, \tag{13}$$

where $\mathbf{W}$ is a 2×2 matrix of operators independent of t. One finds by expanding expression (3) for $\mathbf{H}$ to first order in t, using equations (12a, b), that

$$\mathbf{W} = \begin{bmatrix} \alpha & -\beta \\ \beta & -\alpha \end{bmatrix}. \tag{14}$$

Formal series expansions in powers of t can now be found for the elements of $\mathbf{H}$. However, convergence is generally poor, even for moderate values of t. The series solution is discussed in Appendix A.

It is illuminating and computationally useful to introduce diagonalized forms of $\mathbf{H}$ and $\mathbf{W}$. Diagonalizing $\mathbf{W}$ diagonalizes $\mathbf{H}$ also. If diagonalized matrices are denoted by bars and $\mathbf{S}$ is the diagonalizing matrix for $\mathbf{W}$, then

$$\mathbf{H} = \mathbf{S}^{-1}(\exp - \bar{\mathbf{W}}t)\mathbf{S}$$
$$= \mathbf{S}^{-1}(\exp - \mathbf{S}\mathbf{W}\mathbf{S}^{-1}t)\mathbf{S}. \tag{15}$$

Using Eq. (15) to compute $\mathbf{H}$ has a number of advantages, which are related to the fact that $\mathbf{W}$ and $\mathbf{S}$ are independent of the slab thickness but depend on the medium only, i.e., on the cross sections.

We will assume throughout that $\mathbf{W}$ can be diagonalized. If a cross section minimum occurs in the range of energies of interest, there is some question as to when $\mathbf{W}$ can be diagonalized. This is just the problem of the conditions under which eigendistributions of $\mathbf{W}$ exist. This problem has been considered in certain special cases,[14] but there is no general treatment. We will not pursue it further here, since in the ultimate reduction to finite matrices, one can always diagonalize an approximate $\mathbf{W}$ obtained by perturbing the eigenvalues somewhat. In any case, in principle one does not need to diagonalize $\mathbf{W}$ in order to compute $\mathbf{H}$ (e.g., by power series).

The diagonalization of $\mathbf{W}$ can be reduced to that of the operator $\mathbf{A} = \boldsymbol{\delta}\boldsymbol{\delta}$, where

$$\boldsymbol{\delta} = \alpha + \beta, \tag{16a}$$

$$\boldsymbol{\delta} = \alpha - \beta. \tag{16b}$$

One can write

$$\tilde{\mathbf{A}} = \mathbf{X}^{-1}\mathbf{A}\mathbf{X}, \tag{17}$$

where $\tilde{\mathbf{A}}$ is diagonal. Then defining $\boldsymbol{\Lambda}$ by the operator relation[15]

$$\boldsymbol{\Lambda} = \tilde{\mathbf{A}}^{\frac{1}{2}},$$

one finds

$$\bar{\mathbf{W}} = \begin{bmatrix} \boldsymbol{\Lambda} & 0 \\ 0 & -\boldsymbol{\Lambda} \end{bmatrix} \tag{18}$$

and[16]

$$\mathbf{S}^{-1} = \frac{1}{2}\begin{bmatrix} \mathbf{B}_+ & \mathbf{B}_- \\ \mathbf{B}_- & \mathbf{B}_+ \end{bmatrix}, \tag{19a}$$

$$\mathbf{S} = \frac{1}{2}\begin{bmatrix} \mathbf{C}_+ & \mathbf{C}_- \\ \mathbf{C}_- & \mathbf{C}_+ \end{bmatrix} \tag{19b}$$

with

$$\mathbf{B}_\pm = \mathbf{X} \pm \boldsymbol{\delta}\mathbf{X}\boldsymbol{\Lambda}^{-1}, \tag{20a}$$

$$\mathbf{C}_\pm = \mathbf{X}^{-1} \pm \boldsymbol{\Lambda}^{-1}\mathbf{X}^{-1}\boldsymbol{\delta}. \tag{20b}$$

These expressions for $\bar{\mathbf{W}}$, $\mathbf{S}$, and $\mathbf{S}^{-1}$ can be verified directly.

IV. DETERMINATION OF T AND R

With the above expressions for $\bar{\mathbf{W}}$ and $\mathbf{S}$, one finds on equating expressions (3) and (15) for $\mathbf{H}$ that

$$\mathbf{U} = \tfrac{1}{4}(\mathbf{B}_- e^{-\boldsymbol{\Lambda}t}\mathbf{C}_- + \mathbf{B}_+ e^{\boldsymbol{\Lambda}t}\mathbf{C}_+), \tag{21a}$$

$$\mathbf{R}\mathbf{U} = \tfrac{1}{4}(\mathbf{B}_+ e^{-\boldsymbol{\Lambda}t}\mathbf{C}_- + \mathbf{B}_- e^{\boldsymbol{\Lambda}t}\mathbf{C}_+). \tag{21b}$$

Then

$$\mathbf{T} = \mathbf{U}^{-1} \tag{22a}$$

and

$$\mathbf{R} = (\mathbf{R}\mathbf{U})\mathbf{T}. \tag{22b}$$

For purposes of numerical calculation, the various operators must be represented by matrices. The

[14] For an infinite medium, the isotropic scattering problem has been discussed by K. M. Case, Ann. Phys. (N.Y.) 9, 1 (1960). This treatment has been extended to anisotropic scattering by F. Shure and M. Natelson, Ann. Phys. (N.Y.) 26, 274 (1964).

[15] Those eigenvalues of $\mathbf{A}^{\frac{1}{2}}$ with *positive* real part are assigned to $\boldsymbol{\Lambda}$.

[16] Note that $\mathbf{S}$ is defined in the same sense as $\mathbf{X}^{-1}$, not as $\mathbf{X}$. That is, $\tilde{\mathbf{A}} = \mathbf{X}^{-1}\mathbf{A}\mathbf{X}$ while $\bar{\mathbf{W}} = \mathbf{S}\mathbf{W}\mathbf{S}^{-1}$.

matrix formulas (21a)–(22b) for **T** and **R** are not suited for numerical calculations as they stand because the significant information in the elements of the matrix **U** is carried in the least-significant digits and can be lost in roundoff error. In physical terms, it is the most slowly attenuating modes in an attenuating medium which dominate the transmission for moderately thick slabs. Thus the smallest rather than the largest eigenvalue of **Λ** determines the attenuation asymptotically. One would therefore like to obtain expressions for **T** and **R** which avoid the explicit appearance of the increasing exponential $e^{\mathbf{\Lambda} t}$. It is easily verified that if one defines

$$\mathbf{D} = \mathbf{C}_+^{-1}, \tag{23a}$$

$$\mathbf{F} = \mathbf{C}_-\mathbf{D}, \tag{23b}$$

then

$$\mathbf{T} = 4\mathbf{D}e^{-\mathbf{\Lambda} t}(\mathbf{B}_+ + \mathbf{B}_-e^{-\mathbf{\Lambda} t}\mathbf{F}e^{-\mathbf{\Lambda} t})^{-1}, \tag{24a}$$

$$\mathbf{R} = (\mathbf{B}_- + \mathbf{B}_+e^{-\mathbf{\Lambda} t}\mathbf{F}e^{-\mathbf{\Lambda} t})(\mathbf{B}_+ + \mathbf{B}_-e^{-\mathbf{\Lambda} t}\mathbf{F}e^{-\mathbf{\Lambda} t})^{-1}. \tag{24b}$$

This set of formulas is algebraically equivalent to Eqs. (21a)–(22b) and has the desired properties. An alternative way of modifying formulas (21a)–(22b) to avoid the numerical difficulty just discussed is described in Appendix B.

V. INTEGRAL REPRESENTATION OF α AND β FOR RADIATION TRANSPORT

The foregoing discussion is quite general, with application to various types of physical problems. We now specialize to the transport of radiation, with polarization effects neglected. The variables will be

x = depth in slab;
ω = cosine of angle between direction of radiation and normal to slab;
V = function of energy, to be specified later.

The transport equation can be written[17]

$$\omega \frac{\partial \Phi(x, \omega, V)}{\partial x} = -\mu(V)\Phi(x, \omega, V) + n\int_{-1}^{1} d\omega' \int dV' \int_0^{2\pi} d\phi\, \sigma(\hat{\Omega}\cdot\hat{\Omega}'; V, V') \times \Phi(x, \omega', V'), \qquad -1 \le \omega \le 1, \tag{25}$$

where

$\Phi(x, \omega, V)$ = flux at (x, ω, V) per unit V per unit solid angle;
$\mu(V)$ = total macroscopic cross section at V;
$\hat{\Omega}'$ = unit vector in direction of particle velocity before scattering;
$\hat{\Omega}$ = unit vector in direction of particle velocity after scattering;
ϕ = change of azimuth on scattering;
n = density of scatterers;
$\sigma(\hat{\Omega}\cdot\hat{\Omega}'; V, V')$ = differential cross section for scattering from $(V', \hat{\Omega}')$ to $(V, \hat{\Omega})$ per unit V and per unit solid angle at $\hat{\Omega}$.

The transmission and reflection operators for this problem, as well as α and β, can be written as integral operators. Differentiating Eq. (1) and using Eqs. (12a, b), or alternatively differentiating Eq. (13) and using Eqs. (2) and (14), one finds

$$d\chi_1/dx = -\alpha\chi_1 + \beta\chi_2, \qquad d\chi_2/dx = -\beta\chi_1 + \alpha\chi_2.$$

More explicitly, since Φ is to be identified with χ_1 for $\omega > 0$ and χ_2 for $\omega < 0$, the first equation becomes[18]

$$\frac{d\Phi(x, \omega, V)}{dx} = -\int dV' \int_0^1 d\omega'\, K_\alpha(\omega, \omega'; V, V')\Phi(x, \omega', V') + \int dV' \int_{-1}^{0} d\omega'\, K_\beta(\omega, -\omega'; V, V') \times \Phi(x, \omega', V'), \qquad \omega > 0. \tag{26}$$

$K_\alpha(\omega, \omega'; V, V')$ and $K_\beta(\omega, \omega'; V, V')$ are, respectively, the kernels of α and β. The equation involving $d\chi_2/dx$ gives an identical equation for Φ with all directions reversed, i.e., $\omega \to -\omega$, $\omega' \to -\omega'$.

The scattering geometry is represented by the spherical triangle in Fig. 4. $\hat{\imath}$ represents a unit vector in the x direction. The law of cosines gives

$$\hat{\Omega}\cdot\hat{\Omega}' = \cos\psi = \omega\omega' + (1 - \omega^2)^{\frac{1}{2}}(1 - \omega'^2)^{\frac{1}{2}}\cos\phi. \tag{27}$$

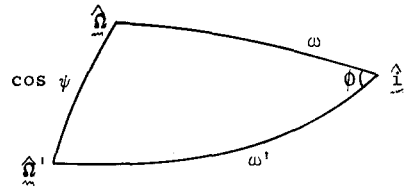

FIG. 4. Scattering geometry.

[17] Equation (25) is written for a single species of scatterer. If there are several such species, n is replaced by $\sum_i n_i\sigma_i$, where the index i refers to the ith species. The necessary generalization of the ensuing results is straightforward.

[18] The convention that ω and ω' are always positive will be used. This means that the direction for incident radiation is measured with respect to the inward normal on the incident face, and that after scattering the direction is measured with respect to the outward normal on the exit face.

Substituting this in Eq. (25), one finds by comparison with Eq. (26) that

$$K_\alpha(\omega, \omega'; V, V') = \frac{1}{\omega}\Big\{\mu(V)\delta(\omega - \omega')\delta(V - V') - n\int_0^{2\pi} d\phi\, \sigma(\omega\omega' + (1-\omega^2)^{\frac{1}{2}}(1-\omega'^2)^{\frac{1}{2}}\cos\phi; V, V')\Big\} \quad (28a)$$

$$K_\beta(\omega, \omega'; V, V') = \frac{n}{\omega}\int_0^{2\pi} d\phi\, \sigma(-\omega\omega' + (1-\omega^2)^{\frac{1}{2}}(1-\omega'^2)^{\frac{1}{2}}\cos\phi; V, V'). \quad (28b)$$

Equations (28a, b) hold for $\omega, \omega' > 0$ according to our convention.

There are the two cases of particular interest for which K_α and K_β can be found easily.

Case 1. *Isotropic scattering, one-velocity problem.*

In this case the variable V does not appear.

$$K_\alpha(\omega, \omega') = (1/\omega)\{\mu\delta(\omega - \omega') - (n\sigma_s/2)\}, \quad (29a)$$

$$K_\beta(\omega, \omega') = (1/\omega)(n\sigma_s/2). \quad (29b)$$

σ_s is the total microscopic scattering cross section.

Case 2. *Elastic scattering of neutrons and gamma rays.*

For elastic collisions of neutrons and gamma rays there is a relation of the form

$$\cos\psi = \gamma(V - V'), \quad (30)$$

i.e., the scattering angle is a function of $V - V'$.

In neutron problems, V is identified with the lethargy. For a scatterer of atomic weight A,

$$\gamma(V - V') = \cosh\tfrac{1}{2}(V - V') - A\sinh\tfrac{1}{2}(V - V'). \quad (31)$$

For gamma rays, V is identified with the wavelength. If V is given in compton units,

$$\gamma(V - V') = 1 - (V - V'). \quad (32)$$

Whatever the radiation, the differential cross section per unit V is

$$\sigma(V, V') = \int \sigma(\hat{\Omega}\cdot\hat{\Omega}'; V, V')\, d\hat{\Omega} \qquad [-1 \le \gamma(V - V') \le 1]$$
$$= 0 \quad \text{otherwise.}$$

The angle-energy correlation (30) implies that

$$\sigma(\hat{\Omega}\cdot\hat{\Omega}'; V, V') = \sigma(V, V')\delta[\cos\psi - \gamma(V - V')]/2\pi. \quad (33)$$

Using (27), introducing $\cos\psi$ instead of ϕ as the integration variable, and setting $\omega = \cos\theta$, $\omega' = \cos\theta'$, one has for the integral in Eqs. (28)

$$\int_0^{2\pi} d\phi\, \sigma(\hat{\Omega}\cdot\hat{\Omega}'; V; V') = \frac{\sigma(V, V')}{\pi}\,\mathrm{Re}\,\frac{1}{S}, \quad (34)$$

where

$$S^2 = [\cos(\theta - \theta') - \gamma(V - V')] \times [\gamma(V - V') - \cos(\theta + \theta')] = 1 - \omega^2 - \omega'^2 - \gamma^2 + 2\omega\omega'\gamma. \quad (35)$$

S is symmetric in ω, ω', and γ. It is real when the triangular relations

$$|\theta - \theta'| < \psi < \theta + \theta',$$
$$|\theta - \psi| < \theta' < \theta + \psi,\ |\theta' - \psi| < \theta < \theta' + \psi$$

are obeyed.

VI. INTRODUCTION OF MATRIX REPRESENTATION

So far everything has been exact. We must now choose a discrete representation in angle and energy in order to deal with matrices rather than integral operators. The choice is by no means unique. We will use a group representation in energy and an angular expansion in a complete set of functions $F^n(\omega)$ of ω. Consider first the angular representation. Expand the flux separately in each hemisphere as[19]

$$\Phi(\omega) = \sum_{n=0}^{\infty} \phi^n F^n(\omega). \quad (36)$$

The coefficients ϕ^n may be combined into a vector $\boldsymbol{\phi}$.

We now wish to obtain the elements α^{mn} of the matrix $\boldsymbol{\alpha}$, which is the representation of the operator $\boldsymbol{\alpha}$ using the basis $F^n(\omega)$. To do this, we determine the elements of the vector $\boldsymbol{\alpha\phi}$. In terms of the integral kernel K_α,

$$[\boldsymbol{\alpha\phi}](\omega) = \int_0^1 K_\alpha(\omega, \omega')\phi(\omega')\, d\omega'. \quad (37)$$

To determine the expansion of the left side of (37) for an arbitrary weight function $g(\omega)$, we multiply the equation by $g(\omega)F^k(\omega)$ and integrate over ω, obtaining

[19] We will reserve superscripts for angular indices, subscripts for energy indices, except as noted.

$$\int_0^1 g(\omega)F^k(\omega)[\alpha\phi](\omega)\,d\omega$$

$$= \sum_m \int_0^1 g(\omega)F^k(\omega)(\alpha\phi)^m F^m(\omega)\,d\omega$$

$$= \sum_{mn} z^{km}\alpha^{mn}\phi^n = (z\alpha\phi)^k, \tag{38}$$

where

$$z^{km} = \int_0^1 g(\omega)F^k(\omega)F^m(\omega)\,d\omega. \tag{39a}$$

If we define

$$y^{kn} = \int_0^1 g(\omega)F^k(\omega)\,d\omega \int_0^1 d\omega'\, K_\alpha(\omega, \omega')F^m(\omega'), \tag{39b}$$

then from (37), the elements of α can be computed from the matrix relation

$$\alpha = z^{-1}y. \tag{40}$$

If the F^k are orthonormal with respect to the weight function $g(\omega)$, $z^{km} = \delta_{km}$; so $\alpha = y$.

Because K_α and K_β contain factors ω^{-1}, all the elements of y do not exist unless $g(\omega)F^k(\omega) \to 0$ for all k as $\omega \to 0$. To avoid this difficulty, one multiplies by ω before integrating over ω in the equations above.[20] Then using (28a) and (34), one finds in analogy with (38),

$$\sum_{mn} e^{km}\alpha^{mn}\phi^n(V) = \sum_n \mu(V)z^{kn}\phi^n(V)$$

$$- n\sum_n \int_0^V dV'\,\sigma(V, V')\phi^n(V')$$

$$\times \int_0^1 g(\omega)F^k(\omega)\,d\omega \int_0^1 d\omega'\,F^m(\omega')$$

$$\times \operatorname{Re}\frac{1}{\pi S[\omega, \omega', \gamma(V - V')]} \tag{41}$$

where

$$e^{km} = \int_0^1 \omega g(\omega)F^k(\omega)F^m(\omega)\,d\omega. \tag{42}$$

Now assume that F^m and g can be expanded in power series in ω:

$$F^m(\omega) = \sum_n q^{mn}\omega^n, \tag{43}$$

$$g(\omega) = \sum_n g^i\omega^i, \tag{44}$$

and define

$$D^{mn}(\gamma) = D^{nm}(\gamma) = \int_0^1 \int_0^1 \omega^m\omega'^n$$

$$\times \operatorname{Re}\frac{1}{\pi S(\omega, \omega', \gamma)}\,d\omega\,d\omega'. \tag{45}$$

The functions D^{mn} are evaluated in Appendix C.

Defining a matrix $\mathbf{G}$ by

$$\begin{aligned} G^{ij} &= g^{j-i}, \qquad j \geq i \\ &= 0, \qquad j < i, \end{aligned} \tag{46}$$

and matrices $\mathbf{q}$ and $\tilde{\mathbf{q}}$ those (m, n) elements are, respectively, q^{mn} and q^{nm}, we have formally the matrix equation

$$\mathbf{e}\alpha\phi(V) = \mu(V)\mathbf{z}\phi(V)$$

$$- n\int_0^V dV'\,\sigma(V, V')\mathbf{q}\mathbf{G}\mathbf{D}[\gamma(V - V')]\tilde{\mathbf{q}}\phi(V'). \tag{47}$$

If the F^m are the full-range Legendre polynomials and $g(\omega) = 1$, then the matrix $\bar{\mathbf{D}} = \mathbf{q}\mathbf{D}\tilde{\mathbf{q}}$ can be computed directly (Appendix D).

Note that the physics enters only through $\mu(V)$, $\sigma(V, V')$, and the argument $\gamma(V - V')$ of $\mathbf{D}$.

An expression analogous to (47) holds for β:

$$\mathbf{e}\beta\phi(V) = n\int_0^V dV'\,\sigma(V, V')\mathbf{q}\mathbf{G}$$

$$\times \mathbf{D}^*[\gamma(V - V')]\tilde{\mathbf{q}}\phi(V'), \tag{48}$$

where

$$(D^*)^{mn}(\gamma) = (-1)^m \int_{-1}^0 d\omega \int_0^1 d\omega'\,\omega^m\omega'^n \operatorname{Re}\frac{1}{\pi S}. \tag{49}$$

Because of the symmetry of S in γ and ω,

$$(D^*)^{mn}(\gamma) = D^{mn}(-\gamma). \tag{50}$$

With F^m of the form (43) and g of the form (44), one has from (39a) and (42)

$$\mathbf{z} = \mathbf{q}\mathbf{G}\mathbf{h}\tilde{\mathbf{q}} \tag{51}$$

and

$$\mathbf{e} = \mathbf{q}\mathbf{G}\bar{\mathbf{h}}\tilde{\mathbf{q}}, \tag{52}$$

where $\mathbf{h}$ is the Hilbert matrix[21], with elements

$$h_{ij} = 1/(i + j + 1) \tag{53}$$

and

$$\bar{h}_{ij} = 1/(i + j + 2). \tag{54}$$

In view of (51) and (52), a factor $\mathbf{q}\mathbf{G}$ appears on the left of every term in (47) and (48), and so one

[20] One can avoid this problem by approximating ω^{-1} as a polynomial in ω, but this leads to complex eigenvalues for the resulting approximation to $\mathbf{W}$. There is a more complete discussion in Ref. 11.

[21] I. R. Savage and E. Lukacs in *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, edited by O. Taussky, p. 105, National Bureau of Standards, Applied Mathematics Series No. 39 (1954).

can multiply from the left by $(\mathbf{qG})^{-1}$, which in general exists, to get expressions for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in which $\mathbf{G}$ does not appear. One expects this to be so, *a priori*, since in effect all complete sets of polynomials are equivalent.

The cancelation of the common factor $\mathbf{qG}$ is no longer possible when the expansion is truncated. To determine the elements of the finite matrix approximations $\boldsymbol{\alpha}'$ and $\boldsymbol{\beta}'$ to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ corresponding to arbitrary truncated expansions of the form $\psi_N(\omega) = \sum_{k=0}^{N} a^k F^k(\omega)$, we require that for the kth element of the vector $\boldsymbol{\alpha}'\boldsymbol{\psi}_N$, the integral

$$\int_{-1}^{1} g(\omega)\, |(\boldsymbol{\alpha}\boldsymbol{\psi}_\infty)^k - (\boldsymbol{\alpha}\boldsymbol{\psi}_N)^k|^2\, d\omega$$

be a minimum, $k = 0, 1, \cdots, N$, for the arbitrary weight function $g(\omega)$. It can be proved that this least squares criterion gives $\boldsymbol{\alpha}' = {}_N\boldsymbol{\alpha}_N$ and similarly, $\boldsymbol{\beta}' = {}_N\boldsymbol{\beta}_N$. Here ${}_M\mathbf{A}_N$ denotes the matrix obtained by retaining only the first $(M+1)$ rows and $(N+1)$ columns[22] of a larger matrix $\mathbf{A}$.

The truncated scattering term in (47) or (48) then involves a matrix of the form ${}_N(\mathbf{qGD\tilde{q}})_N$. This is not in general equal to[23] ${}_N\mathbf{q}_N\mathbf{G}_N\mathbf{D}_N\tilde{\mathbf{q}}_N$. If $g(\omega)$ is restricted to be a polynomial of order K, then by reference to the defining Eq. (46) for $\mathbf{G}$, one sees that

$${}_N(\mathbf{qGDq})_N = {}_N\mathbf{q}_N\mathbf{G}_{N+K}\mathbf{D}_N\tilde{\mathbf{q}}_N, \tag{55}$$

$${}_N\mathbf{z}_N = {}_N\mathbf{q}_N\mathbf{G}_{N+K}\mathbf{h}_N\tilde{\mathbf{q}}_N, \tag{56}$$

$${}_N\mathbf{e}_N = {}_N\mathbf{q}_N\mathbf{G}_{N+K}\bar{\mathbf{h}}_N\tilde{\mathbf{q}}_N. \tag{57}$$

That is, to represent $g(\omega)$, one needs the rectangular matrix ${}_N\mathbf{G}_{N+K}$. A rectangular matrix has no inverse, so that ${}_N\mathbf{G}_{N+K}$ cannot be cancelled from (47) and (48), and thus the approximations to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ depend on the choice of a weight function.

Equations (56) and (57) may give ${}_N\mathbf{z}_N$ and ${}_N\mathbf{e}_N$ as small differences of large numbers, so that extreme care must be taken in computation. The infinite-order matrix $\mathbf{e}$ has no inverse if $g(0) \neq 0$. However, ${}_N\mathbf{e}_N$ has an inverse for most forms of $g(\omega)$. Certain elements of the inverse increase sufficiently rapidly with increasing N, though, so that there are problems of obtaining sufficient precision in the computation of the inverse even when N is as small as 5 or 6.

Since in most problems of interest the radiation never gains energy in a collision, the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ matrices (and therefore, it will be shown, all relevant matrices) are triangular in structure in these energy ranges, having no nonvanishing elements above the main energy diagonal. This simplification is obviously not possible for thermal neutrons. The elements in energy are of course matrices in the angular variable. For matrices the size of $\mathbf{W}$ or $\mathbf{H}$, the elements in the angular variable are themselves 2×2 matrices in the direction index, which can take on two values, "forward" and "backward".

Let the lth energy group range from V_{l-1} to V_l. Equation (47) becomes in an energy group formulation, when the angular expansion is truncated after $n = N$, and $N + K$ is written as M,

$$\begin{aligned} {}_N(\mathbf{e}\boldsymbol{\alpha}_{ll'})_N &= \delta_{ll'}\, {}_N\mathbf{z}_N \int_{V_{l'-1}}^{V_{l'}} dV'\, \mu(V')_N[\mathbf{f}_{l'}(V')]_N \\ &- n_N\mathbf{q}_N\mathbf{G}_M \int_{V_{l'-1}}^{V_{l'}} dV' \int_{V_{l-1}}^{V_l} dV\, \sigma(V, V') \\ &\times {}_M\mathbf{D}[\gamma(V - V')]_N\tilde{\mathbf{q}}_N[\mathbf{f}_{l'}(V')]_N, \end{aligned} \tag{58}$$

where we have taken for the vector $\boldsymbol{\phi}_l$ the total flux vector in the lth group, i.e.,

$$\phi_l^n = \int_{V_{l-1}}^{V_l} \phi^n(V)\, dV. \tag{59}$$

$\mathbf{f}_{l'}(V)$ is a diagonal matrix whose elements are the normalized coefficients in the expansion in angle for the flux, that is,

$$[\mathbf{f}_{l'}(V)]^{nm} = [\phi^n(V)/\phi_{l'}^n]\delta_{nm}, \qquad V_{l'-1} < V \leq V_{l'}.$$

It serves as a shape function. One approximates it in the usual way by assuming some reasonable distribution of flux within the group. Then from (58) and from a similar relation for ${}_N\mathbf{e}\boldsymbol{\beta}_N$ one computes ${}_N\boldsymbol{\alpha}_N$ and ${}_N\boldsymbol{\beta}_N$ for $1 \leq l' \leq l \leq L$, where L is the number of energy groups, by multiplying on the left by $({}_N\mathbf{e}_N)^{-1}$. In practice one would probably choose the same shape for each element of $\mathbf{f}_l$.

VII. RECURSIVE DIAGONALIZATION IN ENERGY

The block-triangular structure in energy of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ matrices permits simplification in the diagonalization of $\mathbf{A} = \boldsymbol{\delta}\boldsymbol{\delta}$ [Eqs. (16)–(20)]. The block-triangular property of $\boldsymbol{\alpha}$ can be written $\boldsymbol{\alpha}_{ll'} = 0$ for $l' > l$.

The sum or product of two block-triangular matrices is also block-triangular in the same way, so $\mathbf{A}_{ll'} = 0$ for $l' > l$. Let the mth column of $\mathbf{A}_{ll'}$ be denoted by $A_{l'}^{m}$ and the mth row by A_{l}^{m}. Assume that the blocks on the diagonal are diagonalized. The set of eigenvalues γ_l^m of $\mathbf{A}$ is just the collection of sets of eigenvalues of the blocks on the diagonal.

[22] It is assumed that the functions of interest are best expanded in the *first* $N+1$ of the F^k, i.e., those with $0 \leq k \leq N$.

[23] Note that adjacent duplicate subscripts are suppressed in a product.

Assume further[24] that none of the eigenvalues $\gamma_{l'}^{m'}$ is equal to γ_l^m when $l' \neq l$. The diagonalization condition (17) gives

$$\sum_{k=1}^{l-1} \mathbf{A}_{lk} X_{kl'}^{m'} = (\gamma_{l'}^{m'} \mathbf{I}_{ll} - \mathbf{A}_{ll}) X_{ll'}^{m'}, \qquad (60)$$

where $\mathbf{I}_{ll}$ is a unit matrix in the appropriate angular subspace. For $l = 1$, the left-hand side vanishes. The factor in parentheses is nonsingular when $l' \neq l$, by hypothesis. Thus $\mathbf{X}_{1l'} = 0$ for $l' > 1$. By induction, one has from (60) that $\mathbf{X}_{ll'} = 0$ for $l' > l$, so $\mathbf{X}$ too is block-triangular. Further, if $\mathbf{X}_{kl'}$ is known for $k < l$, then $\mathbf{X}_{ll'}$ is computed from

$$X_{ll'}^{m'} = (\gamma_{l'}^{m'} \mathbf{I} - \mathbf{A}_{ll})^{-1} \sum_{k=1}^{l-1} \mathbf{A}_{lk} X_{kl'}^{m'}. \qquad (61)$$

If there is an upper limit to the change in V that can occur in a single collision, as is often the case, then all the blocks of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ below a certain diagonal also vanish. In that case the summation in Eq. (61) is restricted.

As a first step in the diagonalization, one must diagonalize the blocks on the energy diagonal. But this problem is one of diagonalizing L matrices of order $(N + 1)$ rather than the more difficult original one of diagonalizing one matrix of order $L(N + 1)$.

In a procedure similar to the diagonalization, the inverse $\mathbf{B}$ of an arbitrary block-triangular matrix $\mathbf{A}$ can be computed recursively. $\mathbf{B}$ has the same block-triangular structure as $\mathbf{A}$. Thus $\mathbf{B}_{1l} = 0$ for $l > 1$ and $\mathbf{B}_{ll} = \mathbf{A}_{ll}^{-1}$. Further,

$$\mathbf{B}_{ll'} = \mathbf{B}_{ll}\mathbf{I}_{ll'} - \mathbf{B}_{ll} \sum_{k=l'}^{l-1} \mathbf{A}_{lk}\mathbf{B}_{kl'}. \qquad (62)$$

$\mathbf{B}_{ll'}$ can be computed from (62) once $\mathbf{B}_{kl'}$ is known for $k < l$.

VIII. REMARKS ON EIGENVALUES AND EIGENDISTRIBUTIONS

For simplicity in both computation and interpretation one would like to use an approximation in which $\mathbf{W}$ has real eigenvalues and eigenvectors. One can show[25] that if $g(\omega) = 1$, the eigenvalues of the approximate $\mathbf{A}$ defined as ${}_N\mathfrak{d}_N\mathfrak{d}_N$ are real in a nonregenerating medium.[26] Therefore the eigenvalues of the corresponding approximate $\mathbf{W}$ are either real or pure imaginary. These $\mathbf{W}$ eigenvalues seem, in fact, to be real in an absorbing medium. This is certainly true in the one-velocity case and also in a multigroup situation with very narrow group widths.[27]

As observed previously,[20] there are other methods of approximating $\mathbf{A}$ and $\mathbf{W}$ which produce complex eigenvalues even for an absorbing medium.

The eigendistributions (columns of $\mathbf{B}_+$ and $\mathbf{B}_-$) have considerable theoretical interest, even though they are artifacts in that they depend on the energy grouping. By examining the corresponding eigenvalues, one can see how many eigendistributions will contribute significantly to the transmitted radiation at a given thickness. In many cases only the fundamental eigendistribution will be important for large thicknesses. If the representation is such that the fundamental eigendistribution is similar to the true asymptotic distribution, the results will be good out to very large distances.

IX. SOME NUMERICAL INVESTIGATIONS

Calculations on the IBM-704 were carried out for gamma rays, for which $V = E^{-1}$ in units of mc^2. The material was aluminum. In the test runs to be discussed, the energy grouping was: group 1, 2.75–2.50 MeV; group 2, 2.50–1.75 MeV; group 3, 1.75–1.00 MeV.

In all the computations each $\mathbf{f}_l$ was approximated by $\mathbf{f}_l = f_l\mathbf{I}$. To obtain some idea of the sensitivity of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to the choice of f_l, two forms were used: $f_l^{(1)} = c_l$ and $f_l^{(2)} = c_l V^{-2}$, where c_l is independent of V. The ratios of corresponding elements of the $\boldsymbol{\alpha}$ -matrices for the two forms of f_l were equal to within about one percent within each block (i.e., the ratio was nearly independent of angular indices for each pair of energy indices), but varied more significantly from block to block. The ratios $\alpha_s^{(2)}/\alpha_s^{(1)}$ of corresponding elements of $\boldsymbol{\alpha}_s^{(2)}$ and $\boldsymbol{\alpha}_s^{(1)}$, where $\boldsymbol{\alpha}_s$ is the scattering part of $\boldsymbol{\alpha}$, are given in Table I.

The effect of uncertainties in f_l can be minimized by taking the intervals as small as possible. For the most accurate results, f_l would be determined iteratively, using the output of each stage to determine the form of f_l at the next.

Tables II and III show total transmitted fluxes for 20 and 100 cm of aluminum, respectively, for

[24] One can show, by an argument paralleling the demonstration that the eigenvalues of $\mathbf{W}$ are real, that two eigenvalues corresponding to a single energy group will be different. This assures, incidentally, that $\mathbf{A}$ and therefore $\mathbf{W}$ can be diagonalized. In case of accidental degeneracy involving two eigenvalues corresponding to different energy groups l and l', one of the eigenvalues can be perturbed somewhat.

[25] The argument, given in Ref. 11, is not rigorous, but it seems quite convincing.

[26] Thus the exact $\mathbf{A}$ obtained by letting N go to infinity is seen to have real eigenvalues.

[27] This is shown by considering the problem of diffusion over the sphere of directions, with the diffusion coefficient depending on position.

TABLE I. Ratios of corresponding elements: $[\alpha_s^{(2)}]_{mn}/[\alpha_s^{(1)}]_{mn}$.

Exit group	Incident group 1	2	3
1	1.03	...	...
2	1.00	1.10	...
3	1.00	0.96	1.13

isotropic incidence.[28] The transmission matrix results are given for either three or five energy groups, and either three or seven angular terms of a double-P_n representation.[29] For the five-group calculation, the group from 1.75–2.50 MeV was broken into subgroups. The last two columns contain Monte Carlo results[30] for comparison. In all cases the normalization is to unit incident flux. The results are total transmitted flux in each energy group, integrated over direction.

More detailed results for aluminum slabs are given elsewhere.[11]

A basic problem is to determine how many terms are needed in the angular expansion. The number required depends not only on the type of angular expansion functions but on the width of the energy groups. If the groups are narrow, each eigendistribution will be sharply peaked in angle in its top group and many angular basis functions must be used to represent the peak adequately.

There are several reasons to believe that out at least to ten or 15 mean free paths, it is sufficient to take about seven terms (through P_6) in a double-P_n expansion for the energy grouping used in the sample problems. Seven terms were originally used because of programing limitations.

TABLE II. Transmission through 20 cm of aluminum, in units of 10^{-2} photons cm^{-2} sec^{-1}. Isotropically incident flux normalized to 1 photon cm^{-2} sec^{-1}.

Energy groups / Angle terms	Transfer matrix 3 / 3	3 / 7	5 / 3		Adjusted Monte Carlo	
Unscattered	3.809	3.833		3.809	3.75	4.04
2.75–2.50					0.29	
2.50–2.25			0.639		0.63	
2.25–2.00	1.970	1.971	0.651	1.966	0.78	2.11
2.00–1.75			0.676		0.70	
1.75–1.00	2.456	2.464		2.456		2.49

TABLE III. Transmission through 100 cm of aluminum, in units of 10^{-6} photons cm^{-2} sec^{-1}. Isotropically incident flux normalized to 1 photon cm^{-2} sec^{-1}.

Energy groups / Angle terms	Transfer matrix 3 / 7	5 / 3		Adjusted Monte Carlo	
Unscattered	4.169		3.864	3.83	5.31
2.75–2.50				1.48	
2.50–2.25		2.263		3.10	
2.25–2.00	7.818	2.431	7.348	3.29	9.75
2.00–1.75		2.654		3.36	
1.75–1.00	11.116		10.798		16.88

One argument comes from examining the eigendistributions, which correspond to the columns of $\mathbf{B}_+$. The most peaked of all the angular distributions appearing in any eigendistribution is expected to be that in the fundamental (i.e., least attenuated) distribution for radiation in the highest energy group. Table IV shows the fundamental eigendistribution for the sample problem, broken up by l-value, for a double-P_6 calculation. The $l = 1$ column is the most significant, as just indicated. The last element in the $l = 1$ column is seen to be half the preceding one. Now it can be shown that for a P_n expansion[31] the elements in the top energy group, when arranged in order of increasing angular index, decrease factorially when the index is sufficiently large. It seems reasonable to assume that the sample problem shows the beginning of such a decrease. Another argument is that extensive calculations by the moments method[32] showed[33] that it was sufficient to go to

TABLE IV. Fundamental eigendistribution for aluminum. Initial energy group 2.50–2.75 MeV (arbitrary normalization).

Angle index n	Energy group l 1	2	3
0	0.21	0.91	1.28
1	0.59	2.38	2.54
2	0.87	3.02	1.85
3	1.00	2.83	0.65
4	0.96	2.14	0.00
5	0.75	1.33	−0.13
6	0.40	0.67	0.01

[28] The transfer matrix calculations inadvertently used an erroneous conversion factor which produced a scattering cross section about 5 percent too low, though the total cross section was correct. It is not now possible to repeat the calculation with corrected data, but a corrected computation would presumably improve the agreement in Tables II and III.

[29] The basis functions in each hemisphere are $P_n(2\omega - 1)$.

[30] H. Steinberg and R. Aronson, WADC-TR-59-771 (1959).

[31] See, for instance, B. Davison, *Neutron Transport Theory* (Oxford University Press, New York, 1958).

[32] U. Fano, L. V. Spencer, and M. J. Berger, "Penetration and Diffusion of X-Rays," in *Encyclopedia of Physics*, Part II, edited by S. Flügge (Springer-Verlag, Berlin, 1959), Vol. XXXVIII.

[33] H. Goldstein, *Fundamental Aspects of Reactor Shielding* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1959).

P_6 to get a good representation of total flux and current out to 15 mean free paths. Further, Tables II and III show reasonable agreement of a seven-angle-term calculation with Monte Carlo calculations, especially in the upper energy group.

None of these arguments is conclusive, but together they make it plausible that in problems similar to those discussed here, about seven angular terms should suffice.

X. EVALUATION AS A COMPUTATIONAL METHOD

The angle and energy representations in the calculations done so far were chosen primarily for convenience. Nevertheless, the results indicate that the method is perfectly feasible for large-scale computation and that a proper choice of representation can be expected to produce quite accurate values for angular and energy spectra.

The transfer matrix calculations give the solution of a large number of problems at once, one corresponding to each component of the incident distribution. That is, one obtains the entire **T** and **R** matrices. It has previously been very difficult to get such extensive information, especially angular distributions. Thus if detailed information about angular and energy spectra for a variety of source distributions is wanted, the method is very attractive. On the other hand, if one is interested only in a single problem, for instance, the total dose due to a given monoenergetic source with a given angular distribution, then it gives much more information than is wanted.

A major computational advantage is that various intermediate results are common to problems for different configurations. The bulk of the computing time for a single problem goes into evaluating the matrices $\mathbf{B}_\pm$ and $\mathbf{C}_\pm$ which together form the diagonalizing matrix for **W**. These are specific to the material and do not depend on the slab thickness. The **H**, **T**, **U**, and **R** matrices for a slab are characteristic only of the material and the thickness of the particular layer, and do not depend on the remainder of the configuration. Thus intermediate results such as the $\mathbf{B}_\pm$, $\mathbf{C}_\pm$, and **H** matrices can be stored on cards or tape and need never be recomputed. This indicates that the method is best suited for extensive computation programs.

All the calculations so far performed have used a half-range Legendre polynomial expansion in angle and a multigroup representation in energy. There is no reason to believe that this is the best representation possible, especially in angle. The question is an important one, since the total computation time is roughly proportional to the time required to perform a single matrix multiplication, which in turn goes roughly as the cube of the order of the matrices. There is thus a very strong incentive to represent the distributions as efficiently as possible. The problem is to determine a finite set of functions of angle such that linear combinations of them will give good approximations to the transmitted angular distributions for a large variety of configurations.

XI. COMPARISON AND SUMMARY

The essential principles of the transfer matrix method developed here are:

1. The reflection and transmission matrices can be combined into an **H**-matrix that satisfies a simple composition law (Sec. II).
2. For a slab one has $\mathbf{H} = \exp - \mathbf{W}t$, where **W** is independent of thickness (Sec. III).
3. **W** can be determined in terms of the differential cross sections, that is, from microscopic properties of the medium (Sec. V).

In recent years, Principles 1 and 2 have often been rediscovered and used in various applications. The exponential property of **H** is not always pointed out. In particular, when α and β are 1×1 matrices—that is, numbers—the diagonalization is simple and one need not formulate Principle 2 explicitly. It is instructive to look at the papers mentioned in the Introduction in relation to these three principles.

In his work on the transmission of polarized light, Jones[7] was not concerned with internal reflections, and so in his work there is no distinction between scattering and transfer matrices. Both reduce to transmission matrices for the amplitudes of the electric field in the two polarization states. Principle 1 follows immediately. Jones explicitly states and uses Principle 2.

Bobrowsky[6] derived Principle 1. Principles 2 and 3 did not enter into his work since he did not look at the structure of the individual **H**-matrices. Rather, he made some hand computations using somewhat arbitrary values for the elements of the individual transfer matrices.

The work of Rowe,[2] of Aronson,[3] and of Yarmush[4] on electromagnetic or mechanical vibrations all make use of Property 1. Property 2 occurs only implicitly in their work. Pace and Mittra[5] also use only Principle 1 in their treatment of transmission in a bifurcated wave guide. In that context they find it necessary to discuss the algebra explicitly.

In none of these papers is **W** determined in terms of microscopic or atomic properties (Principle 3);

all the relevant information is already contained in the index of refraction or some other macroscopic quantity. A comparable situation arises in diffusion theory. The **H**-matrix can be found from the diffusion solution by matching boundary conditions. Then **W** is given directly in terms of the diffusion coefficient and the attenuation length, which are considered to be macroscopic parameters.[34]

The transfer-matrix approach is really a whole family of techniques. Each standard approximation for the angle and energy integration of the transport equation corresponds to a particular choice of angular and energy basis functions. Thus, for instance, the discrete ordinate (Wick–Chandrasekhar) method,[9] the spherical harmonic (P_n) method[31] and the S_n method,[35] which are generally regarded as methods for approximating the integrodifferential linear Boltzmann equation, can equally well be used to approximate the explicit solution obtained in Secs. III-V. From this point of view the transfer matrix approach can be regarded as approximating the formal solution of the Boltzmann equation directly rather than first approximating the equation itself. Explicit formulas for a polynomial expansion in the cosine of the angle and a multigroup scheme in energy have been given in Sec. VI. In Appendices D and E the method is applied specifically to a P_n expansion. The sample numerical calculations discussed in Sec. IX indicate that large scale calculations are feasible in a double P_n multigroup approximation. The point of this paper has been, however, not to justify the use of any particular basis but to outline the general procedure with enough detail so that one can see how to set up numerical computations.

Because the method is in principle so versatile, it is tempting to try to apply it to irregular configurations in which position as well as direction and energy is a variable. So far such an extension has proved impractical, since the matrices become too large to handle readily. However, in geometries in which there are certain simplifications, notably in one-dimensional spherical and cylindrical configurations, the method may be attractive. A discussion of media whose bounding surfaces are of the form $\rho_i = r_i f(\theta, \phi)$ where ρ_1 and ρ_2 are, respectively, the radii vectors to the inner and outer surfaces, is given in Appendix F.

[34] We make the distinction here between diffusion theory, which is a macroscopic theory, and the P_1-approximation, in which the diffusion parameters have a meaning in terms of the cross sections.

[35] B. Carlson in *Methods in Computational Physics*, edited by B. Alder, S. Fernbach, and M. Rotenberg (Academic Press Inc., New York, 1963) Vol. 1, p. 1.

The method may also prove useful in multiplying media and in the presence of internal sources, though calculations in such media have not been carried out. One can also determine the conditions for criticality. In the formalism described here, criticality is characterized by the condition[2,3] $\det \mathbf{U} = 0$.

ACKNOWLEDGMENTS

We wish to acknowledge the contribution of Julius Zell, who was responsible for the machine programming and much of the accompanying analysis. We also wish to thank Dr. Herbert Steinberg for many helpful discussions.

APPENDIX A. SERIES SOLUTION FOR H

To find a series expansion for **H** it is most convenient to work with

$$\mathbf{J} = \mathbf{P}^{-1}\mathbf{W}\mathbf{P} \tag{A1}$$

where **W** is given by (14) and

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}. \tag{A2}$$

I is a unit matrix in the energy-direction subspace. **J** has the explicit structure

$$\mathbf{J} = \begin{bmatrix} 0 & \boldsymbol{\delta} \\ \boldsymbol{\delta} & 0 \end{bmatrix}, \tag{A3}$$

where $\boldsymbol{\delta} = \boldsymbol{\alpha} + \boldsymbol{\beta}$ and $\boldsymbol{\delta} = \boldsymbol{\alpha} - \boldsymbol{\beta}$. One can show easily by induction that

$$\mathbf{J}^{2n} = \begin{bmatrix} (\boldsymbol{\delta}\boldsymbol{\delta})^n & 0 \\ 0 & (\boldsymbol{\delta}\boldsymbol{\delta})^n \end{bmatrix},$$
$$\mathbf{J}^{2n+1} = \begin{bmatrix} 0 & \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n \\ \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n & 0 \end{bmatrix}. \tag{A4}$$

Then

$$\mathbf{H} = \mathbf{P}(\exp - \mathbf{J}t)\mathbf{P}^{-1} = \sum_{k=0}^{\infty} \frac{(-1)^k t^k}{k!} \mathbf{P}\mathbf{J}^k\mathbf{P}^{-1}. \tag{A5}$$

Breaking up the summation in (A5) into sums over even and odd values of k, evaluating $\mathbf{P}\mathbf{J}^k\mathbf{P}^{-1}$ using the forms (A4) for $\mathbf{J}_k$, one finds

$$\mathbf{H} = \frac{1}{2}\sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} \begin{bmatrix} (\boldsymbol{\delta}\boldsymbol{\delta})^n + (\boldsymbol{\delta}\boldsymbol{\delta})^n & (\boldsymbol{\delta}\boldsymbol{\delta})^n - (\boldsymbol{\delta}\boldsymbol{\delta})^n \\ (\boldsymbol{\delta}\boldsymbol{\delta})^n - (\boldsymbol{\delta}\boldsymbol{\delta})^n & (\boldsymbol{\delta}\boldsymbol{\delta})^n + (\boldsymbol{\delta}\boldsymbol{\delta})^n \end{bmatrix}$$
$$- \frac{1}{2}\sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!}$$
$$\cdot \begin{bmatrix} \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n + \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n & \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n - \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n \\ -\boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n + \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n & -\boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n - \boldsymbol{\delta}(\boldsymbol{\delta}\boldsymbol{\delta})^n \end{bmatrix}. \tag{A6}$$

Thus $\mathbf{U}$ can be written as

$$\mathbf{U} = \frac{1}{2}\sum_{k=0}^{\infty}(\mathbf{A}_k + \mathbf{B}_k)\frac{t^k}{k!}, \tag{A7}$$

with $\mathbf{A}_0 = \mathbf{B}_0 = \mathbf{I}$ and $\mathbf{A}_k$ and $\mathbf{B}_k$ computed recursively by

$$\begin{aligned}\mathbf{A}_k &= \mathfrak{d}\mathbf{B}_{k-1},\\ \mathbf{B}_k &= \mathfrak{d}\mathbf{A}_{k-1}.\end{aligned} \tag{A8}$$

The recursive generation of the coefficients $\mathbf{A}_k$ and $\mathbf{B}_k$ makes the series solution easily adaptable to machine computation. However, the series converges very slowly, since convergence is determined by the most rapidly attenuating component, whereas physically the least rapidly attenuating component dominates for large thicknesses. The diagonalization method described earlier was found to be preferable.

It is of interest that the series can be formally summed to give

$$\mathbf{H} = \frac{1}{2}\begin{bmatrix}\xi + \mathfrak{n} & \xi - \mathfrak{n}\\ \xi - \mathfrak{n} & \xi + \mathfrak{n}\end{bmatrix} + \frac{1}{2}\begin{bmatrix}-\mathfrak{y} - \mathfrak{v} & -\mathfrak{y} + \mathfrak{v}\\ \mathfrak{y} - \mathfrak{v} & \mathfrak{y} + \mathfrak{v}\end{bmatrix} \tag{A9}$$

where

$$\begin{aligned}\xi &= \cosh(\mathfrak{d}\mathfrak{d})^{\frac{1}{2}}t, & \mathfrak{n} &= \cosh(\mathfrak{d}\mathfrak{d})^{\frac{1}{2}}t,\\ \mathfrak{y} &= \mathfrak{d}(\mathfrak{d}\mathfrak{d})^{\frac{1}{2}}\sinh(\mathfrak{d}\mathfrak{d})^{-\frac{1}{2}}t & \mathfrak{v} &= \mathfrak{d}(\mathfrak{d}\mathfrak{d})^{-\frac{1}{2}}\sinh(\mathfrak{d}\mathfrak{d})^{\frac{1}{2}}t.\end{aligned} \tag{A10}$$

APPENDIX B. CUTBACK PROCEDURE FOR NUMERICAL COMPUTATION OF T AND R

The transmission and reflection matrices are given in terms of decreasing exponentials only, according to Eqs. (23a)–(24b). However, it is possible to compute T and R directly, according to Eqs. (21a)–(22b), without eliminating the increasing exponentials, if sufficient care is taken.

Let us assume that the machine holds floating point numbers to m decimal digits. Define $M = m \log_e 10$. Let λ_0 be the smallest eigenvalue and λ_n some other eigenvalue for which $(\lambda_n - \lambda_0)t > M$, where t is the thickness. $\mathbf{U}$ is a sum of terms of the form $b_i(\exp \lambda_i t)c_i$. We assume that the b_i and c_i are of the order of unity. Then a term $b_n(\exp \lambda_n t)c_n$ will be approximately 10^m times as large as $b_0(\exp \lambda_0 t)c_0$, and no trace of the latter term remains in any element of $\mathbf{U}$.

To avoid this loss of significant information, we replace λ_n by a cutback value λ_n' defined by

$$(\lambda_n' - \lambda_0)t = K.$$

Here K is a cutback constant whose value is optimally chosen, as we shall now show, as about equal to $M/2$. With this value for K, the physically important term $b_0(\exp \lambda_0 t)c_0$ is about $\exp(-K)$ times as large as the physically unimportant term $b_n(\exp \lambda_n' t)c_n$.

Define $k = K \log_{10} e$. The important information is then contained in the $(m - k)$ less significant digits of $\mathbf{U}$. On the other hand, the eigendistribution corresponding to λ_n' is computed to be attenuated by a factor no greater than about $\exp(-K)$ times the attenuation of the fundamental distribution. If a calculated eigendistribution is attenuated more strongly than this, the fractional error in the final result introduced in replacing λ_n by λ_n' is about $\exp(-K)$, so that only the first k digits in the calculated transmission are reliable. Thus the number of meaningful digits is $\min(m - k, k)$, which is a maximum for $k = \frac{1}{2}m$.

APPENDIX C. EVALUATION OF D_{mn}

Equation (45) defines $D_{mn}(\gamma)$ as[36]

$$D_{mn}(\gamma) = \int_0^1 d\omega \int_0^1 d\omega'\, \omega^m \omega'^n \operatorname{Re}\frac{1}{\pi S(\omega, \omega', \gamma)},$$

where

$$S^2 = 1 - \omega^2 - \omega'^2 - \gamma^2 + 2\omega\omega'\gamma.$$

The integrand vanishes when $S^2 < 0$. This condition cuts off the integral at values of ω and ω' less than the nominal upper limit, unity. Let

$$\omega = \rho \sin\phi, \qquad \omega' = \rho\cos\phi,$$

$$r^2 = (1 - \gamma^2)(1 - \gamma \sin 2\phi)^{-1}.$$

Then $S^2 \geq 0$ implies $\rho \leq r$. In that case, $\omega^2 \leq r^2 \sin^2\phi \leq 1$ and $\omega'^2 \leq r^2\cos^2\phi \leq 1$, so the limits on ϕ are $(0, \frac{1}{2}\pi)$. Let $\rho = r\cos\alpha$. Integrating over $0 \leq \phi \leq \frac{1}{2}\pi$, $0 \leq \alpha \leq \frac{1}{2}\pi$, we get

$$D_{mn}(\gamma) = C_{m+n+1}\sin^{m+n+1}\eta K_{mn}(\gamma), \tag{C1}$$

where

$$\begin{aligned}C_k &= \frac{1}{\pi}\int_0^{\frac{1}{2}\pi}\cos^k\alpha\, d\alpha\\ &= \frac{1\cdot 3\cdot 5\cdots(k-1)}{2\cdot 4\cdot 6\cdots k}\frac{1}{2}, \quad k > 0 \text{ and even}\\ &= \frac{2\cdot 4\cdots(k-1)}{3\cdot 5\cdots k}\frac{1}{\pi}, \quad k > 0 \text{ and odd},\end{aligned} \tag{C2}$$

[36] In this appendix the indices of the elements of D will be written as subscripts to avoid confusion with exponents.

$$K_{mn}(\gamma) = K_{nm}(\gamma) = \int_0^{\frac{1}{2}\pi} (1 - \gamma \sin 2\phi)^{-\frac{1}{2}(m+n+2)} \cdot \sin^m \phi \cos^n \phi \, d\phi, \qquad \gamma^2 \leq 1 \tag{C3}$$

and

$$\eta = \arccos \gamma. \tag{C4}$$

From (C3) one finds the recursion relation

$$K_{mn}(\gamma) = \frac{1}{m+n} \frac{d}{d\gamma} K_{m-1,n-1}(\gamma). \tag{C5}$$

By making the substitution

$$\cot \phi = \cos \eta - \sin \eta \cot z$$

in (C3), one finds

$$K_{m0} = \csc^{m+1} \eta \int_\eta^\pi \sin^m z \, dz. \tag{C6}$$

Combining (C1) and (C6), one finds

$$D_{m0}(\gamma) = C_{m+1} \int_\eta^\pi \sin^m z \, dz, \tag{C7}$$

and therefore

$$D'_{m0}(\gamma) = C_{m+1}(1 - \gamma^2)^{\frac{1}{2}(m-1)}, \tag{C8}$$

where a prime is used to denote a derivative with respect to γ.

We will now demonstrate algebraic recurrence relations for computing the D_{mn}. It will turn out that the cases $m + n$ even and $m + n$ odd are completely uncoupled. Since the D_{mn} are symmetric in m and n, they need be computed only for $m \geq n$.

Integrating (C7) by parts and using (C2), one has

$$D_{m0}(\gamma) = \frac{1}{m+1} [C_{m-1}\gamma(1 - \gamma^2)^{\frac{1}{2}(m-1)} + (m - 1)D_{m-2,0}(\gamma)]. \tag{C9}$$

From (C9) and the starting relations

$$D_{00} = 1 - (\eta/\pi), \tag{C10a}$$

$$D_{10} = \tfrac{1}{4}(1 + \gamma), \tag{C10b}$$

we find all the D_{m0}.

From (C1) and (C5) we find

$$D_{mn} = \frac{C_{m+n+1} \sin^{m+n+1} \eta}{(m+n)(m+n-2) \cdots (m-n+2)} \times \frac{d^n}{d\gamma^n} K_{m-n,0}. \tag{C11}$$

Equation (C11) implies directly that

$$D_{m+1,n+1} = [1/(m+n+3)][(1 - \gamma^2)D'_{mn} + (m+n+1)\gamma D_{mn}]. \tag{C12}$$

Specializing to $n = 0$ and using (C8), we obtain

$$D_{m1} = [1/(m+2)][C_m(1 - \gamma^2)^{\frac{1}{2}m} + m\gamma D_{m-1,0}], \tag{C13}$$

which gives all the D_{m1}.

To obtain the other D_{mn}, we first observe that

$$D'_{m1} = [m/(m+2)]D_{m-1,0}, \tag{C14}$$

by explicit differentiation of (C11) for $n = 1$, using (C6). From (C14) and the differentiated form of (C12), we obtain by induction on $m + n$

$$D'_{mn} = [mn/(m+n+1)]D_{m-1,n-1}. \tag{C15}$$

(C12) and (C15) yield a second-order differential equation

$$(1 - \gamma^2)D''_{mn} + (m+n-1)D'_{mn} - mnD_{mn} = 0 \tag{C16}$$

and a purely algebraic recursion relation

$$D_{m+1,n+1} = \frac{1}{(m+n+1)(m+n+3)} \times [(m+n+1)^2\gamma D_{mn} + mn(1 - \gamma^2)D_{m-1,n-1}]. \tag{C17}$$

(C17) is the final relation needed for computing the remainder of the D_{mn}.

We note finally that for $m + n$ even, D_{mn} has the structure

$$D_{mn}(\gamma) = \left(1 + \frac{\eta}{\pi}\right)A_{mn}(\gamma) + \frac{(1 - \gamma^2)^{\frac{1}{2}}}{\pi} B_{mn}(\gamma), \tag{C18}$$

where A_{mn} is a polynomial in γ of degree min (m, n) and B_{mn} a polynomial of degree [max $(m, n) - 1$]. A_{mn} and B_{mn} are either even or odd polynomials. This all follows from the fact that D_{00} [(Eq. (C10a)] is of the form (C18) and the form is preserved by all the relations (C9), (C13), and (C17) from which the other D_{mn} are found. Similarly, for $m + n$ odd, D_{mn} is a polynomial of degree max (m, n).

APPENDIX D. LEGENDRE POLYNOMIALS AS BASIS FUNCTIONS

Consider the full-range Legendre polynomials $P_n(\omega)$ taken as basis functions, with $g(\omega) = 1$. Then certain of the elements $\bar{D}_{mn}$ of the matrix $\bar{\mathbf{D}} = \mathbf{qGDq} = \mathbf{qDq}$ can be evaluated directly, without first evaluating the D_{mn} as in the previous appendix and then taking linear combinations.

We note first that we can write

$$\operatorname{Re}\frac{1}{\pi S(\omega,\omega',\gamma)} = \frac{1}{\pi}\int_0^\pi d\phi\ \delta(\gamma-\omega\omega'-(1-\omega^2)^{\frac{1}{2}}(1-\omega'^2)^{\frac{1}{2}}\cos\phi). \tag{D1}$$

$\bar{D}_{mn}$ can be written explicitly as

$$\bar{D}_{mn}(\gamma) = \int_0^1 d\omega \int_0^1 d\omega'\ \operatorname{Re}\frac{P_m(\omega)P_n(\omega')}{\pi S(\omega,\omega',\gamma)}. \tag{D2}$$

We introduce the Legendre expansion

$$\bar{D}_{mn}(\gamma) = \sum_{k=0}^{\infty} \tfrac{1}{2}(2k+1)a_{k,mn}P_k(\gamma). \tag{D3}$$

Then

$$a_{k,mn} = \int_{-1}^1 \bar{D}_{mn}(\gamma)P_k(\gamma)\,d\gamma = \int_{-1}^1 P_k(\gamma)\,d\gamma \int_0^1 P_m(\omega)\,d\omega \int_0^1 P_n(\omega')\,d\omega' \times \frac{1}{\pi}\int_0^\pi d\phi\ \delta(\gamma-\omega\omega' - (1-\omega^2)^{\frac{1}{2}}(1-\omega'^2)^{\frac{1}{2}}\cos\phi), \tag{D4}$$

where we have used (D1) and (D2). We integrate first over γ and then over ϕ, and use the addition theorem for spherical harmonics. Only one term survives after integration and we find

$$\bar{D}_{mn}(\gamma) = \sum_{k=0}^{\infty} \tfrac{1}{2}(2k+1)I_{km}I_{kn}P_k(\gamma) \tag{D5}$$

where

$$I_{km} = \int_0^1 P_k(\omega)P_m(\omega)\,d\omega. \tag{D6}$$

If m and k have the same parity,

$$I_{km} = \delta_{km}/(2m+1).$$

Thus, if m and n have opposite parity,

$$\bar{D}_{mn}(\gamma) = \tfrac{1}{2}[P_m(\gamma)+P_n(\gamma)]I_{mn}. \tag{D7}$$

If m and n have the same parity, there may yet be some point in using the series (D5) if convergence is sufficiently rapid, rather than first computing the D_{mn}. In that case,

$$\bar{D}_{mn}(\gamma) + (-1)^m\bar{D}_{mn}(-\gamma) = \frac{1}{2m+1}P_m(\gamma)\delta_{mn}. \tag{D8}$$

Now for the Legendre polynomials, $q_{mn} = 0$ for $m+n$ odd. Then from the form (C18) of the D_{mn} we deduce

$$\bar{D}_{mn}(\gamma) = \bar{A}_{mn}(\gamma)\left(1-\frac{\eta}{\pi}\right) + \bar{B}_{mn}(\gamma)\frac{(1-\gamma^2)^{\frac{1}{2}}}{\pi}, \tag{D9}$$

where $\bar{\mathbf{A}} = \mathbf{q}\mathbf{A}\tilde{\mathbf{q}}$, $\bar{\mathbf{B}} = \mathbf{q}\mathbf{B}\tilde{\mathbf{q}}$. $\bar{A}_{mn}$ has the same parity as m and $\bar{B}_{mn}$ has opposite parity. It follows, by evaluating the left-hand side of (D8) with the help of (D9), that

$$\bar{A}_{mn}(\gamma) = \frac{1}{2m+1}P_m(\gamma)\delta_{mn}. \tag{D10}$$

No simple form for the $\bar{B}_{mn}$ has been found. They are apparently related to the second solution of Legendre's equation.

Finally, it is of interest to note that (D1) implies the expansion

$$\operatorname{Re}\frac{1}{\pi S(\omega,\omega',\gamma)} = \sum_{k=0}^{\infty}\frac{2k+1}{2}P_k(\omega)P_k(\omega')P_k(\gamma), \qquad S^2>0. \tag{D11}$$

APPENDIX E. INTERIOR FLUX BY SPHERICAL HARMONIC EXPANSION

There is nothing in the transfer matrix formalism that specifies whether one is computing the flux in the interior or at the surface of a medium. The only difference is in the boundary conditions. It is of interest to consider a spherical harmonic expansion of the interior flux for a one-velocity problem. We will show how one gets the standard differential equations for the spherical harmonic coefficients for the transport problem from the differential form of (9). To see this, one rewrites (9) for a differential thickness dx as

$$\mathbf{\Phi}(x+dx) = \mathbf{H}(dx)\mathbf{\Phi}(x) + \mathbf{Q}(dx). \tag{E1}$$

By virtue of (12a, b) and (13), one has the differential equation

$$\mathbf{\Phi}'(x) = -\mathbf{W}\mathbf{\Phi}(x) + \mathbf{Q}'(x), \tag{E2}$$

where

$$\mathbf{Q}'(x) = \begin{bmatrix} S_+(x) \\ -S_-(x)\end{bmatrix}. \tag{E3}$$

In this appendix a prime will denote a derivative with respect to x. $S_+(x)dx$ is the unscattered flux at $x+dx$ due to sources in dx and $S_-(x)dx$ is that at x due to sources in dx. $\mathbf{\Phi}$ can be written as

$$\mathbf{\Phi} = \begin{bmatrix}\chi_1 \\ \chi_2\end{bmatrix}.$$

We now introduce the angular coordinate explicitly. Let the flux $\Phi(x, \omega)$ be expanded in Legendre polynomials with respect to the positive x axis:

$$\Phi(x, \omega) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} a_l(x) P_l(\omega), \qquad -1 \leq \omega \leq 1. \tag{E4}$$

Now by our convention, χ_1 and S_+ are measured with respect to the positive x axis and χ_2 and S_- are measured with respect to the negative x axis. In the latter case, $-\omega \to \omega$. Thus

$$\chi_1(x, \omega) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} a_l(x) P_l(\omega), \qquad 0 \leq \omega \leq 1 \tag{E5a}$$

$$\chi_2(x, \omega) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} (-1)^l a_l(x) P_l(\omega), \qquad 0 < \omega \leq 1. \tag{E5b}$$

Let $S(x, \omega)$ be the source strength at x and ω per unit volume and solid angle. S can also be expanded:

$$S(x, \omega) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} s_l P_l(\omega), \qquad -1 \leq \omega \leq 1. \tag{E6}$$

Then

$$S_+(x, \omega) = \frac{1}{\omega} \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} s_l(x) P_l(\omega), \qquad 0 < \omega \leq 1, \tag{E7a}$$

$$S_-(x, \omega) = \frac{1}{\omega} \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} (-1)^l s_l(x) P_l(\omega), \qquad 0 < \omega \leq 1. \tag{E7b}$$

One should note that there is a redundancy in the elements of Φ. For each l, a_l appears as a component of χ_1 and $(-1)^l a_l$ as a component of χ_2. It is convenient to consider even and odd l components of Eq. (E2) separately. One finds

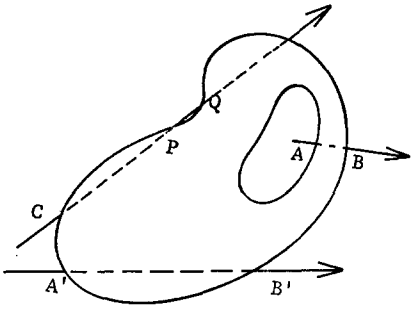

FIG. 5. Schematic general geometry.

$$\mathbf{F}' = -\mathfrak{d}\mathbf{G} + (\mathbf{S}_+ - \mathbf{S}_-), \tag{E8a}$$

$$\mathbf{G}' = -\mathfrak{d}\mathbf{F} + (\mathbf{S}_+ + \mathbf{S}_-), \tag{E8b}$$

where $\mathbf{F} = \chi_1 + \chi_2$, $\mathbf{G} = \chi_1 - \chi_2$, and $\mathfrak{d}$ and $\mathfrak{d}$ are given by (16a, b). In terms of the integral kernels,

$$\mathfrak{d}\mathbf{G} \to \int_0^1 [K_\alpha(\omega, \omega') + K_\beta(\omega, \omega')] \times G(x, \omega')\, d\omega', \qquad 0 < \omega \leq 1,$$

with a parallel expression for $\mathfrak{d}\mathbf{F}$. Using the appropriate one-velocity form for (28a, b), one finds after some manipulation

$$a_l' = \sum_{k=0}^{\infty} \frac{2k+1}{2} [s_k - (\mu - n\sigma_s f_k) a_k] \times \int_{-1}^{1} \frac{P_k(\omega) P_l(\omega)}{\omega}\, d\omega, \tag{E9}$$

where σ_s is the scattering cross section and $\sigma_s f_k$ is the normalized kth partial scattering cross section:

$$f_k = \frac{2\pi}{\sigma_s} \int_{-1}^{1} \sigma(\omega) P_k(\omega)\, d\omega. \tag{E10}$$

The integral in (E9) is to be taken in a principal value sense, so that it vanishes when k and l have the same parity.

From (E9),

$$\frac{n+1}{2n+1} a_{n+1}' + \frac{n}{2n+1} a_{n-1}' = s_n - (\mu - n\sigma_s f_n) a_n, \tag{E11}$$

which is the usual form of the P_n equations.[31]

APPENDIX F. GENERAL GEOMETRIES

Consider a body of arbitrary general geometry in which the surface is divided into two nonoverlapping complementary faces. A particle is said to be reflected if it emerges from the same face at which it enters. If it emerges from the other face, it is transmitted. Thus, the particles incident at A′ and C in Fig. 5 are reflected from the outer face. The particle emerging at P is assumed to be absorbed in some way and not to reenter at Q even though Q lies on the straight line CP. The particle incident at A is transmitted. If the medium is not a slab, an infinite cylindrical shell, or a spherical shell, then $\mathbf{T}$, $\mathbf{R}$, $\mathbf{T}^*$, and $\mathbf{R}^*$ are integral operators whose kernels depend on the entrance and exit points of the rays as well as on direction and energy.

We now specialize to a family of concentric similar shells whose inner and outer faces are given, respectively, by

$$\rho_{in} = r_1 f(\theta, \phi), \qquad \rho_{out} = r_2 f(\theta, \phi) \tag{F1}$$

where (ρ, θ, ϕ) are spherical coordinates with some point in the interior as origin. r is the parameter which indicates which member of the family of bounding surfaces is under consideration and can be regarded as the radius of that surface.

The **H**-matrix for the shell whose first face has radius r_1 and whose second face has radius r_2 will be written $\mathbf{H}(r_2, r_1)$. The formula (5) becomes

$$\mathbf{H}(r_3, r_2)\mathbf{H}(r_2, r_1) = \mathbf{H}(r_3, r_1). \tag{F2}$$

This relation holds whether or not r_2 is between r_1 and r_3.

In order to eliminate the inverse r^2 effect, it is convenient to define intensities as the ordinary flux per unit area and solid angle multiplied by r^2. With this convention a simultaneous scaling up of the density by a given factor and scaling down of distances by the same factor will leave **H** unchanged.

Consider now $\mathbf{H}(r + \epsilon, r)$, with ϵ infinitesimally small. Clearly, one can write

$$\mathbf{H}(r + \epsilon, r) - \mathbf{I} = \epsilon(\mathbf{C}_r + \mathbf{E}_r), \tag{F3}$$

where $\epsilon\mathbf{C}_r$ is the change in **H** from the unit matrix **I** due to geometrical factors and $\epsilon\mathbf{E}_r$ is that due to collision. To first order in ϵ, the change due to collision is just $-\epsilon\mathbf{W}$. To this order there is no interaction between collision and geometrical effects. The form of $\mathbf{C}_r$ is determined by noting that in the absence of collision, $\mathbf{H}(r + \epsilon, r)$ is a function of ϵ/r only, for any ϵ. Thus

$$\mathbf{C}_r = \mathbf{C}/r, \tag{F4}$$

$$\mathbf{E}_r = -\mathbf{W} \tag{F5}$$

and **C** and **W** are independent of r.

It is convenient to measure all ratios from some standard radius $r = c$. Then defining

$$\mathbf{Y}(r) = \mathbf{H}(r, c),$$

(F3) becomes

$$\mathbf{Y}'(r) = [(\mathbf{C}/r) - \mathbf{W}]\mathbf{Y}(r). \tag{F6}$$

To find the structure of **C**, we examine the operational meaning of the transmission and reflection operators in the absence of collision, i.e., for a transparent medium. Consider a ray impinging on a face (Fig. 5) at a point A and leaving the other face at a point B. The operator for such processes is **T**. Now consider a ray which enters at B and leaves at A. The operator for such processes is $\mathbf{T}^*$. But for a transparent medium this ray just retraces the path of the first one, and we must have $\mathbf{TT}^* = \mathbf{I}$, or $\mathbf{T} = \mathbf{T}^{*-1} = \mathbf{U}^*$.

Consider now $\mathbf{U}^*\mathbf{R} = \mathbf{TR}$. **R** is the operator corresponding to a ray entering the first face at A′ and leaving the same face at B′. To evaluate **TR**, consider such an arbitrary ray exiting at B′, reverse its direction, and then determine what is transmitted (i.e., crosses the second face). This is clearly zero, since the reversed ray emerges at A'. Hence $\mathbf{U}^*\mathbf{R} = \mathbf{TR} = 0$. If the inner surface is convex, the result is immediately obvious since then $\mathbf{R} = 0$. Similarly, $\mathbf{R}^*\mathbf{U}^* = \mathbf{R}^*\mathbf{T} = 0$. Thus for a transparent medium,

$$\mathbf{H} = \begin{bmatrix} \mathbf{T} & 0 \\ 0 & \mathbf{U}^* \end{bmatrix} = \begin{bmatrix} \mathbf{T} & 0 \\ 0 & \mathbf{T} \end{bmatrix}.$$

From (F6) and the fact that $\mathbf{Y}(c) = \mathbf{I}$, it follows that for a transparent medium, $\mathbf{C} = c\mathbf{Y}'(c)$. Thus **C** has the form

$$\mathbf{C} = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix} = c \begin{bmatrix} \gamma_c & 0 \\ 0 & \gamma_c \end{bmatrix} \tag{F7}$$

where

$$\gamma_c = \lim_{\epsilon \to 0} (d/d\epsilon)\mathbf{T}_{tr}(c + \epsilon, c). \tag{F8}$$

We have used the symbol $\mathbf{T}_{tr}$ to emphasize that the transmission operator in (F8) is that for a transparent medium. (F6) then becomes

$$\mathbf{Y}'(r) = \begin{bmatrix} -\alpha + \frac{\gamma}{r} & \beta \\ -\beta & \alpha + \frac{\gamma}{r} \end{bmatrix} \mathbf{Y}(r). \tag{F9}$$

Cartan Frames and the General Relativistic Dirac Equation*

DIETER R. BRILL AND JEFFREY M. COHEN†
Sloane Physics Laboratory, Yale University, New Haven, Connecticut
(Received 26 April 1965)

The Dirac equation for spin-½ particles in curved space–time is formulated using Cartan calculus. Unlike previous formulations, this method is easy to use because it expresses the Dirac equation in terms of well known objects like partial derivatives and special relativistic Dirac matrices. It allows a simple and direct treatment of neutrinos in homogeneous nonisotropic universes and in plane-wave geometries. These solutions are compared and contrasted with the corresponding solutions containing electromagnetic radiation.

I. INTRODUCTION

ONE of the most fruitful innovations in modern differential geometry is Cartan's calculus of differential forms and movable frames. Familiar to mathematicians since 1901, this calculus has recently been applied to various physical problems. It is particularly appropriate to formulate Maxwell's theory of electromagnetism and Einstein's theory of gravitation in terms of Cartan's calculus; not only does it allow a unified description of the various differential operations and thus facilitate many computations, but it also has opened the door to discovery of many new features and relationships of these theories.[1]

In the case of the Dirac equation in curved space, a formalism involving movable frames (tetrads, Vierbeine) is not only useful but virtually unavoidable. It is therefore natural to use the identical frames to describe the Dirac field in curved space and the curved space itself. In other words, *the Dirac equation can be formulated more simply and directly in a space–time structure described by Cartan frames* than in any other space–time description. Whereas the curved-space Dirac equation, written in holonomic coordinates, has been known to physicists since 1928,[2] its Cartan form[3] has seldom been applied. A short review of the formalism and the derivation of the Dirac equation in curved space is given in Sec. II. In Sec. III we discuss the Dirac equation in homogeneous, nonisotropic universes. Section IV contains a study of the plane wave geometry in the presence of scalar, spinor and electromagnetic fields. In the concluding Sec. V we point out some of the implications of this work for cosmological models and already unified field theory.

II. DIRAC EQUATION IN CARTAN'S MOVING FRAMES

1. Outline of the Cartan Formalism

Cartan's formulation of differential geometry makes it possible to describe independently two aspects of tensor fields which appear inseparable in the usual ("holonomic") tensor calculus: (1) A coordinate system x^μ ($\mu = 0, 1, 2, 3$, in this paper) *labels* the points in space–time, (2) the ("nonholonomic") *components* of vectors and tensors at each point are expressed in orthonormal frames ω^μ which originate at the point, and may change orientation from point to point. These basic "one-forms" ω^μ may be specified, e.g., by expanding them in terms of the gradients dx^ν of the coordinates:

$$\omega^\mu = a^\mu{}_\nu\, dx^\nu. \tag{1}$$

In some cases a preferred set of ω^μ follows naturally from some group property of the space–time.[4]

Since the ω^μ are orthonormal, the nonholonomic metric has the diagonal Minkowskian form

$$g_{\mu\nu} = \text{diag}\,(-1, 1, 1, 1)$$

or

$$ds^2 = -(\omega^0)^2 + \sum_i (\omega^i)^2. \tag{2}$$

* This research has been supported in part by the National Aeronautics and Space Administration.

† National Science Foundation Predoctoral Fellow.

[1] By way of example, see the discussion of electromagnetism in general relativity by C. W. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) 2, 525 (1957).

[2] H. Tetrode, Z. Physik 50, 336 (1928); V. A. Fock and D. Ivanenko, Z. Physik 57, 261 (1929); E. Schrödinger, Sitzber. Preuss. Akad. Wiss., Phys. Math. Kl., 105 (1932); V. Bargmann, Sitzber. Preuss. Akad. Wiss., Phys. Math. Kl., 346 (1932); R. Penrose, Ann. Phys. (N.Y.) 10, 171 (1960). Also see W. L. Bade and H. Jehle, Rev. Mod. Phys., 25, 714 (1953) and the references cited there.

[3] A. Lichnerowicz, Bull. Soc. Math. France, 92, 11 (1964); A. Lichnerowicz, Ann. Inst. Henri Poincaré 1, 233 (1964); A. Lichnerowicz, *Relativity Groups and Topology* (Gordon and Breach Science Publishers, New York, 1964), p. 823. One of us (J.M.C.) independently formulated the Dirac equation in Cartan frames before the above references became available. Also, see Ref. 12 for a discussion of the use of Cartan calculus in physics.

[4] See, for example, A. H. Taub, Ann. Math. 53, 472 (1951).

The frame components of tensors change from point to point for two reasons: (a) change in the tensor itself (b) differential rotation of the frames. The latter is described by the antisymmetric tensor ("two-form"):

$$d\omega^{\mu} = \tfrac{1}{2}(a^{\mu}{}_{\nu,\kappa} - a^{\mu}{}_{\kappa,\nu})\, dx^{\kappa} \wedge dx^{\nu} = -\omega^{\mu}{}_{\kappa} \wedge \omega^{\kappa}, \quad (3)$$

where $dx^{\kappa} \wedge dx^{\nu} = -dx^{\nu} \wedge dx^{\kappa}$. The operator d used here and in the following denotes exterior (antisymmetric) differentiation. Of course, the exterior differential of a scalar is just the ordinary differential from calculus as the notation suggests.

The one-forms $\omega^{\mu}{}_{\kappa}$ occurring in the expansion (3) of the two-forms $d\omega^{\mu}$ are the connection one-forms. Their components $\gamma_{\mu\kappa\lambda}$,

$$\omega_{\mu\kappa} = \gamma_{\mu\kappa\lambda}\omega^{\lambda},$$

are the familiar Ricci rotation coefficients. The orthonormality of the frames implies that $\omega_{\mu\kappa} = -\omega_{\kappa\mu}$; with this condition, Eq. (3) can be solved for a unique set of $\omega^{\mu}{}_{\kappa}$.

The covariant differential of a vector, $a = a^{\alpha}\omega_{\alpha}$, is defined in terms of these connection forms[5]:

$$Da^{\alpha} = da^{\alpha} + \omega^{\alpha}{}_{\beta}a^{\beta}. \quad (4)$$

The covariant derivative $a^{\alpha}{}_{;\gamma}$ is the component of the covariant differential along the ω_{γ} direction:

$$a^{\alpha}{}_{;\gamma} \equiv Da^{\alpha}\cdot\omega_{\gamma} = da^{\alpha}\cdot\omega_{\gamma} + \omega^{\alpha}{}_{\beta}a^{\beta}\cdot\omega_{\gamma}; \quad (5)$$

hence,

$$a^{\alpha}{}_{;\gamma} = \omega_{\gamma}(a^{\alpha}) + \gamma^{\alpha}{}_{\beta\gamma}a^{\beta}.$$

Here ω_{γ}, the vector dual to the form ω^{γ}, is a linear mapping of forms into numbers defined by

$$\omega^{\alpha}\cdot\omega_{\gamma} = \delta^{\alpha}{}_{\gamma}, \quad (6a)$$

which in holonomic frames becomes the well known relation[6] on base vectors e_{γ}:

$$dx^{\alpha}\cdot e_{\gamma} = e_{\gamma}(x^{\alpha}) = \partial_{\gamma}(x^{\alpha}) = \delta^{\alpha}{}_{\gamma}. \quad (6b)$$

2. Dirac Matrices in Moving Frames

The 4×4 matrices γ^{μ} necessary for the formulation of the Dirac equation satisfy the anticommutation relations

$$\gamma^{\mu}\gamma^{\nu} + \gamma^{\nu}\gamma^{\mu} = 2g^{\mu\nu}. \quad (7)$$

Since we are working in orthonormal frames, the metric $g^{\mu\nu}$ has the Minkowskian form (2) of special relativity; therefore, one solution of (7) is given by the special relativity Dirac matrices $\tilde{\gamma}^{\mu}$. The general solution of (7) differs from these by a position-dependent similarity transformation,

$$\gamma^{\mu}{}_{\text{general}} = S^{-1}(x^{\mu})\tilde{\gamma}^{\mu}S(x^{\mu}). \quad (8)$$

Such a more general solution can always be transformed back to the special relativity solution by suitable position-dependent rotations of the Cartan frames. All physical quantities formed from the general solution (8) are the same as those obtained from the special solution $\tilde{\gamma}^{\mu}$. Without loss of generality we may therefore confine attention to the particular solution $\tilde{\gamma}^{\mu}$. The invariance of the usual formulation of Dirac's equation in general relativity under similarity transformations is here replaced by invariance under rotations of the Cartan frames.[7]

3. Spinor Connection and Covariant Differentiation of Spinors

In addition to the matrices γ^{μ}, a spinor ψ (vector in spin space) appears in the Dirac equation. Just as in special relativity, we choose the transformation properties of the spinor ψ so that the Dirac equation (and the $\tilde{\gamma}^{\mu}$ appearing in it) is invariant under change of reference frames. Therefore, the covariant derivative of ψ must contain a term due to differential rotation of the spinor frames[8]:

$$\nabla_{\mu}\psi = \psi_{;\mu} - \Gamma_{\mu}\psi. \quad (9)$$

As in the previous section this reduces to

$$\nabla_{\mu}\psi = \omega_{\mu}(\psi) - \Gamma_{\mu}\psi \quad (10)$$

for orthonormal frames.

[5] The contravariant component with respect to the basis ω_i of the vectoral 0-form $a = \omega_i a^i$ is the projection of this vector along a given direction in exterior differential form space, i.e., $a^k = \omega^k\cdot\omega_i a^i$. ($\omega^k$ is a real-valued linear functional which maps a vector a into its kth component with respect to a basis ω_i.) Similarly, the contravariant component of the vectoral 1-form da is $\omega^k\cdot da$. This component of the change in a, denoted by Da^k, is known as the covariant differential of a^k. (Da^k is a real-valued linear functional which maps a basis vector ω_l into the components $a^k{}_{;l}$ of Da^k with respect to ω^l.) The change in a is $da = d(\omega_i a^i) = \omega_i da^i + d\omega_i a^i$; $d\omega_i$ describes the motion of the basis vectors and is not an exact differential unless space is flat. It is determined by the antisymmetric tensoral 1-form ω_{ij} through the relation $d\omega_i = \omega_j\omega^j{}_i$ which leads to the corresponding equation (3) for the dual space by the demand that the torsion vanish. Hence, one obtains $da = \omega_i(da^i + \omega^i{}_j a^j)$, and thus $Da^k = \omega^k\cdot da = da^k + \omega^k{}_j a^j$. Similarly, the covariant derivative $a^k{}_{;l}$ of a vector with components a^k is the covariant component (with respect to the basis ω^i) of the covariant differential in a given direction in vector space,

$$\begin{aligned} a^k{}_{;l} &= Da^k\cdot\omega_l = \omega^k\cdot da\cdot\omega_l \\ &= da^k\cdot\omega_l + \gamma^k{}_{ji}\omega^i a^j\cdot\omega_l \\ &= \omega_l(a^k) + \gamma^k{}_{jl}a^j. \end{aligned}$$

[6] C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), p. 81.

[7] This idea may also be expressed in a different form: In the usual formulation of Dirac's equation, it is convenient to introduce "Vierbeine" in order to solve Eq. (7). Our choice of solution above implies that we have identified the Cartan frames which we use to express components of a tensor, and the Vierbeine which are used to express the γ matrices.

[8] See, for example, Table II in Ref. 11.

To determine the spinor connection Γ_μ we observe that, by the definition in the previous section, the matrices $\tilde{\gamma}^\mu$ are the same in every reference system (with orthonormal frames). Therefore, their covariant derivative must vanish;

$$\nabla_\mu\tilde{\gamma}^\nu = \tilde{\gamma}^\nu{}_{;\mu} - \Gamma_\mu\tilde{\gamma}^\nu + \tilde{\gamma}^\nu\Gamma_\mu = 0. \quad (11)$$

The expression for this covariant derivative follows from the vector nature of γ^μ in coordinate space and its matrix nature in spin space. In Cartan frames this equation becomes, using (5),

$$\omega_\mu(\tilde{\gamma}^\nu) + \gamma^\nu{}_{\alpha\mu}\tilde{\gamma}^\alpha + \tilde{\gamma}^\nu\Gamma_\mu - \Gamma_\mu\tilde{\gamma}^\nu = 0. \quad (12)$$

Since the $\tilde{\gamma}^\nu$ are the (constant) special relativity Dirac matrices, we have the relation $\omega_\mu(\tilde{\gamma}^\nu) = 0$. Consequently, the condition determining Γ_μ becomes

$$\gamma^\nu{}_{\alpha\mu}\tilde{\gamma}^\alpha + [\tilde{\gamma}^\nu, \Gamma_\mu] = 0. \quad (13)$$

One solution of this condition is

$$\Gamma'_\mu = -(\tfrac{1}{4})\gamma^\alpha{}_{\nu\mu}\tilde{\gamma}_\alpha\tilde{\gamma}^\nu. \quad (14)$$

Let Γ_μ denote the general solution of (13). From (13) it then follows that

$$[\tilde{\gamma}^\nu, \Gamma'_\mu - \Gamma_\mu] = 0. \quad (15)$$

Since $\Gamma'_\mu - \Gamma_\mu$ commutes with the four $\tilde{\gamma}^\nu$, it also commutes with all the matrices in the group of Dirac matrices. Since these matrices form an irreducible representation of the group, we can apply Schur's Lemma: $\Gamma'_\mu - \Gamma_\mu$ must be a multiple of the unit matrix **1**, and the general Γ_μ is given by

$$\Gamma_{\mu\,\text{general}} = -\tfrac{1}{4}\gamma^\alpha{}_{\nu\mu}\tilde{\gamma}_\alpha\tilde{\gamma}^\nu + a_\mu 1, \quad (16)$$

where a_μ is an arbitrary vector. For charged particles, a_μ is identified with the vector potential. For neutrinos we annul a_μ. The explicit form (16) for the Γ_μ holds only in orthonormal Cartan frames[1]; however, (16) is much simpler and more readily applied than the corresponding expression in either the Vierbein formalism[2] or the more general representation in holonomic frames.[9]

4. Dirac Equation in Cartan Frames

In moving orthonormal frames, the curved-space Dirac equation

$$(\gamma^\mu\nabla_\mu + m)\psi = 0 \quad (17)$$

becomes (as also found by Lichnerowicz[3] in somewhat different notation and by a somewhat different route)

$$\tilde{\gamma}^\mu\omega_\mu(\psi) - \tilde{\gamma}^\mu\Gamma_\mu\psi + m\psi = 0. \quad (18)$$

This formulation quickly and easily yields the Dirac equation in terms of well known objects like partial derivatives and special relativity Dirac matrices. This simplicity is an important reason for using the Cartan method, rather than holonomic frames, for the formulation of the general relativistic Dirac equation.

III. DIRAC EQUATION IN A HOMOGENEOUS NONISOTROPIC UNIVERSE

It was recently shown[10] that the homogeneous, nonisotropic purely gravitational universe of Taub[4] can be extended to include a homogeneous electromagnetic field ("standing electromagnetic wave of maximal wavelength") which is strong in the sense that its contribution to the curvature of space is nonnegligible; in the limiting case the curvature of space into closure is caused entirely by the electromagnetic field, in contrast to the Taub universe where the required content of effective energy arises entirely from a standing gravitational wave. It is therefore natural to ask, can one similarly incorporate the other zero-rest-mass field known in nature, the neutrino field, into a homogeneous universe.

The metric of a homogeneous universe of the type investigated by Taub has the form (here and in the following, i, j, k, will denote a cyclic permutation of 1, 2, 3):

$$ds^2 = -dt^2 + (A_1\sigma_1)^2 + (A_2\sigma_2)^2 + (A_3\sigma_3)^2 \quad (19a)$$

with

$$d\sigma_i = \sigma_j \wedge \sigma_k, \quad (19b)$$

$$A_i = A_i(t). \quad (19c)$$

Such a metric has been characterized[10] as describing a closed universe which (for a finite time) has spacelike surfaces t = const of three-sphere topology, and contains gravitational waves in the lowest possible mode (maximum symmetry).

In order to discuss the Dirac equation in this metric, it is convenient to choose the orthonormal frames

$$\omega^0 = dt, \qquad \omega^i = A_i\sigma_i \qquad \text{(no summation)}.$$

From Eqs. (3) and (19b) the nonzero connection forms follow immediately;

$$\omega^i{}_j = -\omega^j{}_i = \omega^k(A_i{}^2 + A_j{}^2 - A_k{}^2)/2A_iA_jA_k,$$
$$\omega^i{}_0 = \omega^0{}_i = \omega^i\dot{A}_i/A_i \qquad \text{(no summation)}. \quad (20)$$

The connection coefficients are identified in Eq. (3a),

[9] J. G. Fletcher, Nuovo Cimento **8**, 451 (1958).

[10] D. R. Brill, Phys. Rev. **133**, B845 (1964).

and the spinor connection is found by simple substitution into Eq. (14),

$$\Gamma_0 = a_0 1, \tag{21}$$

$$\Gamma_i = a_i 1 + \tilde{\gamma}_i\tilde{\gamma}_0(\dot{A}_i/2A_i) - \tilde{\gamma}_i\tilde{\gamma}_k(A_i^2 + A_k^2 - A_i^2)/4A_1A_2A_3 \quad \text{(no summation)}.$$

The Dirac equation (18) can now be written down. We give the form appropriate for a homogeneous neutrino field ($m = 0$, $a_\mu = 0$, $\psi = \psi(t)$):

$$\partial_t\psi + \sum_i [(\dot{A}_i/2A_i) + \gamma_5(A_i^2/4A_1A_2A_3)]\psi = 0. \tag{22}$$

Use new measures of the wavefunction and of time,

$$\psi' = (A_1A_2A_3)^{\frac{1}{2}}\psi \quad \text{and}$$
$$dt' = \sum (A_i^2/2A_1A_2A_3)\, dt. \tag{23}$$

Then the equation takes the simpler form

$$\partial\psi'/\partial t' = -\tfrac{1}{2}\tilde{\gamma}_5\psi'. \tag{24}$$

It can be explicitly integrated,

$$\psi = (A_1A_2A_3)^{-\frac{1}{2}}\psi' = (A_1A_2A_3)^{-\frac{1}{2}}e^{-\tilde{\gamma}_5 t/2}\psi_0. \tag{25}$$

Here ψ_0 is a constant spinor of integration, which is determined (except for an overall, nonmeasurable phase) by the normalization condition

$$\int_{t=\text{const}} \psi^*\psi\, d^3\sigma = 1. \tag{26}$$

Except for the change in amplitude necessary to maintain the normalization as the volume of the universe varies, the wavefunction therefore only undergoes a "duality rotation,"[11] as shown by Eq. (25). Thus, as in the electromagnetic case[10] the features of the time dependence affecting the stress-energy tensor can be determined by a very simple physical argument.

Equation (25) solves the problem of the response of a neutrino field *to* the metric and is a complete solution for a weak (test) field of negligible stress-energy. However, to discuss the question whether the curvature generated *by* the neutrino field is consistent with the high degree of symmetry of the universe, we compute the components of the stress-energy tensor from the expression (92) given in Ref. 11:

$$\begin{aligned} T_{00} &= i\psi_0^*\tilde{\gamma}_5\psi_0 \sum_i A_i^2/4A_1^2A_2^2A_3^2, \\ T_{ii} &= i\psi_0^*\tilde{\gamma}_5\psi_0(A_j^2 + A_k^2 - A_i^2)/4A_1^2A_2^2A_3^2, \\ T_{0i} &= \psi_0^*\sigma_i\psi_0(A_j^2 + A_k^2)/4A_1^2A_2^2A_3^2, \\ T_{ij} &= (\psi_0^*\sigma_k\psi_0/4A_1A_2A_3)\partial_t \ln (A_i/A_j). \end{aligned} \tag{27}$$

Thus we see that the stress-energy tensor is not diagonal. In particular, one of the components of the energy flux vector must differ from zero if $\psi_0 \neq 0$. However, the Ricci tensor of the metric (19a) is diagonal.[12] Therefore no choice of the A_i (so far arbitrary) will permit us to satisfy Einstein's equations with the source tensor (27). We conclude that, unlike the electromagnetic field, the neutrino field generates a curvature which in *not* consistent with the high degree of symmetry of a universe of the Taub type.

IV. PLANE-WAVE SOLUTIONS

1. Plane-Wave Metric

In this section further examples are given of the use of Cartan frames for the case of a nondiagonal metric. In particular, we obtain a complete solution in which the nongravitational fields are not treated as weak.

We consider metrics of the type discussed by Takeno,[13]

$$ds^2 = A\, dx^2 + B\, dy^2 + C\, dz^2 - D\, dt^2 + E\, dx\, dy,$$

which admit a five-parameter group of motions, three translations and two rotations in the hypersurface orthogonal to the propagation direction. A, B, C, D, and E are functions of $z + t$. In a suitable coordinate system the metric takes the form

$$ds^2 = \varphi^2\, dx^2 + 2\alpha\varphi\, dx\, dy + (\alpha^2 + \beta^2)\, dy^2 + dz^2 - dt^2. \tag{28}$$

Here $(-z)$ is the propagation direction, and φ, α and β are functions of $z + t$ only.

A convenient set of orthonormal Cartan frames is

$$\omega^0 = dt; \quad \omega^1 = \varphi dx + \alpha\, dy; \quad \omega^2 = \beta\, dy; \quad \omega^3 = dz. \tag{29}$$

From Eq. (3) one finds the nonvanishing connection forms (here $'$ denotes differentiation with respect to $z + t$):

[11] D. R. Brill and J. A. Wheeler, Rev. Mod. Phys. 29, 465 (1957).

[12] C. W. Misner, J. Math. Phys. 4, 924 (1963).

[13] H. Takeno, "The Mathematical Theory of Plane Gravitational Waves in General Relativity", Scientific Reports of the Research Institute for Theoretical Physics, Hiroshima University, No. 1 (1961). Also see H. Takeno, Tensor 10, 34 (1960) and the reference cited there.

$$\omega_3^1 = \omega_0^1 = (\varphi'/\varphi)\omega^1 + (\varphi/2\beta)(\alpha/\varphi)'\omega^2,$$
$$\omega_3^2 = \omega_0^2 = (\beta'/\beta)\omega^2 + (\varphi/2\beta)(\alpha/\varphi)'\omega^1, \quad (30)$$
$$\omega_2^1 = -(\varphi/2\beta)(\alpha/\varphi)\,(\omega^0 + \omega^3).$$

The nonvanishing components of the Ricci tensor are[14]

$$R_0^0 = -R_3^3 = -R_0^3 = (\varphi''/\varphi) + (\beta''/\beta) + [(\alpha/\varphi)'\alpha\varphi/\beta]'/2\beta. \quad (31)$$

2. Neutrinos in Plane-Wave Metric

The spinor connection follows from Eq. (14);

$$\Gamma_0 = (\varphi/4\beta)(\alpha/\varphi)'\tilde{\gamma}^1\tilde{\gamma}^2 + b_0 1,$$
$$\Gamma_1 = -\tfrac{1}{2}[(\varphi'/\varphi)\tilde{\gamma}_1 + (\varphi/2\beta)(\alpha/\varphi)'\tilde{\gamma}_2](\tilde{\gamma}^0 + \tilde{\gamma}^3) + b_1 1,$$
$$\Gamma_2 = -\tfrac{1}{2}[(\varphi/2\beta)(\alpha/\varphi)'\tilde{\gamma}_1 + (\beta'/\beta)\tilde{\gamma}_2](\tilde{\gamma}^0 + \tilde{\gamma}^3) + b_2 1,$$
$$\Gamma_3 = (\varphi/4\beta)(\alpha/\varphi)'\tilde{\gamma}^1\tilde{\gamma}^2 + b_3 1. \quad (32)$$

The Dirac equation for zero rest mass therefore takes the form ($b_\mu = 0$)

$$\{\gamma^0\partial_0 + \gamma^1(1/\varphi)\partial_1 + \gamma^2(\beta^{-1}\partial_2 - (\alpha/\varphi\beta)\partial_1) + \gamma^3\partial_3 + (\gamma^0 + \gamma^3)[\tfrac{1}{2}(\varphi'\varphi^{-1} + \beta'\beta^{-1}) - (\varphi/4\beta)(\alpha/\varphi)'\gamma_5]\}\psi = 0. \quad (33)$$

A solution of this equation can easily be written in the representation for the $\tilde{\gamma}^\mu$ matrices used by Jauch and Rohrlich,[15]

$$\psi = \begin{pmatrix} a(z+t) \\ i\sigma_3 a(z+t) \end{pmatrix}, \quad (34)$$

where $a(z + t)$ is an arbitrary two-component spinor field. The nonzero components of the stress–energy tensor of the neutrino field (34) are

$$T_0^0 = -T_3^3 = -T_0^3 = i[a^*a' - a^{*\prime}a] - (\varphi/2\beta)(\alpha/\varphi)'a^*\sigma_3 a. \quad (35)$$

We see that these components of the stress-energy tensor fulfill the same algebraic relationships as the Ricci tensor in Eq. (31). To satisfy the combined Einstein–Dirac equations, it is only necessary to fulfill one additional equation, e.g., $R_{00} = T_{00}$ (since $R = 0$ for these metrics). The solution of this one equation is obtained in Sec. 4.

To obtain a solution for a Lee–Yang-type neutrino with right-handed helicity, we demand that the projection of ψ onto left-handed states vanish,

$$(1 - i\tilde{\gamma}_5)\psi = 0,$$

which implies, in the representation of the γ matrices used by Jauch and Rohrlich,[15] that the second component of $a(z + t)$ vanishes.

3. Integral Spin Fields in Plane-Wave Metric

For completeness we mention how integral spin fields can be incorporated into the plane-wave geometry. For a scalar field Φ in the metric (28), the conditions

$$T_{11} = T_{22} = 0 \quad \text{and} \quad -R = T = -[m\Phi^2 + (\omega_1\Phi)^2 + (\omega_2\Phi)^2] \quad (36)$$

imply that the field must be massless and constant in the 1 and 2 directions. Again we require an algebraic structure as in Eq. (35) for the stress-energy tensor, and find that only

$$\Phi = \Phi(z + t) \quad (37)$$

satisfies these conditions. The massless Klein–Gordon equation is also satisfied by (37). By proper choice of the amplitude of Φ, the remaining relationship

$$R_{00} = T_{00} = (\partial_0\Phi)^2 \quad (38)$$

can be satisfied.

Electromagnetic fields in a plane-wave metric were investigated by Takeno.[13] His solution, rewritten in the language of Cartan frames,

$$f_{01} = f_{31} = e_1(z + t), \quad f_{02} = f_{32} = e_2(z + t), \quad (39)$$

satisfies the free-space Maxwell equations,

$$df = 0, \qquad d^*f = 0, \quad (40)$$

and the algebraic relationships (35) of the stress-energy tensor. Again Einstein's equations can be satisfied by the proper choice of the amplitude of the electromagnetic fields.

4. Example of a Solution of Einstein–Klein Gordon–Dirac–Maxwell Equation

In order to obtain a simple solution, the fields are chosen such that the contribution to the stress-energy tensor from each field is constant.[16] This is so if we choose

$$\Phi = K_s(z + t), \qquad e_1 = K_e \sin(z + t), \qquad e_2 = K_e \cos(z + t), \quad (41)$$

[14] For discussion of the computation of the Ricci tensor in the Cartan formulation see, e.g., Ref. 12.

[15] J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons*, (Addison–Wesley Publishing Company, Inc., Reading, Massachusetts, 1959).

[16] The existence of such fields in flat space—e.g., circularly polarized electromagnetic waves—is well known. The problem of the infinite *total* energy represented by such waves is common to all types of plane waves, and can be resolved by considering them as a limit of spherical waves very far from the source.

and

$$\psi_{,} = \begin{pmatrix} 1 \\ i\sigma_3 \end{pmatrix} f e^{-i(z+t)}, \quad \text{where} \quad \tilde{f}f = \tfrac{1}{2}K_{,}^{2}.$$

Here $K_{,}$, $K_{,}$, and $K_{,}$ are constants. For simplicity, choose

$$\alpha = 0 \quad \text{and} \quad \beta = B \cosh n(z + t). \tag{42}$$

Then Einstein's equations reduce to

$$(\varphi''/\varphi) + n^2 = -K_{,}^{2} - K_{,}^{2} - K_{,}^{2}, \tag{43}$$

with the solution

$$\varphi = K \cos (n^2 + K_{,}^{2} + K_{,}^{2} + K_{,}^{2})^{\frac{1}{2}}(z + t + \eta), \tag{44}$$

where η and K are integration constants. This solution appears singular for

$$z + t = \pm\infty \quad \text{and} \quad z + t + \eta = \pi(m + \tfrac{1}{2}) \times (n^2 + K_{,}^{2} + K_{,}^{2} + K_{,}^{2})^{-\frac{1}{2}},$$

where m is an integer, but these singularities can be removed by coordinate transformation.[17]

V. CONCLUSION

The Cartan method has long been recognized as an appropriate way of treating electromagnetism. In the present paper we have seen that it is equally successful in describing spin-$\frac{1}{2}$ fields in curved space. By using the *physical* components[18] (components in orthonormal frames) of all quantities one retains many of the simplicities of the flat-space Dirac equation.

The behavior of neutrinos in homogeneous, nonisotropic universes showed similarities to, as well as important differences from, that of electromagnetic radiation. Since the energy flux vector (T_{0i}) of neutrinos never vanishes, the curvature *generated* by them is *not* consistent with the high degree of symmetry in the homogeneous Taub-type universes. This fact is closely related to the absence of "neutrino charge."[19]

The plane-wave solution of the Einstein–Klein Gordon–Dirac–Maxwell equations allows an interesting mathematical conclusion. If only an electromagnetic field is present ($K_{,} = K_{,} = 0$), a knowledge of the geometry alone—i.e., Eqs. (42) and (44)—allows one to reconstruct the field amplitudes. Rainich, Misner, and Wheeler[1] have shown that such a reconstruction is possible for general electromagnetic field distributions and have formulated an "already unified field theory" of electromagnetism and gravitation in purely geometrical terms. A similar geometrical formulation is possible for the case of a scalar field and gravitation.[20] It is natural to ask, is it possible to write an already unified field theory which includes *all three* zero-rest-mass fields considered here, scalar, spinor, and vector. Inspection of Eq. (44) shows that the amplitudes of all the nongravitational fields enter in the same way into the expressions (42), (44) for the metric coefficients. It is therefore not possible to conclude what the separate amplitudes are from a knowledge of geometry alone. Thus the plane-wave solution of Sec. IV.4 is a counterexample to an already unified theory of gravitation, electromagnetism, neutrinos, and scalar fields.

ACKNOWLEDGMENTS

We are grateful to Professor J. A. Wheeler for reading the manuscript and making several helpful suggestions. We are also indebted to Marion D. Cohen for checking some of the calculations.

[17] W. B. Bonner, Ann. Inst. Henri Poincaré 15, 146 (1957).
[18] J. Weber, *General Relativity and Gravitational Waves* (Interscience Publishers, Inc., New York, 1961).

[19] For electromagnetic fields in the Taub-type universe,[10] the outer (NUT) space is charged in the sense that there exists a nonzero electric field flux through a large sphere in the asymptotically flat region. If neutrinos in a Taub-type universe had an identical behavior, then the outer space would exhibit neutrino charge. However, the absence of neutrino charge has been shown by Klauder and Wheeler, Rev. Mod. Phys. 29, 516 (1957).
[20] See, for example, D. R. Brill, Nuovo Cimento Suppl. 2, 1 (1964).

Expansion Theorem for Functions of Operators*

LEON COHEN
Sloane Physics Laboratory, Yale University, New Haven, Connecticut
(Received 13 July 1965)

A method is given for expanding operator functions of q and p, where $\mathbf{p} = \hbar/i(\partial/\partial\mathbf{q})$, such that all q factors are to the left of the p factors. The method is applicable to the rearrangement of creation and annihilation operators.

1. INTRODUCTION

It is sometimes necessary to raise a function of the operators $\mathbf{q}$ and $\mathbf{p}$, where $\mathbf{p} = \hbar/i(\partial/\partial\mathbf{q})$, to the nth power. One may then also want to rearrange the resulting expression in such a way that all the $\mathbf{q}$'s precede the $\mathbf{p}$'s. This can be accomplished by the use of brute force in expanding the function and using the commutation relation $\mathbf{pq} = \mathbf{qp} + (\hbar/i)$ to put the $\mathbf{q}$ factors on the left of the $\mathbf{p}$ factors. But the labor involved prohibits this method in all but the most simple of cases. Thus, for example, in attempting to expand $(\mathbf{q} + \mathbf{p})^n$ in powers of $\mathbf{q}^k\mathbf{p}^l$ by calculation of $(\mathbf{q} + \mathbf{p})(\mathbf{q} + \mathbf{p}) \cdots (\mathbf{q} + \mathbf{p})$ one would very quickly be entangled in long and unwieldy manipulations. We describe here a rather simple method for doing this. Our main result is this: Given a function $F(\mathbf{q}, \mathbf{p})$ of the operators $\mathbf{q}$ and $\mathbf{p}$ then

$$F^n(\mathbf{q}, \mathbf{p}) = \sum_{k=0}^{\infty} \alpha_k^n u_k(\mathbf{q}) \int_{-\infty}^{\infty} u_k^*(\mathbf{q} + \theta) e^{i/\hbar\theta\mathbf{p}}\, d\theta, \qquad (1)$$

where α_k and $u_k(\mathbf{q})$ are the eigenvalues and eigenfunctions of the eigenvalue problem

$$F[q, (\hbar/i)(\partial/\partial q)]u_k(q) = \alpha_k u_k(q).$$

The $u_k(q)$ are assumed to be normalized to one if the spectrum is discrete and to a delta function if the spectrum is continuous. In the continuous case the summation is replaced by an integration.

2. PROOF OF THE GENERAL THEOREM

We prove (1) by showing it to be true for $n = 1$ and then using induction to prove it in general. Let $F(\mathbf{q}, \mathbf{p})$ operate on an arbitrary function $f(q)$ whose expansion in terms of the $u_k(q)$ is

$$f(q) = \sum_{k=0}^{\infty} a_k u_k(q).$$

Then

$$F\left(q, \frac{\hbar}{i}\frac{\partial}{\partial q}\right) f(q) = \sum_{k=0}^{\infty} a_k \alpha_k u_k(q). \qquad (2)$$

Now consider

$$\sum_{k=0}^{\infty} \alpha_k u_k(\mathbf{q}) \int_{-\infty}^{\infty} u_k^*(\mathbf{q} + \theta) e^{i/\hbar\theta\mathbf{p}} f(q)\, d\theta$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} a_l \alpha_k u_k(q) \int_{-\infty}^{\infty} u_k^*(q + \theta) e^{\theta\partial/\partial q} u_l(q)\, d\theta$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} a_l \alpha_k u_k(q) \int_{-\infty}^{\infty} u_k^*(q + \theta) u_l(q + \theta)\, d\theta$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \alpha_k a_l \delta_{kl} u_k(q) = \sum_{k=0}^{\infty} a_k \alpha_k u_k(q),$$

since

$$\int_{-\infty}^{\infty} u_k^*(q + \theta) u_l(q + \theta)\, d\theta = \delta_{kl}.$$

Thus (1) is proved for $n = 1$. To prove it for any n we assume it to be true for n and show that it holds for $n + 1$.

$$F^{n+1}f = F\left(q, \frac{\hbar}{i}\frac{\partial}{\partial q}\right) F^n\left(q, \frac{\hbar}{i}\frac{\partial}{\partial q}\right) f(q)$$

$$= \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \alpha_l u_l(q) \int_{-\infty}^{\infty} u_l^*(q + \theta) e^{i/\hbar\theta\mathbf{p}} \alpha_k^n u_k(q)$$

$$\times \int_{-\infty}^{\infty} u_k^*(q + \theta') e^{i/\hbar\theta'\mathbf{p}} f(q)\, d\theta\, d\theta'$$

$$= \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \alpha_l \alpha_k^n u_l(q)$$

$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_l^*(q + \theta) u_k(q + \theta)$$

$$\times u_k^*(q + \theta + \theta') f(q + \theta + \theta')\, d\theta\, d\theta'$$

$$= \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \alpha_l \alpha_k^n u_l(q) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_l^*(q + \theta)$$

$$\times u_k(q + \theta) u_k^*(q + \theta') f(q + \theta')\, d\theta\, d\theta'$$

$$= \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \alpha_l \alpha_k^n u_l(q) \int_{-\infty}^{\infty} u_k^*(q + \theta') f(q + \theta') \delta_{kl}\, d\theta'$$

$$= \sum_{k=0}^{\infty} \alpha_k^{n+1} u_k(q) \int_{-\infty}^{\infty} u_k^*(q + \theta) e^{i/\hbar\theta\mathbf{p}} f(q)\, d\theta,$$

and therefore, (1) is proved in general.

* Research sponsored in part by the Air Force Office of Scientific Research.

If one wants to carry out the integration in (1) before F^n operates, the following procedure can be used.[1] Replace $\mathbf{q}$ and $\mathbf{p}$ by ordinary variables, q and p, and carry through the integration. After the integration is performed, arrange the expression so that all factors of q stand to the left of p and then replace q and p by the operators $\mathbf{q}$ and $\mathbf{p}$. This works since the $\mathbf{q}$ factors preceded the $\mathbf{p}$ factors in the original expression.

3. AN EXAMPLE

As an example we consider the expansion of $(\lambda\mathbf{q} + \mathbf{p})^n$ where λ is a real parameter. The eigenvalue problem

$$\left(\lambda q + \frac{\hbar}{i}\frac{\partial}{\partial q}\right)u_\alpha(q) = \alpha u_\alpha(q)$$

can readily be solved. The eigenfunctions normalized to a delta function are

$$u_\alpha(q) = (2\pi\hbar)^{-\frac{1}{2}}e^{i/\hbar(\alpha q - \frac{1}{2}\lambda q^2)}.$$

In this case the eigenvalues α are continuous. We now replace $\mathbf{q}$ and $\mathbf{p}$ by q and p as discussed at the end of Sec. 2. Thus (1) becomes

$$\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\alpha^n$$

$$\times \exp\left[-i/\hbar(\alpha\theta - \lambda q\theta - \tfrac{1}{2}\lambda\theta^2) + i/\hbar\theta p\right]d\theta\, d\alpha.$$

Considering $\int_{-\infty}^{\infty}\alpha^n e^{-i/\hbar\alpha\theta}d\alpha$ as a distribution, this yields

$$(\tfrac{1}{2}i\lambda\hbar)^{n/2}\mathrm{H}_n[(1/2i\lambda\hbar)^{\frac{1}{2}}(\lambda q + p)]$$

where H_n is the Hermite polynominal of order n. Using the definition

$$\mathrm{H}_n(x) = \sum_{k=0}^{[\frac{1}{2}n]}\frac{(-1)^k n!(2x)^{n-2k}}{k!(n-2k)!},$$

the above expression can be written as

$$\sum_{k=0}^{[\frac{1}{2}n]}\sum_{l=0}^{n-2k}\frac{(-1)^k n!}{k!(n-2k)!}\binom{n-2k}{l}\left(\frac{i\hbar}{2}\right)^k\lambda^{n-k-l}q^{n-2k-l}p^l.$$

We now substitute $\mathbf{q}$ for q and $\mathbf{p}$ for p to obtain

$$(\lambda\mathbf{q} + \mathbf{p})^n = \sum_{k=0}^{[\frac{1}{2}n]}\sum_{l=0}^{n-2k}\frac{(-1)^k n!}{k!(n-2k)!}\binom{n-2k}{l}$$

$$\times (i\hbar/2)^k\lambda^{n-k-l}\mathbf{q}^{n-2k-l}\mathbf{p}^l.$$

4. CONCLUSION

In conclusion some extensions of the above method may be mentioned. Equation (1), for $n = 1$, which expresses an operator in terms of its eigenfunction's and eigenvalues can sometimes be profitably used to bring all $\mathbf{q}$ factors to the left of the $\mathbf{p}$ factors of product of functions. Thus if $F_k(\mathbf{q}, \mathbf{p})$ denotes a set of arbitrary functions,

$$\prod_{k=1}^{l}F_k(\mathbf{q}, \mathbf{p}) = \sum_{n_1 n_2\cdots n_l}\alpha_{n_1}^{(1)}\alpha_{n_2}^{(2)}\cdots\alpha_{n_l}^{(l)}g_{n_1 n_2\cdots n_l}^{(\theta_1)}$$

$$\times u_{n_1}^{(1)}(\mathbf{q})u_{n_1}^{*(1)}(\mathbf{q} + \theta_1)e^{i/\hbar\theta_1 p}\,d\theta_1$$

where

$$g_{n_1 n_2\cdots n_l}^{(\theta_1)} = \int\cdots\int\prod_{k=1}^{l-1}u_{n_k}^{(k)}(\theta_{k+1})u_{n_{k+1}}(\theta_{k+1})\,d\theta_{k+1}$$

and

$$F_k\left(q, \hbar/i\frac{\partial}{\partial q}\right)u_n^{(k)}(q) = \alpha_n^{(k)}u_n^{(k)}.$$

These results can be used in manipulating functions of the creation and annihilation operators, $\mathbf{a}^\dagger$ and $\mathbf{a}$. All that is required is to substitute $\mathbf{a}$ for $\mathbf{q}$ and $ih\mathbf{a}^\dagger$ for $\mathbf{p}$. This follows from the relation $[\mathbf{q}, \mathbf{p}] = ih[\mathbf{a}, \mathbf{a}^\dagger]$. A rigorous derivation of this can be based on the use of Entire functions as developed by Bargmann.[2] In that representation the annihilation operator is "represented" by differentiation.

ACKNOWLEDGMENT

The author wishes to thank Professor Henry Margenau for his help and guidance.

[1] N. H. McCoy, Proc. Natl. Acad. Sci. U.S. 18, 674 (1932).

[2] V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961).

Nonlinear Theory of Elastic Surfaces*

H. COHEN AND C. N. DESILVA
University of Minnesota, Minneapolis, Minnesota
(Received 25 January 1965)

The present paper develops a nonlinear theory for the deformation of an elastic surface by assuming the existence of a strain energy function and postulating a principle of virtual work which governs its mechanical behavior. By considering the strain energy function to depend on the first- and second-order deformation gradients, the field equations and the general constitutive relations are obtained. In addition to the conventional couple stresses, there are shown to exist energetically undetermined double stresses without moment.

1. INTRODUCTION

EARLY investigations in the theory of elasticity were mainly concerned with establishing special theories associated with thin bodies. Following the formulation of the general three-dimensional equations of elasticity, however, theories of thin bodies were derived as limiting cases of this general theory.[1] From that time on, few attempts have been made to develop special theories of thin bodies independently of the general equations of elasticity. One such attempt was carried out by the Cosserats[2] who, following an idea of Duhem,[3] introduced the concept of the directed line and the directed surface. Ericksen and Truesdell[4] have elaborated the ideas of the Cosserats to formulate a nonlinear theory of strain for rods and shells.

Recently there has been considerable interest in developing a consistent nonlinear theory of shells.[5-9] These references all have, as their starting point, the equations of classical, three-dimensional elasticity. There are, however, certain difficulties inherent in this traditional approach which, we believe, may be overcome by treating the deformation of an elastic surface. Here, using the latter procedure, we obtain special theories of elasticity which govern the nonlinear behavior of a surface.

In Sec. 2, the geometry of surfaces is reviewed and the concept of a surface deformation is defined. Deformation tensors **L** and **K** are defined which are associated with changes in the intrinsic and spatial aspects of surface geometry, respectively. In Sec. 3, assuming the existence of a strain energy function, a principle of virtual work is postulated as governing the mechanical behavior of a surface. This principle is applied in Sec. 4 to a Noll-type simple material[10] and a theory of membrane action is derived in which the stress tensor is a symmetric tangential surface tensor. Nonlinear constitutive equations relating this tensor to the deformation tensor **L** are obtained. The special form of these equations when specialized to an isotropic material is a simple two-dimensional analog of the equations of Finger.[11]

Finally in Sec. 5, the principle of virtual work is applied to a restricted material of class two[12] under the action of arbitrary virtual displacements to derive a complete theory of bending. In this theory, both the stress tensor and the couple stress tensor are symmetric surface tensors which depend on the deformation tensors **L** and **K**. The nonlinear constitutive equations are obtained and their simple form in the case of "surface isotropy" is given. It is found, in applying the principle, that it is necessary to define a set of quantities $\bar{\lambda}$ which may be interpreted as double stresses and which do no work during the deformation. This tensor gives rise, in the conventional equilibrium equations, to a term which can be considered as a transverse shear. Moreover, there occur three additional equations of equilibrium involving $\bar{\lambda}$.

In the Appendix, we show the relationship between the equations of equilibrium derived here and those given previously by Ericksen and Truesdell.[4]

* This work was supported by the National Science Foundation under Research Grant NSF-GK99 and by the National Aeronautics and Space Administration under Grant NGR-24-005-059.

[1] A. E. H. Love, *The Mathematical Theory of Elasticity* (Cambridge University Press, New York, 1927), pp. 1–31.

[2] E. and F. Cosserat, *Théorie des Corps Déformables* (Hermann & Cie., Paris, 1909).

[3] P. Duhem, Ann. École Norm. (3) **10**, 187 (1893).

[4] J. L. Ericksen and C. Truesdell, Arch. Ratl. Mech. Anal. **1**, 295 (1958).

[5] R. W. Leonard, "Non-Linear First Approximation Thin Shell and Membrane Theory," Ph.D. thesis, Virginia Polytechnic Institute (1961).

[6] J. L. Sanders, Quart. Appl. Math. **21**, 21 (1963).

[7] P. M. Naghdi and R. P. Nordgren, Quart. Appl. Math. **21**, 49 (1963).

[8] W. L. Wainwright, Intl. J. Eng. Sci. **1**, 339–358 (1963).

[9] C. B. Sensenig, IMM-NYU 313, New York University (1963).

[10] W. Noll, Arch. Ratl. Mech. Anal. **2**, 197 (1958).

[11] J. Finger, Sitzber. Akad. Wiss. Wien (IIa) **103**, 1073 (1894).

[12] R. A. Toupin, Arch. Ratl. Mech. Anal. **17**, 85 (1964).

2. KINEMATICS OF DEFORMABLE SURFACES

Let X^K be rectangular Cartesian coordinates defined in an E_3. We define a surface S or an R_2 imbedded in this E_3 by

$$X^K = X^K(U^\Delta), \tag{2.1}$$

where U^Δ are general curvilinear coordinates in S. Here and in all cases, Latin and Greek indices take on the values 1, 2, 3 and 1, 2, respectively. Equation (2.1) is assumed to be single-valued and continuously differentiable as many times as desired except possibly at certain singular curves or points. This remark applies as well to all transformations with which we shall be concerned.

The base vectors of S are given by $X^K{}_{,\Delta}$ where

$$X^K{}_{,\Delta} = \partial X^K/\partial U^\Delta \tag{2.2}$$

and the surface metric $\mathbf{A}$ is given by

$$A_{\Delta\Gamma} = \delta_{KM} X^K{}_{,\Delta} X^M{}_{,\Gamma}. \tag{2.3}$$

In Eq. (2.2), $(\ )_{,\Delta}$ denotes the total covariant derivative.[13] The unit normal N_K to S is defined by

$$e_{\Gamma\Delta} N_K = \epsilon_{KMP} X^M{}_{,\Gamma} X^P{}_{,\Delta} \tag{2.4}$$

where

$$e_{\Gamma\Delta} = \epsilon_{\Gamma\Delta} A^{\frac{1}{2}}, \qquad A = \det A_{\Sigma\Delta} \tag{2.5}$$

and $\epsilon_{\Gamma\Delta}$, ϵ_{KMP} are the conventional permutation symbols for two- and three-dimensional space, respectively. The vector element of area $d\Sigma_J$ is

$$d\Sigma_J = N_J A^{\frac{1}{2}}\, dU^1\, dU^2. \tag{2.6}$$

The spatial aspects of S are described by the second fundamental form $\mathbf{B}$ defined by:

$$B_{\Gamma\Delta} = N_K X^K{}_{,\Gamma\Delta} = -X^K{}_{,\Gamma} N_{K,\Delta} \tag{2.7}$$

which satisfies the equations of Gauss and Weingarten

$$X^K{}_{,\Gamma\Delta} = B_{\Gamma\Delta} N^K; \qquad N^K{}_{,\Gamma} = -B^\Delta_\Gamma X^K{}_{,\Delta}. \tag{2.8}$$

Conditions of integrability to be satisfied by the fundamental tensors $\mathbf{A}$ and $\mathbf{B}$ are the equations of Mainardi-Codazzi

$$B_{\Gamma\Delta,\Sigma} = B_{\Gamma\Sigma,\Delta} \tag{2.9}$$

and the Gauss equations

$$R_{\Gamma\Delta\Lambda\Sigma} = B_{\Gamma\Lambda} B_{\Delta\Sigma} - B_{\Delta\Lambda} B_{\Gamma\Sigma}, \tag{2.10}$$

where $\mathbf{R}$ is the Riemann curvature tensor of S.

We now consider surfaces S and s given respectively by

$$X^K = X^K(U^\Delta); \qquad x^k = x^k(u^\delta). \tag{2.11}$$

We assume a deformation of S into s by a continuous process through a succession of surfaces $x^k(u^\delta, t)$ where t is the time variable. The mapping is defined by

$$u^\delta = u^\delta(U^\Delta), \tag{2.12}$$

associating points on S and s which will be referred to as the undeformed and deformed surfaces, respectively. Majuscule letters and indices will be associated with S, minuscule letters and indices with s. Hence for the geometry of s, Eqs. (2.1) to (2.10) hold with majuscules replaced by minuscules where appropriate.

From the previous discussion of surface geometry, we may define, as measures of nonlinear strain of a surface, deformation tensors

$$L_{\Delta\Sigma} = x^i{}_{,\Delta} x^i{}_{,\Sigma} = a_{\delta\sigma} u^\delta{}_{,\Delta} u^\sigma{}_{,\Sigma}, \tag{2.13}$$

$$K_{\Delta\Sigma} = n^i\, x^i{}_{,\Delta\Sigma} = b_{\delta\sigma} u^\delta{}_{,\Delta} u^\sigma{}_{,\Sigma}. \tag{2.14}$$

Here $L_{\Delta\Sigma}$, which we call the first Love–Kirchhoff deformation tensor, is the two-dimensional analog of the Cauchy–Green tensor enabling us to calculate length and angle changes. The second Love–Kirchhoff tensor $K_{\Delta\Sigma}$ arises as a natural consequence of surface geometry and allows the calculation of changes in normal curvature.

Clearly corresponding to $\mathbf{L}$ and $\mathbf{K}$ we may define $\mathbf{l}$ and $\mathbf{k}$ by

$$l_{\alpha\beta} = A_{\Delta\Gamma} U^\Delta{}_{,\alpha} U^\Gamma{}_{,\beta}, \tag{2.15}$$

$$k_{\alpha\beta} = B_{\Delta\Gamma} U^\Delta{}_{,\alpha} U^\Gamma{}_{,\beta}. \tag{2.16}$$

If T^Δ is a unit surface vector on S, then the stretch Λ of a surface element originally in the T^Δ direction is

$$\Lambda^2 = L_{\Delta\Sigma} T^\Delta T^\Sigma.$$

The extremum values of Λ^2 with respect to direction at a point on S are the roots of the equation

$$\det (L_{\Delta\Sigma} - \Lambda^2 A_{\Delta\Sigma}) = 0. \tag{2.17}$$

The basic invariants I_L, II_L are the coefficients in Eq. (2.17), i.e.,

$$I_L = A^{\Delta\Sigma} L_{\Delta\Sigma} = L^\Delta_\Delta, \tag{2.18}$$

$$II_L = (\det L_{\Delta\Sigma})/A = (a/A)\, |u/U|^2, \tag{2.19}$$

where $|u/U|$ is the Jacobian of the transformation (2.12). Similarly we define the principal values of $\mathbf{K}$ as solutions of

$$\det (K_{\Delta\Sigma} - \phi A_{\Delta\Sigma}) = 0$$

with the basic invariants given by

$$I_K = K^\Delta_\Delta, \tag{2.20}$$

[13] J. L. Ericksen, "Tensor Fields" in *Handbuch der Physik*, edited by S. Flügge (Springer Verlag, Berlin, 1960), Vol. III/1, pp. 794–858.

$$II_K = (\det K_{\Delta z})/A. \tag{2.21}$$

We next postulate that the surface has the physical property of mass such that there exist a surface mass density $\Gamma(U^\Delta)$ associated with S and $\gamma(u^\delta)$ associated with s. From the principle of conservation of mass

$$\int_S \Gamma \, d\Sigma = \int_s \gamma \, d\sigma. \tag{2.22}$$

If $d\Sigma$ and $d\sigma$ are the magnitudes of $d\Sigma_J$ and $d\sigma_i$, respectively, then from Eq. (2.6)

$$d\sigma/d\Sigma = (a/A)^{\frac{1}{2}} \, |u/U|$$

and hence from Eqs. (2.22) and (2.19)

$$\gamma/\Gamma = II_L^{-\frac{1}{2}}. \tag{2.23}$$

An incompressible or isochoric deformation is defined by

$$II_L = 1. \tag{2.24}$$

In the same manner, we may postulate surface thickness functions H and h leading to the mass per unit volume P and ρ, respectively, defined by

$$P = \Gamma/H, \qquad \rho = \gamma/h.$$

If we now define an isochoric deformation such that

$$\rho = P, \tag{2.25}$$

it follows that

$$h/H = II_L^{-\frac{1}{2}}. \tag{2.26}$$

It is seen from Eq. (2.26) that Eq. (2.25) is implied by Eq. (2.23). It is apparent that the concept of surface thickness will have meaning only if used in conjunction with a generalized definition of a surface which has the ingredients to represent changes in thickness during deformation. The deformation of S into s given by Eqs. (2.11) and (2.12) maps one two-dimensional space into another. We may generalize the concept of a surface to include not only a two-dimensional continuum of points but also the field of normal vectors on S. If we regard these vectors as rigidly attached to their respective tangent planes, then normals to S will map into normals to s. In general, however, we may specify that normals to S map after deformation into some field of vectors not normal to s.

The concept of such a generalized surface was introduced by the Cosserats[2] and amplified by Ericksen and Truesdell.[4] They defined such a generalized space as consisting of a surface with an associated field of vector triads and called it a directed surface. The deformation of such a directed surface is then specified not only by the deformation of its points but also by the deformation of its vector triads. Such a surface is of interest since it may have sufficient structure to lead to an adequate theory of thin shells.

We shall not pursue this concept in its full generality but shall consider a surface with a single field of vectors defined on it. The deformation given by Eqs. (2.11) and (2.12) is then augmented by defining vector fields on S by

$$D^K = D^K(U^\Delta) \tag{2.27}$$

and on s by

$$d^k = d^k(u^\delta) \tag{2.28}$$

and specifying that D^K at some point on S deform or map into d^k at the corresponding point on s. In special cases either both or one of D^K and d^k may be taken as normal to their respective surfaces. In the former case, normals map into normals, while in the latter the deformation of the normal is specified.

3. A PRINCIPLE OF VIRTUAL WORK

We shall postulate a principle of virtual work which is assumed to govern the mechanical behavior of a directed surface. This principle assumes the existence of a strain energy function, thus endowing the surface with hyperelastic material properties. The concept of a "variation" as applied to the deformed state of the surface is exactly that defined by Truesdell and Toupin.[14]

The virtual work $\mathfrak{A}$ associated with an arbitrary virtual displacement is defined by

$$\mathfrak{A} = \oint_c [s_k \, \delta x^k + \bar{m}_k^\sigma \, \delta x^k{}_{,\sigma} + p_k \, \delta d^k] \, dc,$$

$$+ \int_\sigma \gamma [f_k \, \delta x^k + \bar{l}_k^\sigma \, \delta x^k{}_{,\sigma} + q_k \, \delta d^k] \, d\sigma, \tag{3.1}$$

where c is a circuit enclosing a surface σ in s. The quantities s_k and f_k will ultimately be associated with the stress and body force vectors, while $\bar{m}_k^\sigma$ and $\bar{l}_k^\sigma$ will define the double stress and the double body vectors, respectively. The quantities p_k and q_k are a set of generalized forces associated with the internal structure of a directed surface. We also assume the existence of a strain energy function ϵ such that

$$\epsilon = \epsilon(x^k, \, x^k{}_{,\Delta}, \, x^k{}_{/\Delta z}), \tag{3.2}$$

[14] C. Truesdell and R. Toupin, *Classical Field Theories*, Handbuch der Physik, Vol. III/1 (Springer-Verlag, Berlin, 1960).

where

$$x^k{}_{/\Delta\Sigma} = \frac{\partial^2 x^k}{\partial U^\Delta\, \partial U^\Sigma} - \left\{ \begin{matrix} \Gamma \\ \Delta\ \Sigma \end{matrix} \right\}_a \frac{\partial x^k}{\partial U^\Gamma}$$

$$= x^k{}_{,\delta\sigma} u^\delta{}_{,\Delta} u^\sigma{}_{,\Sigma}.$$

Note that this is to be distinguished from

$$x^k{}_{,\Delta\Sigma} = \frac{\partial^2 x^k}{\partial U^\Delta\, \partial U^\Sigma} - \left\{ \begin{matrix} \Gamma \\ \Delta\ \Sigma \end{matrix} \right\}_A \frac{\partial x^k}{\partial U^\Gamma}$$

$$= x^k{}_{,\delta\sigma} u^\delta{}_{,\Delta} u^\sigma{}_{,\Sigma} + x^k{}_{,\delta} u^\delta{}_{,\Delta\Sigma}.$$

In the above definitions, $\left\{ \begin{matrix} \Gamma \\ \Delta\ \Sigma \end{matrix} \right\}$ represent Christoffel symbols and the subscripts a and A indicate that these quantities are evaluated with respect to the metric of the deformed and undeformed surface, respectively. Hence the surface we are investigating is a restricted material of grade two.[12] The energy W stored during the deformation is then given by

$$W = \int_\sigma \gamma\epsilon \, d\sigma. \tag{3.3}$$

The strain energy function is subject to the condition of "material indifference" or "isotropy of space," i.e., ϵ is invariant under arbitrary rigid motions:

$$\delta\epsilon = 0, \tag{3.4a}$$

when

$$\delta x^i = c^i, \qquad \delta x^i = a^i_j x^j, \tag{3.4b, c}$$

where c^i is a constant vector and a^i_j an arbitrary constant skew-symmetric tensor. From Eqs. (3.4a, b), ϵ is independent of x^k and from Eqs. (3.4a, c)

$$\frac{\partial \epsilon}{\partial x^{[i}{}_{,\Delta}} x^{j]}{}_{,\Delta} + \frac{\partial \epsilon}{\partial x^{[i}{}_{/\Delta\Sigma}} x^{j]}{}_{/\Delta\Sigma} = 0 \tag{3.5}$$

where the notation $B_{[ij]}$ is that of Ericksen,[13] i.e., $B_{[ij]} = \frac{1}{2}(B_{ij} - B_{ji})$. We now postulate a principle of virtual work that a necessary condition for the equilibrium of our surface is

$$\mathcal{A} = \delta W \tag{3.6}$$

for arbitrary virtual displacements consistent with whatever constraints may exist. The variation δW in Eq. (3.6) will be subject to the requirements of objectivity given by Eq. (3.5) as well as of conservation of mass, i.e.,

$$\delta(\gamma \, d\sigma) = 0. \tag{3.7}$$

If we give the surface a rigid translation defined by Eq. (3.4b) such that $\delta W = 0$, then from (3.1) and (3.6) we obtain

$$\int_c s_k \, dc + \int_\sigma \gamma f_k \, d\sigma = 0$$

which represents the equation of force equilibrium of the surface. Hence s_k represents a stress vector with dimensions force per unit length, and f_k a body force vector with dimensions force per unit surface mass. In an analogous fashion, if we subject s to a rigid rotation defined by Eq. (3.4c) such that $\delta W = 0$, we obtain

$$\oint_c [s_{[k} x^{j]} + \bar{m}^\delta_{[ik} x^{j]}{}_{,\delta} + p_{[k}\, d^{j]}] \, dc$$

$$+ \int_\sigma \gamma[f_{[k} x^{j]} + \bar{l}^\delta_{[ik} x^{j]}{}_{,\delta} + q_{[k}\, d^{j]}] \, d\sigma = 0,$$

which represents the equation of moment equilibrium. Hence, for example, we may identify $\bar{m}^\sigma_{[ik} x^{j]}{}_{,\sigma}$, $p_{[k} d^{j]}$ as couple stress vectors and $\bar{l}^\delta_{[ik} x^{j]}{}_{,\delta}$, $q_{[k} d^{j]}$ as body couple vectors. (A couple vector can be equivalently represented by an axial vector m^i or an absolute second-order skew-symmetric tensor m^{ij}, see Appendix.) It is apparent that we can define in general

$$\bar{m}^\delta_k x^j{}_{,\delta} = \bar{m}^\delta_{[ik} x^{j]}{}_{,\delta} + \bar{m}^\delta_{(ik} x^{j)}{}_{,\delta},$$

where the parentheses about two indices means that the tensor is symmetrized with respect to these indices.[13] Since $\bar{m}^\delta_{[ik} x^{j]}{}_{,\delta}$ can be interpreted as double force with moment, then $\bar{m}^\delta_{(ik} x^{j)}{}_{,\delta}$ is a double force without moment[15] and $\bar{m}^\delta_k x^j{}_{,\delta}$ is a double force distribution. Similar remarks will apply for $p_k d^j$, $\bar{l}^\delta_k x^j{}_{,\delta}$, and $q_k d^j$.

We state here a form of Green's theorem which will be needed in our future work. Let $c^{i\alpha}$ be an arbitrary double tensor field defined on s. Let c be a circuit enclosing the region σ in s and let ν^α be the unit outward normal to c and tangential to s. Then Green's theorem has the form

$$\int_\sigma c^{i\alpha}{}_{,\alpha} \, d\sigma = \oint_c c^{i\alpha} \nu_\alpha \, dc, \tag{3.8}$$

where the spatial components of $c^{i\alpha}$ are referred to a rectangular Cartesian coordinate system.

4. A MEMBRANE THEORY

In this section the principle of virtual work will be utilized under the assumption that the strain energy function depends only on the deformation gradients $x^i{}_{,\Delta}$, i.e., the material is simple.[10] Application of the principle under the restrictions imposed by the condition of material indifference leads to

[15] R. D. Mindlin, Arch. Ratl. Mech. Anal. 16, 51 (1964).

a system of mechanics in which the stress tensor is a symmetric tangential surface tensor satisfying the membrane equilibrium equations, and is related to the deformation tensor $L_{\Delta\Sigma}$ through a set of nonlinear constitutive relations. The form taken by these constitutive relations under the assumption of material isotropy is investigated.

Under the assumption that

$$\epsilon = \epsilon(x^k{}_{,\Delta}) \tag{4.1}$$

then if we set

$$t^\delta_k = \gamma(\partial\epsilon/\partial x^k{}_{,\Delta})u^\delta{}_{,\Delta} \tag{4.2}$$

the principle of virtual work (3.6), on applying Eqs. (3.1), (3.3), (3.7), and (3.8), takes the form

$$\oint_c [(s_k - t^\delta_k\nu_\delta)\,\delta x^k + \bar{m}^\sigma_k\,\delta x^k{}_{,\sigma} + p_k\,\delta\, d^k]\,dc + \int_\sigma [(t^\sigma_{k,\sigma} + \gamma f_k)\,\delta x^k + \gamma\bar{l}^\sigma_k\,\delta x^k{}_{,\sigma} + \gamma q_k\,\delta\, d^k]\,d\sigma = 0. \tag{4.3}$$

Since Eq. (4.3) must hold for arbitrary variations, we consider a virtual translation $\delta x^k = b^k$, where b^k is a constant and $\delta x^k{}_{,\sigma} = \delta d^k = 0$. It follows that

$$t^\sigma_{k,\sigma} + \gamma f_k = 0 \quad \text{in } s, \tag{4.4}$$

$$s_k = t^\sigma_k\nu_\sigma \quad \text{on } c. \tag{4.5}$$

By now considering arbitrary nonzero virtual displacements $\delta x^k{}_{,\sigma}$ and δd^k it follows that

$$\bar{m}^\sigma_k = \bar{l}^\sigma_k = 0, \tag{4.6}$$

$$p_k = q_k = 0. \tag{4.7}$$

Here s_k and f_k are the stress and body force vectors, respectively, t^σ_k is the stress tensor, and Eqs. (4.4) and (4.5) correspond to equations of force equilibrium and boundary conditions, respectively. The constitutive relations are given by Eq. (4.2). It follows from Eq. (4.6) that the double stress and double body vectors are zero. We now wish to consider the implications of material indifference on the form of the constitutive relations. From Eqs. (3.5) and (4.1), this condition is expressed by

$$(\partial\epsilon/\partial x^{[i}{}_{,\Delta})x^{j]}{}_{,\Delta} = 0. \tag{4.8}$$

Equation (4.8) constitutes a set of three independent first-order homogeneous partial differential equations which must be satisfied by the strain energy function ϵ. Since there are six independent variables $x^i{}_{,\Delta}$, Eq. (4.8) possesses three functionally independent solutions and these may be shown to be given by

$$L_{\Delta\Sigma} = x^i{}_{,\Delta}x^i{}_{,\Sigma}. \tag{4.9}$$

It thus follows that for material indifference

$$\epsilon = \epsilon(L_{\Delta\Sigma}) \tag{4.10}$$

and we recall that $L_{\Delta\Sigma}$ has been defined as the first Love–Kirchhoff deformation tensor.

If we now substitute Eq. (4.2) into Eq. (4.8), the condition of material indifference is seen to require

$$t^\delta_{[i}x^{j]}{}_{,\delta} = 0, \tag{4.11}$$

or equivalently

$$\epsilon^{ij}_k t^\delta_j x^k{}_{,\delta} = 0. \tag{4.12}$$

Taking the tangential and normal components of Eq. (4.12) results in

$$\epsilon^{ij}_k x_{i,\gamma}x^j{}_{,\delta}t^\delta_k = 0, \tag{4.13a}$$

$$\epsilon^{ij}_k n_i x^k{}_{,\delta}t^\delta_j = 0, \tag{4.13b}$$

where

$$t^{i\delta} = t^{\alpha\delta}x^i{}_{,\alpha} + t^\delta n^i. \tag{4.14}$$

Noting that

$$\epsilon_{ijk}x^j{}_{,\lambda}n^k = -a^{\alpha\delta}e_{\lambda\alpha}x_{i,\delta}, \tag{4.15}$$

then on using Eqs. (4.14), (4.15), and (2.4), we obtain from Eq. (4.13a)

$$t^\delta = 0 \tag{4.16}$$

and from Eq. (4.13b),

$$t^{[\alpha\delta]} = 0. \tag{4.17}$$

Thus the stress tensor t^δ_i, under the restrictions imposed by Eqs. (4.1) and (4.8), is a symmetric surface tensor $t^{\alpha\beta}$. The equations of equilibrium and boundary conditions as given by Eqs. (4.4) and (4.5) reduce to the well-known set[14]

$$t^{\alpha\beta}{}_{,\beta} + \gamma f^\alpha = 0, \tag{4.18a}$$

$$b_{\alpha\beta}t^{\alpha\beta} + \gamma f = 0, \tag{4.18b}$$

and

$$s^\alpha = t^{\alpha\beta}\nu_\beta, \qquad s = 0 \tag{4.19}$$

where $s^k = s^\alpha x^k{}_{,\alpha} + sn^k$.

The constitutive relations, from Eqs. (4.2), (4.9), and (4.10), are given by

$$t^{\alpha\beta} = 2\gamma(\partial\epsilon/\partial L_{\Delta\Sigma})u^\alpha{}_{,\Delta}u^\beta{}_{,\Sigma}, \tag{4.20}$$

where ϵ is regarded as a function of the four components of $L_{\Delta\Sigma}$ treated as independent variables under the restriction that

$$\epsilon_{\Delta\Sigma}(\partial\epsilon/\partial L_{\Delta\Sigma}) = 0, \tag{4.21}$$

Eqs. (4.9), (4.17), (4.18), (4.19), (4.20), and (4.21) comprise the basic equations of membrane surface theory.

If the material is isotropic at every point, then since the intrinsic surface geometry in the neighborhood of an arbitrary point may be regarded as locally Euclidean,[16] the strain energy function will at every point be a function of the basic invariants of $L_{\Delta\Sigma}$. We have, therefore;

$$\epsilon = \epsilon(I_L, II_L). \tag{4.22}$$

From Eqs. (4.20), (4.22), (2.18), and (2.19), the constitutive equations for an isotropic surface may be reduced to the form

$$t^{\alpha\beta} = 2\gamma\left[\frac{\partial\epsilon}{\partial I_L} l^{-1\alpha\beta} + II_L \frac{\partial\epsilon}{\partial II_L} a^{\alpha\beta}\right], \tag{4.23}$$

where

$$l^{-1\alpha\beta} = A^{\Delta\Gamma} u^{\alpha}{}_{,\Delta} u^{\beta}{}_{,\Gamma}, \qquad l^{-1\alpha\beta} l_{\beta\gamma} = \delta^{\alpha}_{\gamma}, \tag{4.24}$$

and $l_{\alpha\beta}$ is given by Eq. (2.15). Equation (4.23) may be regarded as the simple two-dimensional analog of Finger's relations.[11]

Since the assumption of an elastic material presupposes the existence of a natural state, then in the special case when this state is stress-free, the expressions for the constitutive relations given by Eqs. (4.20) and (4.23) will be subject to the conditions

$$(\partial\epsilon/\partial L_{\Delta\Gamma})_0 = 0$$

for the anisotropic case, and

$$(\partial\epsilon/\partial I_L)_0 + (\partial\epsilon/\partial II_L)_0 = 0$$

for the isotropic case, where the subscript 0 indicates that the derivatives are evaluated in the natural state.

5. A BENDING THEORY

In this section, the principle of virtual work is utilized under the assumption that the strain energy function depends on $x^i{}_{,\Delta}$ and $x^i{}_{/\Delta\Sigma}$. In order to allow arbitrary virtual displacements, the principle is suitably modified to include certain kinematic constraints.

We have, therefore,

$$\epsilon = \epsilon(x^i{}_{,\Delta}, x^i{}_{/\Delta\Sigma}), \tag{5.1}$$

where the kinematic variables $x^i{}_{,\Delta}$, $x^i{}_{/\Delta\Sigma}$ are subject to the constraints imposed by the surface geometry and given by

$$x^i{}_{,\Delta} x^i{}_{/\Sigma\Gamma} = 0. \tag{5.2}$$

The constraints expressed by Eq. (5.2) are introduced into the principle of virtual work by writing the variation $\delta\epsilon$ in the form

$$\delta\epsilon = \frac{\partial\epsilon}{\partial x^i{}_{,\Delta}} \delta x^i{}_{,\Delta} + \frac{\partial\epsilon}{\partial x^i{}_{/\Delta\Sigma}} \delta x^i{}_{/\Delta\Sigma} + \bar{\lambda}^{\Delta\Sigma\Gamma}\, \delta(x^k{}_{,\Delta} x^k{}_{/\Sigma\Gamma}), \tag{5.3}$$

where $\bar{\lambda}^{\Delta\Sigma\Gamma}$ is an arbitrary surface tensor which is symmetric in the last two indices, i.e., $\bar{\lambda}^{\Delta\Sigma\Gamma} = \bar{\lambda}^{\Delta\Gamma\Sigma}$. If we set

$$t^{\delta}_{k} = \gamma(\partial\epsilon/\partial x^k{}_{,\Delta}) u^{\delta}{}_{,\Delta}, \tag{5.4}$$

$$\bar{\mu}^{\delta\sigma}_{k} = \gamma(\partial\epsilon/\partial x^k{}_{/\Delta\Sigma}) u^{\delta}{}_{,\Delta} u^{\sigma}{}_{,\Sigma}, \tag{5.5}$$

$$\bar{l}^{\alpha} = b_{\sigma\gamma}\bar{\lambda}^{\alpha\sigma\gamma}, \tag{5.6}$$

then from Eqs. (3.1), (3.6), and (3.8), the principle of virtual work may be reduced to the form

$$\oint_c [\{s_k - (t^{\sigma}_{k} + \bar{l}^{\sigma} n_k)\nu_\sigma\}\, \delta x^k + \{\bar{m}^{\beta}_{k} - (\bar{\mu}^{\beta\sigma}_{k} + x^k{}_{,\lambda}\bar{\lambda}^{\lambda\beta\sigma})\nu_\sigma\}\, \delta x^k{}_{,\beta} + p_k\, \delta\, d^k]\, dc$$

$$+ \int_\sigma [\{(t^{\beta}_{k} + \bar{l}^{\beta} n_k)_{,\beta} + \gamma f_k\}\, \delta x^k + \{(\bar{\mu}^{\beta\sigma}_{k} + x^k{}_{,\alpha}\bar{\lambda}^{\alpha\beta\sigma})_{,\sigma} + \gamma\bar{l}^{\beta}_{k}\}\, \delta x^k{}_{,\beta} + \gamma q_k\, \delta\, d^k]\, d\sigma = 0. \tag{5.7}$$

For arbitrary variations δx^k, $\delta x^k{}_{,\beta}$, and δd^k, we find that on c

$$s_k = (t^{\delta}_{k} + \bar{l}^{\delta} n_k)\nu_\delta, \tag{5.8a}$$

$$\bar{m}^{\delta}_{k} = (\bar{\mu}^{\delta\sigma}_{k} + x^k{}_{,\beta}\bar{\lambda}^{\beta\delta\sigma})\nu_\sigma, \tag{5.8b}$$

$$p_k = 0, \tag{5.8c}$$

and in s

$$(t^{\alpha}_{k} + \bar{l}^{\alpha} n_k)_{,\alpha} + \gamma f_k = 0, \tag{5.9a}$$

$$(\bar{\mu}^{\alpha\sigma}_{k} + x_{k,\beta}\bar{\lambda}^{\beta\alpha\sigma})_{,\sigma} + \gamma\bar{l}^{\alpha}_{k} = 0, \tag{5.9b}$$

$$q_k = 0. \tag{5.9c}$$

Equations (5.8) express the boundary conditions, Eqs. (5.9) the equations of equilibrium.

If we multiply Eq. (5.9b) by $x^i{}_{,\alpha}$ we obtain

$$(\bar{\mu}^{\alpha\sigma}_{k} + x_{k,\beta}\bar{\lambda}^{\beta\alpha\sigma})_{,\sigma} x^i{}_{,\alpha} + \gamma\bar{l}^{\alpha}_{k} x^i{}_{,\alpha} = 0 \tag{5.10}$$

and this expresses the equilibrium of the double forces as defined in Sec. 3. The double stresses are given by the quantities $(\bar{\mu}^{\alpha\sigma}_{k} + x_{k,\beta}\bar{\lambda}^{\beta\alpha\sigma}) x^i{}_{,\alpha}$. The part of Eq. (5.10) which is antisymmetric in k and j is the equation of moment equilibrium, i.e.,

$$(\bar{\mu}^{\alpha\sigma}_{[k} + x_{[k,\beta}\bar{\lambda}^{\beta\alpha\sigma})_{,\sigma} x^{j]}{}_{,\alpha} + \gamma\bar{l}^{\alpha}_{[k} x^{j]}{}_{,\alpha} = 0.$$

We note that the quantities $\bar{\mu}^{\alpha\sigma}_{k} x^i{}_{,\alpha}$ have energetic significance by virtue of Eq. (5.5), whereas the quantities $x_{k,\beta}\bar{\lambda}^{\beta\alpha\sigma} x^i{}_{,\alpha}$ are double stresses which are

[16] J. L. Synge and A. Schild, *Tensor Calculus* (The University of Toronto, Toronto, Canada, 1949).

undetermined in terms of the strain energy function.

We recall that ϵ, when specified by Eq. (5.1), must satisfy the condition of material indifference expressed by Eq. (3.5), i.e.,

$$(\partial\epsilon/\partial x^{[i}{}_{,\Delta})x^{j]}{}_{,\Delta} + (\partial\epsilon/\partial x^{[i}{}_{/\Delta\Sigma})x^{j]}{}_{/\Delta\Sigma} = 0. \tag{5.11}$$

If we regard $x^i{}_{/\Delta\Sigma}$ as constituting 12 independent variables, ϵ must be further subject to the constraining equations

$$\partial\epsilon/\partial x^i{}_{/\Delta\Sigma} = \partial\epsilon/\partial x^i{}_{/\Sigma\Delta}. \tag{5.12}$$

Equations (5.11) and (5.12) comprise a set of six independent first-order homogeneous partial differential equations in the 18 independent variables $x^i{}_{,\Delta}$ and $x^i{}_{/\Delta\Sigma}$. There are, therefore, 12 functionally independent solutions of this set given by

$$\begin{aligned} &L_{\Delta\Sigma} = x^i{}_{,\Delta}x^i{}_{,\Sigma}; \\ &K_{\Delta\Sigma} = n^i x^i{}_{/(\Delta\Sigma)}, \qquad C_{\Delta\Sigma\Gamma} = x^i{}_{,\Delta}x^i{}_{/(\Sigma\Gamma)}. \end{aligned} \tag{5.13}$$

The scalar function ϵ must be expressible in terms of these solutions which are, however, subject to the geometrical constraints of Eq. (5.2), i.e., $C_{\Delta\Sigma\Gamma}=0$. It follows that ϵ is expressible in terms of the Love–Kirchhoff tensors **L** and **K**, i.e.,

$$\epsilon = \epsilon(L_{\Delta\Sigma}, K_{\Delta\Sigma}). \tag{5.14}$$

On substituting Eqs. (5.4) and (5.5) into Eq. (5.11), the condition of material indifference requires

$$\epsilon_{ijk}t^{i\delta}x^k{}_{,\delta} + \epsilon_{ijk}\bar{\mu}^{i\delta\sigma}x^k{}_{,\delta\sigma} = 0. \tag{5.15}$$

Decomposing Eq. (5.15) into its surface representation yields

$$\epsilon_{ijk}x^i{}_{,\gamma}x^k{}_{,\delta}t^{i\delta} + \epsilon_{ijk}x^i{}_{,\gamma}n^k b_{\delta\sigma}\bar{\mu}^{i\delta\sigma} = 0,$$

$$\epsilon_{ijk}n^i x^k{}_{,\beta}t^{i\beta} = 0.$$

These equations lead at once to the following results:

$$t^\delta = b_{\alpha\sigma}\bar{\mu}^{\delta\alpha\sigma}, \tag{5.16}$$

$$t^{[\delta\sigma]} = 0, \tag{5.17}$$

when we use the surface representation of $t^{i\alpha}$ as given by Eq. (4.14) and that of $\bar{\mu}^{k\delta\sigma}$ given by

$$\bar{\mu}^{k\delta\sigma} = \bar{\mu}^{\alpha\delta\sigma}x^k{}_{,\alpha} + \bar{\mu}^{\delta\sigma}n^k.$$

We note that Eq. (5.16) has the same form as Eq. (5.6). Moreover, from Eqs. (5.4), (5.5), and (5.14) there follows immediately

$$t^\delta = \bar{\mu}^{\alpha\delta\sigma} = 0. \tag{5.18}$$

If we now take the surface decomposition of Eq. (5.9) and use the results of Eq. (5.18), we obtain the equations of force equilibrium in the form

$$t^{\beta\alpha}{}_{,\alpha} - \bar{l}^\alpha b^\beta_\alpha + \gamma f^\beta = 0, \tag{5.19a}$$

$$\bar{l}^\alpha{}_{,\alpha} + b_{\beta\alpha}t^{\beta\alpha} + \gamma f = 0. \tag{5.19b}$$

The corresponding boundary conditions from Eq. (5.8a), are

$$s^\alpha = t^{\alpha\beta}\nu_\beta, \tag{5.20a}$$

$$s = \bar{l}^\alpha\nu_\alpha. \tag{5.20b}$$

Similarly taking the surface decomposition of the skew symmetric part of Eq. (5.10) leads to the equation of moment equilibrium in the form

$$e_{\alpha\delta}\bar{\mu}^{\delta\sigma}{}_{,\sigma} - 2e_{\alpha\delta}b_{\gamma\sigma}\bar{\lambda}^{[\delta\gamma]\sigma} + e_{\alpha\delta}\bar{l}^\delta + \gamma e_{\alpha\delta}\bar{l}^\delta = 0, \tag{5.21a}$$

$$e_{\alpha\beta}\bar{\lambda}^{\alpha\beta\sigma}{}_{,\sigma} + e_{\alpha\beta}b^\beta_\sigma\bar{\mu}^{\alpha\sigma} + \gamma e_{\alpha\beta}\bar{l}^{\alpha\beta} = 0. \tag{5.21b}$$

The symmetric part of Eq. (5.10) yields

$$\bar{\lambda}^{(\gamma\delta)\sigma}{}_{,\sigma} - b^{(\gamma}_\sigma\bar{\mu}^{\delta)\sigma} + \gamma\bar{l}^{(\gamma\delta)} = 0. \tag{5.22}$$

The corresponding boundary conditions from Eq. (5.8b) are

$$\bar{m}^\alpha = \bar{\mu}^{\alpha\beta}\nu_\beta, \tag{5.23a}$$

$$\bar{m}^{\alpha\delta} = \bar{\lambda}^{\alpha\delta\sigma}\nu_\sigma, \tag{5.23b}$$

where the surface representation of $\bar{m}^{k\beta}$ is given by Eq. (A5). Equations (5.19) correspond to the equations of force equilibrium given by Ericksen and Truesdell[4] with $t^{[\alpha\beta]} = 0$. As shown in the Appendix, Eq. (5.21) may also be brought into correspondence with the equations of moment equilibrium of Ref. 4 [Eqs. (26.8) and (26.9) p. 320]. Equation (5.22) corresponds to equilibrium of double stresses without moment and is new.

The quantities $\bar{\lambda}^{\alpha\beta\gamma}$ may be interpreted as reactions which force the surface to conform to the kinematic constraints. Since these constraints do no work during deformation, they are undetermined by the strain energy function. It is apparent from the arbitrariness of the variations considered, that these constraints prevent the surface base vectors $x^i{}_{,\alpha}$ from deforming in a manner other than that given by the surface mapping. The quantities $\bar{\lambda}^{\alpha\beta\gamma}$ then correspond to forces which resist the deformation of the base vectors and define not only couples but also new mechanical quantities whose equilibrium is expressed by Eq. (5.22). If we write

$$\bar{\lambda}^{\alpha\beta\gamma} = \bar{\lambda}^{(\alpha\beta)\gamma} + \bar{\lambda}^{[\alpha\beta]\gamma}, \tag{5.24}$$

then from the interpretation of Eq. (5.21) as an equation of moment equilibrium, following Mindlin,[18] $\bar{\lambda}^{[\alpha\beta]\gamma}$ may be interpreted as double stresses with moment, $\bar{\lambda}^{(\alpha\beta)\gamma}$ as double stresses without moment.

The constitutive equations relating the nonzero components $t^{\alpha\beta}$, $\bar{\mu}^{\alpha\beta}$ of the stress and the couple

stress tensors, respectively, to the deformation tensors are, from Eqs. (5.4), (5.5), (5.13), and (5.14), given by

$$t^{\alpha\beta} = 2\gamma(\partial\epsilon/\partial L_{\Delta\Sigma})u^{\alpha}{}_{,\Delta}u^{\beta}{}_{,\Sigma}, \tag{5.25}$$

$$\mu^{\alpha\beta} = 2\gamma(\partial\epsilon/\partial K_{\Delta\Sigma})u^{\alpha}{}_{,\Delta}u^{\beta}{}_{,\Sigma}, \tag{5.26}$$

where ϵ is regarded as a function of the four components of $L_{\Delta\Sigma}$, and $K_{\Delta\Sigma}$, respectively, treated as independent variables under the restriction that

$$\epsilon_{\Delta\Sigma}\,\partial\epsilon/\partial L_{\Delta\Sigma} = \epsilon_{\Delta\Sigma}\,\partial\epsilon/\partial K_{\Delta\Sigma} = 0. \tag{5.27}$$

The constitutive Eqs. (5.25) and (5.26) may also be subject to the condition that they vanish in the undeformed or natural state.

In the previous section it was argued that ϵ would be a form-invariant function of $L_{\Delta\Sigma}$ and hence a function of the invariants of $L_{\Delta\Sigma}$ if the material was assumed initially isotropic. The success of the argument depends upon making the surface metric in the undeformed state an isotropic tensor at any point. This argument fails when considering the second fundamental tensor since, in general, it will depend on the direction of the coordinate axes chosen at the point in question. Hence, the assumption of material isotropy will, in general, require that

$$\epsilon = \epsilon(L_{\Delta\Sigma}, K_{\Delta\Sigma}). \tag{5.28}$$

In certain special cases such as the plane and sphere, however, the second fundamental tensor is independent of coordinates and ϵ will be a form invariant function of $L_{\Delta\Sigma}$ and $K_{\Delta\Sigma}$. We can, however, define our surface as isotropic if ϵ is a form-invariant function of $L_{\Delta\Sigma}$ and $K_{\Delta\Sigma}$. Rivlin[17] has shown that for two symmetric second-order tensors there are five basic invariants. Four of these, I_L, II_L, I_K, II_K, have been defined previously by Eqs. (2.18), (2.19), (2.20), and (2.21). The fifth invariant, I_{LK}, is given by

$$I_{LK} = A^{\Delta\Sigma}A^{\Phi\Gamma}L_{\Delta\Gamma}K_{\Sigma\Phi}. \tag{5.29}$$

For this special case of "surface isotropy", the constitutive relations (5.25) and (5.26) become

$$t^{\alpha\beta} = 2\gamma\Big[\frac{\partial\epsilon}{\partial I_L}\,l^{-1\alpha\beta} + II_L\frac{\partial\epsilon}{\partial II_L}\,a^{\alpha\beta} + \frac{\partial\epsilon}{\partial I_{LK}}\,b_{\sigma\phi}l^{-1\sigma\alpha}l^{-1\phi\beta}\Big], \tag{5.30}$$

$$\mu^{\alpha\beta} = 2\gamma\Big[\frac{\partial\epsilon}{\partial I_K}\,l^{-1\alpha\beta} + II_K\frac{\partial\epsilon}{\partial II_K}\,b^{-1\alpha\beta} + \frac{\partial\epsilon}{\partial I_{LK}}\,a_{\sigma\phi}l^{-1\sigma\alpha}l^{-1\phi\beta}\Big], \tag{5.31}$$

[17] R. S. Rivlin, J. Ratl. Mech. Anal. 4, 681 (1955).

where

$$b^{-1\alpha\beta}b_{\beta\gamma} = \delta^{\alpha}_{\gamma}. \tag{5.32}$$

Equation (5.30) may be regarded as the R_2 analog of Finger's relations. For a simple material, Eq. (5.30) reduces to Eq. (4.23).

In summary, then, the field equations of the bending theory of surfaces are given by Eqs. (5.19), (5.21), and (5.22). The stress boundary conditions are given by Eqs. (5.20) and (5.23). The general constitutive relations are given by Eqs. (5.25) and (5.26). For the case of "surface isotropy" as defined, the constitutive relations are replaced by Eqs. (5.30) and (5.31).

APPENDIX

A couple stress vector m^i and body couple vector l^i are axial vectors which can be represented by equivalent skew symmetric tensors $\bar{m}^{ij}$ and $\bar{l}^{ij}$. This dual representation is given by

$$2\bar{m}^{ij} = \epsilon^{kij}m_k, \tag{A1}$$

$$m^i = \epsilon_{ijk}\bar{m}^{jk}, \tag{A2}$$

with similar expressions relating l^i and $\bar{l}^{ij}$. If we define

$$\bar{m}^{ij} = \bar{m}^{[i\beta}x^{j]}{}_{,\beta} \tag{A3}$$

and set

$$m^i = m^{\alpha}x^i{}_{,\alpha} + mn^i, \tag{A4}$$

$$\bar{m}^{i\alpha} = \bar{m}^{\beta\alpha}x^i{}_{,\beta} + \bar{m}^{\alpha}n^i, \tag{A5}$$

then from Eq. (A2), using Eqs. (2.4) and (4.15), we find

$$m_{\alpha} = e_{\beta\alpha}\bar{m}^{\beta}, \tag{A6}$$

$$m = e_{\alpha\beta}\bar{m}^{\alpha\beta}. \tag{A7}$$

Similarly, the skew-symmetric couple stress tensor $\mu^{[i\alpha\beta}x^{j]}{}_{,\alpha}$ corresponds to a second-order tensor $\mu^{i\alpha}$. Relations analogous to Eqs. (A6) and (A7) are, when we set $\mu^{i\alpha} = \mu^{\beta\alpha}x^i{}_{,\beta} + \mu^{\alpha}n^i$,

$$\mu^{\sigma}_{\beta} = e_{\alpha\beta}\bar{\mu}^{\alpha\sigma}, \tag{A8}$$

$$\mu^{\delta} = e_{\alpha\beta}\bar{\mu}^{\alpha\beta\delta}. \tag{A9}$$

If $\bar{\mu}^{\alpha\beta\delta} = 0$, then from Eq. (A9), the normal component of the couple stress tensor $\mu^{\delta} \equiv \mu^{i\delta}n^i = 0$.

If in Eq. (5.21), we use relations of the form given by Eqs. (A8) and (A9) as well as the identity

$$2e_{\sigma\alpha}b_{\beta\gamma}\bar{\lambda}^{[\sigma\beta]\gamma} = e_{\sigma\beta}b_{\gamma\alpha}\bar{\lambda}^{[\sigma\beta]\gamma}, \tag{A10}$$

we obtain the equations of moment equilibrium in the conventional form[4,14] (with $t^{[\alpha\beta]} = 0$)

$$\mu^{\alpha}_{\beta,\alpha} - b_{\beta\alpha}\lambda^{\alpha} + e_{\beta\alpha}l^{\alpha} + \gamma l_{\beta} = 0, \tag{A11}$$

$$\lambda^{\alpha}{}_{,\alpha} + b^{\beta}_{\alpha}\mu^{\alpha}_{\beta} + \gamma l = 0. \tag{A12}$$

The Generalization of Choh–Uhlenbeck's Method in the Kinetic Theory of Dense Gases

LEOPOLDO S. GARCÍA-COLÍN* AND ASDRÚBAL FLORES†
Escuela de Ciencias Físico-Matemáticas, Universidad Autónoma de Puebla, Puebla, Pue. (México)
(Received 20 July 1965)

The method proposed by Choh and Uhlenbeck to deal with kinetic phenomena in dense gases is generalized to all orders in the density. The set of integral equations for the functions defining the transport coefficients is derived. It is shown that the thermal conductivity and the shear viscosity are independent of the way in which the local temperature is introduced, namely, through the kinetic energy and through the total energy density. However, the bulk viscosity does depend on the particular definition of temperature. The relationship between the corresponding bulk viscosities is explicitly obtained.

INTRODUCTION

In a recent paper[1] we derived the first order in the density corrections to the transport coefficients of a moderately dense gas using a method which starts from a generalized Boltzmann equation which is valid to all orders in the density, and following Choh,[2] we have defined the temperature of the system through the kinetic energy only. Also, we derived an inhomogeneous linear integral equation for the perturbation function $\varphi_k(p)$ which describes the local nonequilibrium state of the system, linear in the macroscopic gradients. This equation is valid to all orders in the density; in fact, it is independent of the existence of an expansion in powers series in the density of the two body distribution functional, which is contained in the equation. Furthermore, we assumed that such an expansion exists, but we restricted ourselves to discuss the solution up to the term linear in the density. In this paper, we want to consider the full integral equation for $\varphi_k(p)$ and, therefore, to extend to all orders in the density the results of Choh. This is interesting from the point of view of finding out how the temperature definition can affect the values of the transport coefficients for such a system.

In Sec. I of this paper we very briefly sketch how the integral equation for the perturbation function $\varphi_k(p)$ was derived in I. In Sec. II we discuss the structure of the solution for such an equation, together with the subsidiary conditions that it must satisfy. Finally, in Sec. III we discuss the nature of the transport coefficients obtained with this solution.

I. THE NONEQUILIBRIUM STATE

The main steps leading to the definition of a nonequilibrium state linear in the macroscopic gradients, for a gas which is assumed to be composed of N molecules enclosed in a container of volume V and interacting among themselves via short-range repulsive forces, in the absence of any external forces except that exerted by the walls of the container, may be summarized as follows:

First, a generalized Boltzmann equation based on the assumption that all molecular distribution functions of order higher than the first, are time-independent functionals of the one particle distribution function, is established.[3] This equation which depends on the nonlocal two-body distribution functional, is approximated to include only those effects which are linear in the gradients.

The resulting equation reads

$$\frac{\partial f_1}{\partial t} + \frac{\mathbf{p}}{m}\cdot\frac{\partial f_1}{\partial \mathbf{q}} = \Phi[x_1 \mid f_1(\mathbf{q})] + \int dx'\, \Phi'[x_1, x' \mid f_1(\mathbf{q})](\mathbf{q}' - \mathbf{q})\cdot\left(\frac{\partial f_1}{\partial \mathbf{q}'}\right)_{\mathbf{q}'=\mathbf{q}} \tag{1.1}$$

where $x_1 = (\mathbf{q}, \mathbf{p})$ and

$$\Phi[x_1 \mid f_1(\mathbf{q})] = \int dx_2\, \theta_{12} f_2[x_1, x_2 \mid f_1(\mathbf{q})], \tag{1.2}$$

$f_2[x_1, x_2 \mid f_1(\mathbf{q})]$ being the two-particle distribution functional evaluated for a local one-particle distribution function $f_1(\mathbf{q}, t)$ and

$$\theta_{12} = \frac{\partial\phi(r_{12})}{\partial \mathbf{q}}\cdot\frac{\partial}{\partial \mathbf{p}} + \frac{\partial\phi(r_{12})}{\partial \mathbf{q}_2}\cdot\frac{\partial}{\partial \mathbf{p}_2}, \tag{1.3}$$

$\phi(r_{12})$ being the intermolecular potential and $r_{12} = |\mathbf{q}_2 - \mathbf{q}|$. [We omit writing explicitly the time de-

* Comisión Nacional de Energía Nuclear.
† Becario del Instituto Nacional de la Investigación Científica.
[1] L. S. García-Colín and A. Flores, Physica (to be published); we here after refer to this paper as I.
[2] S. T. Choh, Ph.D. dissertation, University of Michigan (1958).
[3] L. S. García-Colín, M. S. Green, and F. Chaos, Physica (to be published).

pendence of f_1 in all expressions. Also, $f_1(\mathbf{q}, t) \equiv f_1(\mathbf{q}, \mathbf{p}''; t)$ were $\mathbf{p}''$ indicates that the momentum is not fixed.]

In Eq. (1.1) $\Phi'[x_1, x' \mid f_1(\mathbf{q})]$ is the functional derivative of Φ defined in Eq. (1.2) taken at a point $x' \equiv (\mathbf{q}', \mathbf{p}')$ and evaluate for a local $f_1(\mathbf{q}, t)$.

Secondly, without introducing any assumption about the existence of a power series expansion in the density for $f_2[x_1, x_2 \mid f_1(\mathbf{q})]$, we proceed to solve Eq. (1.1) in the hydrodynamical stage for the gas. This stage is defined by assuming that f_1 is now a time-independent functional of the thermodynamic variables specifying the local equilibrium or nonperturbed state for the gas. These variables are chosen to be the local average particle density $n(\mathbf{q}, t)$, the hydrodynamic velocity $\mathbf{u}(\mathbf{q}, t)$ and the local temperature $\theta(q, t)$. Thus,

$$f_1(x_1, t) = f_1[x_1 \mid n(\mathbf{q}, t), \mathbf{u}(\mathbf{q}, t), \theta(\mathbf{q}, t)]. \quad (1.4)$$

The solution seeked is expressed, following the Chapman–Enskog method,[3] as a power series in a uniformity parameter which we call μ. This parameter is a measure of the macroscopic gradients in the system. Since we are dealing with a linear theory,

$$f_1 = f_1^{(0)} + \mu f_1^{(1)}, \quad (1.5)$$

where $f_1^{(0)}$ is the single-particle distribution function describing the nonperturbed state of the gas and $f_1^{(1)}$ is the function describing its perturbed state linear in the gradients.

Finally, one proceeds to determine $f_1^{(0)}$ and $f_1^{(1)}$ substituting Eq. (1.5) back into Eq. (1.1). The nonperturbed state is found to be defined by a local Maxwellian distribution function with five arbitrary parameters. However, since we want to describe such a state through local thermodynamics, one identifies these five parameters with the five thermodynamic quantities n, $\mathbf{u}$, and θ which are defined through the following relationships:

$$n(\mathbf{q}, t) = \int f_1^{(0)}\, d\mathbf{p}, \quad (1.6a)$$

$$\mathbf{u}(\mathbf{q}, t) = \frac{1}{n}\int f_1^{(0)} \frac{\mathbf{p}}{m}\, d\mathbf{p}, \quad (1.6b)$$

$$\tfrac{3}{2}n\theta(\mathbf{q}, t) = \int d\mathbf{p}\, \frac{\mathfrak{p}^2}{2m} f_1^{(0)}(\mathbf{p}), \quad (1.6c)$$

where $\mathfrak{p} = \mathbf{p} - m\mathbf{u}$ is the thermal momentum and $f_1^{(0)} = n(2\pi m\theta)^{-\frac{3}{2}} \exp(-\mathfrak{p}^2/2m\theta)$.

Therefore, the time rate of change of these variables needed to evaluate $\partial f_1^{(0)}/\partial t$ are given by the hydrodynamic equations linear in the gradients, i.e., Euler's equations. These equations are[1]

$$\partial n/\partial t = -\operatorname{div}(n\mathbf{u}), \quad (1.7a)$$

$$\partial \mathbf{u}/\partial t = -(1/nm)\operatorname{grad}\pi - \mathbf{u}\cdot\operatorname{grad}\mathbf{u}, \quad (1.7b)$$

$$\tfrac{3}{2}n(\partial\theta/\partial t) = -\tfrac{3}{2}n\mathbf{u}\cdot\operatorname{grad}\theta - (\pi - C'^{(0)})\operatorname{div}\mathbf{u} + R_1^{(0)}(\mathbf{q} \mid f_1^{(0)}), \quad (1.7c)$$

where π is the local equilibrium pressure and $C'^{(0)}$ and $R_1^{(0)}(\mathbf{q} \mid f_1^{(0)})$ are given by

$$C'^{(0)} = \frac{1}{3\theta}\iint d\mathbf{p}\, dx_2 \frac{\partial\phi}{\partial\mathbf{q}}\cdot\frac{(\mathbf{p}_2 - \mathbf{p})}{2m} \times \int dx'\, f_2'[x_1, x_2, x' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{p}^1\cdot(\mathbf{q}' - \mathbf{q}), \quad (1.8a)$$

$$R_1^{(0)}(\mathbf{q} \mid f_1^{(0)}) = \iint d\mathbf{p}\, dx_2 \frac{\partial\phi}{\partial\mathbf{q}}\cdot\frac{(\mathbf{p}_2 - \mathbf{p})}{2m} \times \int d\mathbf{p}'\, f_2'(x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q}))f_1^{(1)}(\mathbf{p}'). \quad (1.8b)$$

The determination of $f_1^{(1)}$ is now a matter of tedious algebra and it is given in I. The result is the following one, namely (the index k is introduced to distinguish our results from those obtained in Ref. 3. This distinction was not made in I),

$$f_1^{(1)}(p) = f_1^{(0)}(p)\varphi_k(p), \quad (1.9)$$

where $\varphi_k(p)$ satisfies the following linear inhomogeneous integral equation

$$\mathbf{G}_k(p)\cdot\frac{\partial \ln\theta}{\partial\mathbf{q}} + \mathbf{A}_k(p) : \frac{\partial\mathbf{u}}{\partial\mathbf{q}} + B_k(p)\operatorname{div}\mathbf{u} = \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})]f_1^{(0)}(p')\varphi_k(p')\, d\mathbf{p}' + \left(1 - \frac{\mathfrak{p}^2}{3m\theta}\right)\frac{f_1^{(0)}}{n\theta}\cdot\iint d\mathbf{p}\, dx_2 \frac{\partial\phi}{\partial\mathbf{q}}\cdot\frac{(\mathbf{p}_2 - \mathbf{p})}{2m} \times \int d\mathbf{p}'\, f_2'[x_1, x_2, x' \mid f_1^{(0)}(\mathbf{q})]f_1^{(0)}(p')\varphi_k(p'), \quad (1.10)$$

where

$$\mathbf{G}_k(p) = f_1^{(0)}\frac{\mathfrak{p}}{m}\left(\frac{\mathfrak{p}^2}{2m\theta} - \frac{3}{2} - \frac{\beta}{n\kappa}\right) - \int \Phi'[x_1, x' \mid f_1^{(0)}(\mathbf{q})](\mathbf{q}' - \mathbf{q})f_1^{(0)}(p') \times \left(\frac{\mathfrak{p}'^2}{2m\theta} - \frac{3}{2}\right)dx', \quad (1.11a)$$

$$\mathbf{A}_k(p) = f_1^{(0)}\frac{\mathbf{p}^0\mathbf{p}}{m\theta} - \frac{1}{\theta} \times \int \Phi'[x_1, x' \mid f_1^{(0)}(\mathbf{q})]\mathbf{S}'^0\mathbf{S}' f_1^{(0)}(p')\, dx', \quad (1.11b)$$

and

$$B_k(p) = f_1^{(0)} L_k - \frac{1}{3\theta} \times \int \Phi'[x_1, x' \mid f_1^{(0)}(\mathbf{q})] \mathfrak{p}' \cdot (\mathbf{q}' - \mathbf{q}) f_1^{(0)}(p')\, dx', \tag{1.11c}$$

where

$$\beta = \kappa(\partial\pi/\partial\theta)_n, \qquad \kappa = n^{-1}(\partial n/\partial\pi)_\theta, \tag{1.12}$$

$$\mathsf{P}^0\mathsf{P} = \mathfrak{p}\mathfrak{p} - \tfrac{1}{3}\mathfrak{p}^2 \mathsf{I}, \tag{1.13}$$

$$L_k = [1 - (\mathfrak{p}^2/3m\theta)][(\pi - n\theta - C'^{(0)})/n\theta], \tag{1.14}$$

I being the unit tensor, and $\mathsf{S}^0\mathsf{S}$ is the symmetric traceless part of the tensor $\mathfrak{p}'(\mathbf{q}' - \mathbf{q})$.

The subsidiary conditions which $\varphi_k(p)$ must obey follow directly from the general definition of n, $\mathbf{u}$, and θ^2 plus Eqs. (1.6). One finds indeed that

$$\int \begin{Bmatrix} 1 \\ \mathfrak{p} \\ \mathfrak{p}^2 \end{Bmatrix} f_1^{(0)}(p) \varphi_k(p)\, d\mathbf{p} = 0. \tag{1.15}$$

We now proceed to solve Eq. (1.10) subject to the conditions given by (1.15) without resorting to any density expansion.

II. THE SOLUTION TO THE INTEGRAL EQUATION

The integral equation satisfied by the perturbation function $\varphi_k(p)$ describing the nonequilibrium state of the gas, linear in the gradients, is given by Eq. (1.10). Except for the second term in the right-hand side, this equation is similar to the one discussed by García-Colín, Green and Chaos[3] in their treatment of the same problem. (We here after refer to this paper as II.) Therefore, we proceed to solve Eq. (1.10) in a rather intuitive way. In fact, we postulate a form for the perturbation function $\varphi_k(p)$ and then show that the resulting equation actually obeys the solubility conditions. Then let

$$\varphi_k(p') = \mathcal{G}_k(p') \mathfrak{p} \cdot \frac{\partial \ln \theta}{\partial \mathbf{q}} + \mathcal{A}_k(p) \mathsf{P}^0\mathsf{P} : \frac{\partial \mathbf{u}}{\partial \mathbf{q}} + \mathcal{B}_k(p) \operatorname{div} \mathbf{u}. \tag{2.1}$$

Substituting Eq. (2.1) into Eq. (1.10) and noticing that the last term in the right-hand side vanishes for the first two terms of Eq. (2.1), due to the isotropy of the fluid, we get the following equations for the functions $\mathcal{G}_k$, $\mathcal{A}_k$, and $\mathcal{B}_k$:

$$\mathbf{G}_k(\mathbf{p}) = \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p') \mathcal{G}_k(p') \mathfrak{p}'\, d\mathbf{p}', \tag{2.2a}$$

$$\mathsf{A}_k(p) = \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p') \mathcal{A}_k(p') \mathsf{P}'^0\mathsf{P}'\, d\mathbf{p}', \tag{2.2b}$$

$$B_k(p) = \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p') \mathcal{B}_k(p')\, d\mathbf{p}' + \left(1 - \frac{\mathfrak{p}^2}{3m\theta}\right) \cdot \frac{f_1^{(0)}}{n\theta} \iint d\mathbf{p}\, dx_2 \frac{\partial \phi}{\partial \mathbf{q}} \cdot \frac{(\mathbf{p}_2 - \mathbf{p})}{2m} \times \int d\mathbf{p}'\, f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p') \mathcal{B}_k(p'). \tag{2.2c}$$

Eqs. (2.2a) and (2.2b) are identical to the corresponding ones derived in II for the coefficients of the temperature and velocity gradients, respectively. Therefore, these equations satisfy the following properties:

(i). The right eigenfunctions with eigenvalue zero for the kernel $\Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')$ are 1, $\mathfrak{p}'$, and $\mathfrak{p}'^2$.

(ii). The left eigenfunctions with zero eigenvalue of this kernel are 1, $\mathfrak{p}$, and $E'[x' \mid f_1^{(0)}(\mathbf{q})]$, i.e., the functional derivative of the total energy evaluated for a local Maxwellian distribution function $f_1^{(0)}(\mathbf{q})$.

(iii). These left eigenfunctions with zero eigenvalue are orthogonal to the inhomogeneous parts of the equations.

These properties, which are proved in Appendix B of II, establish the existence of a solution of these integral equations and also, they show that the kernel is not a symmetric one.

However, Eq. (2.2c) is not the same as the corresponding equation in II for the coefficient of the divergence of $\mathbf{u}$. This equation is given by

$$L f_1^{(0)}(\mathbf{p}) - \frac{1}{3\theta} \int \Phi'[x_1, x' \mid f_1^{(0)}(\mathbf{q})] \mathfrak{p}' \cdot (\mathbf{q}' - \mathbf{q}) f_1^{(0)}(p')\, dx' = \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p') \mathcal{B}(p')\, d\mathbf{p}', \tag{2.3}$$

where

$$L = [(\mathfrak{p}^2/3m\theta) - 1][1 - \tfrac{2}{3}(\beta/n\kappa C_v)], \tag{2.4}$$

C_v being the specific heat per particle and β and κ being defined in Eq. (1.12). Comparing Eqs. (2.3) and (2.2c), we find that they differ in their inhomogeneous parts since L_k is not equal to L, and also, the second term in the homogeneous part of Eq. (2.2c) is missing from Eq. (2.3). Nevertheless, it is still possible to show that Eq. (2.2c) satisfies properties similar to (i), (ii), and (iii), mentioned above. To do so we first cast the second term in the right-

hand side of (2.2c) into another form. (The authors are indebted to Dr. M. H. Ernst for enlightening correspondence on this point.) Since $f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})]$ is symmetric in x_1 and x_2, i.e., it does not change under a permutation of the coordinates of particles 1 and 2, and also, since $\mathbf{r} = \mathbf{q}_2 - \mathbf{q}$,

$$(\partial\phi/\partial\mathbf{q})\cdot(\mathbf{p}_2 - \mathbf{p}) = (\partial\phi/\partial r)(\mathbf{p} - \mathbf{p}_2)$$

is also symmetric under this permutation. We thus have that

$$\iint d\mathbf{p}\, dx_2 \frac{\partial\phi}{\partial\mathbf{q}}\cdot(\mathfrak{p}_2 - \mathfrak{p}) \times \int d\mathbf{p}'\, f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{B}_k(p') = -\iint d\mathbf{p}\, dx_2 \frac{\mathfrak{p}_1^2}{2m}\frac{\partial\phi}{\partial r}\cdot\frac{\partial}{\partial\mathbf{p}} \times \int d\mathbf{p}'\, f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{B}_k(p') \quad (2.5)$$

after a partial integration with respect to $\mathbf{p}$ is performed. Furthermore,

$$\iint d\mathbf{p}\, dx_2 \frac{\mathfrak{p}_1^2}{2m}\frac{\partial\phi}{\partial r}\cdot\frac{\partial}{\partial\mathbf{p}_2} \times \int d\mathbf{p}'\, f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{B}_k(p') = 0,$$

so that adding this equation to Eq. (2.5) and using Eq. (1.3), one gets that the left-hand side of Eq. (2.5) equals

$$\int d\mathbf{p}\, \frac{\mathfrak{p}^2}{2m} \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{B}_k(p')\, d\mathbf{p}', \quad (2.6)$$

where use has been made of Eq. (1.2).

Using Eq. (2.6) we may write (2.2c) in the desired form, namely,

$$B_k(p) = \int \Phi'[x_1, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{B}_k(p')\, d\mathbf{p}' + \left(1 - \frac{\mathfrak{p}^2}{3m\theta}\right)\frac{f_1^{(0)}}{n\theta}\int d\mathbf{p}'\, \frac{\mathfrak{p}'^2}{2m} \times \int \Phi'[x', \mathbf{p}'' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p'')\mathfrak{B}_k(p'')\, d\mathbf{p}''. \quad (2.7)$$

It is now a trivial matter to show that Eq. (2.7) has 1, $\mathfrak{p}$, and $\mathfrak{p}^2$ as both left and right eigenfunctions with zero eigenvalue. Also, one may verify that these functions are also orthogonal to the inhomogeneous part $B_k(p)$. Therefore, the solubility conditions for Eq. (2.2c) or (2.7) are satisfied although they are not the same ones that those for Eqs. (2.2a), (2.2b), and (2.3). In fact the kernel of Eq. (2.7) *is symmetric*, whereas the kernel for the other equations is not.

Having established the existence of a solution to Eq. (1.10), namely, that expressed by Eq. (2.1), we must assure its uniqueness. In fact, since this solution is determined up to an arbitrary linear combination of the solutions to the homogeneous equation, say, $\alpha_1 + \alpha_2\cdot\mathfrak{p} + \alpha_3\mathfrak{p}^2$, we still have to determine the five arbitrary constants α_1, α_2, and α_3. This is accomplished through the five subsidiary conditions which $\varphi_k(p)$ has to satisfy, defined by Eq. (1.15). Substituting φ_k into these equations, we find that Eq. (2.1) is unique if the functions $\mathfrak{S}_k$ and $\mathfrak{B}_k$ satisfy the following conditions, namely,

$$\int \mathfrak{B}_k(p) f_1^{(0)}(p)\, d\mathbf{p} = 0, \quad (2.8)$$

$$\int \begin{Bmatrix}\mathfrak{B}_k(p)\\ \mathfrak{S}_k(p)\end{Bmatrix} \mathfrak{p}^2 f_1^{(0)}(p)\, d\mathbf{p} = 0, \quad (2.9)$$

with no condition imposed upon $\mathfrak{A}_k(p)$. Eqs. (2.1) together with Eqs. (2.8) and (2.9) determine $\varphi_k(p)$ uniquely, and this function can now be used to calculate the transport coefficients for the gas.

III. THE TRANSPORT COEFFICIENTS

The calculation of the transport coefficients for the gas proceeds in the usual way. One simply substitutes Eqs. (1.5), (1.9), and (2.1) into the well known expressions for the fluxes (heat current and stress tensor) for the system. This substitution has already been undertaken in II so that we shall not repeat it here. Furthermore, since the integral equations defining the functions $\mathfrak{S}_k$ and $\mathfrak{A}_k$ are identical to those obtained in II and since these functions determine the thermal conductivity and the shear viscosity, respectively, the expressions for these transport coefficients are those quoted in that paper.

Thus we see that different temperature definitions will give rise to different expressions for the bulk viscosity. This coefficient may be calculated with $\varphi_k(p)$ following exactly the same procedure as in II, the result being

$$\eta_2 = -\frac{1}{3m}\int d\mathbf{p}\, \mathfrak{p}^2 f_1^{(0)}(p)\begin{Bmatrix}\mathfrak{B}(p)\\ \mathfrak{B}_k(p)\end{Bmatrix} + \frac{1}{6}\iint dx_2\, d\mathbf{p}\, \gamma\phi'(r)\int d\mathbf{p}'\, f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})] \times f_1^{(0)}(p')\begin{Bmatrix}\mathfrak{B}(p')\\ \mathfrak{B}_k(p')\end{Bmatrix} + \frac{1}{18\theta}\iint dx_2\, d\mathbf{p}\, r\phi'(r) \times \int dx'\, f_2'[x_1, x_2, x' \mid f_1^{(0)}(\mathbf{q})] f_1^{(0)}(p')\mathfrak{p}'\cdot(\mathbf{q}' - \mathbf{q}), \quad (3.1)$$

where the value of η_2 calculated with $\mathcal{B}(p)$ corresponds to a local temperature defined through the energy density, as in II, and the value calculated with $\mathcal{B}_k(p)$ corresponds to the kinetic temperature.

It is interesting to compare what the difference is between the integral equations (2.2c) and (2.3), together with the subsidiary conditions for the functions $\mathcal{B}_k(p)$ and $\mathcal{B}(p)$.[4] This will allow us to establish the difference between the corresponding bulk viscosities. In the first place, one can show that the two integral equations are equivalent.[4] By this we mean that a particular solution of one of them is also a solution to the second one. This is shown in the Appendix. Therefore, we may conclude that the two bulk viscosities will be different because the functions $\mathcal{B}_k$ and $\mathcal{B}$ satisfy different subsidiary conditions. In fact, both functions satisfy Eq. (2.8) but only $\mathcal{B}_k$ satisfies Eq. (2.9). The corresponding condition for $\mathcal{B}(p)$ is given by

$$\int \mathcal{E}'[\mathbf{q}, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})]\mathcal{B}(p')f_1^{(0)}(p')\,d\mathbf{p}' = -\frac{1}{6\theta}\iint d\mathbf{p}\,dx_2\,\phi(r) \times \int dx'\, f_2'[x_1, x_2, x' \mid f_1^{(0)}(\mathbf{q})] \times \mathfrak{p}'\cdot(\mathbf{q}' - \mathbf{q})f_1^{(0)}(p'). \tag{3.2}$$

Further more, since the two integral equations are equivalent, the solutions will differ at most by a linear combination of the solutions to their homogeneous parts. Thus

$$\mathcal{B}(p) = \mathcal{B}_k(p) + C_1 + \mathbf{C}_2\cdot\mathfrak{p} + [(\mathfrak{p}^2/2m\theta) - \tfrac{3}{2}](C_3/\theta). \tag{3.3}$$

Since both $\mathcal{B}(p)$ and $\mathcal{B}_k(p)$ satisfy Eq. (2.8) and because of the scalar character of $\mathcal{B}(p)$, we immediately find that $C_1 = 0$, $\mathbf{C}_2 = 0$. To determine C_3 we subsitute $\mathcal{B}(p)$ into Eq. (3.3) and find the following result, namely,

$$C_3 = -n(\partial\theta/\partial\mathcal{E}^{(0)})_n\Gamma_k', \tag{3.4}$$

where

$$\Gamma_k' = \int \mathcal{E}'[\mathbf{q}, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})]\mathcal{B}_k(p')f_1^{(0)}(p')\,d\mathbf{p}' + \int \mathcal{E}'[\mathbf{q}, x' \mid f_1^{(0)}(\mathbf{p})]\frac{\mathfrak{p}'\cdot(\mathbf{q}' - \mathbf{q})}{3\theta}f_1^{(0)}(p')\,dx'. \tag{3.5}$$

Therefore

$$\mathcal{B}(p) = \mathcal{B}_k(p) - [(\mathfrak{p}'^2/2m\theta^2) - (3/2\theta)]n(\partial\theta/\partial\mathcal{E}^{(0)})_n\Gamma_k', \tag{3.6}$$

where we have called $\mathcal{E}^{(0)} = \mathcal{E}[\mathbf{q} \mid f_1^{(0)}(\mathbf{q})]$ [c.f., Eqs. (3.7c) and (3.11c) of II].

Substituting this result back into Eq. (3.1) and calling η_2 and $\eta_2^{(k)}$ the values for the bulk viscosity as computed with $\mathcal{B}(p)$ and $\mathcal{B}_k(p)$, respectively, we finally get that

$$\eta_2 = \eta_2^{(k)} - \Gamma_k'(\partial\pi/\partial\mathcal{E}^{(0)})_n, \tag{3.7}$$

which is the desired relationship between the two viscosities. This relationship has also been derived by Ernst.[4,5]

In conclusion, we may state that the integral equations defining the bulk viscosity are equivalent for the two different definitions of temperature, but the bulk viscosities differ themselves in the way indicated by Eq. (3.7). There is still the question of analyzing which temperature definition is the correct one but this will be done elsewhere.

ACKNOWLEDGMENT

The authors wish to thank Dr. M. H. Ernst for useful correspondence and for sending us preprints of his work prior to their publication.

APPENDIX

We want to show that if $\mathcal{B}(p)$ is a solution of the integral equation (2.3) then it also satisfies Eq. (2.2c). Let us then substitute $\mathcal{B}_k(p)$ by $\mathcal{B}(p)$ in this equation. Making use of the fact that $\mathcal{B}(p)$ is also a solution to Eq. (2.3), we get the following relationship:

$$B_k(p) = Lf_1^{(0)}(p) - \frac{1}{3\theta} \times \int \Phi'[x_1, x' \mid f_1^{(0)}(\mathbf{q})]\mathfrak{p}'\cdot(\mathbf{q}' - \mathbf{q})f_1^{(0)}(p')\,dx' + \left(1 - \frac{\mathfrak{p}^2}{3m\theta}\right)\frac{f_1^{(0)}}{n\theta}\iint d\mathbf{p}\,dx_2\,\frac{\partial\phi}{\partial\mathbf{q}}\cdot\frac{\mathfrak{p}_2 - \mathfrak{p}_1}{2m} \times \int d\mathbf{p}'\, f_2'[x_1, x_2, \mathbf{p}' \mid f_1^{(0)}(\mathbf{q})]f_1^{(0)}(p')\mathcal{B}(p'). \tag{A1}$$

Using Eqs. (2.6) and (2.3) once more, the last two terms in the right-hand side of Eq. (A.1) can be reduced to the following expression:

$$-Lf_1^{(0)}(p) - \left(1 - \frac{\mathfrak{p}^2}{3m\theta}\right)\frac{f_1^{(0)}}{n\theta}\int d\mathbf{p}'\,\frac{\mathfrak{p}'^2}{2m} \times \int \Phi'[x', x'' \mid f_1^{(0)}(\mathbf{q})]\frac{\mathfrak{p}''\cdot(\mathbf{q}'' - \mathbf{q})}{3\theta}f_1^{(0)}(p'')\,dx'',$$

[4] M. H. Ernst, "Temperature Definitions and Transport Coefficients," Preprint.

[5] M. H. Ernst, Ph.D. dissertation, University of Amsterdam (1965).

where use has been made of Eq. (2.4). Substituting this result back into Eq. (A1), we find that

$$f_1^{(0)}L_k = -\left(1 - \frac{\mathfrak{p}^2}{3m\theta}\right)\frac{f_1^{(0)}}{n\theta}\int d\mathfrak{p}'\,\frac{\mathfrak{p}'^2}{2m} \times \int \Phi'[x', x'' \mid f_1^{(0)}(\mathfrak{q})]\,\frac{\mathfrak{p}''\cdot(\mathfrak{q}'' - \mathfrak{q})}{3\theta}\,f_1^{(0)}(p'')\,dx''. \tag{A2}$$

To show that Eq. (A2) is an identity, we decompose $\Phi'[x', x'' \mid f_1^{(0)}(\mathfrak{q})]$ in its symmetric and antisymmetric parts,[4]

$$\Phi'[x', x'' \mid f_1^{(0)}(\mathfrak{q})] = \Phi_s'[x', x'' \mid f_1^{(0)}(\mathfrak{q})] + \Phi_a'[x', x'' \mid f_1^{(0)}(\mathfrak{q})], \tag{A3}$$

their explicit form being given in Ref. 4. There, it is also shown that

$$\int \Phi_a'[x', x'' \mid f_1^{(0)}(\mathfrak{q})]\,\frac{\mathfrak{p}''\cdot(\mathfrak{q}'' - \mathfrak{q})}{3\theta}\,f_1^{(0)}(p'')\,dx'' = -\left(\frac{\pi - n\theta}{n\theta}\right)\left(\frac{\mathfrak{p}'^2}{3m\theta} - 1\right)f_1^{(0)}(p'). \tag{A4}$$

On the other hand, using Eq. (1.8a), the fact that $(\partial\phi/\partial\mathfrak{q})\cdot(\mathfrak{p}_2 - \mathfrak{p})$ is invariant under the exchange of indices for particles 1 and 2, and an argument similar to the one leading to Eq. (2.6), it is easy to show that

$$C'^{(0)} = \int d\mathfrak{p}\,\frac{\mathfrak{p}^2}{2m}\int \Phi_s[x_1, x' \mid f_1^{(0)}(\mathfrak{q})] \times \frac{\mathfrak{p}'\cdot(\mathfrak{q}' - \mathfrak{q})}{3\theta}\,f_1^{(0)}(p')\,dx'. \tag{A5}$$

Using Eqs. (A4) and (A5), we get that

$$f_1^{(0)}L_k = [1 - (\mathfrak{p}^2/3m\theta)][(\pi - n\theta - C'^{(0)})/n\theta]f_1^{(0)}, \tag{A6}$$

where use has been made of the identity

$$\int d\mathfrak{p}\,f_1^{(0)}(p)\,\frac{\mathfrak{p}^2}{2m}\left(\frac{\mathfrak{p}^2}{3m\theta} - 1\right) = n\theta. \tag{A7}$$

Comparing Eqs. (A6) and (1.14), the definition of L_k, we get that indeed this relationship is an identity and therefore, $\mathfrak{G}(p)$ is a solution to Eq. (2.2c). Following a similar argument one can show that $\mathfrak{G}_k(p)$ is also a solution to Eq. (2.3), thus proving the equivalence between the two integral equations.

The Stability of Many-Particle Systems

MICHAEL E. FISHER* AND DAVID RUELLE
Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette (S. et O.) France
(Received 28 June 1965)

It is shown that a quantal or classical system of N particles of distinct species $\alpha, \beta = 1, 2, \cdots \mu$ interacting through pair potentials $\varphi_{\alpha\beta}(\mathbf{r})$ are *stable*, in the sense that the total energy is always bounded below by $-NB$, provided $\varphi_{\alpha\beta}(\mathbf{r})$ exceeds some $\varphi_{\alpha\beta}{}^{(2)}(\mathbf{r})$ whose Fourier transform $\hat{\varphi}_{\alpha\beta}(\mathbf{p})$ corresponds to a positive semidefinite $\mu \times \mu$ matrix for all $\mathbf{p}$.

This result is applied to discuss "charged" systems and stability is proved for Coulomb interactions if the charges are somewhat smeared rather than concentrated at points. For a large class of potentials it is shown that classical instability implies quantum instability in the case of bosons and, in three or more dimensions, also of fermions. Quantum systems with Coulomb interactions (point charges) are discussed and it is shown in particular that their stability cannot depend on the ratios between the masses of the particles.

I. INTRODUCTION

CONSIDER a classical system of N particles in a ν-dimensional space with total potential energy $U_N = U_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$, where $\mathbf{r}_i$ is the position vector of the i-th particle. In order that the system behave thermodynamically in the limit $N \to \infty$ it is natural to ask that the total configurational energy satisfy the *stability condition*

$$U_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \geq -NB \tag{I.1}$$

for all sets of $\mathbf{r}_i$, where B is a fixed bound. (Otherwise the energy per particle in the thermodynamic limit might not be bounded below.) With the aid of this condition, a further condition on the potentials at large particle separations and suitable restrictions on the shapes of domain containing the system, one can prove rigorously the existence of the thermodynamic limit for the canonical and grand canonical partition functions for both classical and quantum mechanical systems.[1-3]

Suppose the particles interact only through a pair potential $\varphi(\mathbf{r})$ so that

$$U_N = \sum_{i<j} \varphi(\mathbf{r}_j - \mathbf{r}_i). \tag{I.2}$$

It has then been shown[1-3] that stability is assured if the following conditions are satisfied:

(A) The pair potential can be decomposed as

$$\varphi(\mathbf{r}) = \varphi^{(1)}(\mathbf{r}) + \varphi^{(2)}(\mathbf{r}) \tag{I.3}$$

where $\varphi^{(1)}(\mathbf{r})$ may take the value $+\infty$ but is nonnegative, that is,

$$\varphi^{(1)}(\mathbf{r}) \geq 0 \tag{I.4}$$

and

$$\varphi^{(2)}(\mathbf{r}) = \int d\mathbf{p}\, e^{i\mathbf{p}\cdot\mathbf{r}} \hat{\varphi}^{(2)}(\mathbf{p}), \tag{I.5}$$

where the Fourier transform $\hat{\varphi}^{(2)}(\mathbf{p})$ is (absolutely) integrable and satisfies

$$\hat{\varphi}^{(2)}(\mathbf{p}) \geq 0, \tag{I.6}$$

so that, in other words, $\varphi^{(2)}(\mathbf{r})$ is of *positive type*.

With the aid of this theorem one can show[3,4] that the following simple conditions are sufficient for stability:

(B) for $r < a_1$ $\quad \varphi(\mathbf{r}) \geq C/r^{\nu+\epsilon}$, (I.7)

for $a_1 \leq r \leq a_2$ $\quad \varphi(\mathbf{r}) \geq -w$, (I.8)

for $r > a_2$ $\quad \varphi(\mathbf{r}) \geq -C'/r^{\nu+\epsilon'}$, (I.9)

where a_1, a_2, C, ϵ, w, C', and ϵ' are positive constants.

More recently, Dobrushin[5] has shown by an independent method that the following closely related but more general conditions, are also sufficient for stability:

(C) there are monotonic decreasing functions $\xi(r)$ and $\eta(r)$ such that for

$$r < a_1, \quad \varphi(\mathbf{r}) \geq \xi(r), \tag{I.10}$$

while

$$\int_0^{a_1} \xi(r) r^{\nu-1}\, dr = +\infty. \tag{I.11}$$

for $a_1 < r < a_2$, $\quad \varphi(\mathbf{r}) \geq -w$;

for $r > a_2$, $\quad \varphi(\mathbf{r}) \geq -\eta(r)$, (I.12)

* On leave from Wheatstone Physics Laboratory, King's College, London W.C.2, England.

[1] D. Ruelle, Helv. Phys. Acta **36**, 183 (1963).
[2] D. Ruelle, Helv. Phys. Acta **36**, 789 (1963).
[3] M. E. Fisher, Arch. Rat. Mech. Anal. **17**, 377 (1964).
[4] D. Ruelle, Lectures in Theoretical Physics, Vol. VI, Boulder 1963, pp. 93–95 (University of Colorado Press, 1964).
[5] R. L. Dobrushin, Th. Prob. Appl. (U.S.S.R.) **9**, 646 (1964).

while

$$\int_{a_0}^{\infty} \eta(r) r^{\nu-1}\, dr < \infty. \tag{I.13}$$

Dobrushin's proof is rather involved (and proceeds through a complicated inequality) so in Appendix A we present a proof of the sufficiency of conditions (C) which shows that they are, in fact, encompassed by the conditions (A).

Our main purpose in this paper, however, is to extend the conditions (A) to systems in which species of different particles interact with one another or in which the pair potentials depend on some internal—for example, orientational—coordinates. The principal result is to replace the nonnegativity of the Fourier transform $\hat{\varphi}(\mathbf{p})$, Eq. (I.6), by the positive-semidefiniteness of a corresponding *stability matrix* $\Phi(\mathbf{p}) = \hat{\varphi}_{\alpha\beta}(\mathbf{p})$. This theorem is applied to the discussion of charged systems interacting through Coulomb and more general forces. We show that such a system is stable if the charges are slightly "smeared" by some not too singular distribution, for example by a Yukawa-type function, or if the potential is cut off in some more drastic way at small r.

For quantum mechanical systems it seems likely that when account is taken of kinetic energy T_N, stability in the sense

$$\langle \mathcal{H}_N \rangle = \langle (T_N + U_N) \rangle \geq -NB \tag{I.14}$$

would be attained with purely Coulomb interactions. We have been unable to solve this challenging problem but we make some remarks on the relation between the stability of classical and quantal systems. We also show that the long-range part of the Coulomb potential does not cause instability and that stability cannot depend on the mass ratios of the differently charged species (as might perhaps be suggested by the "observed stability" of a system of hydrogen or deuterium atoms and the large ratio of nucleon to electron mass).

II. MULTISPECIES SYSTEMS

Consider a system of N particles made up of μ different species with N_1 particles of species 1, $\cdots$ N_α particles of species α, etc., so that

$$N_1 + N_2 + \cdots + N_\mu = N. \tag{II.2}$$

Let the position of the ith particle of species α be $\mathbf{r}_{i(\alpha)}$ $[i(\alpha) = 1, 2, \cdots N_\alpha]$ and suppose that the total configurational energy is given by

$$\begin{aligned} U_N(\mathbf{r}_{1(\alpha)} \cdots \mathbf{r}_{N_\mu(\mu)}) &= \sum_{\alpha=1}^{\mu} \sum_{i(\alpha)<j(\alpha)} \varphi_{\alpha\alpha}(\mathbf{r}_{j(\alpha)} - \mathbf{r}_{i(\alpha)}) \\ &+ \sum_{\alpha<\beta}^{\mu} \sum_{i(\alpha)} \sum_{j(\beta)} \varphi_{\alpha\beta}(\mathbf{r}_{j(\beta)} - \mathbf{r}_{i(\alpha)}) \end{aligned} \tag{II.2}$$

in which $\varphi_{\alpha\beta}(r)$ is the interaction potential between a particle of species α and one of species β and may take the value $+\infty$ as well as all real values.

We will not assume that the functions $\varphi_{\alpha\beta}(\mathbf{r})$ are rotationally symmetric. For instance we may suppose that the particles are asymmetric molecules but that the orientation of each one in space is held fixed as their positions vary. (Each different orientation may be considered as a distinct species.) It is obvious, however, that we should require

$$\varphi_{\alpha\beta}(\mathbf{r}) = \varphi_{\beta\alpha}(-\mathbf{r}) \tag{II.3}$$

for all α, β, and $\mathbf{r}$. Then we have

Theorem I. Let $\varphi_{\alpha\beta}(\mathbf{r}) = \varphi^{(1)}_{\alpha\beta}(\mathbf{r}) + \varphi^{(2)}_{\alpha\beta}(\mathbf{r})$ be a decomposition of the potentials respecting (II.3), i.e., such that

$$\varphi^{(2)}_{\alpha\beta}(\mathbf{r}) = \varphi^{(2)}_{\beta\alpha}(-\mathbf{r}) \quad \text{for all} \quad \alpha, \beta, \mathbf{r}, \tag{II.4}$$

and suppose

$$\varphi^{(2)}_{\alpha\beta}(\mathbf{r}) = \int d\mathbf{p}\, e^{i\mathbf{p}\cdot\mathbf{r}} \hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p}) \tag{II.5}$$

where the Fourier transform $\hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p})$ is absolutely integrable.

If $\varphi^{(1)}_{\alpha\beta}(\mathbf{r}) \geq 0$ for all α, β, $\mathbf{r}$ and if the $\mu \times \mu$ matrix $\Phi(p) = [\hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p})]$ has no negative eigenvalues for any value of $\mathbf{p}$ then the total potential energy, defined in (II.2), satisfies

$$U_N(\mathbf{r}_{1(\alpha)} \cdots \mathbf{r}_{N_\mu(\mu)}) \geq -\tfrac{1}{2} \sum_{\alpha=1} N_\alpha \varphi^{(2)}_{\alpha\alpha}(0). \tag{II.6}$$

To prove this theorem notice firstly that the matrix Φ is Hermitian because

$$\hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p}) = (2\pi)^{-\nu} \int d\mathbf{r}\, e^{-i\mathbf{p}\mathbf{r}} \varphi^{(2)}_{\alpha\beta}(\mathbf{r}),$$

and so by (II.4)

$$\begin{aligned} \hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p}) &= (2\pi)^{-\nu} \int d\mathbf{r}\, e^{-i\mathbf{p}\cdot\mathbf{r}} \varphi^{(2)}_{\beta\alpha}(-\mathbf{r}) \\ &= (2\pi)^{-\nu} \int d\mathbf{r}\, e^{i\mathbf{p}\cdot\mathbf{r}} \varphi^{(2)}_{\beta\alpha}(\mathbf{r}) = \hat{\varphi}^{(2)}_{\beta\alpha}(\mathbf{p})^*. \end{aligned} \tag{II.7}$$

Thus Φ has real eigenvalues and the statement that these are never negative is equivalent to the assertion that the Hermitian quadratic form $\mathbf{x}^\dagger \Phi \mathbf{x}$, where $\mathbf{x}$ is a $\mu \times 1$ column vector, is positive-

semidefinite. Secondly by (II.5) and the absolute integrability of $\hat{\varphi}^{(2)}_{\alpha\alpha}(\mathbf{p})$, it follows that $\varphi^{(2)}_{\alpha\alpha}(\mathbf{0})$ is finite for each α.

By definition (2.2) and the assumed positivity of the potentials $\varphi^{(1)}_{\alpha\beta}(\mathbf{r})$ we have

$$U_N \geq \sum_{\alpha=1}^{\mu} \sum_{i(\alpha)<j(\alpha)} \varphi^{(2)}_{\alpha\alpha}(\mathbf{r}_{j(\alpha)} - \mathbf{r}_{i(\alpha)}) + \sum_{\alpha<\beta}^{\mu} \sum_{i(\alpha)} \sum_{j(\beta)} \varphi^{(2)}_{\alpha\beta}(\mathbf{r}_{j(\beta)} - \mathbf{r}_{i(\alpha)}) \geq \tfrac{1}{2}W_N - \frac{1}{2}\sum_{\alpha=1}^{\mu} N_\alpha \varphi^{(2)}_{\alpha\alpha}(\mathbf{0}) \tag{II.8}$$

where

$$W_N = \sum_{\alpha=1}^{\mu}\sum_{\beta=1}^{\mu}\sum_{i(\alpha)=1}^{N}\sum_{j(\beta)=1}^{N} \varphi^{(2)}_{\alpha\beta}(r_{j(\beta)} - r_{i(\alpha)}). \tag{II.9}$$

Introducing the Fourier transform through (II.5) yields

$$W_N = \sum_{\alpha=1}^{\mu}\sum_{\beta=1}^{\mu}\sum_{i(\alpha)}\sum_{j(\beta)} \int d\mathbf{p} \times \exp\left[i\mathbf{p}\cdot(\mathbf{r}_{j(\beta)} - \mathbf{r}_{i(\alpha)})\right]\hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p}) = \int d\mathbf{p} \sum_{\alpha=1}^{\mu}\sum_{\beta=1}^{\mu}\left[\sum_{i(\alpha)} \exp\left(i\mathbf{p}\cdot\mathbf{r}_{i(\alpha)}\right)\right]^* \hat{\varphi}^{(2)}_{\alpha\beta}(\mathbf{p}) \times \left[\sum_{j(\beta)} \exp\left(i\mathbf{p}\cdot\mathbf{r}_{j(\beta)}\right)\right].$$

On defining the column vector

$$\hat{\mathbf{n}}(\mathbf{p}) = \left[\sum_{i(\alpha)=1}^{N} \exp\left(i\mathbf{p}\cdot\mathbf{r}_{i(\alpha)}\right)\right], \tag{II.10}$$

this may be written simply as

$$W_N = \int d\mathbf{p}\, \hat{\mathbf{n}}(\mathbf{p})^\dagger \mathbf{\Phi}(\mathbf{p})\hat{\mathbf{n}}(\mathbf{p}). \tag{II.11}$$

Since the integrand is a nonnegative quadratic form for all $\mathbf{p}$ we have $W_N \geq 0$ and the theorem is proved.

Remarks. The theorem may obviously be extended to the case where, in addition to two-body potentials, U_N contains nonnegative many-body potentials. The theorem applies even when $N_\alpha = 1$ (all α) so that each particle belongs to a distinct species; but in that case the "self interaction" potential $\varphi_{\alpha\alpha}(r)$ may be chosen arbitrarily.

III. CHARGED SYSTEMS

Suppose the particles of the system are "charged" so that each particle of species α carries a charge q_α and the interactions are given by

$$\varphi_{\alpha\beta}(\mathbf{r}) = q_\alpha q_\beta \chi(\mathbf{r}) \tag{III.1}$$

where $\chi(\mathbf{r})$ is a fixed "shape factor" satisfying

$$\chi(\mathbf{r}) = \chi(-\mathbf{r}) \quad \chi(\mathbf{r}) \to 0 \text{ as } r \to \infty. \tag{III.2}$$

We have principally in mind, of course, Coulomb systems for which $\chi(r) = 1/r$, screened Coulomb or Yukawa systems with $\chi(\mathbf{r}) = e^{-\kappa r}/r$, etc. The charges q_α may be positive, negative, or zero. [Complex charges $q_\alpha = q'_\alpha + iq''_\alpha$ may be included equally if the interactions are taken proportional to $\frac{1}{2}(q^*_\alpha q_\beta + q_\alpha q^*_\beta) = (q'_\alpha q'_\beta + q''_\alpha q''_\beta)$.]

To test the stability of such a system consider the stability matrix

$$\mathbf{\Phi}(\mathbf{p}) = [q_\alpha q_\beta \hat{\chi}(\mathbf{p})] \tag{III.3}$$

where $\hat{\chi}(\mathbf{p})$ is the Fourier transform of $\chi(\mathbf{r})$. Since the rows of the matrix are proportional to one another it may be factorized as

$$\mathbf{\Phi} = \hat{\chi}(p)\mathbf{q}\mathbf{q}^\dagger \quad \text{where} \quad \mathbf{q} = [q_\alpha]. \tag{III.4}$$

It follows that $\mathbf{\Phi}$ has one eigenvalue

$$\lambda_1(\mathbf{p}) = \hat{\chi}(\mathbf{p}) \sum_{\alpha=1}^{\mu} |q_\alpha|^2 \quad \text{and } \mu - 1 \text{ zero eigenvalues.}$$

Consequently the conditions of Theorem I will be satisfied (with $\varphi^{(1)}_{ij} \equiv 0$) if

$$\hat{\chi}(\mathbf{p}) = (2\pi)^{-\nu} \int d\mathbf{r}\, e^{-i\mathbf{p}\cdot\mathbf{r}}\chi(\mathbf{r}) \geq 0 \quad \text{all } \mathbf{p} \tag{III.5}$$

and

$$\chi(\mathbf{0}) = \int d\mathbf{p}\, \hat{\chi}(p) < \infty. \tag{III.6}$$

In that case we have

$$U_N \geq -\tfrac{1}{2}\chi(\mathbf{0}) \sum_{\alpha=1}^{N\mu} |q_\alpha|^2. \tag{III.7}$$

For the Coulomb and Yukawa potentials in ν dimensions we have

$$\hat{\chi}(p) = C_\nu/p^2, \qquad C_\nu/(p^2 + \kappa^2), \tag{III.8}$$

respectively, where the C_ν are constants ($C_3 = \frac{1}{2}\pi^{-2}$). When we try to apply the above result we find that (III.5) is satisfied, but that for $\nu \geq 2$ the condition (III.6) is violated owing to a divergence of the integral at the upper limit ($p \to \infty$)[6]. This corresponds, of course, simply to the divergence of the potential itself as $r \to 0$. Accordingly let us suppose more generally (and more realistically!) that the charges are distributed rather than concentrated

[6] For $\nu \leq 2$ the integral (3.6) for the pure Coulomb case would diverge also at its lower limit owing to the divergence of the potential as $r \to \infty$ in violation of (3.2). This prevents the unambiguous determination of the zero of potential energy and is the reason for restricting attention to potentials which converge to zero at infinity, however slowly.

at points. If $\rho_\alpha(\mathbf{r})$ is the charge density of a particle of the α species with respect to its position as origin, the interaction between particles becomes

$$\varphi_{\alpha\beta}(\mathbf{r}) = \int d\mathbf{x}_\alpha \int d\mathbf{x}_\beta \, \rho_\alpha(\mathbf{x}_\alpha) \times \rho_\beta(\mathbf{x}_\beta)\chi(\mathbf{r} + \mathbf{x}_\beta - \mathbf{x}_\alpha). \quad \text{(III.9)}$$

We assume that the charge distributions of the particles are not perturbed by their mutual interaction but rather are "frozen" so that, in particular, Van der Waals or dispersion forces cannot arise. Permanent charges (ions) and dipoles are, however, adequately represented.

Introducing the Fourier transforms of the densities by

$$\hat{\rho}_\alpha(\mathbf{p}) = \int d\mathbf{r}\, e^{-i\mathbf{p}\cdot\mathbf{r}}\rho_\alpha(\mathbf{r}) \quad \text{(III.10)}$$

and noting that $\hat{\rho}_\alpha(-\mathbf{p}) = \hat{\rho}_\alpha(\mathbf{p})^*$ since $\rho_\alpha(\mathbf{r})$ is real, gives

$$\hat{\rho}_{\alpha\beta}(\mathbf{p}) = \rho_\alpha(\mathbf{p})\rho_\beta(\mathbf{p})^*\hat{\chi}(\mathbf{p}). \quad \text{(III.11)}$$

We thus, as before, find that the matrix Φ has only the single nonzero eigenvalue $\lambda_1(\mathbf{p}) = \hat{\chi}(\mathbf{p}) \sum_\alpha |\hat{\rho}_\alpha(\mathbf{p})|^2$. For stability we require again that $\hat{\chi}(p)$ is nonnegative, but also, in place of (III.6), that

$$\varphi_{\alpha\alpha}(\mathbf{0}) = \int d\mathbf{p}\, |\hat{\rho}_\alpha(\mathbf{p})|^2\hat{\chi}(\mathbf{p}) < \infty \quad \text{(III.12)}$$

which justifies (III.11). From Theorem I we then obtain

$$U_N \geq -\sum_{\alpha=1}^{\mu} N_\alpha E_\alpha \quad \text{(III.13)}$$

where

$$E_\alpha = \tfrac{1}{2}\varphi_{\alpha\alpha}(\mathbf{0}) = \frac{1}{2}\int d\mathbf{x} \int d\mathbf{x}' \times \rho_\alpha(\mathbf{x})\rho_\alpha(\mathbf{x}')\chi(\mathbf{x}' - \mathbf{x}). \quad \text{(III.14)}$$

Evidently E_α is just the self-energy of the charge distribution $\rho_\alpha(\mathbf{r})$ (which might even have zero total charge). Notice that if $\chi(\mathbf{r})$ is spherically symmetric and $\rho_\alpha(\mathbf{r})$ and $\rho_\beta(\mathbf{r})$ differ only by a rotation then $E_\alpha = E_\beta$. Furthermore if we add some positive potentials to the charge interactions, (III.13) remains unmodified. Accordingly we have proved the following result.

Theorem II. Suppose $U_N = U_N^{(1)} + U_N^{(2)}$ where $U_N^{(1)}$ is positive and $U_N^{(2)}$ is the potential energy of N_1 particles of species 1 with charge distribution $\rho_1(\mathbf{r})$, $\cdots$, N_μ particles of species μ with charge distribution $\rho_\mu(\mathbf{r})$, where point charges q and q' interact with a pair potential $qq'\,\chi(\mathbf{r}) = qq'\,\chi(-\mathbf{r})$ tending to zero as $r \to \infty$. [If $\chi(\mathbf{r})$ is spherically symmetric the distributions $\rho_\alpha(\mathbf{r})$ may be identified up to a rotation.] Then

$$U_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \geq -\sum_{\alpha=1}^{\mu} N_\alpha E_\alpha \quad \text{(III.15)}$$

provided the self-energy of a particle of species α,

$$E_\alpha = \frac{1}{2}\int d\mathbf{x} \int d\mathbf{x}'\, \rho_\alpha(\mathbf{x})\rho_\alpha(\mathbf{x}')(\mathbf{x}' - \mathbf{x}),$$

$$= \frac{1}{2}\int d\mathbf{p}\, |\hat{\rho}_\alpha(\mathbf{p})|^2\,\hat{\chi}(\mathbf{p}), \quad \text{(III.16)}$$

where $\hat{\rho}_\alpha(\mathbf{p})$ and $\hat{\chi}(\mathbf{p})$, are the Fourier transforms of $\rho_\alpha(\mathbf{r})$ and $\chi(\mathbf{r})$ is finite for all α.

Remark. For pure Coulomb forces another derivation of the theorem is as follows.[7,8] Let $\phi(\mathbf{R})$ be the (total) electrostatic potential due to a charge distribution $\rho(\mathbf{R})$. Then it is well known that the electrostatic self-energy of $\rho(\mathbf{R})$ may be expressed in terms of the electric field $\mathcal{E} = \nabla\phi$ since

$$\frac{1}{2}\int d\mathbf{R} \int d\mathbf{R}' \frac{\rho(\mathbf{R})\rho(\mathbf{R}')}{|\mathbf{R}' - \mathbf{R}|} = \frac{1}{2}\int d\mathbf{R}\, \rho(\mathbf{R})\phi(\mathbf{R}) \quad \text{(III.17)}$$

and so by Poisson's equation (taking $\nu = 3$)

$$= -(8\pi)^{-1}\int d\mathbf{R}\, [\nabla^2\phi(\mathbf{R})]\phi(\mathbf{R})$$

$$= (8\pi)^{-1}\int d\mathbf{R}\, [\nabla\phi(\mathbf{R})]^2,$$

$$= (8\pi)^{-1}\int d\mathbf{R}\, \mathcal{E}^2. \quad \text{(III.18)}$$

Now if $\phi_i(R)$ is the electrostatic field due to a distribution of charge $\rho_i(\mathbf{R} - \mathbf{r}_i)$ centered at $\mathbf{r}_i$ we may write, distinguishing between all particles,

$$U_N^{(2)} = \sum_{i<j} \varphi_{ij}^{(2)}(\mathbf{r}_j - \mathbf{r}_i)$$

$$= \tfrac{1}{2}\sum_i \sum_j \varphi_{ij}^{(2)}(\mathbf{r}_j - \mathbf{r}_i) - \tfrac{1}{2}\sum_i \varphi_{ii}^{(2)}(\mathbf{0})$$

$$= \frac{1}{2}\int d\mathbf{R} \int d\mathbf{R}' \,[\sum_i \rho_i(\mathbf{R} - \mathbf{r}_i)] \times [\sum_i \rho_i(\mathbf{R} - \mathbf{r}_i)]/|\mathbf{R}' - \mathbf{R}|$$

$$- \sum_{i=1}^{N} \frac{1}{2}\int d\mathbf{x} \int d\mathbf{x}'\, \rho_i(\mathbf{x})\rho_i(\mathbf{x}')/|\mathbf{x}' - \mathbf{x}|$$

[7] The stability of an electrostatic system seems first to have been considered by L. Onsager, J. Phys. Chem. **43**, 189 (1939) who supposed the particles interacted in addition with an infinite hard core (see also below).

[8] A proof of stability somewhat similar to that presented here has been communicated privately (to M.E.F.) by O. Penrose.

and so by (III.17)

$$U_N^{(2)} = (8\pi)^{-1} \int d\mathbf{R}\, [\nabla \sum_i \phi_i(R)]^2 - \sum_{i=1}^{N} (8\pi)^{-1} \int d\mathbf{x}\, [\nabla\phi_i(\mathbf{x})]^2 \geq -\sum_{i=1}^{N} (8\pi)^{-1} \int d\mathbf{x}\, [\nabla\phi_i(\mathbf{x})]^2. \quad \text{(III.19)}$$

This is just (III.15) with another expression for the electrostatic self-energy, namely,

$$E_i = (8\pi)^{-1} \int dx\, [\nabla\phi_i(\mathbf{x})]^2. \quad \text{(III.20)}$$

Applications

The condition that the self-energy (III.16) be finite means physically that the density $\rho_\alpha(\mathbf{r})$ must be sufficiently "smooth" in relation to the shape factor $\chi(\mathbf{r})$. It may, nonetheless, be quite singular. As an example consider the Coulomb potential in three dimensions. We will show that stability is assured if the charge is distributed over the surface of a sphere so that

$$\rho_\alpha(\mathbf{r}) = \sigma(\theta, \psi)\delta(r - a) \quad \text{(III.21)}$$

where r, θ, and ψ are polar coordinates, provided the surface density is bounded, say by σ_0. For by (III.16)

$$|E_\alpha| \leq \frac{1}{2} \int d\mathbf{x} \int d\mathbf{x}'\, \delta(x - a)\, |\sigma(\theta, \psi)| \times \delta(x' - a)\, |\sigma(\theta', \psi')| \cdot |x' - x|^{-1} \leq \tfrac{1}{2}\sigma_0 \int d\mathbf{x}\, \delta(x - a)\phi_0(\mathbf{x}), \quad \text{(III.22)}$$

where $\phi_0(\mathbf{x})$ is simply the potential due to a uniform surface distribution of total charge $4\pi a^2\sigma_0$ on a sphere of radius a. This has the finite value $4\pi a^2\sigma_0/a$ *on* the sphere so that

$$E_\alpha \leq 8\pi^2\sigma_0^2 a^3. \quad \text{(III.23)}$$

With a surface distribution such as (III.21) one may, for example, reproduce outside the sphere the field of a point dipole.

By the same argument it is clear that stability will be obtained for any volume distribution of bounded density, vanishing outside a bounded region. In fact one may even allow singular distributions, such as the Yukawa or its square, provided the divergence is not worse than $1/|\mathbf{r} - \mathbf{r}_0|^\delta$ with $\delta < \frac{5}{2}$ (or, in ν dimensions, $\delta < \frac{1}{2}\nu + 1$). This may be seen by studying the Fourier transforms for large p. On the other hand it is easily seen that a linear distribution of charge does not suffice for stability.

Let us further establish stability for truncated Coulomb interactions defined (for $\nu = 3$) by

$$\varphi_{\alpha\beta}(r) \geq \sup\,(q_\alpha q_\beta/|\mathbf{r}|, -|q_\alpha q_\beta|/a) \quad \text{(III.24)}$$

for all α, β, $\mathbf{r}$ and some fixed a. A simple example is provided by supposing the particles have hard cores of diameter a in addition to pure Coulomb interactions.[7] Consider the Coulomb interaction $\varphi^{(2)}_{\alpha\beta}(\mathbf{r})$ between two charges q_α and q_β each distributed uniformly through a sphere of diameter a (and hence radius $\frac{1}{2}a$). Since the convolution of two distributions vanishing outside a sphere of radius $\frac{1}{2}a$, vanishes outside a sphere of radius a, this interaction satisfies

$$\varphi^{(2)}_{\alpha\beta}(\mathbf{r}) = |q_\alpha q_\beta|/|\mathbf{r}|, \quad \text{for} \quad |\mathbf{r}| \geq a \quad \text{(III.25)}$$

and

$$|\varphi^{(2)}_{\alpha\beta}(r)| \geq |q_\alpha q_\beta|/a, \quad \text{for} \quad |\mathbf{r}| \geq a. \quad \text{(III.26)}$$

Thus we have a stable potential always lying below the truncated Coulomb potential (III.24) which is hence also stable.

Finally consider a system of point dipoles interacting through the usual dipole–dipole (or tensor) forces, namely,

$$\varphi_{\alpha\beta}(\mathbf{r}; \mathbf{m}_\alpha, \mathbf{m}_\beta) = \frac{\mathbf{m}_\alpha \cdot \mathbf{m}_\beta}{\mathbf{r}^3} - 3\,\frac{(\mathbf{m}_\alpha \cdot \mathbf{r})(\mathbf{m}_\beta \cdot \mathbf{r})}{r^5}, \quad \text{(III.27)}$$

where $\mathbf{m}_\alpha$ and $\mathbf{m}_\beta$ are the dipole moments. The Fourier transform is

$$\hat{\varphi}_{\alpha\beta}(p; \mathbf{m}_\alpha, \mathbf{m}_\beta) = C_3(\mathbf{m}_\alpha \cdot \mathbf{p})/|\mathbf{p}|^2, \quad \text{(III.29)}$$

so that, as before, the stability matrix Φ will be nonnegative-definite for all $\mathbf{m}_\alpha$, $\mathbf{m}_\beta$. As it stands, however, the self-energy is divergent but we may clearly obtain a stable system if the dipole moment $\mathbf{m}_\alpha$ is distributed with some density $\boldsymbol{\mu}_\alpha(\mathbf{r})$. The smoothness conditions on $\boldsymbol{\mu}_\alpha(\mathbf{r})$ are, however, more stringent than on the charge density as is natural since a distribution $\rho_\alpha(\mathbf{r})$ proportional to the divergence of $\boldsymbol{\mu}(\mathbf{r})$ will have the same electrostatic field. A Yukawa distribution of dipole moments would yield stability but a surface distribution on a sphere would not.

IV. QUANTUM MECHANICAL SYSTEMS

For the stability of a quantum mechanical system of N particles of masses m_i it is sufficient to require only that

$$\langle \mathcal{H}_N \rangle = \sum_{i=1}^{N} (\hbar^2/2m_i) \int |\nabla_i \Psi_N|^2 \, d\mathbf{r}_1 \cdots d\mathbf{r}_N + \int U_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \, |\Psi_N|^2 \, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \geq -NB, \tag{IV.1}$$

for fixed B and all N-body functions $\Psi_N = \Psi_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ in the domain of $\mathcal{H}_N$. (We assume the wavefunction vanishes on the boundary of the domain and on any hard cores.[2]) Since the kinetic energy is evidently positive for any Ψ_N the stability of a classical system, that is the assertion $U_N \geq -NB$ for all $\mathbf{r}_i$ and N, implies the stability of the corresponding quantal systems. It is of interest to inquire, however, into potentials that might be stable for quantum systems of given statistics while being unstable for classical systems or for different statistics. In particular, the experimentally observed stability of systems of electrons and protons (hydrogen) or electrons and deuterons (deuterium) suggest strongly that a charged quantum mechanical system should be stable even with *pure* Coulomb forces. We take this specific question up in Sec. 5 and consider firstly the converse general problem of establishing instability.

For a classical system of identical particles, one may show that certain potentials are *catastrophically unstable* in the sense that for N indefinitely large there are sets of configurations for which

$$U_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \leq -w^* N^{1+\eta}, \tag{IV.2}$$

where w^* and η are positive constants, and that the grand canonical partition function does not exist (the series being divergent). In particular for pair interactions this is so if[9]:

(D) $\varphi(\mathbf{r})$ is finite-valued and piecewise continuous,[10] and for some k and configuration $\mathbf{r}_i$ one has

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \varphi(\mathbf{r}_j - \mathbf{r}_i) < 0. \tag{IV.3}$$

It is quite straightforward to generalize the proof of this result[9] to a multi-species system provided that N_α/N remains bounded below for all α as $N = \sum_\alpha N_\alpha \to \infty$.

For a quantum mechanical system we may prove,

Theorem III. If the potential energy of a quantum mechanical system of N identical particles is the sum of pairwise interactions with potential $\varphi(\mathbf{r})$ which is (i) bounded for large $|\mathbf{r}|$, (ii) finite-valued upper semicontinuous,[11] and (iii) which for some k and $\mathbf{r}_i^0 (i = 1, 2, \cdots k)$ satisfies

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \varphi(\mathbf{r}_j^0 - \mathbf{r}_i^0) = -U_0 < 0, \tag{IV.4}$$

then for a sufficiently large domain there is a $w^* > 0$ and an N^* such that the ground-state energy satisfies

$$E_0(N) \leq -w^* N^2 \quad \text{for} \quad N \geq N^*, \tag{IV.5}$$

provided either (a) the particles obey Bose–Einstein or Boltzmann statistics (and ν is arbitrary) or (b) they obey Fermi–Dirac statistics and $\nu \geq 3$ (or $\nu = 2$ and U_0 is sufficiently large).

Remarks. When the conditions of the theorem are satisfied it is evident that the canonical free energy per particle cannot approach a finite thermodynamic limit since, in a diagonal representation, the canonical partition function contains a term $\exp[-\beta E_0(N)] = \exp(\beta W^* N^2)$. For the same reason the grand canonical partition function does not exist.

Notice that for Bose–Einstein and Boltzmann statistics the theorem is as strong as in the classical case. For fermions on the other hand the theorem proves instability for all U_0 only in three or more dimensions. In two dimensions the system will be unstable if U_0 is sufficiently large [or more generally if (IV.4) holds as an inequality over a sufficiently large region of configuration space] but might perhaps be stable otherwise. Indeed the arguments in the proof, which depend on the way the total kinetic energy of a system confined in a domain increases with N, suggest that one and two-dimensional Fermi–Dirac systems probably are stable for certain forces that would be classically unstable. We have however, not established any such counterexamples.

Theorem III is proved by constructing a trial wavefunction Ψ_N, and hence a variational upper bound for $E_0(N)$, which corresponds to superimposing closely many replicas of the configuration satisfying (IV.4). As in the classical case there

[9] D. Ruelle, Ref. 4 pp. 86–88.

[10] At a discontinuity $\varphi(\mathbf{r}_0)$ should be assigned the value $\limsup \varphi(\mathbf{r})$ as $\mathbf{r} \to \mathbf{r}_0$. More generally, see Theorem III, we only require *upper semi-continuity* which means that for any $\mathbf{r}_0$ and $\epsilon > 0$ there is a $\delta > 0$ such that $\varphi(\mathbf{r}) < \varphi(\mathbf{r}_0) + \epsilon$ if $|\mathbf{r} - \mathbf{r}_0| < \delta$.

[11] See footnote 10.

are no obstacles to generalizing the results to multispecies systems if N_α/N is bounded below. The proof occupies the remainder of this section; the reader uninterested in the technical details is advised to proceed directly to Sec. V.

Proof. We first observe, following the classical argument,[9] that the function

$$U(\mathbf{r}_1 \cdots \mathbf{r}_k; \mathbf{r}'_1 \cdots \mathbf{r}'_k) = \sum_{i=1}^{k} \sum_{j=1}^{k} \varphi(\mathbf{r}'_j - \mathbf{r}_i) \qquad \text{(IV.6)}$$

is upper semicontinuous in the space of the $2k\nu$ coordinates $\mathbf{r}_1 = (x_{1,1}, \cdots x_{\nu,1})$ to $\mathbf{r}'_k = (x'_{1,k}, \cdots x'_{\nu,k})$. Consequently there exists a positive length d and a set of k cubes Γ^0_i of edge d_i namely

$$\Gamma^0_i : \quad 0 \leq x_\gamma - x^0_{\gamma,i} \leq d, \quad (\gamma = i, 2, \cdots \nu), \qquad \text{(IV.7)}$$

such that

$$U(\mathbf{r}_1, \cdots, \mathbf{r}'_k) < -\tfrac{1}{2}U_0 \qquad \text{for} \quad \mathbf{r}_i \in \Gamma^0_i \quad \text{and} \quad \mathbf{r}'_j \in \Gamma^0_j. \qquad \text{(IV.8)}$$

Now if $N = hk + c$ where h is an integer and $0 \leq c < k$, we may write the total potential energy as

$$U_N = \frac{1}{2} \sum_{f=1}^{h} \sum_{g=1}^{h} \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} \varphi(\mathbf{r}_{gk+j} - \mathbf{r}_{fk+i}) \right\} - \tfrac{1}{2}hk\varphi(\mathbf{0}) + \sum_{j=1}^{c} \sum_{i=1}^{N} \varphi(\mathbf{r}_i - \mathbf{r}_{hk+j}) - c\varphi(\mathbf{0}). \qquad \text{(IV.9)}$$

If the coordinates of k of the first h k particles lie within each cube Γ^0_i, that is

$$0 \leq x_{\gamma,fk+i} - x^0_{\gamma,i} \leq d \quad \text{all} \quad \gamma, f, i, \qquad \text{(IV.10)}$$

then we have by (IV.6) and (IV.8)

$$U_N \leq -\tfrac{1}{4}h^2U_0 - \tfrac{1}{2}hk\varphi(0) + Y_{N,c} - c\varphi(0) \qquad \text{(IV.11)}$$

where $Y_{N,c}$ denotes the penultimate term in (IV.9). Since $\varphi(r)$ is bounded for large $|\mathbf{r}|$ there is a distance a_2 and a fixed η such that

$$|\varphi(\mathbf{r})| < \eta, \quad \text{for} \quad |\mathbf{r}| \geq a_2. \qquad \text{(IV.12)}$$

The k cubes Γ^0_i are fixed and so in a sufficiently large domain we can find c further similar cubes Γ^1_j ($j = 1, \cdots c$) such that if $\mathbf{r}$ is in cube Γ^1_j and r' is in *any* other cube then $|\mathbf{r}' - \mathbf{r}| > a_2$, i.e., the mutual distances exceed a_2. Let Γ^1_j be defined by $0 \leq x_\gamma - x'_{\gamma,j} \leq d$ for all γ. If we impose

$$0 \leq x_{\gamma,hk+j} - x^1_{\gamma,j} \leq d \quad \text{all} \quad \gamma, j = 1, \cdots c, \qquad \text{(IV.13)}$$

we therefore have

$$Y_{N,c} \leq Nc\eta \leq Nk\eta. \qquad \text{(IV.14)}$$

Thus under conditions (IV.10) and (IV.13) we may write (for $N \geq k$)

$$U_N \leq -(N-k)^2(U_0/4k^2) + N(\tfrac{1}{2}|\varphi(0)| + k\eta) + \tfrac{1}{2}k|\varphi(\mathbf{0})|, \qquad \text{(IV.15)}$$

which in fact restablishes instability for the classical case under slightly wider conditions.

Let $\xi_\gamma = x_\gamma - x^0_{\gamma,i}$ or $x_\gamma - x^1_{\gamma,j}$ and consider the single-particle wavefunction which vanishes outside the cube Γ: $0 \leq \xi_\gamma \leq d$ (all γ) but is given internally by

$$\psi_{l_1 \cdots l_\nu}(\xi_1, \cdots \xi_\nu) = (\tfrac{1}{2}d)^{-\frac{1}{2}\nu} \prod_{\gamma=1}^{\nu} \sin(l_\gamma \pi \xi_\gamma / d) \qquad \text{(IV.16)}$$

where the l_γ are positive integers. For this wavefunction the kinetic energy has the expectation value

$$t_{l_1 \cdots l_\nu} = (\hbar^2\pi^2/2md^2)(l_1^2 + \cdots + l_\nu^2). \qquad \text{(IV.17)}$$

For the Bose-Einstein and Boltzmann case (a) now take as a trial wavefunction a product of N single particle functions $\psi_{11\ldots1}$ with h functions based on each of the cubes Γ^0_i and one based on each of the Γ^1_j. For this Ψ_N we clearly have

$$\langle T_N \rangle = N(\nu\hbar^2\pi^2/2md^2), \qquad \text{(IV.18)}$$

while $\langle U_N \rangle$ satisfies the inequality (IV.15). By the variational principle the sum of these terms exceeds the ground-state energy $E_0(N)$ and so (IV.5) follows.

For the case (b) of Fermi–Dirac statistics we can allow only totally antisymmetric trial wave functions. Thus in each cube Γ^0_i take h different $\psi_{l_1 \ldots l_\nu}$ and antisymmetrize the product wave function with respect to their arguments. (We may assume the cubes do not overlap.) The expectation value of U_N will clearly still satisfy (IV.15) but for the kinetic energy we have

$$\langle T_N \rangle = (\hbar^2\pi^2/2md^2)[kD_\nu(h) + c] \qquad \text{(IV.19)}$$

where

$$D_\nu(h) = \sum_{f=1}^{h} |\mathbf{l}_{(f)}|^2 \qquad \text{(IV.20)}$$

in which the $\mathbf{l}_{(f)}$ are h *distinct* vectors with positive integral coordinates. This function will be a minimum when the vectors fill out, to best approximation the positive 2^ν-ant of a ν-dimensional hypersphere. Let $A_\nu L^\nu$ be the volume of such a hypersphere of radius L and let $B_\nu L^{\nu+2}$ be its moment of inertia about one axis. If we choose L so that

$$2^{-\nu} A_\nu L^\nu = h \qquad \text{(IV.21)}$$

then for large h we will have

$$D_\nu(h) \approx 2^{-\nu} B_\nu L'^{\nu+2} = 4\nu B_\nu A_\nu^{-1-(2/\nu)} h^{1+(2/\nu)}. \quad \text{(IV.22)}$$

Consequently for sufficiently large N there are positive numbers ζ_1 and ζ_2 depending only on k and ν, such that

$$\zeta_2(\hbar^2/md^2)N^{1+(2/\nu)} \leq \langle T_N \rangle \leq \zeta_1(\hbar^2/md^2)N^{1+(2/\nu)}. \quad \text{(IV.23)}$$

Hence for $\nu \geq 3$ the kinetic energy increases no faster than $N^{5/3}$ and so is dominated by the potential energy which diverges as N^2 thus proving the theorem. For $\nu = 2$ the kinetic energy increases as N^2 and the theorem follows only if

$$U_0 > 4k^2\zeta_2(\hbar^2/2md^2), \quad \text{(IV.24)}$$

that is if U_0 is sufficiently large. This completes the proof.

For $\nu = 1$ the method always fails. Indeed since (IV.23) will hold for any antisymmetric wavefunction vanishing outside a bounded domain of dimensions of order d it appears that the conditions on the potential might in general be insufficient for instability.

V. QUANTUM SYSTEMS WITH COULOMB INTERACTIONS

The Hamiltonian of a system of N point charges q_i with Coulomb interactions is given by

$$\mathcal{H}_N = T_N + U_N, \quad \text{(V.1)}$$

$$T_N = \sum_{i=1}^{N} \left(\frac{-\hbar^2}{2m_i} \nabla_i^2 \right), \quad \text{(V.2)}$$

$$U_N = \sum_{i<j} \frac{q_i \cdot q_j}{|\mathbf{r}_j - \mathbf{r}_i|}, \quad \text{(V.3)}$$

where m_i is the mass of the ith particle. Notice that we do not enclose the particles in a box; $\mathcal{H}_N$ acts thus in the Hilbert space of square-integrable functions of $3N$ real variables.

We shall restrict ourselves to the case where q_i takes only the values $+1$ and -1 and m_i two values m_+ (if $q_i = +1$) and m_- (if $q_i = -1$). We suppose that there are N_+ particles with charge $+1$ and mass m_+ and N_- particles with charge -1 and mass m_-. The remarks which we shall make could, however, be extended to the situation of different values of q_i and more than two different masses.

We first prove a *scaling property* of the Hamiltonian $\mathcal{H}_N$. Let $\lambda > 0$ and, if ψ is a square-integrable function of $\mathbf{r}_1, \cdots, \mathbf{r}_N$, let $\psi^{(\lambda)}$ be defined by

$$\psi^{(\lambda)}(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \lambda^{3N/2}\psi(\lambda\mathbf{r}_1, \cdots, \lambda\mathbf{r}_N). \quad \text{(V.4)}$$

The transformation $\psi \to \psi^{(\lambda)}$ conserves the scalar products (it is a unitary transformation of Hilbert space). Furthermore

$$(\lambda^{-2}T_N + \lambda^{-1}U_N)\psi^{(\lambda)} = [(T_N + U_N)\psi]^{(\lambda)}. \quad \text{(V.5)}$$

From this it follows that *the spectrum of* $\mathcal{H}_N^{(\lambda)} = \lambda^{-2}T_N + \lambda^{-1}U_N$ *is the same as that of* $\mathcal{H}_N$.

Let $E_0(N)$ be the greatest lower bound (glb) to the spectrum of $\mathcal{H}_N$, or "lowest eigenvalue" of $\mathcal{H}_N$. By the scaling property it is also the glb to the spectrum of $\mathcal{H}_N^{(\lambda)}$. The glb. $E_0(2)$ to the spectrum of

$$\mathcal{H}_2 = -\frac{\hbar^2}{2m_+}\nabla_1^2 - \frac{\hbar^2}{2m_-}\nabla_2^2 - \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \quad \text{(V.6)}$$

is just the ground-state energy of a "hydrogen atom" with masses m_+ and m_-. We now compute an upper and a lower bound of $E_0(N)$ in terms of $E_0(2)$. For simplicity we take $N_+ = N_- = N_0 = \frac{1}{2}N$.

We may obtain an upper bound to $E_0(N)$ by taking the expectation value of $\mathcal{H}_N$ for a normalized test function ψ. If we take as ψ the wavefunction of N_0 "hydrogen atoms" formed by pairing the positive and negative charges, and if we assume that the mutual distances between these "hydrogen atoms" is large, we find

$$E_0(N) \leq -N_0E_0(2). \quad \text{(V.7)}$$

To find a lower bound to $E_0(N)$ we write

$$\begin{aligned}
\mathcal{H}_N &\geq \sum_{i=1}^{N_0} \frac{-\hbar^2}{2m_+}\nabla_i^2 + \sum_{i=1}^{N_0} \frac{-\hbar^2}{2m_-}\nabla_i^2 \\
&\quad + \sum_{i=1}^{N_0}\sum_{j=1}^{N_0} \frac{-1}{|\mathbf{r}_j - \mathbf{r}_i|} \\
&= \sum_i \sum_j \left[N_0^{-1}\left(-\frac{\hbar^2}{2m_+}\nabla_i^2 \right.\right. \\
&\quad \left.\left. - \frac{\hbar^2}{2m_-}\nabla_j^2 \right) - \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \right] \\
&= N_0 \sum_i \sum_j \left[N_0^{-2}\left(-\frac{\hbar^2}{2m_+}\nabla_i^2 \right.\right. \\
&\quad \left.\left. - \frac{\hbar^2}{2m_-}\nabla_j^2 \right) + N_0^{-1} \cdot \frac{-1}{|\mathbf{r}_j - \mathbf{r}_i|} \right].
\end{aligned} \quad \text{(V.8)}$$

Using the scaling property for $\mathcal{H}_2$ we see that the operator in square brackets on the r.h.s. of (V.8) is bounded below by $E_0(2)$, therefore (V.8) gives

$$E_0(N) \geq -N_0^3E_0(2). \quad \text{(V.9)}$$

We have thus found for $E_0(N)$ an upper bound (V.7) which is linear in the number of particles, and a lower bound (V.9) which is cubic. As discussed in Section 4 one would expect that a linear lower

bound should exist, with perhaps the restriction that either the positive or the negative particles obey Fermi statistics. What we can at least show is that the existence of a linear lower bound

$$E_0(N) \geq -N_0 B \tag{V.10}$$

for any one choice of the masses m_+, m_- implies its existence for any other choice m_+', m_-'. This statement follows readily from the relations (V.11) and (V.12) below. The identity

$$E_0(N, m_+, m_-) = \lambda E_0(N, m_+', m_-')$$
$$\text{if} \quad \frac{m_+'}{m_+} = \frac{m_-'}{m_-} = \lambda \tag{V.11}$$

follows from the scaling property if one notices that the Hamiltonian for the masses m_+', m_-' is just $\lambda \mathcal{H}^{(\lambda)}$. The inequality

$$E_0(N, m_+, m_-)$$
$$\leq E_0(N, m_+', m_-') \quad \text{if} \quad m_\pm \geq m_\pm' \tag{V.12}$$

is obtained by remarking that to change $m_\pm$ to $m_\pm'$ amounts to adding the positive term

$$\left(\frac{\hbar^2}{2m_+'} - \frac{\hbar^2}{2m_+}\right) \sum_{i=1}^{N_0} (-\nabla_i^2)$$
$$+ \left(\frac{\hbar^2}{2m_-'} - \frac{\hbar^2}{2m_-}\right) \sum_{i=1}^{N_0} (-\nabla_i^2) \tag{V.13}$$

to the Hamiltonian.

As a last remark we notice that the difficulty in proving (V.10) does not come from the long-range part of the Coulomb potential. Indeed let

$$\frac{1}{r} = \varphi_1(\mathbf{r}) + \varphi_2(\mathbf{r}), \tag{V.14}$$

with

$$\varphi_1(\mathbf{r}) = \frac{e^{-\alpha r}}{r}, \qquad \varphi_2(\mathbf{r}) = \frac{1 - e^{-\alpha r}}{r}. \tag{V.15}$$

We may write

$$\mathcal{H}_N = \mathcal{H}_N^{(1)} + U_N^{(2)}, \tag{V.16}$$

$$\mathcal{H}_N^{(1)} = \sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m_i} \nabla_i^2\right)$$
$$+ \sum_{i<j} q_i q_j \varphi_1(\mathbf{r}_i - \mathbf{r}_i), \tag{V.17}$$

$$U_N^{(2)} = \sum_{i<j} q_i q_j \varphi_2(\mathbf{r}_i - \mathbf{r}_i). \tag{V.18}$$

In $\mathcal{H}_N^{(1)}$ we have removed the long-range part of the Coulomb potential, replacing $1/r$ by a Yukawa potential. Our remark is that the long-range part of the Hamiltonian, namely $U_N^{(2)}$, is bounded below by a multiple of N. This follows from theorem I if one notices that $\varphi_2(0)$ is finite and that the Fourier transform of φ_2 is nonnegative, being proportional to

$$\frac{1}{p^2} - \frac{1}{p^2 + \alpha^2} = \frac{\alpha^2}{p^2(p^2 + \alpha^2)}. \tag{V.19}$$

ACKNOWLEDGMENTS

The authors are pleased to thank Dr. J. Ginibre for a helpful discussion. One of us (M.E.F.) would like to acknowledge correspondence with Dr. O. Penrose and to thank M. Motchane, the Director of the Institut des Hautes Etudes Scientifiques, for his hospitality during a visit in the course of which the work reported here was done.

APPENDIX A

The object of this appendix is to prove the following result:

Theorem IV. Let $0 < a_1 < a_2$ and let $\xi(r)$, $\eta(r)$ be monotonic non-increasing non-negative functions, defined on the intervals $(0, a_1)$and $(a_2, +\infty)$ respectively, such that

$$\int_0^{a_1} \xi(r) r^{\nu-1}\, dr = +\infty, \tag{A.1}$$

$$\int_{a_2}^{+} \eta(r) r^{\nu-1}\, dr < +\infty. \tag{A.2}$$

If the pair potential $\varphi(\mathbf{r})$ satisfies

$$\varphi(\mathbf{r}) \geq \xi(r) \qquad \text{for} \quad r \leq a_1, \tag{A.3}$$

$$\varphi(\mathbf{r}) \geq -\eta(r) \quad \text{for} \quad r \geq a_2, \tag{A.4}$$

and if there exists a constant $w \geq 0$ such that

$$\varphi(\mathbf{r}) \geq -w \quad \text{for all} \quad \mathbf{r}, \tag{A.5}$$

then there exists a constant $B \geq 0$ such that

$$U(\mathbf{r}_1, \cdots, \mathbf{r}_n) = \sum_{1 \leq i \leq j \leq n} \varphi(\mathbf{r}_j - \mathbf{r}_i) \geq -nB \tag{A.6}$$

for all n, $\mathbf{r}_1, \cdots, \mathbf{r}_n$.

To prove the theorem we show that we may write

$$\varphi(\mathbf{r}) = \varphi^{(1)}(\mathbf{r}) + \varphi^{(2)}(r) \tag{A.7}$$

where $\varphi^{(1)}(\mathbf{r}) \geq 0$ and $\varphi^{(2)}$ has an integrable nonnegative Fourier transform $\hat{\varphi}^{(2)}(p)$, i.e. $\varphi^{(2)}$ is of positive type[1-4]. The fact that φ admits a positive-type minorant $\varphi^{(2)}$ follows from lemmas 1 and 2 below as the reader will immediately check by writing

$$\varphi^{(2)}(r) = \xi_1(r) - \eta_3(r). \tag{A.8}$$

Lemma 1. *There exists a nonnegative function* η_3 *such that*

$$\varphi(\mathbf{r}) \geq -\eta_3(\mathbf{r}) \quad \text{for all} \quad \mathbf{r} \tag{A.9}$$

and the Fourier transform of η_3 *satisfies*

$$|\hat{\eta}_3(p)| \leq C(p^2 + 1)^{-\nu} \tag{A.10}$$

for some positive constant C.

We define a function η_1 in the interval $(0, +\infty)$ as follows:

$$\eta_1(r) = \begin{cases} w & \text{if} \quad r \leq a_2, \\ \inf\{w, \eta(r)\} & \text{if} \quad r \geq a_2. \end{cases} \tag{A.11}$$

Then our hypotheses imply that

$$\varphi(r) \geq -\eta_1(r). \tag{A.12}$$

Given b such that $0 < b < a_2$ we introduce also a function η_2 on $(0, +\infty)$ by

$$\eta_2(r) = \begin{cases} w & \text{if} \quad r \leq b \\ \eta_1(r - b) & \text{if} \quad r \geq b. \end{cases} \tag{A.13}$$

The functions η_1 and η_2 are nonnegative, nonincreasing and integrable because

$$\int_0^\infty \eta_1(r) r^{\nu-1}\, dr = w \int_0^{a_2} r^{\nu-1}\, dr + \int_{a_2}^\infty \eta(r) r^{\nu-1}\, dr < +\infty, \tag{A.14}$$

$$\begin{aligned} \int_0^\infty \eta_2(r) r^{\nu-1}\, dr &= w \int_0^b r^{\nu-1}\, dr + \int_b^\infty \eta_1(r-b) r^{\nu-1}\, dr \\ &= w \int_0^b r^{\nu-1}\, dr + \int_0^\infty \eta_1(r)(r+b)^{\nu-1}\, dr < +\infty, \end{aligned} \tag{A.15}$$

where we have used the fact that $(r + b)^{\nu-1}/r^{\nu-1}$ tends to unity as $r \to \infty$.

From the definition (A.13) of η_2 and the monotonicity of η_1 and η_2 it follows that

$$\eta_2(r') \geq \eta_1(r) \quad \text{if} \quad |r' - r| \leq b. \tag{A.16}$$

If η_1 and η_2 are considered as functions of position vectors we therefore also have

$$\eta_2(\mathbf{r}') \geq \eta_1(\mathbf{r}) \quad \text{if} \quad |\mathbf{r}' - \mathbf{r}| \leq b. \tag{A.17}$$

Now let ψ be a non-negative function on R^ν vanishing outside a sphere of radius b centered at the origin and such that

$$\int d\mathbf{r}\, \psi(\mathbf{r}) = 1. \tag{A.18}$$

We suppose that ψ has continuous derivatives of all orders and define η_3 as

$$\eta_3(\mathbf{r}) = \int d\mathbf{r}'\, \psi(\mathbf{r} - \mathbf{r}')\eta_2(r'). \tag{A.19}$$

Then, (A.16) gives

$$\eta_3(r) \geq \eta_1(r). \tag{A.20}$$

Furthermore the Fourier transform $\hat{\eta}_3$ of η_3 is proportional to the product of $\hat{\eta}_2$ (which is continuous and bounded) and $\hat{\psi}$ (which is continuous and decreases at infinity faster than any inverse polynomial). Therefore $\hat{\eta}_3$ is continuous and decreases faster than any inverse polynomial at infinity. Lemma 1 follows from this fact and (A.12), (A.20).

Lemma 2. *There exists a (nonnegative) function* ξ_1 *such that*

$$\xi_1(r) \begin{cases} \leq \varphi(r) & \text{for} \quad r \leq a_1 \\ = 0 & \text{for} \quad r \geq a_1 \end{cases} \tag{A.21}$$

and the Fourier transform of ξ_1 *satisfies*

$$\hat{\xi}_1(r) \geq C(p^2 + 1)^{-\nu} \tag{A.22}$$

with the same constant C *as in Lemma 1.*

Let $\chi(\mathbf{r})$ be continuous, nonnegative and satisfy $\chi(0) \geq 0$ and $\chi(\mathbf{r}) = 0$ for $r \geq \frac{1}{2}$. Define χ_1 as

$$\chi_1(\mathbf{r}) = \int d\mathbf{r}'\, \chi(\mathbf{r} - \mathbf{r}')\chi(\mathbf{r}'). \tag{A.23}$$

Then χ_1 is continuous, nonnegative and $\chi_1(\mathbf{r}) = 0$ if $r \geq 1$. Furthermore the Fourier transform $\hat{\chi}_1$ of χ_1 is continuous, nonnegative (being proportional to the square of $\hat{\chi}$) and nonzero in some neighborhood of the origin.

Consider the function χ_2 defined by

$$\chi_2(\mathbf{r}) = \chi_1(\mathbf{r}) \int d\mathbf{p}\, e^{i\mathbf{p}\cdot\mathbf{r}}(p^2 + 1)^{-\nu}. \tag{A.24}$$

Then χ_2 is continuous, nonnegative (the integral in (A.24) is a K-function of the theory of Bessel functions[12]) and $\chi_2(\mathbf{r}) = 0$ if $r \geq 1$. Dividing χ_2 by $\max_{r \leq 1} \chi_2(r)$ we obtain a function χ_3 with properties given as follows:

Lemma 3. *The function* χ_3 *is continuous, nonnegative, bounded above by* 1 *and* $\chi_3(r) = 0$ *if* $r \geq 1$.

[12] See Ref. 4. p. 94 or Formula 7.12 (20) in A. Erdelyi, Magnus, Oberhettinger, Tricomi. Higher Transcendental Functions, Vol. 2. McGraw–Hill, New York (1953).

The Fourier transform of χ_3 satisfies

$$\hat{\chi}_3(p) \geq C'(p^2 + 1)^{-\nu} \tag{A.25}$$

for some positive C'.

To prove (A.25) we remark that $\hat{\chi}_3$ is proportional to

$$\int d\mathbf{p}'\, \hat{\chi}_1(\mathbf{p}')[(\mathbf{p} - \mathbf{p}')^2 + 1]^{-\nu}, \tag{A.26}$$

where $\hat{\chi}_1$ is continuous and nonnegative. Since $\hat{\chi}_1(0) \neq 0$, (A.25) follows from (A.26) by restricting the integration to a small neighborhood of the origin.

We now use Lemma 3 to prove Lemma 2. We may suppose that ξ is a strictly decreasing function of r. Then, for all sufficiently large positive integers n ($n \geq n_0$), let $\alpha_n < 1$ be defined by $\xi(\alpha_n) = n$.

The step function $\xi^*(r) = n$ for $\alpha_{n+1} < r \leq \alpha_n$, clearly does not exceed $\xi(r)$ when $r < \alpha_{n_0}$. This step function is simply the sum of the unit step functions $\theta_n(r) = 1$ for $r \leq \alpha_n$ but zero otherwise. But the properties of $\chi_3(r)$ stated in Lemma 3 imply that $\chi_3(r/\alpha_n) \leq \theta_n(r)$. In total we thus find

$$\sum_{n \geq n_0} \chi_3(r/\alpha_n) \leq \xi(r). \tag{A.27}$$

On the other hand, because of (A.1) we have

$$\sum_{n \geq n_0} \alpha_n^\nu = +\infty. \tag{A.28}$$

Since $\alpha_n < 1$, (A.25) yields for the Fourier transform of $\sum_{n_0}^{n_1} \chi_3(\alpha_n^{-1}\mathbf{r})$ the inequality

$$\sum_{n_0}^{n_1} \alpha_n^\nu \hat{\chi}_3(\alpha_n \mathbf{p}) \geq \left[\sum_{n_0}^{n_1} \alpha_n^\nu\right] C'(p^2 + 1)^{-\nu}. \tag{A.29}$$

Lemma 3 follows from (A.27), (A.28), (A.29) if we write

$$\xi_1(\mathbf{r}) = \sum_{n_0}^{n_1} \chi_3(\alpha_n^{-1} r) \tag{A.30}$$

for n_1 sufficiently large.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 2 FEBRUARY 1966

Axiomatic Foundations of Quantum Theories*

MARCEL GUENIN
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 12 April 1965)

A formalism is developed on an axiomatic basis and is shown to contain as special cases classical mechanics, the usual quantum mechanics, as well as the quantum theory of systems with continuous superselection rules. The structure of the symmetries of a general physical theory is studied and a classification of observables is exhibited.

INTRODUCTION

ONE can ask the question whether an axiomatization of quantum mechanics is indeed needed. In fact, quantum mechanics is "axiomatized" if one says that observables are "represented" by self-adjoint operators on a separable Hilbert space; states are generated by the vectors (or the ray) of the Hilbert space.

The only trouble with these postulates, is that we cannot call them physically intuitive, and as one is looking towards formulating more sophisticated theories, for instance quantum field theories in the spirit of Wightman,[1] it is not immediately clear that the above-mentioned requirements are the minimal ones which ensure that the theory is a quantum theory.

Our aim is not to propose a radical departure from quantum mechanics; what we are looking for, are the physically simple intuitive axioms, from which one could deduce the usual postulates of quantum mechanics. We present here what we think to be a solution to this problem. Our formulation will turn out to be general enough to contain as a particular case not only the usual quantum mechanics, but also the classical point or statistical mechanics, and we shall be able to treat rigorously the case of continuous superselection rules.

We claim however, that such an axiomatization is needed before one discusses any attempt of generalizing quantum mechanics, and that any such attempt should start by explaining which axiom is to be rejected, by which one it has to be replaced and then by working out a similar construction to ours.

As far as the axioms are concerned, the present work is based on the thesis of Piron,[2] and it is in its development very much in the spirit of the work of Segal[3-5] and Mackey.[6]

We shall get algebraic structures out of our axioms, and many authors have studied them in quantum mechanical system. For instance Segal[3-5] who assumes the algebra of observables is a C^*-algebra and Jauch[7,8] who assumes that is a discrete von Neumann algebra with Abelian commutant. In quantum field theories, Haag introduced algebraic structures in the Lille conference,[9] and the first systematic treatment was given by Araki in its Zurich lectures[10] followed by a series of papers.[11-14] Slightly different approaches may be found in Haag,[15,16] Guenin and Misra,[17] and Kadison.[18]

It is not possible to give a complete discussion of all axioms, definitions and theorems here nor to discuss all their implications, because this paper would grow much too long. For the basic axioms we refer to the thesis of Piron[2] which is anyway our starting point. We think that most of the discussion which could be made and most of the corollaries which could be deduced from our theorems are more

* Research supported in part by U. S. Air Force Office Research, ARDC.

[1] See, for instance, R. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964).

[2] C. Piron, Helv. Phys. Acta **37**, 439 (1964).

[3] I. Segal, Kgl Dansk. Videnskab. Selskab Mat-Fys. Medd. **31**, No. 12 (1959).

[4] I. Segal, Can. J. Math. **13**, 1 (1961).

[5] I. Segal, Illinois J. Math. **6**, 500 (1962).

[6] G. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc. New York, 1963).

[7] J. Jauch, Helv. Phys. Acta **33**, 711 (1960).

[8] J. Jauch and B. Misra, Helv. Phys. Acta **34**, 699 (1961).

[9] *Les Problemes mathematiques de la théorie quantique des champs* (CNRS, Paris, 1959).

[10] H. Araki, Lecture notes Eidgenossische Technische Hochschule, Zurich, 1961 (to be published).

[11] H. Araki, J. Math. Phys. **4**, 1343 (1963).

[12] H. Araki, J. Math. Phys. **5**, 1 (1964).

[13] H. Araki, Progr. Theoret. Phys. (Kyoto) **32**, 956 (1964).

[14] H. Araki, Progr. Theoret. Phys. (Nyoto) **32**, 844 (1964).

[15] R. Haag and B. Schroer, J. Math. Phys. **3**, 248 (1962).

[16] D. Kastler and R. Haag, J. Math. Phys. **5**, 848 (1964).

[17] M. Guenin and B. Misra, Nuovo Cimento **30**, 1272 (1963).

[18] R. Kadison, "Transformation of States in Operator Theory and Dynamics," preprint.

or less obvious so that we shall leave them to the imagination of the reader.

I. CALCULUS OF PROPOSITIONS

We start from the lattice theoretical formulation of quantum theories in the form given by Piron,[2] based on the old idea of Birkhoff and Von Neumann,[19] and in this first part, we follow Piron very closely.

Among all possible observables on a family of physical system, we shall consider those for which the result of measurement can be expressed by yes or no and we shall call them *propositions*. For a given system, a proposition is said to be *true* if the answer is yes with certainty. If this definition is to have a sense, if a is true, it should be possible to measure a without perturbing the system. We shall admit this.

If a and b are two propositions, it may happen that one implies the other, i.e., that every time a is true then b is also true. We shall write this $a \subseteq b$ and $a = b$ means $a \subseteq b$, $b \subseteq a$.

On one set τ of propositions, we shall require the following axioms:

Axiom O: The relation $\subseteq$ is an order relation; i.e.,

$$O_1 : a \subseteq a \qquad \forall a \in \tau$$

$$O_2 : a \subseteq b, \qquad b \subseteq c \Rightarrow a \subseteq c$$

From this axiom we can deduce that $a = b, b = c \Rightarrow a = c$; $a = a\ \forall a \in \tau$, thus "$=$" is an equivalence relation and we shall identify a and b whenever $a = b$.

Axiom T: $\exists$ a greatest lower bound for any family of propositions; i.e., given a_i, $i \in J$, $\exists$ a proposition denoted by $\bigcap_J a_i$ such that

$$T_1 : x \subseteq a_i \qquad \forall i \in J \Rightarrow x \subseteq \bigcap_J a_i.$$

This axiom expresses the "and" of the logic. It implies the existence of a proposition ϕ such that $\phi \subseteq a$, $\forall a \in \tau$. It is the absurd proposition which cannot be true unless all propositions of τ are true.

Axiom C: $\exists$ an orthocomplementation in τ; i.e., to each proposition $a \in \tau$, corresponds a proposition $a' \in \tau$ such that

$$C_1 : (a')' = a,$$

$$C_2 : a' \cap a = \phi,$$

$$C_3 : a' \subseteq b' \Leftrightarrow b \subseteq a.$$

Physically speaking, this correspondence is obtained by exchanging "yes" and "no" in the use of measuring device. This of course has only a sense for an apparatus which during a measurement does not perturb the system if one of the answer yes or no is certain.

C_2 corresponds to what in logic is called the law of the excluded middle and C_3 is known as de Morgan's law. In logic always, the law of double negation is $(a')' \subseteq a$; applying it of a' and using C_3 we get C_1.

In τ not only a greatest lower bound exists, but also a least upper bound for any family of propositions. Indeed, define

$$\bigcup_J a_i = (\bigcap_J a_i')';$$

then

$$a_i \subseteq x \quad \forall i \in J$$

$$\Leftrightarrow x' \subseteq a_i' \quad \forall i \in J \Leftrightarrow x' \subseteq \bigcap_J a_i' \Leftrightarrow \bigcup_J a_i \subseteq x.$$

We shall call $\phi' = I$. I is the maximal proposition, $a \subseteq I\ \forall a \in \tau$. I is true as soon as one proposition is true.

The axioms OTC imply that τ is an orthocomplemented complete lattice.

The lattice associated with a classical system, for instance to the point mechanics, has other additional properties. First it satisfies the axiom:

Axiom D:

$$D_1 : a \cap (b \cup c) = (a \cap b) \cup (a \cap c) \qquad \forall a, b, c \in \tau$$

which implies

$$D_2 : a \cup (b \cap c) = (a \cup b) \cap (a \cup c) \qquad \forall a, b, c \in \tau$$

known as distributivity.

Further, if we define an *atom* as an element $p \in \tau$ different from ϕ and such that $\phi \subseteq x \subseteq p$ implies $x = \phi$ or $x = p$, we have the

Axiom A: A_1: $\forall a \in \tau$, $\exists p$ such that p is an atom and $p \subseteq a$.

A_2: if p is an atom, $a \subseteq x \subseteq a \cup p \Rightarrow a = x$ or $x = a \cup p$.

A lattice which satisfies the axioms O, T, C, D, A is called an atomic Boolean lattice.

We want to make our theory general enough to contain quantum theory, and it is well known that the lattice of quantum mechanics cannot be Boolean (cf. Refs. 2, 20). We have thus to drop at least one axiom. It is not sufficient to drop A only, because a lattice satisfying OTCD may always be imbedded in a lattice satisfying OTCDA (the classical

[19] G. Birkhoff and J. von Neumann, Ann. Math. 37, 823 (1936).

[20] J. Jauch and C. Piron, Helv. Phys. Acta 37, 439 (1964).

statistical mechanics is an example). Thus D cannot be maintained.

To guide our choice, we may remark that the simple quantum mechanical systems in finite Hilbert space (for instance spin systems) satisfy the

Property M: $M: x \subseteq z \Rightarrow x \cup (y \cap z) = (x \cup y) \cap z \; \forall y \in \tau$,

which is called modularity and is weaker than distributivity (axiom *D*). It is not possible to retain property *M* together with OTCA (with *D*) for a quantum theory, because one can show that one should then restrict oneself to the lattice of projectors of a family of finite-dimensional vector spaces (which is the case of finite quantum systems). For quantum system of the infinite type, we have to release one condition at least. Von Neumann and Birkhoff[19] have suggested to keep OTCM and drop *A*. The lattice of proposition is then isomorphic to the lattice of projectors of a family of type II_1 factors. On the other hand, the lattice of projectors generated by the spectral decomposition of p and q, $[p, q] \subset 1$ is not modular. We shall say that a set of propositions is compatible if the system behaves under the corresponding measurements like a classical system. As we know that the lattice of propositions for a classical system is a Boolean one, we are lead to the following

Definition: a and b are said to be *compatible*, $a \leftrightarrow b$, if the sublattice generated by a and b is isomorphic to a Boolean lattice. (This lattice is the set obtained by combining arbitrarily unions, intersections and orthocomplementations, in this case at most 16 elements.)

If $a \subseteq b$, the lattice generated by a and b is:

$$\phi, a, a' \cap b, b, b', a \cup b', a', 1.$$

A necessary and sufficient condition for this lattice to be distributive, and thus Boolean is

$$a \cup (a' \cap b) = b \cup (b' \cap a)$$

We shall make the following

Axiom P: $a \subseteq b \Rightarrow a \leftrightarrow b$.

We shall call proposition system, any set τ satisfying O, T, C, A, P. In everything that follows, classical physics is obtained by adding to this the axiom *D*. We have the following fundamental result, due to Piron[2]:

Theorem 1.1. τ (satisfying O, T, C, A, P) is isomorphic to the lattice of all projectors from a family of Hilbert spaces $\mathfrak{H}_\xi$, $\xi \in Z$, over a field. Theorem 1.1 is the strongest result one can get from the axioms; in particular we cannot say anything on Z which can be as well reduced to a single point as be a many-dimensional space. Its exact structure should be given by the physical system considered. In order to be able to carry on our analysis, we have to make the

Postulate Z: Z is a locally compact Hausdorff space.

To simplify our work we shall add

Postulate Z': Z is countably dimensional.

but most of the following proofs could be carried on without this last postulate. The proof of Theorem 1.1 is constructive and in principle determines the field, although not in a unique way. It is not possible to prove from the axioms that all $\mathfrak{H}_\xi$ have to be defined on the same field. As for easily understandable physical reasons, we want the field to be continuous and connected to the unit, we are left to choose between the real, complex, and quaternions. One can make quantum theory over these three fields, but the works of Stueckelberg *et al.*[21] on one side, and Emch[22] on the other side, have shown that they are "essentially" equivalent to a quantum theory on the complex. We shall thus make the mathematically very convenient

Postulate F: The Hilbert spaces $\mathfrak{H}_\xi$ are defined on the complex.

2. ALGEBRA ASSOCIATED WITH A LATTICE OF PROPOSITIONS

The first idea which comes to mind is that all that is needed is to imbed all $\mathfrak{H}_\xi$ in one single (eventually nonseparable) Hilbert space $\mathfrak{H}$ and to call observable any self-adjoint operator which lies in the C^*-algebra, or in the von Neumann algebra (as one prefers!), generated by the projectors representing τ. But there are many troubles with this simple way of doing the things; for example Z is in general not countable and thus even if we have some reason of supposing the $\mathfrak{H}_\xi$ separable, $\mathfrak{H}$ would not be separable; it is known that nonseparable Hilbert spaces are not necessarily equivalent, so we do not know which one to choose; other difficulties arise as one tries to define states on the system.

Let us call $\mathfrak{H}$ the sum of the $\mathfrak{H}_\xi$; i.e. an element $x \in \mathfrak{H}$ is given by a family $\{x(\xi)\}$ of elements $x(\xi)$, $x(\xi) \in \mathfrak{H}_\xi$. (N.B. we do not restrict ourselves

[21] E. C. G. Stueckelberg, Helv. Phys. Acta **33**, 727 (1960).
[22] G. Emch, Helv. Phys. Acta **36**, 739, 770 (1963).

to countable families.) The sum of two elements of $\mathfrak{H}$ is given by

$$x + y = \{x(\zeta)\} + \{y(\zeta)\} = \{(x + y)(\zeta)\}$$

and $\lambda x = \{\lambda x(\zeta)\}$; the zero element of $\mathfrak{H}$ is the family of the zero elements of the $\mathcal{H}_\zeta$. $\mathfrak{H}$ is not necessarily an Hilbert space, but we could make it a Banach space by defining

$$|x|_s = \sup_{\zeta \in Z} |x(\zeta)|.$$

On each $\mathcal{H}_\zeta$, we consider the algebra $\mathfrak{B}(\mathcal{H}_\zeta) \equiv \mathfrak{B}(\zeta)$ of all bounded operators on $\mathcal{H}_\zeta$, and in the same way as for the sum of the $\mathcal{H}_\zeta$ we can define $\mathfrak{T}$ as being the sum of all $\mathfrak{B}(\zeta)$. $a \in \mathfrak{T}$ is given by the family

$$\{a(\zeta)\}, \quad a(\zeta) \in \mathfrak{B}(\zeta), \quad a + b = \{a(\zeta)\} + \{b(\zeta)\},$$
$$ab = \{a(\zeta)b(\zeta)\}, \quad a^* = \{a(\zeta)^*\}.$$

As we want $\mathfrak{T}$ to contain τ among its projectors, we cannot restrict ourself to countable families only. $\mathfrak{T}$ can be made a normed algebra if we define

$$||a||_s = \sup_{\zeta \in Z} ||a(\zeta)||$$

and understand that $\mathfrak{T}$ contains only operators of finite s-norm in this case. One easily verifies that

$$||a + b||_s \leq ||a||_s + ||b||_s,$$
$$||a + b||_s \leq ||a||_s \cdot ||b||_s,$$
$$||I||_s = 1 \quad \text{where 1 is the identity of } \mathfrak{T},$$
$$||\lambda a||_s = |\lambda| \cdot ||a||_s,$$
$$||a||_s \geq 0, \qquad ||a||_s = 0 \Leftrightarrow a = 0$$

and from $||a^*(\zeta)\, a(\zeta)|| = ||a(\zeta)||^2$ follows $||a^*a||_s = ||a||_s^2$.

We have thus the

Lemma 2.1: $\mathfrak{T}$ is a B^*-algebra

if $\mathfrak{T}$ is closed under the norm $||\cdot||_s$ (cf. Def. B^*, Rickart,[23] p. 180). That $\mathfrak{T}$ is closed follows from the fact that the $\mathfrak{B}(\zeta)$ are closed under the norm $||\cdot||$ (cf. Dixmier[24] 1.3.3 p. 6).

Lemma 2.2: $\mathfrak{T}$ is a Bear *-ring of type I. It is homogeneous if all $\mathcal{H}_\zeta$ have the same dimension.

Proof: (i) $\mathfrak{T}$ is a Bear-ring from Kaplansky[25] Ex. 5, p. 3.

(ii) $\mathfrak{T}$ is type I (Def. cf. Kaplansky,[25] Def. 5, p. 6) because if we take a one-dimensional projector $p(\zeta)$ in each $\mathcal{H}_\zeta$, $p = \{p(\zeta)\}$ has I for central support and is Abelian.

(iii) $\mathfrak{T}$ is homogeneous (Kaplansky, Def. 1 p. 32) if all $\mathcal{H}_\zeta$ have the same dimension; we can take $p_i(\zeta)$ one dimension in $\mathcal{H}_\zeta$, $p_i(\zeta) \perp p_k(\zeta)$ for $i \neq k$ and $\bigcup_i p_i(\zeta) = I(\zeta)$ then the $p_i = \{p_i(\zeta)\}$ are mutually orthogonal, *-equivalent, Abelian, and lub $p_i = I$.

Lemma 2.3: $\mathfrak{T}$ is a AW^*-algebra.

Proof: Use Lemma 2.1 and 2.2 and the Def. p. 289 of Rickart.[23]

Theorem 2.4. There exists a *-representation $\mathfrak{T}_c$ of $\mathfrak{T}$ on an Hilbert space $\mathcal{H}_c$, such that it is faithful, norm preserving, and the norm on $\mathfrak{T}_c$ is induced by the norm of $\mathcal{H}_c$. $\mathfrak{T}_c$ is a C^* algebra.

Proof: Lemma 2.1 and Theorem 4.8.11, p. 244 and Corollary 4.8.12, p. 244 of Rickart.[23]

The algebra $\mathfrak{T}$ and its representation $\mathfrak{T}_c$ provides the starting point for the investigations of Segal[3-5] and we can thus say that we have in this direction extended Segal's work by proving one of his main postulate from what seems to us simpler and more fundamental axioms; in fact we get even a stronger result, as we can prove not only that the algebra of all observables is a C^*-algebra, but even that it is an AW^*-algebra, which is also a representation-invariant feature. In what follows however, we shall not adopt Segal's point of view about the definition of states.

3. STATES

Definitions of states for systems of propositions have been given by Piron[2] and by Emch and Piron.[26] In the C^*-algebra approach[3,16] the states are defined as being linear positive functionals on the algebra. What is common to these different points of view is that the states are required to be defined of all elements of the lattice or the algebra. We reject this requirement and we shall try to give a more general definition of a state, i.e., we shall not require it to be defined on all elements of the lattice or of the algebra, and the reader will be able at the end of this paper to convince himself that this generalization is indeed needed, if we want to be able to represent certain kinds of observables and certain nontrivial symmetry groups.

Definition R: τ' is the definition lattice of a state if τ' is a sublattice of τ such that

R_1: $a \in \tau', \quad b \in \tau' \Rightarrow a \cap b \in \tau'$,
R_2: $a \in \tau' \Rightarrow a' \in \tau'$,
R_3: $\phi \in \tau'$.

[23] C. Rickart, *Banach Algebras* (D. van Nostrand Company, Inc., Princeton, New Jersey, 1960).
[24] J. Dixmier, *Les C*-algèbres* (Gauthier-Villars, Paris, 1964).
[25] I. Kaplansky, "Rings of Operators," Chicago Lecture notes, 1955.
[26] G. Emch and C. Piron, J. Math. Phys. 4, 469 (1963).

R_σ: $a_i \in \tau' \Rightarrow \bigcap_i a_i \in \tau'$ for any countable family $\{a_i\}$. (This definition of τ' is very similar to the def. of Boolean algebra cf. p. 21 Halmos[27] and implies that τ' is a σ-lattice in the sense of Birkhoff[28] p. 52.)

Definition E. A state is a mapping ω of a sublattice $\tau' \subset \tau$, (τ' satisfying R_1, R_2, R_3, R_σ) in the closed interval [0, 1] such that

E_1: $\omega(a) \geq 0, \quad \forall a \in \tau'$,

E_2: $\omega(\phi) = 0$,

E_3: $\omega(I) = 1$,

E_4: $\omega(a) = \omega(b) = 1 \Rightarrow \omega(a \cap b) = 1$,

E_5: $a \leftrightarrow b\ \omega(a) + \omega(b) = \omega(a \cup b) + \omega(a \cap b)$.

That τ is a system of propositions implies[2]:

Lemma 3.1.

(a) $a \subseteq b \Rightarrow \omega(a) \leq \omega(b)$.

(b) If ω_1 and ω_2 are two states on the same τ', and if λ_1, and λ_2 are real positive numbers such that $\lambda_1+\lambda_2=1$, then $\lambda_1\omega_1+\lambda_2\omega_2$ is a state on τ'. (This expresses the superposition principle. If two states ω_1 and ω_2 are defined on τ_1' and τ_2' respectively, they only may be superposed on $\tau_1' \cap \tau_2'$.)

(c) E_5 is equivalent to: $a \leftrightarrow b$ and $a \cap b = \phi \Rightarrow \omega(a \cup b) = \omega(a) + \omega(b)$.

(d) E_4 is equivalent to: $\omega(a) = \omega(b) = \omega(a \cup b) \Rightarrow \omega(a \cap b) = \omega(a \cup b)$.

Definition E_σ: We shall say that ω is σ-continuous if $a_i \in \tau'$,

$$i \neq j, \quad a_i \leftrightarrow a_j, \quad a_i \cap a_j = \phi \Rightarrow \omega\left(\bigcup_i^\infty a_i\right) = \sum_i^\infty \omega(a_i), \quad i = 1, 2, \cdots .$$

Definition: E_N. ω is normal if $\omega(\bigcup_{i\in J} a_i) = \sum_{i\in J} \omega(a_i)$

Segal does not make the postulate of σ-continuity for states because this is not a *-representation-invariant concept. We make it by analogy the probability calculus where it is always made and also because we would not be able to make our construction without it. Moreover it will turn out that all mapping that we are concerned with are normal, i.e., will preserve σ-continuity and normalcy.

It can seem unnecessary to restrict the definition of a state to a sublattice τ' instead of defining it on τ directly. In fact there is no difficulty in supposing that a state is defined on all propositions if Z is countable, even if the $\mathfrak{H}_\zeta$ are nonseparable. In the case Z has the power of the continuum however, the things are no longer so simple. We can illustrate this fact by the following example. Let Z be the interval [0, 1] and for sake of simplicity, suppose that all $\mathfrak{H}_\zeta$ are one-dimensional. Thus τ is the lattice of all subsets of the interval [0, 1] (we have *all* subsets because of axiom T). A (σ-additive) state on τ will be a bounded measure on this interval, and if we require the state to be defined on all elements of τ, the measure has to be defined on all subsets of the interval. This requirement is very strong because it excludes such measures as the Lebesgue measure of this interval, since it is well known that there exist subsets of this interval which are not Lebesgue measurable. We do not want to impose restrictions which would eliminate the possibility of considering such measures in our theory,[28a] and this is the reason why we require the states to be only defined on a sublattice. In this example, τ' will be the collection of all Lebesgue-measurable sets.

If we consider the sublattice τ_0 of τ obtained by arbitrary unions and intersections of all $I(\zeta)$, τ_0 form an atomic Boolean lattice which isomorphic to the lattice of all subsets of Z. Because of postulate Z, we can define the Borel sets in Z, and we shall make the following reasonable

Postulate B. If τ' is the definition-lattice of a state, τ' contains all Borel sets of Z.

And from the definitions follows the

Lemma 3.2. A σ-additive state induces a positive bounded measure μ on Z such that all Borel sets in Z are μ-measurable, $\mu(Z) = 1$.

We can now try to figure for ourselves what are the elements of τ which belong to τ'. On the projectors in each $\mathfrak{H}_\zeta$, the state functional, were defined, will induce a form $f_\zeta(p(\zeta))$, so that for $p = \{p(\zeta)\} \in \mathfrak{T}$ a projector of $\mathfrak{T}$,

$$\omega(p) = \int_Z f_\zeta(p(\zeta))\, d\mu(\zeta)$$

and ω will be defined on all $p = \{p(\zeta)\}$ such that $f_\zeta(p(\zeta))$ is a μ-integrable function (of ζ). Summarizing, we can define:

Definition S: We call physical state, or simply

[27] P. Halmos, *Measure Theory* (D. van Nostrand Company, Inc., Princeton, New Jersey, 1950).

[28] G. Birkhoff, *Lattice Theory* (American Mathematical Society, Providence, Rhode Island, 1948).

[28a] In fact if the continuum hypothesis is admitted, this would amount to eliminate all measures which are not the sum of at most a countable number of point measures with suitable coefficients, as follows from Theorem 13, p. 187 of Birkhoff.[28]

state, a positive real functional ω of the form

$$\omega(p) = \int_Z f_\zeta(p(\zeta))\, d\mu(\zeta),$$

where μ is a positive (σ-additive) bounded [$\mu(I) = 1$] measure on Z, and f_ζ a state defined on all projectors on $\mathfrak{H}_\zeta$, satisfying the properties E_1 through E_5 and E_N. ω is defined on all elements $p = \{p(\zeta)\}$ such that $f_\zeta(p(\zeta))$ is a μ-integrable function. These elements are called ω-measurable.

Up to now, we have spoken of states as defined only on projectors; the connection with the algebraic structure can be made using the theorem of Gleason.[29]

Theorem 3.3. If $\mathfrak{P}$ denotes the set of all projectors on an Hilbert space $\mathfrak{H}$, and if a functional $f(p)$ is defined for all $p \in \mathfrak{P}$ and satisfies E_1, E_2, E_3, E_5, and E_N, then f can be considered as the restriction to $\mathfrak{P}$ of a linear normal positive functional f' defined on $\mathfrak{B}(\mathfrak{H})$ and satisfying E_4. Further, f' can be written as

$$f'(a) = \sum_i (x_i, ax_i) \qquad \text{with} \quad \sum_i (x_i, x_i) < \infty$$

and $x_i \in \mathfrak{H}$.

This theorem has been proven by Gleason in the case of a separable Hilbert space. We have extended the proof to the nonseparable case, and the proof will appear somewhere else. Using this theorem, we can thus extend ω to all operators $a \in \mathfrak{T}$ such that $f'_\zeta(a(\zeta))$ is a μ-integrable function. In the following section, we shall add some further specifications and restrictions to the f'.

We conclude this section by two lemmas which are trivial, but necessary for the consistence of our definitions.

Lemma 3.4.[30] $\forall p \in \tau$, $p \neq \phi$, $\exists \omega$ such that $\omega(p) = 1$.

Proof. $p = \{p(\zeta)\}$. Let $x \in \mathfrak{H}_{\zeta_0}$, such that $p(\zeta_0)\, x = x$,

$$|x| = 1 \cdot f_{\zeta_0}(p(\zeta_0)) = (x, p(\zeta_0)x) = 1.$$

Take the Dirac measure on the point ζ_0 with weight $+1$.

Lemma 3.5.[30]

Let p_1, $p_2 \in \tau$; $p_1 \neq p_2$. Then $\exists \omega$ such that $\omega(p_1) \neq \omega(p_2)$.

Proof. $p_1 \neq p_2$ implies that $\exists$ at least one ζ_0 for which $p_1(\zeta_0) \neq p_2(\zeta_0)$, and in an Hilbert space, if two projectors are different, there always exist x, such that $|p_1(\zeta_0)x| \neq |p_2(\zeta_0)x|$ All that is needed is to normalize $|x| = 1$, and take the same measure as for Lemma 3.4.

Remark: One of the reasons for Postulate B is to have a principle of superposition of states a bit less restrictive that the principle proven in Lemma 3.1.(b), which requires the states to be defined on the same τ', the definitions of lattices of two arbitrary states have at least the Borel subsets of Z in common, and the superposition principle is valid on the common domain.

4. ALGEBRA OF MEASURABLE OBSERVABLES

The aim of this section is to use the properties of the fields of von Neumann algebras in order to get an explicit representation of the algebra generated by the measurable propositions. Because of postulate Z, we may consider the family $\mathfrak{H}_\zeta$ as a field of complex Hilbert spaces on Z (Dixmier[31] p. 139). Thus, $\mathfrak{F} = \prod_{\zeta \in Z} \mathfrak{H}_\zeta$ is a complex vector space, an element $x \in \mathfrak{F}$ is a mapping $\zeta \to x(\zeta)$ defined on Z such that $x(\zeta) \in \mathfrak{H}_\zeta$, $\forall \zeta \in Z$; such a mapping is called a field of vectors on Z. If Y is a subset of Z, an element of $\prod_{\zeta \in Y} \mathfrak{H}_\zeta$ is called a field of vectors on Y.

Definition (Dixmier, Ref. 31, Def. 1. p. 141) One says that the $\mathfrak{H}_\zeta$ constitute a μ-measurable field of Hilbert spaces if one has given a vectorial subspace $\mathfrak{G}$ of $\mathfrak{F}$ such that

(i) $\forall x \in \mathfrak{G}$, the function $\zeta \to ||x(\zeta)||$ is μ-measurable,

(ii) if $y \in \mathfrak{F}$ is such that, $\forall x \in \mathfrak{G}$, the complex numerical function $\zeta \to (x(\zeta), y(\zeta))$ is μ-measurable, then $y \in \mathfrak{G}$.

(iii) $\exists$ a sequence $\{x_1, x_2, \cdots\}$ of elements of $\mathfrak{G}$ such that $\forall \zeta \in Z$, the $x_n(\zeta)$ are a total sequence in $\mathfrak{H}_\zeta$.

The field of vectors belonging to $\mathfrak{G}$ are called μ-measurable field of vectors. A sequence $x_1, x_2, \cdots$ of μ-measurable fields of vectors with property (iii) is called a fundamental sequence of μ-measurable field of vectors.

Unfortunately, this definition of measurable fields of vectors only apply to the case where all $\mathfrak{H}_\zeta$ are separable, as follows from (iii). We cannot simply drop (iii) because Dixmier[31] (p. 143) has shown that we have in this case also to drop (ii). This is

[29] A. Gleason, J. Ratl. Mech. and Anal. 6, 885 (1957).

[30] The states constructed in Lemmas 3.4 and 3.5 satisfy in an evident way a postulate L which will be added later.

[31] J. Dixmier, *Les algèbres d'operateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1957).

a problem which is not yet solved and so the lack of available mathematical techniques obliges us to abandon a part of the generality we wanted to have in treating our problem, we shall make from now on the

Postulate N. All $\mathfrak{H}_\zeta$ are separable.

One can show, however, that one always can reduce the case where the $\mathfrak{H}_\zeta$ are nonseparable to the separable case if one does not request that operators with zero expectation value in the considered state to be represented by nonzero operators in the representation space.

Definition. A measurable field of vectors $x(\zeta)$ on Z is said to be square integrable if

$$\int_Z |x(\zeta)|^2 \, d\mu(\zeta) < \infty .$$

Lemma 4.1. The set of all square integrable fields is a complex vector space $\mathfrak{K}$. For $x \in \mathfrak{K}$, $y \in \mathfrak{K}$, $(x(\zeta), y(\zeta))$ is an integrable function of ζ. If we write

$$(x, y) = \int_Z (x(\zeta), y(\zeta)) \, d\mu(\zeta)$$

the space $\mathfrak{K}$ is provided with a pre-Hilbertian structure.

Proof. See discussion p. 145, Dixmier.[31]
We have, for $x \in \mathfrak{K}$, $|x|^2 = \int |x(\zeta)|^2 \, d\mu(\zeta)$ and the $x \in \mathfrak{K}$ for which $|x| = 0$ are the x which vanish almost everywhere. Call $\mathfrak{H}_f$, $\mathfrak{K}$ modulo the elements x with $|x| = 0$.

Lemma 4.2. $\mathfrak{H}_f$ is a separable Hilbert space.

Proof. That $\mathfrak{H}_f$ is complete follows from Proposition 5, p. 146, Dixmier.[31] That it is separable is a consequence of Postulate Z' and Corollary, p. 149 Dixmier.[31] $\mathfrak{H}_f$ is usually written $\int^{\oplus} \mathfrak{H}_\zeta \, d\mu(\zeta)$. We shall write $\mathfrak{H}$ instead of $\mathfrak{H}_f$ whenever no confusion is possible.

We have now to look on our problem from the physical point of view. What is given to us is the field $\mathfrak{H}_\zeta$ of Hilbert spaces and Z. We have seen in the preceeding section that each state induces a measure on Z. But this measure is not sufficient to determine uniquely $\mathfrak{G}$ if Z is not countable. We shall sharpen our definition of a physical state so that to get rid of this indeterminancy and give a canonical way of constructing $\mathfrak{G}$ and, therefore, $\mathfrak{H}$.

In our definition of the state we have imposed conditions on $f_\zeta(.)$, but no condition on the relations between the $f_\zeta(.)$ for different ζ. The following postulate can thus be considered as being part of our definition of a physical state:

Postulate L. Let $f_\zeta(.) = \sum_i (x_i(\zeta), .x_i(\zeta)$. There should be, in each $\mathfrak{H}_\zeta$, an ordering of the $x_i(\zeta)$ such that if $x_i = \{x_i(\zeta)\}$, $x_i \in \mathfrak{F}$,

(i) x_i is μ-measurable $\forall i$ and square-integrable.
(ii) $(x_i(\zeta), x_j(\zeta)$ is μ-measurable $\forall i, j$.

Let $\mathfrak{M}(\zeta)$ be in each $\mathfrak{H}_\zeta$ the subspace spanned by the $x_i(\zeta)$, $\mathfrak{M}^\perp(\zeta)$ the orthogonal complement (in $\mathfrak{H}_\zeta$). We choose $y_i(\zeta)$ in each $\mathfrak{M}^\perp(\zeta)$:

Lemma 4.3. The $y_i(\zeta)$ may be choosen such that they are mutually orthonormal, μ-measurable, square-integrable, and are total in each $\mathfrak{M}^\perp_\zeta$.

Proof. Consider the field $\mathfrak{M}^\perp(\zeta)$ and apply Proposition 1, p. 143 (i) and (ii), and Ex. 2, p. 154, Dixmier.[31]

Lemma 4.4. There exists one and only one $\mathfrak{G}$ such that $x_i, y_i \in \mathfrak{G}$; i.e., the fundamental sequence $\{x_i, y_i\}$ determines the structure of measurable field.

Proof. Proposition 4, p. 144, Dixmier.[31]

Let now $a \in \mathfrak{T}$, i.e., a is a family $\{a(\zeta)\}$, $a(\zeta) \in \mathfrak{B}(\zeta)$ which we call a field of operators and a can be considered as a mapping of the field $\mathfrak{H}_\zeta$ into itself. It is clear that a does not in general transform $\mathfrak{G}$ into itself, neither $\mathfrak{K}$. So we define:

Definition (Dixmier,[31] Def. 1, p. 157). A field $a(\zeta)$ of operators is called μ-measurable, if for any measurable field of vectors $x(\zeta)$, the field of vectors $a(\zeta)x(\zeta)$ is μ-measurable.

Definition. A μ-measurable field $a(\zeta)$ is called essentially bounded, if the essential upper bound λ of the function $||a(\zeta)||$ is finite.

Lemma 4.5. If $a(\zeta)$ is an essentially bounded measurable field, and $x(\zeta) \in \mathfrak{K}$, then $a(\zeta)x(\zeta) \in \mathfrak{K}$.

Proof. Dixmier,[31] bottom of p. 159.

$a(\zeta)$ thus induces a linear operator on $\mathfrak{K}$ and as the reader can easily verify, an operator on $\mathfrak{H}$.

If we denote by $||a||_f$ or, if no confusion is possible, simply by $||a||$ the norm induced by $\mathfrak{H}$ on $\mathfrak{B}(\mathfrak{H})$, a is a bounded operator on $\mathfrak{H}$ [thus belongs to $\mathfrak{B}(\mathfrak{H})$] and

Lemma 4.6. $||a|| = \lambda$.

Proof. Dixmier[31] Proposition 2, p. 160.

Lemma 4.7. If two essentially bounded fields of operators $a(\zeta)$ and $b(\zeta)$ define the same element of $\mathfrak{B}(\mathfrak{H})$, then $a(\zeta) = b(\zeta)$ μ-almost everywhere.

Proof. Dixmier,[31] corollary, p. 160.

Definition. (Dixmier,[31] Definition 2, p. 160): An operator $a \in \mathfrak{B}(\mathfrak{H})$ is called decomposable, if it is defined by a measurable field $a(\zeta)$ essentially bounded.

One then writes

$$a = \int^{\oplus} a(\zeta)\, d\mu(\zeta)$$

Lemma 4.8.

If a_1 and a_2 are decomposable operators and

$$a_1 = \int^{\oplus} a_1(\zeta)\, d\mu(\zeta), \qquad \int^{\oplus} a_2(\zeta)\, d\mu(\zeta) = a_2$$

then

$$a_1 + a_2 = \int^{\oplus} (a_1(\zeta) + a_2(\zeta))\, d\mu(\zeta);$$

$$a_1 a_2 = \int^{\oplus} a_1(\zeta) a_2(\zeta)\, d\mu(\zeta)$$

$$\lambda a_1 = \int^{\oplus} \lambda a_1(\zeta)\, d\mu(\zeta), \qquad a_1^* = \int^{\oplus} a_1^*(\zeta)\, d\mu(\zeta).$$

Proof. Dixmier,[31] proposition 3, p. 161.

Lemma 4.9. There exist in $\mathfrak{B}(\mathfrak{H})$ a sequence $\{b_i\}$, $b_i = \int^{\oplus} b_i(\zeta)\, d\mu(\zeta)$, of decomposable operators such that locally almost everywhere, $\mathfrak{B}(\zeta)$ is the von Neumann algebra generated by the $b_i(\zeta)$.

Proof. Dixmier,[31] proposition 5, p. 163.

This lemma is important for us because it insures that we still have retained "enough" operators from $\mathfrak{T}$. We shall call $\mathfrak{T}_f$ the algebra of all decomposable operators. The purpose of the end of this section is to show that $\mathfrak{T}_f$ is a von Neumann algebra of discrete type.

$\mathfrak{T}_f$ is the von Neumann algebra generated in $\mathfrak{H}$ by the lattice of all measurable propositions by respect to the states f which induce the measure μ on Z and $\mathfrak{G}$. We call shortly $\mathfrak{T}_f$ the algebra of measurable observables, although we shall define observables only later.

Let $L_c^\infty(Z, \mu)$ be the set of all numerical complex, measurable, essentially bounded functions on Z, in which one identify two functions almost everywhere equal. Clearly $L_c^\infty(Z, \mu)$ is a *-algebra. If $g \in L_c^\infty(Z, \mu)$, the field of operators $g(\zeta) \cdot I(\zeta) \in \mathfrak{B}(\zeta)$ is measurable and essentially bounded. We shall denote by a_g the corresponding operator of $\mathfrak{B}(\mathfrak{H})$.

Definition. (Dixmier,[31] Def. 3 p. 165): The operators of the form a_g where $g \in L_c^\infty(Z, \mu)$ are called diagonalizable. Let $\mathfrak{Z}$ be the algebra of all diagonalizable operators.

Lemma 4.10. $\mathfrak{Z}$ is an Abelian von Neumann algebra.

Proof. Dixmier,[31] proposition 8, p. 166 (i).

Theorem 4.11. $\mathfrak{T}_f = \mathfrak{Z}'$, and thus is a von Neumann algebra of discrete type with Abelian commutant.

Proof. Apply corollary p. 169, together with Def. 1 p. 120 of Dixmier.[31]

We have used the postulate Z' only in the proof of Lemma 4.2, in order to show that $\mathfrak{H}$ is separable. In this case it is trivial that $\mathfrak{T}_f$ is of denumerable type. For completeness, we shall show that it is still the case if we drop postulate Z'.

Lemma 4.12. Even if postulate Z' is not fulfilled, $\mathfrak{T}_f$ is a discrete von Neumann algebra of denumerable type with Abelian commutant.

Proof. Use the fact that μ is a bounded measure and proposition 8 (iii), p. 166 of Dixmier.[31]

Moreover we have the following lemma which is interesting in the case of classical physics:

Lemma 4.13. $\mathfrak{T}_f$ is Abelian if and only if the Hilbertian dimension of the $\mathfrak{H}_\zeta$ is zero or one locally almost everywhere.

Proof. Ex. 1, p. 175 Dixmier.[31]

Remark: The attention of the reader is called upon the fact that in general $\mathfrak{T}_f$ is not the homomorphic image of $\mathfrak{T}$ (and thus not a representation of $\mathfrak{T}$ in the algebraic sense.) One can easily show that a necessary (and sufficient) condition is that the σ-lattice of all μ-measurable subsets of Z is the homomorphic image of the lattice of all subsets of Z.

5. REPRESENTATION OF STATES

The following theorem follows from our construction and Theorem 3.3. together with postulate L:

Theorem 5.1. A state $f(.)$ which induces the measure μ on Z and $\mathfrak{G}$, is a linear positive normal functional on $\mathfrak{T}_f$, which can be written as

$$f(a) = \int \sum_i (x_i(\zeta), a(\zeta) x_i(\zeta))\, d\mu(\zeta)$$

$$= \sum_i (x_i, a x_i), \qquad a \in \mathfrak{T}_f,$$

$$a = \int^{\oplus} a(\zeta)\, d\mu(\zeta); \qquad x_i(\zeta) \in \mathfrak{H}_\zeta; \qquad x_i \in \mathfrak{H}.$$

What we have done is that given a state to provide a canonical construction of the algebra of all measurable observables (by respect to this state). The natural point to investigate now, is what has to be the relation between two states so that they are both represented nontrivially (i.e., by nonidentically zero functional) as linear functional on an algebra $\mathfrak{T}_f$ for a certain f.

Definition. Given a state

$$f(.) = \int \sum_i (x_i(\zeta), \cdot x_i(\zeta))\, d\mu(\zeta),$$

the induced measure μ on Z and $\mathfrak{G}$, and the corresponding $\mathfrak{T}_f$, another state

$$g(.) = \int \sum_i (y_i(\zeta), \cdot x_i(\zeta))\, d\rho(\zeta)$$

is said to be faithfully representable on $\mathfrak{T}_f$ if

(i) any ρ-measurable field of operators is also μ-measurable;

(ii) any ρ-measurable field of operators which corresponds to a nonzero operator of $\mathfrak{T}_g$ is also corresponding to a nonzero operator of $\mathfrak{T}_f$;

(iii) The $y_i(\zeta)$ are μ-measurable, square-integrable field of vectors.

Following immediately from the definition is

Theorem 5.2.

$$g(\cdot) = \int \sum_i (y_i(\zeta), \cdot y_i(\zeta))\, d\mu(\zeta)$$

is faithfully representable on $\mathfrak{T}_f$ only if any ρ-measurable subset $Y \subset Z$ is μ-measurable and $\mu(Y) \neq 0$ whenever $\rho(Y) \neq 0$.

Corollary 5.3. This implies that $\tau'_f \supset \tau'_g$, and if any μ-measurable subset Y of Z is ρ-measurable, μ is absolutely continuous by respect to ρ (Halmos,[22] p. 124).

Definition. The states $f_1(.), f_2(.), \cdots$ are said to be together f-representable if there exists a state inducing the measure μ on Z and to which corresponds $\mathfrak{T}_f$, such that $f_1(.), f_2(.), \cdots$ are all faithfully represented on $\mathfrak{T}_f$.

Theorem 5.4. If Z contains a noncountable set of points, there exists no state inducing a measure μ on Z and to which would corresponds $\mathfrak{T}_f$ such that all states on the system are faithfully represented on $\mathfrak{T}_f$.

Proof. Consider the family of states which induce the Dirac measure on every point of Z. Because of Theorem 5.2, μ should be nonzero for every point of Z, but the condition $\mu(Z) = 1$ cannot be satisfied.

6. SYMMETRIES

We call symmetry of our physical system an automorphism (or antiautomorphism) of the lattice τ. A group (resp. a semigroup) of symmetry is a group (resp. a semigroup) of automorphisms of τ.

This definition may be found in Ref. 26 and it is inspired by Segal's[3] work in which the evolution is considered as a group of automorphisms. We shall treat here the evolution as among the symmetries.

For the sake of simplicity, we shall not consider here anti-automorphisms, but the reader interested in them may see for instance Dixmier,[31] pp. 8–10.

It is easy to see that any automorphism of τ induces an automorphism of $\mathfrak{T}$, using a result of Emch and Piron[26]; the converse is of course trivial. We shall thus speak indifferently of automorphisms of τ or of $\mathfrak{T}$.

There are two characteristic classes of automorphisms of $\mathfrak{T}$. First an automorphism may just be a collection of automorphisms in each $\mathfrak{B}(\zeta)$; we have then a mapping $\zeta \to \zeta$ of Z identically onto itself; $\mathfrak{B}(\zeta)$ is mapped onto $\mathfrak{B}(\zeta)$ for the *same* ζ. We call such an automorphism a b-automorphism. Another possibility is to consider a mapping of Z onto itself so that $\mathfrak{H}_\zeta \to \mathfrak{H}_{\zeta'}$. This requires, of course, that $\mathfrak{H}_\zeta$ is isomorphic to $\mathfrak{H}_{\zeta'}$. This mapping induces also a mapping of $\mathfrak{B}(\zeta)$ onto $\mathfrak{B}(\zeta')$, but we do not allow for a supplementary automorphism in $\mathfrak{B}(\zeta')$, i.e., we consider the isometric mappings between the $\mathfrak{H}_\zeta$ to be fixed once for all. An automorphism of this second kind we shall call a z-automorphism.

Theorem 6.1.
Any automorphism of $\mathfrak{T}$ may be written as the product of a b- and a z-automorphism.

Proof. Follows from a similar theorem proved in Ref. 26.

Lemma 6.2. A b-automorphism is generated by a field $u(\zeta)$ of unitary operators.

Proof. Use Dixmier[31] corollary to proposition 4, p. 256.

Theorem 6.3. Let f be a state which induces the measure μ on Z and to which corresponds the algebra $\mathfrak{T}_f$. If U is a b-automorphism of $\mathfrak{T}$ and if the field of operators $u(\zeta)$ from Lemma 6.2 is a measurable field of operators, U is represented by an

automorphism of $\mathfrak{T}_f$ generated by the unitary operator

$$u = \int^{\oplus} u(\zeta)\, d\mu(\zeta), \qquad u \in \mathfrak{T}_f$$

The proof is trivial remembering the construction of $\mathfrak{T}_f$. We remark that by respect to $\mathfrak{T}$ or $\mathfrak{T}_f$, U is what is usually called an inner automorphism. We remark further that we cannot represent the b-automorphisms in $\mathfrak{T}_f$ unless the field $u(\zeta)$ corresponding to U is measurable. This means that physically speaking, we have to retain only the states for which $u(\zeta)$ is a measurable field of operators. Only for these states has the symmetry to which the automorphism corresponds, a meaning. These states are called compatible with the given b-automorphism, and the corresponding algebra is called compatible with it.

Let us now consider z-automorphisms W. Such an automorphism will map a μ-measurable field of operators $a(\zeta)$ on the field $a(W\zeta)$. But the field $a(W\zeta)$ needs no longer be μ-measurable. Thus a z-automorphism only has a meaning for the states which induce measures on Z such that a measurable field of operators is mapped onto a measurable field.

This is, however, still not enough. It may happen that W maps a field of operator $a(\zeta)$ to which corresponds the zero operator of $\mathfrak{T}_f$ onto a field $a(W\zeta)$ to which corresponds a nonzero operator in $\mathfrak{T}_f$, or vice versa. In such a case W will obviously not be represented by an automorphism of $\mathfrak{T}_f$.

Definition. A state $f(.) = \int f_\zeta(.)\, d\mu(\zeta)$ to which corresponds $\mathfrak{T}_f$, is said to be compatible with a z-automorphism W if W induces an automorphism of $\mathfrak{T}_f$.

Following immediately from the above discussion, we have

Theorem 6.4. A state f is compatible with a z-automorphism W if and only if any measurable field of operators is mapped on a measurable field and if the measures μ and $\mu^W[\mu^W(y) \equiv \mu(Wy)]$ are equivalent.

Theorem 6.5. Let V be an automorphism of $\mathfrak{T}$ and $f(.)$ a state compatible with V(i.e., compatible both with the b- and z-automorphism parts of V). Let $\mathfrak{T}_f$ be the algebra corresponding to $f(.)$. Then V induces an automorphism of $\mathfrak{T}_f$ which is implemented by a unitary operator $v \in \mathfrak{B}(\mathfrak{H})$. $v \in \mathfrak{T}_f$ if and only if the z-automorphism part of V reduces to the identity.

Proof. Use Lemma 6.2. for the b-automorphism part. For the z-automorphism part use th. 6.4. and Dixmier,[31] corollary p. 253. The last statement follows from Theorem 4.11 and Proposition p. 255 and corollary, p. 256 of Dixmier,[31] remarking that a z-automorphism does not leave the center of $\mathfrak{T}_f$ elementwise-invariant, as does a b-automorphism.

We shall show in another paper[32] that there exist physical models in field theory which provide examples of continuous group of z-automorphisms, i.e., of symmetries which do not leave the center elementwise-invariant. But it is a somehow surprising fact that the class of possible groups of z-automorphisms is rather restricted. Araki[16] has proven

Theorem 6.6. Let $T(x)$ be the unitary operators representing the translation part of the Poincaré group and assume that

$$T(x) = \int e^{i(x,p)}\, E(dp)$$

satisfies the spectrum condition, i.e., that for outside the forward cone

$$V_+ = \{p; (p,p) \geq 0, p^0 \geq 0\}, \; E(\Delta) = 0,$$

$$(x,p) = x^0p^0 - \sum_{i=1}^{3} x^ip^i$$

and $E(\Delta)$ is a projector for any four-dimensional Borel set Δ. Then if $T(x)$ generate a group of automorphisms of a von Neumann algebra $\mathfrak{N}$, and if there exists a cyclic state by respect to $\mathfrak{N}$, invariant under $T(x)$, then $T(x)$ will leave the center of $\mathfrak{N}$ elementwise-invariant, i.e.,

$$T(x)aT(x)^{-1} = a, \qquad \forall a \in \mathfrak{N} \cap \mathfrak{N}'.$$

Corollary 6.7. On a classical system, the translation part of the Poincaré group cannot satisfy the spectrum condition of Theorem 6.6. without implying that the translation part of the group is represented by the identity automorphism.

Corollary 6.8. Let $A(t)$ be the family of unitary operators representing an Abelian group of automorphisms of a von Neumann algebra $\mathfrak{N}$, and let $A(t) = e^{iHt}$, $H = H^*$. Suppose further that there exists a cyclic state for $\mathfrak{N}$, invariant under $A(t)$. If the group does not leave the center $\mathfrak{N}' \cap \mathfrak{N}$ of $\mathfrak{N}$ elementwise-invariant, H is not an operator with nonnegative spectrum.

We shall end this section with a remark on automorphisms of a subset of $\mathfrak{T}$ only.

Let $\mathfrak{S}$ be a *-subalgebra of $\mathfrak{T}$. Given a state $f(.)$

[32] G. Emch and M. Guenin, "Gauge Invariant formulation of the BCS-model," preprint.

and the corresponding algebra $\mathfrak{T}_f$, we shall get a *-subalgebra $\mathfrak{S}_f$ of $\mathfrak{T}_f$. It is not true that any automorphism of $\mathfrak{T}$ will induce an automorphism of $\mathfrak{S}$, neither is it true that any automorphism of $\mathfrak{S}$ can always be extended to an automorphism of $\mathfrak{T}$.

Definition. We call partial symmetry an automorphism of a *-subalgebra $\mathfrak{S}$ of $\mathfrak{T}$ which cannot be extended to an automorphism of $\mathfrak{T}$.

In the case where the algebra $\mathfrak{T}$ is the algebra $\mathfrak{B}(\mathfrak{H})$ of all bounded operators on a Hilbert space $\mathfrak{H}$ and $\mathfrak{S}$ a *-subalgebra of $\mathfrak{B}(\mathfrak{H})$, then an automorphism of $\mathfrak{S}$ which is not unitarily implementable, i.e., is only an algebraic and not a spatial automorphism, will provide such an example of partial symmetry. In our mind, however, what is important is not the question whether the automorphism is unitarily implementable or not, which is important, but the fact that the automorphism can be extended to an automorphism of $\mathfrak{T}$ or not. One can give examples of two von Neumann algebras $\mathfrak{N}$ and $\mathfrak{A}$ such that $\mathfrak{N} \subset \mathfrak{A}$ (one can even admit that $\mathfrak{A}$ is a factor of type I_∞) and an automorphism of $\mathfrak{N}$ which is implemented by a unitary operator but cannot be extended to an automorphism of $\mathfrak{A}$.

We think that one has yet to make a full theory of the partial symmetries, and to study their possible connections to the broken symmetries observed in physics.

7. OBSERVABLES

Up to now, we have only defined propositions, but we have not yet assigned certain operators of $\mathfrak{F}$ to be observables. In fact the techniques employed in Secs. 5 and 6 are more general, and can be used for cases where the algebras are not generated by propositions, but for instance by the field operators in a field theory; the only point is that in this case we get in general an algebra $\mathfrak{A}(\zeta)$ which can be a subalgebra of $\mathfrak{B}(\zeta)$ only.

The purpose of this section is to define observables, and to introduce a classification among them.

The rough idea is to call observable any mathematical object which can be constructed out of propositions.

Using the spectral theorem, it is possible to construct in this way all self-adjoint operators of $\mathfrak{T}$, and as any operator of $\mathfrak{T}$ is the sum of two self-adjoint ones with suitable coefficients, we can cover $\mathfrak{T}$ in this way. It is customary to consider only self-adjoint operators as being observable, because in any state they will have real expectation value.

Let a be a field $a(\zeta)$ of self-adjoint operators. Each $a(\zeta)$ can be written, using the spectral theorem, $a(\zeta) = \int \lambda \, dE_\zeta(\lambda)$. For any Borel set Δ of the real line, $E_\zeta(\Delta)$ is a projector of $\mathfrak{B}(\zeta)$. If we now consider the field $a(\zeta)$, to each Borel set Δ will correspond a field of projector $E_\zeta(\Delta)$. This field is in fact a proposition. The collection of all fields $E_\zeta(\Delta)$ corresponding to a field of self-adjoint operators we shall call shortly the spectral family of the field.

Definition. We call *class-A-observable,* a field $a(\zeta)$ of operators such that there exists a state f inducing a measure μ on Z and $\mathfrak{G}$ such that the spectral family of the field is a collection of measurable fields of projectors.

An important subset of the class of all class-A observables is introduced by the

Definition. A *class-B observable,* is a class-A observable such that there exists a state f and the corresponding algebra $\mathfrak{T}_f$ with the property that any nonzero proposition of the spectral family of the observable is represented by a nonzero projector of $\mathfrak{T}_f$. We have the

Theorem 7.1. A proposition is a class-B observable.

Proof. Apply the definition and Lemma 3.4.

Theorem 7.2. If a is a class-B observable, then there exists a state f on a $\tau'_f \subset \tau$ such that it takes different values for all propositions of its spectral family. This state is called the separating state of the given observable.

Proof. $\mathfrak{T}_f$ and $\mathfrak{H}$ from the definition of a class-B observable. Let $\mathfrak{A}$ be the Abelian von Neumann algebra generated by a. Being a subalgebra of $\mathfrak{T}_f$, $\mathfrak{A}$ is of denumerable type of Lemma 4.12. Thus by Dixmier,[31] corollary, p. 20, there exists a separating element $x \in \mathfrak{H}$ for $\mathfrak{A}$. The state defined by x on $\mathfrak{T}_f$ fulfills all conditions of the theorem.

We do not think one should assign any physical value to fields of self-adjoint operators which are not class-A observables. But, of course, we cannot prevent anybody from combining a noncountable number of class-A observables and eventually getting out of the class A. We shall thus refer to any field of self-adjoint operators which is not a class-A observable as a class-C observable.

There exists a more curious class of possible observables which appear when we consider theories invariant under a symmetry transformation which is a z-automorphism. For the sake of simplicity, let us make the discussion in a special representation. Choose a state f compatible with the given auto-

morphism, and construct $\mathfrak{T}_f$. By Theorem 6.5, a z-automorphism of $\mathfrak{T}_f$ will be generated by a unitary operator w which belongs to $\mathfrak{B}(\mathfrak{H})$ but not to $\mathfrak{T}_f$. This operator w can be written as e^{ih}, with h a self-adjoint operator. One has the feeling that it is not "fair" to deny to an h constructed in this way the name of observable, because according to our general principle they are constructed, although in an indirect way, from the structure of $\mathfrak{T}$. On the other hand, as h does not belong to $\mathfrak{T}_f$, it is not possible that the spectral projections of h belong to $\mathfrak{T}_f$, and one easily sees that they cannot be propositions. The question is, what is then their physical meaning? We can turn the problem the other way round and ask: are there observables in physics which cannot be decomposed in a family of propositions? We think the answer is yes. Piron[2] has already remarked that τ should not be considered as the lattice corresponding to a new logic. In particular, he says that the statement: "proposition a is true," is not a proposition. We shall go further and remark that, in general, we should not expect that expressions like the sign of the ratio of the value of a state on two specified sets of observables to correspond to a proposition, and neither, more generally, any expression obtained by combining expectation values of sets of observables.

We think that parity measurement provides an example of observable which cannot be expressed by propositions; the statement: "the system has odd (or even) parity" has no meaning. The only meaningful one is: "the system A has odd (or even) parity by respect to the system B." But this involves precisely the combination of expectation values for two sets of observables.

They may be other examples, for instance para-statistics, or some gauge transformations of the type met in the BCS model,[31] but we have no clear idea of them now.

Anyway, we shall call class-Z observables, self-adjoint operators like h, derived as generators of z-automorphisms.

8. POSSIBLE GENERALIZATIONS

The first possible generalization would be to drop Axiom A. If one does this, the difficulty is to prove the analog of the theorem of Piron (Theorem 1.1.) The natural result to expect is that τ is then isomorphic to the lattice of all projectors in a family $\mathfrak{A}(\zeta)$ of von Neumann algebras in a family $\mathfrak{H}_\zeta$ of Hilbert spaces. With this result, it would be then easy to do the same construction as ours, with $\mathfrak{T}_f$ being again a von Neumann algebra, but of course no longer necessarily of discrete type. In particular, we expect Theorem 6.5 to be then not necessarily still valid.

However, we want to call the attention of the reader upon the fact that constructing a model for which the algebra of observable is not of discrete type, is not a contradiction to Axiom A. In the spirit of our work, it is merely an indication that this algebra does not contain *all* observables.

Another axiom which can be attacked, is the Axiom T. One could object that it is not very physical to assume the intersection of arbitrary sets of propositions and that one should only postulate it for countable sets. τ would then be only a σ-lattice instead of a complete lattice. We think we have at least partly taken care of this objection by defining our states only on a σ-sublattice of τ, but, of course, dropping Axiom T, and replacing it by intersection for countable families only, is a possible generalization.

A much deeper problem is the question of normalcy of states (E_N). To be honest, we have to say that we do not know at all if this restriction have any physical meaning; we have made it only to be able to solve the problem mathematically. The main objection to this postulate is that it is only invariant under normal homomorphisms of the algebra. Its main support is that at least σ-additivity is always postulated for probabilities, and in a separable Hilbert space, normalcy is equivalent to σ-additivity.

Still another possible generalization would be to drop E_3 [and then to replace E_4 by (d) of Lemma 3.1.]; that is, not to normalize the states, which is, of course, only of interest in the case $\omega(1) = +\infty$. Such a generalization could be useful for some models and we do not think that there is any fundamental difficulty which would prevent our formalism from being adapted to also contain this case.

ACKNOWLEDGMENTS

The idea that measures on Z should determine in some way the representations of observables appeared in numerous discussions between the author and Dr. C. Piron in Geneva, spring and summer 1964, and I would like to express my warm thanks to Piron for his collaboration. I am also very much indebted to Professor V. Bargman and Professor A. S. Wightman for many comments and suggestions and thanks are due to Professor J. M. Jauch, and Professor R. V. Kadison, as well as to Dr. G. Emch and Dr. B. Misra for discussions. Finally, I want to express my gratitude to the Fond Birkigt of the University of Geneva for a travel grant.

On the Reduction of Direct Products of the Irreducible Representations of $SU(n)$

H. HARARI
Israel Atomic Energy Commission, Soreq Nuclear Research Centre, Yavne, Israel
(Received 13 July 1965)

Explicit simple formulas are given for the reduction of some direct products of the irreducible representations of $SU(n)$. All the interesting products from the point of view of the various symmetry models in particle physics are included as special cases of our results. In particular, all products of the totally symmetric and the regular representations of $SU(n)$ are evaluated.

THE study of the symmetry properties of elementary particles has recently led to various symmetry schemes, most of which are based on the groups $SU(n)$ of unitary unimodular matrices of order n. In addition to the "old" $SU(2)$ description of ordinary spin and isospin, extensive use is made of $SU(3)$, $SU(4)$, $SU(6)$, $SU(8)$, and $SU(12)$. In all these models particles are classified into different irreducible representations (IR's) of the relevant group and the analysis of reactions or symmetry breaking interactions involves the reduction of direct products of IR's. The problem of finding all the IR's which appear in the product of any two representations of $SU(n)$ is usually treated according to the Littlewood theorem.[1] However, the use of higher and higher symmetry groups, whose IR's are of large dimensionalities, makes the explicit calculation somewhat tedious in many cases. The purpose of the present note is to give explicit simple formulas for the reduction of various direct products which are often encountered in particle physics. In particular, all the physically interesting products of $SU(3)$, $SU(4)$, $SU(6)$, $SU(8)$, and $SU(12)$ IR's are obtained as special cases of our formulas.

IR's of $SU(n)$ (or of the Lie algebra A_{n-1}) are usually denoted either by a Young diagram $[f_1, f_2, \cdots, f_{n-1}]$ where $f_i \geq f_{i+1}$, $f_i \geq 0$, or by the maximal weight notation $(\lambda_1, \lambda_2, \cdots, \lambda_{n-1})$. The two notations are related by

$$\lambda_i = f_i - f_{i+1} \quad \text{for} \quad i = 1, \cdots, n-2; \qquad \lambda_{n-1} = f_{n-1}. \tag{1}$$

We shall use the second notation. The dimensionality of $(\lambda_1, \cdots, \lambda_{n-1})$ is given by

$$d(\lambda_1, \cdots, \lambda_{n-1}) = \frac{P_1 \cdot P_2 \cdots P_{n-1}}{1!2! \cdots (n-1)!}$$

where

$$\begin{aligned} P_1 &= \mu_1\mu_2\mu_3 \cdots \mu_{n-1} \\ P_2 &= (\mu_1 + \mu_2)(\mu_2 + \mu_3) \cdots (\mu_{n-2} + \mu_{n-1}) \\ P_3 &= (\mu_1+\mu_2+\mu_3)(\mu_2+\mu_3+\mu_4)\cdots(\mu_{n-3}+\mu_{n-2}+\mu_{n-1}) \\ &\vdots \\ P_{n-2} &= (\mu_1 + \mu_2 + \cdots + \mu_{n-2})(\mu_2 + \mu_3 + \cdots + \mu_{n-1}) \\ P_{n-1} &= (\mu_1 + \mu_2 + \cdots + \mu_{n-1}) \end{aligned} \tag{2a}$$

and

$$\mu_i = \lambda_i + 1 \quad \text{for} \quad i = 1, \cdots, n-1. \tag{2b}$$

For $\lambda_1 = k$, $\lambda_2 = l$, $\lambda_3 = \lambda_4 = \cdots \lambda_{n-1} = 0$, Eq. (2) reduces to

$$d(k, l, 0, \cdots, 0) = \frac{k+1}{k+l+1}\binom{l+n-2}{l}\binom{k+l+n-1}{k+l}. \tag{3}$$

If $l = 0$ we obtain

$$d(k, 0, \cdots, 0) = \binom{n+k-1}{k}, \tag{4}$$

while if $k = 0$

$$d(0, l, 0, \cdots 0) = \frac{1}{l+1}\binom{n+l-1}{l}\binom{n+l-2}{l}. \tag{5}$$

Another important family of IR's is that of $\lambda_1 = \lambda_{n-1} = k$; $\lambda_2 = \lambda_3 = \cdots = \lambda_{n-2} = 0$. In this case we obtain

$$d(k, 0, \cdots, 0, k) = \frac{n+2k-1}{n-1}\binom{n+k-2}{k}^2. \tag{6}$$

For $k = 1$ we obviously get the adjoint representation

$$d(1, 0, \cdots 0, 1) = n^2 - 1. \tag{7}$$

We now deal separately with four sets of products of representations.

[1] D. E. Littlewood, *Theory of Group Characters* (Oxford University Press, London, 1950).

(1) The product of two adjoint representations:

$$(1, 0, \cdots, 0, 1) \otimes (1, 0, \cdots, 0, 1)$$
$$= (2, 0, \cdots, 0, 2)_s \oplus (2, 0, \cdots, 0, 1, 0)_a$$
$$\oplus (0, 1, 0, \cdots, 0, 2)_a \oplus (0, 1, 0, \cdots, 0, 1, 0)_s$$
$$\oplus (1, 0, \cdots, 0, 1)_s \oplus (1, 0, \cdots, 0, 1)_a \oplus (0, \cdots, 0)_s. \tag{8}$$

If we denote the dimensionalities of the representations by $[d]$ we obtain in this case

$$[n^2 - 1] \otimes [n^2 - 1] = [\tfrac{1}{4}n^2(n+3)(n-1)]_s$$
$$\oplus [\tfrac{1}{4}(n^2-1)(n^2-4)]_s \oplus [\overline{\tfrac{1}{4}(n^2-1)(n^2-4)}]_a$$
$$\oplus [\tfrac{1}{4}n^2(n-3)(n+1)]_s \oplus [n^2-1]_s$$
$$\oplus [n^2-1]_a \oplus [1]_s. \tag{9}$$

$[\bar{d}]$ is the conjugate representation to $[d]$. The subscripts s and a denote representations which are, respectively, symmetric and antisymmetric with respect to a permutation of the two multiplied equivalent IR's. Note that for $n \geq 4$ there are always seven IR's in the product, and the adjoint representation appears twice. For $n = 3$ the dimensionalities obtained are: $[27]_s + [10]_a + [\overline{10}]_a + [0]_s + [8]_s + [8]_a + [1]_s$. Hence, only six IR's appear in the reduction, the "$[0]_s$" being excluded.

(2) Another interesting case is that of multiplying two completely symmetric IR's of $SU(n)$. This product reduces according to (we assume $l \leq k$)

$$(k, 0, \cdots, 0) \otimes (l, 0, \cdots, 0) = \sum_{p=0}^{l} (k + l - 2p, p, 0, \cdots, 0) \tag{10}$$

or, in terms of the dimensionalities,

$$\left[\binom{n+k-1}{k}\right] \otimes \left[\binom{n+l-1}{l}\right]$$
$$= \sum_{p=0}^{l} \left[\frac{k+l-2p+1}{k+l-p+1} \times \binom{n+k+l-p-1}{k+l-p}\binom{n+p-2}{p}\right]. \tag{11}$$

For $SU(2)$ we naturally obtain

$$[k+1] \otimes [l+1] = \sum_{p=0}^{l} [k+l+1-2p]. \tag{12}$$

In the special case of $k = l$, (10) and (11) are reduced to

$$(k, 0, \cdots, 0) \otimes (k, 0, \cdots, 0) = \sum_{p=0}^{k} (2k - 2p, p, 0, \cdots, 0), \tag{13}$$

$$\left[\binom{n+k-1}{k}\right] \otimes \left[\binom{n+k-1}{k}\right]$$
$$= \sum_{p=0}^{k} \left[\frac{2k-2p+1}{2k-p+1} \times \binom{n+2k-p-1}{2k-p}\binom{n+p-2}{p}\right]. \tag{14}$$

Even (odd) values of p represent symmetric (antisymmetric) IR's with respect to permutation of the factors. The case $k = 3$ is especially interesting in symmetry models such as $SU(6)$ and $SU(12)$, since the baryons are described by the totally symmetric state of three n-component quarks. With $k = 3$ we get from (13)

$$[\tfrac{1}{6}n(n+1)(n+2)] \otimes [\tfrac{1}{6}n(n+1)(n+2)]$$
$$= \left[\binom{n+5}{6}\right]_s \oplus \left[5 \cdot \binom{n+4}{6}\right]_a \oplus$$
$$\oplus \left[\frac{3n}{2}\binom{n+3}{5}\right]_s \oplus \left[\frac{1}{4}\binom{n+2}{3}\binom{n+1}{3}\right]_a. \tag{15}$$

These are the possible independent channels for a baryon–baryon scattering process in these models.

(3) We proceed now to the product of a totally symmetric IR with its conjugate representation. We find

$$(k, 0, \cdots, 0) \otimes (0, \cdots, 0, k) = \sum_{p=0}^{k} (p, 0, \cdots, 0, p) \tag{16}$$

or

$$\left[\binom{n+k-1}{k}\right] \otimes \left[\overline{\binom{n+k-1}{k}}\right]$$
$$= \sum_{p=0}^{k} \left[\frac{n+2p-1}{n-1}\binom{n-p-2}{p}^2\right]. \tag{17}$$

Note that the dimensionalities of the IR's do not depend on k. For higher values of k we just get additional terms.

For $n = 2$

$$[k+1] \otimes [k+1] = \sum_{p=0}^{k} (2p+1), \tag{18}$$

while for $SU(3)$

$$(k, 0) \otimes (0, k) = \sum_{p=0}^{k} (p, p) \tag{19}$$

or

$$[\tfrac{1}{2}(k+1)(k+2)] \otimes [\overline{\tfrac{1}{2}(k+1)(k+2)}] = \sum_{p=0}^{k} [(p+1)^2]. \tag{20}$$

The baryon–antibaryon system is described in some symmetry schemes by the product (16) or (17), for $k = 3$. This is the case for the $SU(6)$ and $SU(12)$ theories. Substituting $k = 3$ in (17) we get

$$[\tfrac{1}{6}n(n+1)(n+2)] \otimes [\overline{\tfrac{1}{6}n(n+1)(n+2)}] = [1] \oplus \oplus [n^2-1] \oplus [\tfrac{1}{4}n^2(n+3)(n-1)] \oplus \oplus [\tfrac{1}{36}n^2(n+1)^2(n-1)(n+5)]. \quad (21)$$

Combining this with Eq. (9) we conclude that in these theories the baryon–antibaryon annihilation into two mesons will always proceed via four different channels: $(0, \cdots, 0)$, $(1, 0, \cdots, 0, 1)_s$, $(1, 0, \cdots, 0, 1)_a$, and $(2, 0, \cdots, 0, 2)$, whose dimensionalities are $[1]$, $[n^2-1]_s$, $[n^2-1]_a$, and $[\tfrac{1}{4}n^2(n+3)(n-1)]$, respectively.

(4) Finally we present a formula for reducing the product of the adjoint representation and a totally symmetric one. This is encountered in calculating meson–baryon reactions in some higher symmetry models. We obtain

$$(k, 0, \cdots, 0) \otimes (1, 0, \cdots, 0, 1) = (k+1, 0, \cdots, 0, 1) \oplus (k-1, 1, 0, \cdots, 0) \oplus \oplus (k, 0, \cdots, 0) \oplus (k-2, 1, 0, \cdots, 0) \quad (22)$$

or

$$\left[\binom{n+k-1}{k}\right] \otimes [n^2-1] = \left[(n+k+1)\binom{n+k-1}{k+1}\right] \oplus \oplus \left[\frac{kn(k+n)}{n-1}\binom{k+n-2}{k+1}\right] \oplus \oplus \left[\binom{n+k-1}{k}\right] \oplus \left[(k-1)\binom{k+n-2}{k}\right]. \quad (23)$$

For $k = 3$, Eq. (23) leads to

$$[\tfrac{1}{6}n(n+1)(n+2)] \otimes [n^2-1] = \left[(n+4)\binom{n+2}{4}\right] \oplus [\tfrac{1}{8}n^2(n+1)(n-2)(n+3)] \oplus \oplus [\tfrac{1}{6}n(n+1)(n+2)] \oplus [\tfrac{1}{3}n(n^2-1)]. \quad (24)$$

The author would like to acknowledge helpful discussions with M. Kugler, H. J. Lipkin, and Y. Ne'eman.

Monomer Pair Correlations

ROBERT E. HARTWIG
Mathematics Department, University of Adelaide, Adelaide, South Australia
(Received 20 September 1965)

In this paper we evaluate the monomer pair correlation along the diagonals $(p, p+1)$ and $(p+1, p)$ of a square lattice otherwise packed with dimers. Using the perturbed Pfaffian technique the correlation can be expressed as a Toeplitz determinant $i^p\,|b_{i-j+1}|/2x$ generated by the function

$$B(\theta) = \sum_{-\infty}^{\infty} b_k e^{ik\theta} = ire^{i\theta}\left[\frac{\operatorname{sgn}(\sin\theta) + ire^{i\theta}}{1 + ire^{i\theta}}\right] + \sum_{k=0}^{\infty}\left(\frac{i}{r}\right)^k e^{-ik\theta},$$

where $r = x/y$ and x and y are the activities of x and y dimers. We will calculate the determinant exactly and prove that the correlation decays with increasing monomer separation as $B/4r^{\frac{1}{2}}$; where B is simply related to the decay constant of the diagonal spin correlation at the critical point of a square ferromagnetic Ising lattice. The exact value is found to be $B = 0.989487291$.

1. INTRODUCTION

IT has been shown recently[1] by Fisher and Stephenson (FS) that the perturbed Pfaffian technique could be used to solve certain problems related to planar lattices in which all but a few lattice sites were covered by dimers such as to form monomers or holes in a sea of dimers. This technique uses the aid of a Green's function to evaluate perturbed Pfaffians in terms of lower-order Pfaffians. On using this, the interaction between two monomers can in certain cases be expressed in terms of a Toeplitz determinant $|b_{i-j+1}|$ whose entries are derived from the basic Green's function for the problem. The elements of the Green's function are double trigonometric integrals which give rise to elementary functions and thus reflect the fact that the correlation shows no singularity, as will be seen from the exact results.

Taking the paper by FS as our starting point, we show that at the center of an infinite lattice, the monomer pair correlation can be written as a perturbed determinant which in turn can be evaluated in terms of Legendre functions whose order depends on the radial separation distance r. A close analogy with the critical point spin correlation becomes apparent from the perturbed determinant.

From the exact results the asymptotic expansions in terms of the separation r will be found.

In the general nonsymmetric case $(r \neq 1)$ the decay is found to be

$$4\omega(x, y \mid p-1, p) \sim \frac{2B_0^2}{r^{\frac{1}{2}}}\left[\frac{2}{x^2+y^2}\right]^{\frac{1}{2}}\left[1 - \frac{1}{2r\sqrt{2}}\left(\frac{x^2-y^2}{x^2+y^2}\right)\right].$$

When $r = 1$ the second term drops out and we obtain

$$4\omega(x, x \mid p-1, p) \sim (2B_0^2/xr^{\frac{1}{2}})[1 - (4r^2)^{-1}]$$

as conjectured by FS on the basis of numerical calculations. The constant B_0 is the coefficient of the leading term in the asymptotic series of the diagonal spin correlation at the critical point as a function of radial distance.

2. THE GREEN'S FUNCTION

In this paragraph we will for the sake of clarity reproduce some of the results derived in FS.[2]

The Green's function matrix is the inverse G of the basic skew-symmetric counting matrix D_0, whose Pfaffian gives the dimer partition function. At the center of an infinite lattice the elements of the Green's function are given and denoted by

$$G(r, s \mid r', s') = [r' - r \mid s' - s] = [t, u] = \frac{1}{(2\pi)^2}\iint_0^{2\pi}\frac{M(t, u \mid \alpha, \beta)}{\Delta(\alpha, \beta)}\,d\alpha\,d\beta, \tag{2.1}$$

where

$$\Delta(\alpha, \beta) = x^2(1 - \cos\alpha) + y^2(1 - \cos\beta) \tag{2.2}$$

and[3]

$$M(t, u \mid \alpha, \beta) = \begin{cases} 0 & (t, u \text{ both even}) \\ \tfrac{1}{2}x\left[\cos(t+1)\frac{\alpha}{2} - \cos(t-1)\frac{\alpha}{2}\right]\cos\left(\frac{\beta u}{2}\right) & (t \text{ odd}, u \text{ even}) \\ -\tfrac{1}{2}iy\cos\left(\frac{\alpha t}{2}\right)\left[\cos(u+1)\frac{\beta}{2} - \cos(u-1)\frac{\beta}{2}\right] & (t \text{ even}, u \text{ odd}). \end{cases} \tag{2.3}$$

[1] M. E. Fisher and J. Stephenson, Phys. Rev. 132, 1411 (1963).

[2] See Ref. 1, Sec. 4, p. 1415.

[3] We note that there is a misprint sign in the last line of the definition of $M(t, u \mid \alpha, \beta)$ in FS; Sec. 4, p. 1416.

Two important properties of the Green's function are

(i) the symmetry relations

$$[t, u] = (-1)^t[-t, u] = (-1)^u[t, -u] = -[-t, -u], \quad (2.4)$$

(ii) the recurrence relation $GD = I = DG$, i.e.,

$$x[t-1, u] + iy[t, u-1] - x[t+1, u] - iy[t, u+1] = \delta_{0t}\delta_{0u}. \quad (2.5)$$

The determinantal expression for the monomer pair correlation along the diagonals $(p, p+1)$ and $(p+1, p)$ is given by[4]

$$\omega(x, y \mid p-1, p) = \frac{i^p}{2x} |b_{i-j+1}| \quad i, j = 1, \cdots, p,$$

where

$$b_0 = 1 + 2x[1, 0], \; b_l = -2x[l-1, l]$$

and

$$b_{-l} = (i/\tau)^l + 2x[l+1, l] \quad l = 1.2, \cdots. \quad (2.6)$$

3. AN OUTLINE OF THIS PAPER

We first prove a relation between b_k and b_{k+1} for general k, which enables us to reduce the correlation determinant to a new perturbed determinant of simple block structure. Then after inverting the relevant Cauchy matrix the perturbation will be expressed as an integral involving the Legendre polynomial $P_n(\cos\theta)$. The integral is then evaluated exactly for the symmetric and nonsymmetric cases using the theory of the Legendre functions. From the exact perturbation, the asymptotic values of the correlations are found, which check with the numerical estimates made in FS.

4. IDENTITIES

Because of the special structure of the Green's function, it is possible to show for general τ and positive and negative k that

$$b_k - \frac{i}{\tau} b_{k+1} = \begin{cases} 0 & \text{(k even or zero)} \\ \frac{2}{\pi i} \cdot \frac{1}{k} & \text{(k odd)} \end{cases}, \quad (4.1)$$

where the b_k's are given as in (2.6). The fact that this difference is independent of τ enables us to evaluate the monomer correlation exactly. Furthermore, it enables us to find the b_k's and their generating function $B(\theta)$ such that $B(\theta) = \sum_{-\infty}^{\infty} b_k e^{ik\theta}$. When $\tau = 1$ the values of b_k reduce to those given in FS.[5]

If we take for even and odd k the values $k = 2r$ and $2r-1$, respectively, then we have to show

(i) $$2x[2r-1, 2r] = 2iy[2r, 2r+1] \quad (k = 2r), \quad (4.2)$$

(ii) $$-2x[2r-2, 2r-1] + 2iy[2r-1, 2r] = \frac{2}{\pi i} \cdot \frac{1}{2r-1} \quad (k = 2r-1), \quad (4.3)$$

(iii) $$2x[2r+1, 2r] = 2iy[2r, 2r-1] \quad (k = -2r), \quad (4.4)$$

(iv) $$2x[2r, 2r-1] - 2iy[2r-1, 2r-2] = \frac{2}{\pi i} \frac{-1}{2r-1} \quad (k = -2r+1). \quad (4.5)$$

First of all, we verify the identity for the case when $k = 0$ because this is an exceptional case for which the following analysis does not go through.

Now from definition (2.6),

$$\begin{aligned} b_0 - ib_1/\tau &= 1 + 2x[1, 0] - 2iy[0, 1] \\ &= 1 - 2\rho_x - 2\rho_y \\ &= 1 - 2(\rho_x + \rho_y) \\ &= 0, \end{aligned}$$

where

$$\rho_x = (1/\pi) \tan^{-1}(\tau)$$

and

$$\rho_y = (1/\pi) \tan^{-1}(1/\tau) = \tfrac{1}{2} - \rho_x$$

as in FS.

And so we can proceed to the general case, in which we have from the definition (2.1)

$$[2r+1, 2s] = \frac{1}{(2\pi)^2} \iint_0^{2\pi} \tfrac{1}{2}x[\cos(r+1)\alpha - \cos r\alpha] \cos(s\beta) \frac{d\alpha\, d\beta}{\Delta(\alpha, \beta)} \quad (4.6)$$

and

$$[2s, 2r+1] = \frac{1}{(2\pi)^2} \iint_0^{2\pi} -\frac{iy}{2} \cos(s\alpha) \times [\cos(r+1)\beta - \cos r\beta] \frac{d\alpha\, d\beta}{\Delta(\alpha, \beta)}.$$

We will use these two integrals in their exponential form to prove the four identities.

[4] See Ref. 1, Sec. 10, p. 1427. The factor $\pm i$ should read i^p.

[5] See Ref. 1, Appendix B, p. 1430.

(i) $k = 2r$.

Starting with the case when k is positive and even we have that

$$2x[2r-1, 2r] = \frac{x^2}{(2\pi)^2}\iint_0^{2\pi} e^{ir(\alpha+\beta)}[1-e^{-i\alpha}]\frac{d\alpha\,d\beta}{\Delta(\alpha,\beta)} \tag{4.7}$$

and

$$2iy[2r, 2r+1] = \frac{y^2}{(2\pi)^2}\iint_0^{2\pi} e^{ir(\alpha+\beta)}[e^{i\beta}-1]\frac{d\alpha\,d\beta}{\Delta(\alpha,\beta)}, \tag{4.8}$$

where

$$\Delta(\alpha,\beta) = (x^2+y^2) - x^2\cos\alpha - y^2\cos\beta = a - \gamma_1\cos\alpha - \gamma_2\cos\beta.$$

We note here the similarity to the random-walk integrals on the square Ising lattice.[6] The only difference lies in the simple relation between the parameters a, γ_1, and γ_2 which causes the integrals to become elementary.

Forming the difference of (4.7) and (4.8) we have

$$2x[2r-1, 2r] - 2iy[2r, 2r+1] = \frac{1}{(2\pi)^2}\iint_0^{2\pi} e^{ir(\alpha+\beta)}\{x^2(1-e^{-i\alpha}) + y^2(1-e^{i\beta})\}\frac{d\alpha\,d\beta}{\Delta(\alpha,\beta)}. \tag{4.9}$$

To evaluate this we change the variables such that

$$\left.\begin{aligned}\alpha &= \theta+\phi\\ \beta &= \theta-\phi\end{aligned}\right\} \quad\text{and}\quad \begin{aligned}\theta &= \tfrac{1}{2}(\alpha+\beta)\\ \phi &= \tfrac{1}{2}(\alpha-\beta).\end{aligned} \tag{4.10}$$

The Jacobian of the transformation equals -2 and the unit cell $[-\pi, \pi \mid -\pi, \pi]$ of the (α, β) space is doubled in the (θ, ϕ) space. (4.11)

Whence

$$2x[2r-1, 2r] - 2iy[2r, 2r+1] = \frac{1}{2\pi}\int_0^{2\pi} e^{2ri\theta}\frac{d\theta}{2\pi}\int_0^{2\pi}[x^2(1-e^{-i(\theta+\phi)}) + y^2(1-e^{i(\theta-\phi)})]\frac{d\theta\,d\phi}{\Delta(\theta,\phi)}. \tag{4.12}$$

Now

$$\begin{aligned}\Delta(\theta,\phi) &= x^2+y^2-x^2\cos(\theta+\phi) - y^2\cos(\theta-\phi)\\ &= x^2+y^2-[(x^2+y^2)\cos\theta]\cos\phi + [(x^2-y^2)\sin\theta]\sin\phi\\ &= A + B\cos\phi + C\sin\phi,\end{aligned} \tag{4.13}$$

where

$$\begin{aligned}A &= x^2+y^2,\\ B &= -(x^2+y^2)\cos\theta,\\ C &= (x^2-y^2)\sin\theta.\end{aligned} \tag{4.14}$$

Also we have that

$$\begin{aligned}&x^2[1-e^{-i(\theta+\phi)}] + y^2[1-e^{i(\theta-\phi)}]\\ &= x^2+y^2-x^2\cos(\theta+\phi) - y^2\cos(\theta-\phi)\\ &\quad + i[x^2\sin(\theta+\phi) - y^2\sin(\theta-\phi)].\end{aligned} \tag{4.15}$$

The imaginary part of the integral is identically zero since we started off with real integrals in (4.6). The real part of (4.15) equals $\Delta(\theta, \phi)$ and hence gives as required

$$\frac{1}{2\pi}\int_0^{2\pi} e^{2ri\theta}\,d\theta \equiv 0 \quad (r \neq 0). \tag{4.16}$$

(ii) To prove the second identity we will first evaluate $(-i/\tau)b_{2r}$ and then use the first identity $b_{2r} = (i/\tau)b_{2r+1}$. Now

$$(-i/\tau)b_{2r} = 2iy[2r-1, 2r] = \frac{ixy}{(2\pi)^2}\iint_0^{2\pi} e^{ir(\alpha+\beta)}[1-e^{-i\alpha}]\frac{d\alpha\,d\beta}{\Delta(\alpha,\beta)} \tag{4.17}$$

which on changing the variables becomes

$$\frac{ixy}{2\pi}\int_0^{2\pi} e^{2ri\theta}\frac{d\theta}{2\pi} \times \int_0^{2\pi}\frac{1-e^{-i\theta}e^{-i\phi}}{A+B\cos\phi+C\sin\phi}\,d\phi. \tag{4.18}$$

This can be calculated with the aid of the standard integrals

$$I_n = \frac{1}{2\pi}\int_0^{2\pi}\frac{e^{in\phi}\,d\phi}{A+B\cos\phi+C\sin\phi} = \lambda^n(A^2-B^2-C^2)^{-\frac{1}{2}} \tag{4.19}$$

and its conjugate

$$\bar{I}_n = \bar{\lambda}^n(A^2-B^2-C^2)^{-\frac{1}{2}}, \tag{4.20}$$

[6] E. W. Montroll, R. B. Potts, and J. C. Ward, J. Math. Phys. 4, 317 (1963).

where

$$A^2 > B^2 + C^2$$

and

$$\lambda = \frac{(A^2 - B^2 - C^2)^{\frac{1}{2}} - A}{B - iC}. \tag{4.21}$$

Substituting in (4.18) we get

$$\begin{aligned}\bar{I}(\theta) &= \frac{1}{2\pi}\int_0^{2\pi} \frac{1 - e^{-i\theta}e^{-i\phi}}{A + B\cos\phi + C\sin\phi}\,d\phi \\ &= \frac{1 - e^{-i\theta}\bar{\lambda}}{(A^2 - B^2 - C^2)^{\frac{1}{2}}} \\ &= \frac{(B + iC) + Ae^{-i\theta}}{(B + iC)(A^2 - B^2 - C^2)^{\frac{1}{2}}} - \frac{e^{-i\theta}}{B + iC}.\end{aligned} \tag{4.22}$$

Now from (4.14)

$$\begin{aligned} B + iC &= -y^2e^{-i\theta}[\tau^2 + e^{2i\theta}](\tau = x/y), \\ B + iC + Ae^{-i\theta} &= -2iy^2\sin\theta, \\ B^2 + C^2 &= x^4 + y^4 + 2x^2y^2\cos 2\theta, \\ A^2 - B^2 - C^2 &= 4x^2y^2\sin^2\theta, \end{aligned} \tag{4.23}$$

and finally

$$(A^2 - B^2 - C^2)^{\frac{1}{2}} = 2xy\,|\sin\theta|.$$

Equation (4.22) gives

$$\begin{aligned}\bar{I}(\theta) &= \frac{1 - e^{-i\theta}\bar{\lambda}}{(A^2 - B^2 - C^2)^{\frac{1}{2}}} \\ &= \frac{i}{xy}\frac{\sin\theta}{|\sin\theta|}\cdot\frac{e^{i\theta}}{[\tau^2 + e^{2i\theta}]} + \frac{1}{y^2[\tau^2 + e^{2i\theta}]}\end{aligned} \tag{4.24}$$

and its conjugate, which we will need later,

$$\begin{aligned}I(\theta) &= \frac{1 - e^{i\theta}\lambda}{(A^2 - B^2 - C^2)^{\frac{1}{2}}} \\ &= \frac{-i}{xy}\frac{\sin\theta}{|\sin\theta|}\cdot\frac{e^{i\theta}}{[\tau^2e^{2i\theta} + 1]} + \frac{e^{2i\theta}}{y^2[\tau^2e^{2i\theta} + 1]}.\end{aligned} \tag{4.25}$$

Hence for $r > 0$,

$$\begin{aligned}-\frac{i}{\tau}b_{2r} &= 2iy[2r - 1, 2r] \\ &= \frac{-1}{2\pi}\int_0^{2\pi}\frac{e^{(2r+1)i\theta}}{[\tau^2 + e^{2i\theta}]}\frac{\sin\theta}{|\sin\theta|}\,d\theta \\ &\quad + \frac{i\tau}{2\pi}\int_0^{2\pi}\frac{e^{2ri\theta}}{[\tau^2 + e^{2i\theta}]}\,d\theta.\end{aligned} \tag{4.26}$$

The second integral can be written as

$$\frac{\tau}{2\pi}\int_C \frac{z^{2r-1}}{\tau^2 + z^2}\,dz, \tag{4.27}$$

where C is the unit circle and $z = e^{i\theta}$.

Using the residue theorem it becomes

$$\begin{cases}(-1)^{r-1}\tau^{2r-1}i & (\tau < 1) \\ \frac{1}{2}(-1)^{r-1}i & (\tau = 1) \\ 0 & (\tau > 1).\end{cases} \tag{4.28}$$

In the first integral, the factor $\sin\theta/|\sin\theta|$ changes sign at $\theta = \pi$ so if we split the range up into the intervals $[0, \pi]$ and $[\pi, 2\pi]$, and change the variable in the second range such that $\theta = \psi + \pi$, we arrive at

$$-\frac{1}{\pi i}\int_{C_1}\frac{z^{2r}\,dz}{\tau^2 + z^2}, \tag{4.29}$$

where C_1 is the top half of the unit circle and C_2 the real interval $(-1, 1)$. On application of the residue theorem to the closed contour $C_1 + C_2$, it is easily seen to be equal to

$$\frac{1}{\pi i}\int_{-1}^{1}\frac{x^{2r}}{\tau^2 + x^2}\,dx + \begin{cases}(-1)^r\tau^{2r-1}i & (\tau < 1) \\ \frac{1}{2}(-1)^r i & (\tau = 1) \\ 0 & (\tau > 1).\end{cases} \tag{4.30}$$

The last term clearly cancels with (4.28) and so for all τ

$$\begin{aligned}-(i/\tau)b_{2r} &= 2iy[2r - 1, 2r] \\ &= \frac{2}{\pi i}\int_0^1\frac{x^{2r}}{\tau^2 + x^2}\,dx.\end{aligned} \tag{4.31}$$

On changing the variable to $x = \tau\tan\mu$ it can be seen that the integral satisfies the recurrence relation

$$X_{2r} = [\tau^{1-2r}/(2r - 1)] - X_{2r-2} \tag{4.32}$$

which can be solved to give

$$\begin{aligned}-\frac{i}{\tau}b_{2r} &= \frac{2}{\pi i}(-1)^r\tau^{2r-1}\left[\arctan\left(\frac{1}{\tau}\right)\right. \\ &\quad \left. + \sum_{q=1}^{r}\frac{(-1)^q\tau^{1-2q}}{2q - 1}\right].\end{aligned} \tag{4.33}$$

The first term clearly diverges with r if $\tau > 1$. We are now in a position to prove the second identity using (4.33) and the first identity. Now

$$b_{2r-2} = (i/\tau)b_{2r-1} \tag{4.34}$$

and thus

$$b_{2r-1} - \frac{i}{\tau}b_{2r} = \tau^2\left[-\frac{i}{\tau}b_{2r-2}\right] - \frac{i}{\tau}b_{2r}. \tag{4.35}$$

From (4.33) it can be seen that all but the last term cancel and so

$$b_{2r-1} - \frac{i}{\tau}b_{2r} = \frac{2}{\pi i}\cdot\frac{1}{2r - 1}, \tag{4.36}$$

proving the second identity. We note that this is independent of τ.

(iii) For negative indices $b_{-k} = (i/\tau)^k + 2x[k+1, k]$ and on cancelling the constant term we can write

$$b_{-k} - (i/\tau)b_{-k+1} = 2x[k+1, k] - 2iy[k, k-1] \qquad k = 1, 2, \cdots . \tag{4.37}$$

In the even case this becomes

$$b_{-2r} - \frac{i}{\tau} b_{-2r+1} = \frac{1}{(2\pi)^2} \iint_0^{2\pi} e^{-ir(\alpha+\beta)} \{x^2(e^{-i\alpha} - 1) + y^2(e^{i\beta} - 1)\} \frac{d\alpha\, d\beta}{\Delta(\alpha, \beta)}. \tag{4.38}$$

Here we used the negative exponential representation of (4.6). On changing the variables to (θ, ϕ) it is clear that the ϕ-integral is the same as in the previous case (4.12) and since

$$-\int_0^{2\pi} e^{-2ri\theta}\, d\theta \equiv 0 \qquad (r \neq 0)$$

the identity follows.

(iv) In order to prove the last identity we will first evaluate

$$-\frac{i}{\tau} b_{-2r} + \left(\frac{i}{\tau}\right)^{2r+1}$$

in a similar way as was done in (ii).

$$(i/\tau)^{2r+1} - (i/\tau)b_{-2r} = -2iy[2r+1, 2r] = \frac{ixy}{(2\pi)^2} \iint_0^{2\pi} e^{ri(\alpha+\beta)}[1 - e^{i\alpha}] \frac{d\alpha\, d\beta}{\Delta(\alpha, \beta)} = \frac{ixy}{2\pi} \int_0^{2\pi} e^{2ri\theta} \frac{d\theta}{2\pi} \int_0^{2\pi} \frac{1 - e^{i(\theta+\phi)}}{A + B\cos\phi + C\sin\phi} d\phi. \tag{4.40}$$

The second integral is the conjugate of (4.22) and by (4.25) we get

$$-2iy[2r+1, 2r] = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{(2r+1)i\theta}}{[\tau^2 e^{2i\theta} + 1]} \frac{\sin\theta}{|\sin\theta|} d\theta + \frac{i\tau}{2\pi} \int_0^{2\pi} \frac{e^{(2r+2)i\theta}}{\tau^2 e^{2i\theta} + 1} d\theta, \tag{4.41}$$

which can be written by the same argument as used before as

$$\frac{1}{\pi i} \int_{C_1} \frac{z^{2r}}{\tau^2 z^2 + 1} dz + \frac{\tau}{2\pi i} \int_C \frac{z^{2r+1}\, dz}{\tau^2 z^2 + 1}. \tag{4.42}$$

Application of the residue theorem shows again that the contributions from the possible poles cancel and we are left with

$$2iy[2r+1, 2r] = \frac{2}{\pi i} \int_0^1 \frac{x^{2r}}{\tau^2 x^2 + 1} dx. \tag{4.43}$$

Taking a factor of $1/\tau^2$ out of (4.31) and replacing τ by $1/\tau$ in (4.33) we obtain

$$2iy[2r+1, 2r] = (2/\pi i)\cdot(-1)^r \tau^{-1-2r} \times \left[\arctan(\tau) + \sum_{q=1}^{r} \frac{(-1)^q \tau^{2q-1}}{2q-1}\right]. \tag{4.44}$$

From Part (iii) we have

$$2x[2r+1, 2r] = 2iy[2r, 2r-1] \tag{4.45}$$

and using this it follows that

$$b_{-2r+1} - (i/\tau)b_{-2r+2} = 2x[2r, 2r-1] - 2iy[2r-1, 2r-2] = -\tau^2 2iy[2r+1, 2r] - 2iy[2r-1, 2r-2], \tag{4.46}$$

which by (4.44) reduces to

$$(2/\pi i)\cdot[-1/(2r-1)],$$

and thus completes the proof of (4.1).

When $\tau = 1$, the results simplify somewhat and (4.43) can be written in terms of the psi function $\psi(z) = \Gamma'(z)/\Gamma(z)$.

5. THE GENERATING FUNCTION

The calculations of the previous paragraph enable us to deduce for $\tau > 1$ the generating function $B(\theta)$ of the b_k's in a closed form

$$B(\theta) = \sum_{k=-\infty}^{\infty} b_k e^{ik\theta}. \tag{5.1}$$

From (4.17) we see that

$$b_{2r} = \frac{1}{2\pi} \int_0^{2\pi} e^{-2ri\theta}\, d\theta\, [-\tau xy I(\theta)], \tag{5.2}$$

where $I(\theta)$ is given by (4.22) and (4.25). Equation (4.39) gives

$$b_{-2r} = \left(\frac{i}{\tau}\right)^{2r} + \frac{1}{2\pi} \int_0^{2\pi} e^{2ri\theta}\, d\theta\, [-\tau xy I(\theta)] \tag{5.3}$$

and from (4.34) we have

$$b_{2r+1} = \frac{1}{2\pi} \int_0^{2\pi} e^{-(2r+1)i\theta}\, d\theta\, [i\tau^2 xy e^{i\theta} I(\theta)]. \tag{5.4}$$

Finally Eq. (4.45) shows that

$$b_{-2r+1} = \left(\frac{i}{\tau}\right)^{2r-1} + \frac{1}{2\pi} \int_0^{2\pi} e^{(2r-1)i\theta}\, d\theta\, [i\tau^2 xy e^{i\theta} I(\theta)].$$

Hence the generating function for the even terms is, apart from the terms $(i/\tau)^{2r}$,

$$B_e(\theta) = -\tau xy \cdot \frac{1}{2\pi}\int_0^{2\pi} \frac{1 - e^{i(\theta+\phi)}}{\Delta(\theta,\phi)}\, d\phi, \qquad (5.6)$$

and similarly for the odd terms,

$$B_0(\theta) = ixy\tau^2 e^{i\theta} \cdot \frac{1}{2\pi}\int_0^{2\pi} \frac{1 - e^{i(\theta+\phi)}}{\Delta(\theta,\phi)}\, d\phi. \qquad (5.7)$$

Now from (4.25) we have that

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{1 - e^{i(\theta+\phi)}}{\Delta(\theta,\phi)}\, d\phi = -\frac{ie^{i\theta}}{xy}\left[\frac{\operatorname{sgn}(\sin\theta) + i\tau e^{i\theta}}{1 + \tau^2 e^{2i\theta}}\right]. \qquad (5.8)$$

Adding terms we get

$$B_e(\theta) + B_0(\theta) = i\tau e^{i\theta}\left[\frac{\operatorname{sgn}(\sin\theta) + i\tau e^{i\theta}}{1 + i\tau e^{i\theta}}\right]. \qquad (5.9)$$

From the terms b_{-k} we also have the contribution

$$\sum_{k=0}^{\infty}\left(\frac{i}{\tau}\right)^k e^{-ik\theta} = \begin{cases} \dfrac{i\tau e^{i\theta}}{1 + i\tau e^{i\theta}} & (\tau > 1) \\ \text{divergent} & (\tau \le 1). \end{cases} \qquad (5.10)$$

And so the generating function is undefined for $\tau \le 1$. For $\tau > 1$

$$B(\theta) = i\tau e^{i\theta}\left[\frac{1 + \operatorname{sgn}(\sin\theta) + i\tau e^{i\theta}}{1 + i\tau e^{i\theta}}\right]. \qquad (5.11)$$

Here we thus have an example in which some of the entries in a determinant diverge but the value of the determinant still decays to zero. The fact that some of the entries do diverge is a property inherent in this type of problem, and the divergence cannot be removed by elementary transformations, as will be seen in the next paragraph.

We will see that for $\tau > 1$ the Toeplitz representation of the determinant is the most convenient one, whereas for $\tau \le 1$ the perturbed representation of (6.5) is the most satisfactory one.

6. REDUCTION OF THE CORRELATION DETERMINANT

For convenience we will define a new variable $f_k = \frac{1}{2}\pi i b_k$ so that the correlation becomes

$$\omega(x, y \mid p-1, p) = \frac{1}{2x}\left[\frac{2}{\pi}\right]^p |f_{i-j+1}| \qquad i, j = 1, \cdots p. \qquad (6.1)$$

Equation (4.1) becomes

$$f_k - (i/\tau)f_{k+1} = \begin{cases} 0 & (k \text{ even}) \\ 1/k & (k \text{ odd}), \end{cases} \qquad (6.2)$$

and so

$$f_k + (1/\tau^2)f_{k+2} = 1/k \qquad (k \text{ odd}). \qquad (6.3)$$

To reduce the determinant $|F_p| = |f_{i-j-1}|$, we use (6.2) and perform in succession the corresponding operations:

column (k) − (i/τ) column $(k-1)$,

$$k = p, p-1, \cdots, 2. \qquad (6.4)$$

After these operations we are left with a determinant whose structure depends on p.

$$|F_p| = \begin{vmatrix} f_1 & 0 & -1 & 0 & -\frac{1}{3} & 0 & \frac{-1}{p-2} \\ f_2 & 1 & \cdot & & & & 0 \\ f_3 & 0 & & \cdot & & & \cdot \\ \vdots & \vdots & & & \cdot & & \cdot \\ \vdots & \vdots & & & & \cdot & \cdot \\ & \frac{1}{p-2} & & & & \cdot 1 & 0 \\ f_p & 0 & \frac{1}{p-2} & \cdots & & 0 & 1 \end{vmatrix}_{p \text{ odd}}$$

or

$$\begin{vmatrix} f_1 & 0 & -1 & 0 & -\frac{1}{3} & \cdots & \frac{-1}{p-2} & 0 \\ f_2 & 1 & \cdot & & & & & \frac{-1}{p-2} \\ f_3 & 0 & & \cdot & & & & \\ \vdots & & & & \cdot & & & \vdots \\ \vdots & & & & & \cdot & & \vdots \\ & 0 & & & & & \cdot 1 & 0 \\ f_p & \frac{1}{p-1} & \cdot & \cdot & \cdot & & 0 & 1 \end{vmatrix}_{p \text{ even}} \qquad (6.5)$$

It is clear from (6.2) and (6.4) that apart from the first column, the entries in the determinant are either zero or the reciprocal of an odd integer and that along any one diagonal the entries are the same. Furthermore from (4.33) it follows that the entries in the first column diverge with p if $\tau > 1$. By symmetrically made row and column interchanges we can transform the determinant into block form with all the zero entries off the diagonal.

If we denote the following matrix by D_n and its determinant by Δ_n,

$$D_n = \begin{bmatrix} 1 & -1 & -\frac{1}{3} & \cdots & \frac{-1}{2n-3} \\ \frac{1}{3} & 1 & -1 & \cdots & \cdot \\ & \cdot & \cdot & & \\ \frac{1}{5} & & \cdot & \cdot & \cdot \\ & & & \cdot & -1 \\ \frac{1}{2n-1} & \cdot & \cdot & \frac{1}{3} & 1 \end{bmatrix}, \text{ i.e.,}$$

$$d_{ij} = [1 + 2(i - j)]^{-1}, \tag{6.6}$$

then Δ_n represents the spin–spin correlation at the critical point along the diagonal of a square Ising lattice. The close analogy between the two correlations becomes apparent. After the row and column interchanges the determinant $|F_p|$ has the following structure:

$$|F_{2n}| = \begin{vmatrix} (D_n + Y_n) & \mathbf{0} \\ A_n & D_n \end{vmatrix} \text{ or}$$

$$|F_{2n-1}| = \begin{vmatrix} (D_n + Y_n) & \mathbf{0} \\ A_{n-1} & D_{n-1} \end{vmatrix}, \tag{6.7}$$

where

$$n = \left[\frac{p+1}{2}\right], \quad Y_n = \begin{bmatrix} (f_1 - 1) & & \\ (f_3 - \tfrac{1}{3}) & & \mathbf{0} \\ & \vdots & \\ \left(f_{2n-1} - \frac{1}{2n-1}\right) & & \end{bmatrix}$$

and

$$A_n = \begin{bmatrix} f_{2n} & & \\ f_{2n-2} & \mathbf{0} & \\ \vdots & & \\ f_2 & & \end{bmatrix}. \tag{6.8}$$

From (6.3) we see that

$$Y_n = \frac{1}{\tau^2}\begin{bmatrix} -f_3 & & \\ -f_5 & & \mathbf{0} \\ \vdots & & \\ -f_{2n+1} & & \end{bmatrix}. \tag{6.9}$$

Hence

$$|F_p| = |D_n + Y_n| \cdot \begin{cases} \Delta_n & (p = 2n) \\ \Delta_{n-1} & (p = 2n-1) \end{cases} \tag{6.10}$$

$$= |I + D_n^{-1} Y_n| \cdot \begin{cases} \Delta_n^2 & (p = 2n) \\ \Delta_n \Delta_{n-1} & (p = 2n-1). \end{cases} \tag{6.11}$$

If we denote the inverse matrix of D_n by $T_n = [t_{ij}]$ then the matrix product $D_n^{-1} Y_n$ becomes:

$$\begin{bmatrix} -\eta_1 & & \\ -\eta_2 & & \mathbf{0} \\ \vdots & & \\ -\eta_n & & \end{bmatrix}, \tag{6.12}$$

where

$$\eta_1 = [f_3 t_{11} + f_5 t_{12} + \cdots + f_{2n+1} t_{1n}]/\tau^2. \tag{6.13}$$

It follows that

$$|I + D_n^{-1} Y_n| = 1 - \eta_1 = 1 - \sum_{k=1}^{n} \frac{f_{2k+1} t_{1k}}{\tau^2} \tag{6.14}$$

and that

$$\omega(x, y \mid p-1, p) = \frac{1}{2x} \cdot \left[\frac{2}{\pi}\right]^p \begin{Bmatrix} \Delta_n^2 \\ \Delta_n \Delta_{n-1} \end{Bmatrix} \cdot (1 - \eta_1). \tag{6.15}$$

We can get a compact expression for η (dropping the subscript) if we use the integral representation of f_{2k+1} for positive k.

Now from (4.31) and (4.34) it can be seen that

$$f_{2k+1} = \tfrac{1}{2}\pi i b_{2k+1} = \tfrac{1}{2}\pi i \cdot \tau^2 \left[-\frac{i}{\tau} b_{2k}\right]$$

$$= \tfrac{1}{2}\pi i \tau^2 \cdot 2iy[2k-1, 2k]. \tag{6.16}$$

Thus

$$f_{2k+1} = \tau^2 \int_0^1 \frac{x^{2k}}{\tau^2 + x^2}\, dx \tag{6.17}$$

and so

$$\eta = \int_0^1 \sum_{k=1}^{n} \frac{x^{2k} t_{1k}}{\tau^2 + x^2}\, dx. \tag{6.18}$$

However this is not the most convenient form of the integral and we transform it using partial fractions to

$$f_{2k+1} = -\frac{\tau}{4i} \int_{-1}^{1} x^{2k-1} \left[\frac{x - i\tau}{x + i\tau}\right] dx$$

$$= \frac{\tau}{4} \int_{C_1} z^{2k} \left[\frac{z - i\tau}{z + i\tau}\right] \frac{dz}{iz}$$

$$= \frac{\tau}{4} \int_0^{\pi} e^{2ki\theta} \left[\frac{e^{i\theta} - i\tau}{e^{i\theta} + i\tau}\right] d\theta, \tag{6.19}$$

where C_1 is the top half of the unit circle. When $\tau = 1$ this reduces to

$$f_{2k+1} = \frac{1}{4}\int_0^{\pi} \sin(2k\theta)\left[\frac{1-\sin\theta}{\cos\theta}\right] d\theta; \qquad (6.20)$$

since f_{2k+1} is real the imaginary part vanishes.

In (6.19) we write

$$\frac{z-i\tau}{z+i\tau} = 1 + \frac{2(\tau/i)}{z+i\tau}$$

and since $\int_0^{\pi} e^{2ri\theta}\, d\theta \equiv 0$ we have

$$\eta = \frac{1}{2i}\int_0^{\pi} \sum_{k=1}^{n} \frac{e^{2ki\theta} t_{1k}}{e^{i\theta} + i\tau}\, d\theta. \qquad (6.21)$$

Before we can proceed with the evaluation of η we have first to find the inverse matrix $T_n = D_n^{-1}$, which is done in the next paragraph.

7. THE CAUCHY MATRIX INVERSE

A Cauchy matrix is defined to be a matrix which has the form

$$C_n = \left[\frac{1}{a_p + b_q}\right]_1^n p, q = 1, \cdots, n.$$

Firstly we note that both D_n and its "reflection" RD_n, where

$$R = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot & & \\ & 1 & & \\ \cdot & & & \\ 1 & & 0 & \end{bmatrix},$$

are Cauchy matrices and that RD_n is in fact a symmetric Hankel matrix.

The latter is clearly so for all Toeplitz matrices. Furthermore one finds that D_n has complex eigenvalues whereas RD_n has real positive eigenvalues. Although it is very difficult to obtain the general eigenvalues of either one we can still find their determinants by either

(i) triangulation, which is possible because the matrices are real and have non-zero principal minors; however which is rather cumbersome,

(ii) or by the reduction method for Cauchy determinants.

Once we have found the general Cauchy determinant,[7] it is a simple matter to find its inverse because all its minors are again Cauchy determinants.

[7] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis II* (Springer-Verlag, Berlin, 1925), Chap. VII, p. 98.

Let

$$\Delta_n = |C_n| = \left|\frac{1}{a_p + b_q}\right|_1^n$$

$$= \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_q} & \cdots & \frac{1}{a_1+b_n} \\ \vdots & & \vdots & & \vdots \\ \frac{1}{a_p+b_1} & \cdots & \frac{1}{a_p+b_q} & \cdots & \frac{1}{a_p+b_n} \\ \vdots & & \vdots & & \vdots \\ \frac{1}{a_n+b_1} & \cdots & \frac{1}{a_n+b_q} & \cdots & \frac{1}{a_n+b_n} \end{vmatrix}. \qquad (7.1)$$

On performing the row operations: row (k) − row (n), $k = 1, \cdots, n-1$, and taking out the common factors followed by the column operations: column (k) − column (n), $k = 1, \cdots, n-1$, and taking out the common factors we arrive at the recurrence relation

$$\Delta_n = \frac{\prod_{i=1}^{n-1}[a_n - a_i] \prod_{j=1}^{n-1}[b_n - b_j]}{\prod_{i=1}^{n}[a_n + b_i] \prod_{j=1}^{n-1}[a_j + b_n]} \Delta_{n-1}, \qquad (7.2)$$

which gives

$$\Delta_n = \frac{\prod_{i>j}^{1\cdots n}[a_i - a_j] \prod_{i>j}^{1\cdots n}[b_i - b_j]}{\prod_{i=1}^{n}\prod_{j=1}^{n}[a_i + b_j]}, \qquad (7.3)$$

where the denominator is the product over all the denominators in Δ_n.

To find the inverse matrix C_n^{-1}, we use the fact that

$$[C_n^{-1}]_{qp} = (-1)^{p+q}\Delta_{pq}/\Delta_n, \qquad (7.4)$$

where Δ_{pq} is the (p, q) minor in Δ_n, i.e., one that is obtained by omitting the pth row and the qth column in Δ_n or equivalently all the terms which contain a_p and/or b_q. In the numerator these terms are

$$\prod_{r=1}^{p-1}[a_p - a_r] \prod_{r=p+1}^{n}[a_r - a_p] \qquad (7.5)$$

and

$$\prod_{r=1}^{q-1}[b_q - b_r] \prod_{r=q+1}^{n}[b_r - b_q]$$

as can be seen from the triangular array.

In the denominator we have to omit

$$\prod_{r=1}^{n}[a_p + b_r]\prod_{r=1}^{n}[a_r + b_q][a_p + b_q]^{-1}, \tag{7.6}$$

where the term $[a_p + b_q]$ is counted only once.

Hence

$$[C_n^{-1}]_{qp} = \frac{(-1)^{p+q}\prod_{r=1}^{n}[a_p + b_r]\prod_{r=1}^{n}[a_r + b_q][a_p + b_q]^{-1}}{\prod_{r=1}^{p-1}[a_p - a_r]\prod_{r=p+1}^{n}[a_r - a_p]\prod_{r=1}^{q-1}[b_q - b_r]\prod_{r=q+1}^{n}[b_r - b_q]}. \tag{7.7}$$

To apply this formula to D_n we must put

$$\begin{aligned} a_p &= 2p - 1 \quad p, q = 1, \cdots, n. \\ b_q &= 2 - 2q. \end{aligned} \tag{7.8}$$

Substituting these values in (7.2) and (7.7) we see that

$$\frac{\Delta_{n-1}}{\Delta_n} = (2n - 1)[C_{n-1}^{-\frac{1}{2}}]^2 \tag{7.9}$$

and that

$$t_{qp} = \frac{1}{2^{4(n-1)}}\frac{[2p - 1][2n - 2q + 1]}{[2p - 2q + 1]} \times C_{p-1}^{2p-2}C_{q-1}^{2q-2}C_{n-p}^{2n-2p}C_{n-q}^{2n-2q}. \tag{7.10}$$

This shows that all the entries in the inverse matrix with $p \geq q$ are positive. We rewrite this to bring out the character of the generating function of the entries in D_n as

$$t_{qp} = \frac{[2p - 1][2n - 2q + 1]}{[2p - 2q + 1]}C_{p-1}^{-\frac{1}{2}}C_{q-1}^{-\frac{1}{2}}C_{n-p}^{-\frac{1}{2}}C_{n-q}^{-\frac{1}{2}}, \tag{7.11}$$

where $C_n^{-\frac{1}{2}} = C_n^{2n}\cdot(-1)^n/2^{2n}$ which tends to zero as $(-1)^n(\pi n)^{-\frac{1}{2}}$. We are mainly concerned with the first row of the matrix T_n, i.e., with

$$t_{1p} = (2n - 1)C_{n-1}^{-\frac{1}{2}}\cdot C_{p-1}^{-\frac{1}{2}}C_{n-p}^{-\frac{1}{2}} \tag{7.12}$$

which shows that $t_{1p} = t_{1,n-p+1}$. Furthermore both D_n and T_n are symmetric about their second main diagonal.

8. EVALUATION OF THE PERTURBATION INTEGRAL

Substituting Eq. (7.12) in (6.21), we have

$$\eta = (2n - 1)C_{n-1}^{-\frac{1}{2}} \times \frac{1}{2i}\int_0^{\pi}\sum_{k=1}^{n}e^{2ki\theta}C_{k-1}^{-\frac{1}{2}}C_{n-k}^{-\frac{1}{2}}\frac{1}{e^{i\theta} + i\tau}\,d\theta. \tag{8.1}$$

To investigate the sum, we lower the index by 1 so that

$$\sum_{k=1}^{n}e^{2ki\theta}C_{k-1}^{-\frac{1}{2}}C_{n-k}^{-\frac{1}{2}} = \sum_{l=0}^{N}e^{(2l+2)i\theta}C_l^{-\frac{1}{2}}C_{N-l}^{-\frac{1}{2}} = e^{(N+2)i\theta}\sum_{l=0}^{N}e^{(2l-N)i\theta}C_l^{-\frac{1}{2}}C_{N-l}^{-\frac{1}{2}}, \tag{8.2}$$

where $N = n - 1$ and $l = k - 1$. This can be identified as $(-1)^N e^{(N+2)i\theta}P_N(\cos\theta)$; where $P_N(\cos\theta)$ is the Legendre polynomial which is defined by

$$\left[(1 - hz)\left(1 - \frac{h}{z}\right)\right]^{-\frac{1}{2}} = \sum_{N=0}^{\infty}P_N\left[\frac{z + 1/z}{2}\right]h^N, \tag{8.3}$$

so that

$$P_N(\cos\theta) = (-1)^N\sum_{l=0}^{N}e^{(2l-N)i\theta}C_l^{-\frac{1}{2}}C_{N-l}^{-\frac{1}{2}} = (-1)^N 2^{\frac{1}{2}N,\frac{1}{2}(N-1)}\sum_{p=0}\cos(N - 2p)\theta C_p^{-\frac{1}{2}}C_{N-p}^{-\frac{1}{2}} \tag{8.4}$$

with a last term[8]

$$\text{and} \quad \begin{cases} \tfrac{1}{2}[C_{\frac{1}{2}N}^{-\frac{1}{2}}]^2 & \text{when } N = 2s \\ C_{\frac{1}{2}(N-1)}^{-\frac{1}{2}}C_{\frac{1}{2}(N+1)}^{-\frac{1}{2}} & \text{when } N = 2s + 1. \end{cases} \tag{8.5}$$

Using these results, we see that (8.1) gives

$$\eta = [(2n - 1)(-1)^{n-1}C_{n-1}^{-\frac{1}{2}}] \times \frac{1}{2i}\int_0^{\pi}P_N(\cos\theta)\frac{e^{i(N+2)\theta}}{e^{i\theta} + i\tau}\,d\theta, \tag{8.6}$$

which in the symmetric case ($\tau = 1$) with the aid of (6.20) can be written as

$$\eta = [(2n - 1)(-1)^{n-1}C_{n-1}^{-\frac{1}{2}}] \times \frac{1}{4}\int_0^{\pi}P_N(\cos\theta)\sin(N + 2)\theta\left[\frac{1 - \sin\theta}{\cos\theta}\right]d\theta. \tag{8.7}$$

For convenience we denote

$$[(2n - 1)(-1)^{n-1}C_{n-1}^{-\frac{1}{2}}]^{-1}\eta \tag{8.8}$$

[8] L. Robin, *Fonctions sphériques de Legendre et fonctions spheroidales* (Gauthier-Villars, Paris, 1957), Chap. I, Sec. 6, pp. 13–14.

by ζ_{n-1}. From (6.17) and (8.3), an alternative form for ζ_{n-1} is found to be

$$\zeta_{n-1} = \int_0^1 \frac{x^{n+1}}{\tau^2 + x^2} P_{n-1}\left[\frac{x + x^{-1}}{2}\right] dx, \qquad (8.9)$$

which could be used together with the fact that

$$P_N\left[\frac{x + x^{-1}}{2}\right] = \frac{\Gamma(N + \frac{1}{2}).x^N}{\Gamma(N + 1)\Gamma(\frac{1}{2})} F\left[\tfrac{1}{2}, -N, \tfrac{1}{2} - N, \frac{1}{x^2}\right].$$

However we will concentrate on the integrals (8.6) and (8.7). It is characteristic of the problems involving monomers that the behavior of the system is different for finite odd and even separation distance N. In the asymptotic limit, however, they must have the same behavior as will be observed in the case of the monomer pair correlation.

We now prove that

$$1 - \eta = [n(-1)^n C_n^{-\frac{1}{2}}]\left[\frac{\tau}{i}\right]^n \times \begin{cases} Q_{n-1}[-i \sinh \xi] & (e^\xi = \tau > 1) \\ Q_{n-1}[0 - i0] & (\tau = 1) \\ Q_{n-1}[i \sinh \xi] + \pi i P_{n-1}[i \sinh \xi] \cdot & (e^{-\xi} = \tau < 1), \end{cases} \qquad (8.10)$$

where $Q_{n-1}(\mu)$ is the Legendre function of the second kind defined on the cut as[9]

$$Q_{n-1}[0 \pm i0] = Q_{n-1}(0) \mp \tfrac{1}{2}\pi i P_{n-1}(0). \qquad (8.11)$$

Starting with the integral from (8.6),

$$\zeta_N = \frac{1}{2i} \int_0^\pi P_N(\cos\theta) \frac{e^{i(N+2)\theta}}{e^{i\theta} - (\tau/i)} d\theta, \qquad (8.12)$$

we see that when $\tau \neq 1$ the denominator of the integrand can be expanded by the binomial theorem in series which will differ for $\tau > 1$ and $\tau < 1$.

We will use these expansions combined with the theory of the Legendre functions to evaluate ζ_N for the three cases $\tau > 1$, $\tau < 1$ and $\tau = 1$.

(i) When $\tau > 1$ we expand the denominator

$$\frac{1}{e^{i\theta} - \tau/i} = -\sum_{p=0}^{\infty} \left(\frac{i}{\tau}\right)^{p+1} e^{ip\theta} \qquad (8.13)$$

so that

$$\zeta_N = -\frac{1}{2i} \sum_{p=0}^{\infty} \left(\frac{i}{\tau}\right)^{p+1} \times \int_0^\pi P_N(\cos\theta) e^{i(N+p+2)\theta} d\theta. \qquad (8.14)$$

The sum and the integral may be interchanged by comparison with a geometric series.

Now from the theory of Legendre functions[10] we have the two standard integrals

$$I(m) = \int_0^\pi P_N(\cos\theta) \cos m\theta \, d\theta = \begin{cases} 0 & (m > N) \\ 0 & (N - m \text{ odd}) \\ \pi(-1)^N C_p^{-\frac{1}{2}} C_{N-p}^{-\frac{1}{2}} & (m = N - 2p) \end{cases} \qquad (8.15)$$

and

$$J(m) = \int_0^\pi P_N(\cos\theta) \sin m\theta \, d\theta = \begin{cases} 0 & (m < N) \\ 0 & (m + N \text{ even}) \\ \dfrac{2(m - N + 1)(m - N + 3) \cdots (m + N - 1)}{(m - N)(m - N + 2) \cdots (m + N)} & (m = N + 2p + 1), \end{cases} \qquad (8.16)$$

and the Neumann integral representation of $Q_N(\mu)$,[11]

$$Q_N(\mu) = \frac{1}{2} \int_{-1}^{1} \frac{P_N(\nu)}{\mu - \nu} d\nu \quad \mu \notin (-1, 1) = z \int_0^\pi \frac{P_N(\cos\phi) \sin\phi \, d\phi}{1 - 2z\cos\phi + z^2}, \qquad (8.17)$$

where $2\mu = z + 1/z$ and μ does not lie on the cut $(-1, 1)$, i.e., $|z| > 1$. It follows that for $|z| > 1$, on expanding $(1 - 2z\cos\phi + z^2)^{-1}$,

$$Q_N(\mu) = \sum_{p=1}^{\infty} z^{-p} \int_0^\pi P_N(\cos\phi) \sin p\phi \, d\phi = \frac{2^{N+1}\Gamma(N+1)}{\Gamma(2N+2)} z^{-(N+1)} F\left(\tfrac{1}{2}, N + 1, N + \frac{3}{2}, \frac{1}{z^2}\right). \qquad (8.18)$$

For values on the cut one has to put $\mu = \cos\theta + i.o$ and $z = e^{i\theta}$, so that by an extension of Abel's theorem[12] the relation holds for all $|z| \geq 1$, $z \neq 1$, i.e., on the unit circle for all $0 < \theta < 2\pi$.

[9] See Ref. 8, Chap. I, Sec. 12, p. 34. See Ref. 15.
[10] See Ref. 8, Chap. I, Sec. 11, pp. 27–29.
[11] See Ref. 8, Chap. I, Sec. 14, p. 41.
[12] See Ref. 8, Chap. I, Sec. 14, p. 43.

Thus on putting $z = i$ in (8.17) we get

$$Q_N(0) = -\frac{1}{2}\int_0^\pi \frac{P_N(\cos\phi)\sin\phi}{\cos\phi}\,d\phi. \qquad (8.19)$$

Using the standard integrals in (8.14) we see that the cos $(N + p + 2)\theta$ terms drop out and we are left with

$$\zeta_N = -\frac{1}{2}\sum_{p=0}^{\infty}\left(\frac{i}{\tau}\right)^{p+1}\int_0^\pi P_N(\cos\theta) \times \sin(N + p + 2)\theta\,d\theta$$

$$= -\frac{1}{2}\sum_{q=0}^{\infty}\left(\frac{i}{\tau}\right)^{q}\int_0^\pi P_N(\cos\theta) \times \sin(N + q + 1)\theta\,d\theta + \tfrac{1}{2}J(N+1),$$

where

$$J(N+1) = \int_0^\pi P_N(\cos\theta)\sin(N+1)\theta\,d\theta$$

$$= 2[(2n-1)(-1)^{n-1}C_{n-1}^{-\frac{1}{2}}]^{-1} \quad (N = n-1), \qquad (8.20)$$

and thus

$$\zeta_N = -\tfrac{1}{2}\left(\frac{\tau}{i}\right)^{N+1}\sum_{r=0}^{\infty}\left(\frac{i}{\tau}\right)^{N+2r+1}\int_0^\pi P_N(\cos\theta) \times \sin(N + 2r + 1)\theta\,d\theta + \tfrac{1}{2}J(N+1)$$

which by (8.18) becomes

$$\zeta_N = \tfrac{1}{2}J(N+1) - \frac{1}{2}\left(\frac{\tau}{i}\right)^{N+1}Q_N\left[\frac{\tau/i + i/\tau}{2}\right]. \qquad (8.21)$$

Substituting this in (8.6) together with the fact that

$$(2n-1)(-1)^{n-1}C_{n-1}^{-\frac{1}{2}} = (2n)(-1)^n C_n^{-\frac{1}{2}}, \qquad (8.22)$$

we arrive at

$$\eta = 1 - [n(-1)^n C_n^{-\frac{1}{2}}] \times \left(\frac{\tau}{i}\right)^n Q_{n-1}[-i\sinh\xi] (e^\xi = \tau > 1), \qquad (8.23)$$

which gives the first relation of (8.10).

(ii) For $\tau < 1$, we expand

$$\frac{e^{i(N+2)\theta} - (\tau/i)^{N+2}}{e^{i\theta} - (\tau/i)} = \sum_{r=0}^{N+1} e^{i(N-r+1)\theta}\cdot\left(\frac{\tau}{i}\right)^r \qquad (8.24)$$

and so by (8.12),

$$2i\zeta_N = \left(\frac{\tau}{i}\right)^{N+2}\int_0^\pi \frac{P_N(\cos\theta)}{e^{i\theta} - (\tau/i)}\,d\theta + \sum_{r=0}^{N+1}\left(\frac{\tau}{i}\right)^r\int_0^\pi P_N(\cos\theta)e^{i(N-r+1)\theta}\,d\theta. \qquad (8.25)$$

Considering the sum first we see that all the terms involving sin $(N - r + 1)\theta$ vanish apart from the first one which gives $iJ(N+1)$. The integrals involving cos $(N - r + 1)\theta$ give

$$\sum_{r=1}^{N+1}\left(\frac{\tau}{i}\right)^r\int_0^\pi P_N(\cos\theta)\cos(N - r + 1)\theta\,d\theta = \left(\frac{\tau}{i}\right)^{N+1}\left\{\frac{\pi}{2}P_N\left[\frac{\tau/i + i/\tau}{2}\right] + \frac{1}{2}\int_0^\pi P_N(\cos\theta)\,d\theta\right\}. \qquad (8.26)$$

The additional term comes from the definition of the last term in $P_N(\cos\theta)$ and vanishes when N is odd.

In the first integral of (8.25) we expand the denominator as

$$\frac{1}{e^{i\theta} - \tau/i} = \sum_{p=0}^{\infty}\left(\frac{\tau}{i}\right)^p e^{-(p+1)i\theta} \qquad (\tau < 1) \qquad (8.27)$$

and so get

$$\sum_{p=0}^{\infty}\left(\frac{\tau}{i}\right)^{N+p+2}\int_0^\pi P_N(\cos\theta)e^{-(p+1)i\theta}\,d\theta.$$

On splitting this up and using the standard integrals the sum involving the cos $(p+1)\theta$ terms becomes

$$\left(\frac{\tau}{i}\right)^{N+1}\left\{\frac{\pi}{2}P_N\left[\frac{\tau/i + i/\tau}{2}\right] - \frac{1}{2}\int_0^\pi P_N(\cos\theta)\,d\theta\right\}, \qquad (8.28)$$

and the remaining sum reduces to

$$-i\left(\frac{\tau}{i}\right)^{N+1}Q_N\left[\frac{\tau/i + i/\tau}{2}\right]. \qquad (8.29)$$

Hence

$$\int_0^\pi \frac{P_N(\cos\theta)\,d\theta}{e^{i\theta} - \tau/i} = \frac{1}{\tau}\left\{Q_N[i\sinh\xi] + \frac{\pi i}{2}P_N[i\sinh\xi] - \frac{i}{2}\int_0^\pi P_N(\cos\theta)\,d\theta\right\}, \qquad (8.30)$$

where $e^{-\xi} = \tau < 1$. Adding (8.26), (8.28), and (8.29) one obtains directly with the aid of (8.22) that

$$\eta = 1 - [n(-1)^n C_n^{-\frac{1}{2}}]\left(\frac{\tau}{i}\right)^n\{Q_{n-1}[i\sinh\xi] + \pi i P_{n-1}[i\sinh\xi]\}. \qquad (8.31)$$

The case $\tau = 1$ can be considered as the limiting case of the other two when $\tau \downarrow 1$ or $\tau \uparrow 1$. We can either let $\tau \to 1$ in the final results or put $\tau = 1$ in the original integral expression for η and evaluate it as such.

Taking the limit $\tau \downarrow 1$ in (8.23) will in fact give the value of η at $\tau = 1$ because

(i) In (8.12) the limit $\tau \downarrow 1$ may be taken inside the integral since there are no singularities on the contour and the integrand is well behaved (e.g., bounded).

(ii) In (8.14) the limit $\tau \downarrow 1$ may be taken inside the summation. This is permissible since the series representation of $Q_N[0 - i \sinh \xi]$ is still valid at $\tau = 1$, i.e., $\xi = 0$. Clearly the same holds if we let $\tau \uparrow 1$ in (8.31). This gives on using (8.11) that $\mathbf{Q}_{n-1}[0 + i0] + \pi i P_{n-1}[0] = \mathbf{Q}_{n-1}[0 - i0]$, which is the same as was obtained from (8.23). As expected, the asymptotic series will show the same behavior near $\tau = 1$. To evaluate ζ_N directly at $\tau = 1$, without referring to series expansions, we make use of the Eq. (8.7),

$$4\zeta_N = \int_0^\pi P_N(\cos\theta) \times \sin(N+2)\theta\left[\frac{1-\sin\theta}{\cos\theta}\right] d\theta, \tag{8.32}$$

and consider the even and odd cases separately.

This integral is essentially the sum of elementary integrals since either $\sin(N+2)\theta$ or $P_N(\cos\theta)$ contains a factor of $\cos\theta$ depending on the parity of N. When N is even, say $N = 2s$, then the identity[13]

$$\frac{\sin 2(s+1)\theta}{\cos\theta} = 2(-1)^s \sum_{k=1}^{s+1} (-1)^{k+1} \sin(2k-1)\theta \tag{8.33}$$

can be used to give

$$4\zeta_{2s} = 2(-1)^s \sum_{k=1}^{s+1} (-1)^{k+1} \int_0^\pi P_N(\cos\theta) \times \sin(2k-1)\theta(1-\sin\theta)\, d\theta = S_1 + S_2.$$

The only contribution to the first sum S_1 comes from the term with $k = s+1$ and gives $2J(N+1)$. The second sum S_2 can be calculated using the addition formula for cosines and the standard integrals together with (8.4). Collecting terms gives

$$\zeta_{2s} = [(2n-1)(-1)^n C_{n-1}^{-\frac{1}{2}}]^{-1} + \tfrac{1}{4}\pi(-1)^{s+1} P_{2s}(0). \tag{8.34}$$

For odd N, say $N = 2s+1$, the term $P_N(\cos\theta)$ contains a factor of $\cos\theta$ so that

$$4\zeta_{2s+1} = \int_0^\pi \frac{P_N(\cos\theta)}{\cos\theta} [\sin(N+2)\theta - \tfrac{1}{2}\{\cos(N+1)\theta - \cos(N+3)\theta\}]\, d\theta. \tag{8.35}$$

The last two integrals vanish; which follows if we put

$$\frac{P_N(\cos\theta)}{\cos\theta} = \sum_{p=0}^{s} b_p(-1)^{s-p} \times \left[1 + 2\sum_{k=1}^{s-p} (-1)^k \cos 2k\theta\right] \tag{8.36}$$

and then use the addition formula for cosines and integrate. The first integral becomes

$$4\zeta_{2s+1} = J(N+1) + \int_0^\pi P_N(\cos\theta) \frac{\sin\theta}{\cos\theta} \cos(N+1)\theta\, d\theta.$$

This can be calculated with the aid of the identity[14]

$$\frac{\cos 2(s+1)\theta}{\cos\theta} = (-1)^{s+1} \times \left[\frac{1}{\cos\theta} - 2\sum_{k=1}^{s+1} (-1)^{k+1} \cos(2k-1)\theta\right] \tag{8.37}$$

and the standard integrals to give

$$4\zeta_{2s+1} = 2J(N+1) + 2(-1)^s Q_{2s+1}(0).$$

Hence when $\tau = 1$,

$$\eta = 1 + [n(-1)^n C_n^{-\frac{1}{2}}] \times \begin{cases} (-1)^s Q_{2s+1}(0) & (N = 2s+1) \\ (-1)^{s+1} \tfrac{1}{2}\pi P_{2s}(0) & (N = 2s). \end{cases} \tag{8.38}$$

Now if we use the identities[15]

$$P_{n-1}(0) = \pi^{-\frac{1}{2}} \cos\left[(n-1)\frac{\pi}{2}\right] \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}[n+1])}$$
$$Q_{n-1}(0) = -\tfrac{1}{2}\pi^{\frac{1}{2}} \sin\left[(n-1)\frac{\pi}{2}\right] \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}[n+1])}, \tag{8.39}$$

then it is clear that (8.38) reduces to (8.10).

[13] I. Ryshik and I. Gradstein, *Tables* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1957), p. 31.

[14] See Ref. 13.

[15] *The Bateman Manuscript Project*, edited by A. Erdelyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 1, Sec. 3.4, p. 145.

Furthermore we get

$$1-\eta=[n(-1)^nC_n^{-\frac{1}{2}}]\frac{\pi^{\frac{1}{2}}}{2}\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}[n+1])}$$
$$=\frac{n}{2}\cdot\frac{\Gamma(n+\frac{1}{2})}{\Gamma(n+1)}\cdot\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}[n+1])}, \tag{8.40}$$

from which the asymptotic decay of $1-\eta$ will be found at $\tau=1$. This completes the proof of the exact result (8.10) from which the exact result for the pair correlation follows:

$$\omega(x,y\mid p-1,p)=\frac{1}{2x}\left(\frac{2}{\pi}\right)^p\Delta_n\Delta_{p-n}[n(-1)^nC_n^{-\frac{1}{2}}]\left(\frac{\tau}{i}\right)^n$$
$$\times\begin{cases}Q_{n-1}[-i\sinh\xi] & (e^\xi=\tau>1)\\ Q_{n-1}[0-i0] & (\tau=1)\\ Q_{n-1}[i\sinh\xi]+i\pi P_{n-1}[i\sinh\xi] & (e^{-\xi}=\tau<1),\end{cases} \tag{8.42}$$

where

$$\Delta_n^{-1}=\prod_{k=1}^{n}(2k-1)[C_{k-1}^{-\frac{1}{2}}]^2$$
$$=\prod_{s=1}^{n}\frac{\Gamma(s-\frac{1}{2})\Gamma(s+\frac{1}{2})}{\Gamma(s)\Gamma(s)}\text{ and }n=\tfrac{1}{2}[p+1]. \tag{8.43}$$

9. ASYMPTOTIC VALUES

In this paragraph we will find the first two terms in the expansion of the correlation as a function of radial distance r. We consider the symmetric case first and take as our starting point Eq. (8.40) and use the asymptotic series for $\Gamma(n)$,

$$\Gamma(n)\sim e^{-n}n^{n-\frac{1}{2}}(2\pi)^{\frac{1}{2}}$$
$$\times\left[1+\frac{1}{12n}+\frac{1}{288n^2}+O\left(\frac{1}{n^3}\right)\right]. \tag{9.1}$$

Using the duplication formula we find that

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma\frac{1}{2}[n+1]}=\frac{2^{n-1}\Gamma^2(\frac{1}{2}n)}{\pi^{\frac{1}{2}}\Gamma(n)}$$
$$\sim\left(\frac{2}{n}\right)^{\frac{1}{2}}\left[1+\frac{1}{4n}+\frac{1}{32n^2}+O\left(\frac{1}{n^3}\right)\right]. \tag{9.2}$$

A similar analysis on $C_n^{-\frac{1}{2}}=(-1)^nC_n^{2n}\cdot 2^{-2n}$ shows that

$$C_n^{-\frac{1}{2}}\sim\frac{(-1)^n}{(\pi n)^{\frac{1}{2}}}\left[1-\frac{1}{8n}+\frac{1}{128n^2}+O\left(\frac{1}{n^3}\right)\right]. \tag{9.3}$$

Combining these relations in (8.40) gives

$$1-\eta\sim(1/\sqrt{2})\left[1+\frac{1}{8n}+\frac{1}{128n^2}\right]. \tag{9.4}$$

The next step is to find the determinant $|B_p|$ for large p; however we must consider the cases of even and odd p separately.

If $p=2n$ then

$$|B_p|=(2/\pi i)^p\,\Delta_n^2(1-\eta). \tag{9.5}$$

Now since

$$\left(\frac{2}{\pi}\right)^n\Delta_n\sim\frac{E}{n^{\frac{1}{4}}}\left[1-\frac{1}{64n^2}\right], \tag{9.6}$$

where $E=.70338\ldots$ and which can be obtained exactly from (8.43) after a detailed Euler–Maclaurin analysis, we have on combining (9.4) with (9.5)

$$i^p\,|B_p|\sim\frac{E^2}{p^{\frac{1}{2}}}\left[1+\frac{1}{4p}-\frac{3}{32p^2}\right]. \tag{9.7}$$

For $p=2n-1$ we arrive at the same expression; however the computations are a bit more involved.

The last step is to express (9.7) in terms of the radial distance r from the origin to the monomer at $(p,p-1)$. Clearly $r^2=p^2+(p-1)^2$. (9.8)

Solving the quadratic for p and taking the positive root gives

$$p=\tfrac{1}{2}[1+(2r^2-1)^{\frac{1}{2}}]\sim\tfrac{1}{2}r\sqrt{2}[1+1/r\sqrt{2}] \tag{9.9}$$

and also

$$p^{-\frac{1}{2}}\sim(2^{\frac{1}{2}}/r^{\frac{1}{2}})[1-(1/2r\sqrt{2})] \tag{9.10}$$

and

$$p^{-1}\sim(\sqrt{2}/r)[1-(1/r\sqrt{2})]. \tag{9.11}$$

Substituting in (9.7) gives

$$i^p\,|B_p|\sim(2^{\frac{1}{2}}E^2/r^{\frac{1}{2}})[1-(1/4r^2)] \tag{9.12}$$

and hence

$$4\omega(x,x\mid p-1,p)\sim(2B_0^2/xr^{\frac{1}{2}})[1-(1/4r^2)], \tag{9.13}$$

where

$$B_0=2^{1/8}E\quad\text{and}\quad B=2B_0^2. \tag{9.14}$$

This result verifies the numerical calculations done by FS which gave $-0.26/r^2$ as their second term.

In the nonsymmetric case, we start with (8.10) and use the asymptotic expansions for the Legendre functions[16]:

$$Q_n[\cosh\zeta]\sim\left(\frac{\pi}{n}\right)^{\frac{1}{2}}\cdot\frac{1}{\sqrt{2\sinh\zeta}}\cdot e^{-(n+\frac{1}{2})\zeta}$$
$$\times\left[1-\frac{1}{8n}\left(\frac{3e^{2\zeta}-1}{e^{2\zeta}-1}\right)\right] \tag{9.15}$$

[16] See Ref. 8, Chap. V, Sec. 83, p. 228 and Sec. 84, p. 231.

and

$$P_n[\cosh\zeta] \sim [\pi n(e^{2\zeta}-1)]^{-\frac{1}{2}} \times \left\{e^{(n+1)\zeta}\left[1-\frac{1}{8n}\left(\frac{e^{2\zeta}-3}{e^{2\zeta}-1}\right)\right] + e^{-n\zeta}e^{\frac{1}{2}i\pi}\left[1-\frac{1}{8n}\left(\frac{3e^{2\zeta}-1}{e^{2\zeta}-1}\right)\right]\right\}. \quad (9.16)$$

For $\tau > 1$ we put $\zeta = \xi - \frac{1}{2}i\pi$ so that $\cosh\zeta = -i\sinh\xi$, $\sinh\zeta = -i\cosh\xi$ and $e^{\xi} = \tau > 1$. Substituting in (9.15) gives

$$Q_{n-1}[-i\sinh\xi] \sim \left(\frac{\pi}{n}\right)^{\frac{1}{2}} \frac{i^n}{\tau^n(1+\tau^{-2})^{\frac{1}{2}}}\left[1+\frac{1}{8n}\left(\frac{\tau^2+3}{\tau^2+1}\right)\right] \quad (9.17)$$

and thus

$$1-\eta \sim (1+\tau^{-2})^{-\frac{1}{2}}[1+1/4n(\tau^2+1)]. \quad (9.18)$$

As $\tau \downarrow 1$, this clearly reduces to (9.4). For $\tau < 1$ we put $\zeta = \xi + \frac{1}{2}i\pi$ and see that $\cosh\zeta = i\sinh\xi$, $\sinh\zeta = i\cosh\xi$, where $e^{-\xi} = \tau$ and $\xi > 0$. Whence

$$Q_{n-1}[i\sinh\xi] \sim \left(\frac{\pi}{n}\right)^{\frac{1}{2}} \frac{\tau^{n-1}}{(1+\tau^{-2})^{\frac{1}{2}}}\frac{1}{i^n}\left[1+\frac{1}{8n}\left(\frac{1+3\tau^2}{1+\tau^2}\right)\right] \quad (9.19)$$

and

$$\pi i P_{n-1}[i\sinh\xi] \sim \left(\frac{\pi}{n}\right)^{\frac{1}{2}}(1+\tau^{-2})^{-\frac{1}{2}}\left\{\frac{i^n}{\tau^n}\left[1+\frac{1}{8n}\left(\frac{3+\tau^2}{1+\tau^2}\right)\right] + \frac{\tau^{n-1}}{i^{n-2}}\left[1+\frac{1}{8n}\left(\frac{1+3\tau^2}{1+\tau^2}\right)\right]\right\}. \quad (9.20)$$

In the expansion of $Q_{n-1}[i\sinh\xi] + \pi i P_{n-1}[i\sinh\xi]$, it is clear that the second part in the expansion of P_{n-1} must cancel the term coming from Q_{n-1} since at $\tau = 1$ the two terms are identical. Combining (8.10), (9.3), (9.19), and (9.20), we arrive again at (9.18) but this time for $\tau < 1$. Going back to the correlation we finally have for all τ that

$$4\omega(x, y \mid p-1, p) \sim \frac{2E^2}{xp^{\frac{1}{2}}}\left[\frac{2\tau^2}{1+\tau^2}\right]^{\frac{1}{2}}\left[1+\frac{1}{2p(1+\tau^2)}\right] = \frac{2B_0^2}{r^{\frac{1}{2}}}\left[\frac{2}{x^2+y^2}\right]^{\frac{1}{2}}\left[1+\frac{1}{2r\sqrt{2}}\left(\frac{y^2-x^2}{y^2+x^2}\right)\right], \quad (9.21)$$

in which the second term drops out when $\tau = 1$.

We note that the sign of the second term changes as we move across the diagonal from $(p-1, p)$ to $(p, p-1)$, i.e., if we interchange the x and y activities.

ACKNOWLEDGMENTS

I would like to thank Dr. J. Stephenson for suggesting this problem and for his advice and helpful criticism in the preparation of this paper.

The Commonwealth Post-Graduate Award is gratefully acknowledged by the author.

On the Space–Time Behavior of Schrödinger Wavefunctions*

W. HUNZIKER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey†
(Received 28 June 1965)

For the Schrödinger equation in R^l, with a potential $V(x_1 \cdots x_l)$ of the type considered by Kato, the following problem is solved: Given a monomial $M(x_1 \cdots x_l)$ of degree n in the coordinates, find sufficient conditions on the initial state u such that $Me^{-iHt}u$ is continuous in t and increasing in norm not faster than $|t|^n$ as $|t| \to \infty$. In the special case where $V(x_1 \cdots x_l)$ is a bounded C^∞-function with bounded derivatives, the result implies that $(u, t) \to e^{-iHt}u$ is a continuous mapping of $\mathcal{S}(R^l) \times R$ onto $\mathcal{S}(R^l)$, $\mathcal{S}(R^l)$ being the Schwartz space of rapidly decreasing functions in the usual topology.

1. INTRODUCTION

THE problem considered in this paper came up in connection with a discussion of the cluster properties of nonrelativistic multiparticle systems.[1] In trying to apply methods similar to those which proved effective in the relativistic case,[2] one faces at once the difficulty that nonrelativistic particles can move with arbitrarily large velocities. The question then arises what conditions may be imposed on the initial state u of the system such that $e^{-iHt}u$ describes, in a sense yet to be defined, particles which are localized at any time t and move with finite velocities. For the Schrödinger equation it is futile to describe localization in terms of supports of wavefunctions. What we can use, instead, are the expectation values of the coordinates or, for more detailed information, those of arbitrary monomials in the coordinates. The problem then takes the following form: Given a monomial M of degree n in the coordinates, find sufficient conditions on the initial state u such that $e^{-iHt}u$ is in the domain of (multiplication by) M for all t and such that $Me^{-iHt}u$ is bounded in norm by const $(1 + |t|)^n$.

Illustrative is the example of the free particle in one dimension, with coordinate x, momentum $p = -id/dx$, and energy $H = p^2/2$. Formally,

$$x^n e^{-ip^2t/2}u = e^{-ip^2t/2}(x + pt)^n u,$$

and it suffices, apparently, to require that

$$u \in \bigcap_{k+m \leq n} D(x^k p^m),$$

where k, m are integers ≥ 0, and where $D(A)$ denotes the domain of the operator A. In the case of interacting particles we shall see that it is enough, essentially, to replace in this condition p by H.

2. KATO POTENTIALS

Let R^l be the real Euclidean space of l dimensions with Cartesian coordinates $x = (x_1 \cdots x_l)$. Let p_j, $j = 1 \cdots l$, be the usual self-adjoint operators on $L^2(R^l)$ representing the momenta (formally, $p_j = -i\partial/\partial x_j$) and $H_0 = \sum_{j=1}^{l} p_j^2$. Let $V(x)$ be a real measurable function on R^l and V the operator of multiplication by $V(x)$, defined on all functions $u \in L^2(R^l)$ for which this product is again in $L^2(R^l)$. Furthermore, we require V to satisfy the Kato condition[3]:

$D(H_0) \subset D(V)$, and there exist constants $a < 1$, $b < \infty$, such that, for all $u \in D(H_0)$,

$$||Vu|| \leq a\,||H_0 u|| + b\,||u||. \tag{1}$$

Then $H \equiv H_0 + V$ has the domain $D(H_0)$ and is self-adjoint.[3] These are the properties of H which will be tacitly assumed in the rest of this paper.

An immediate consequence of (1) is that the norms $||Hu|| + ||u||$ and $||H_0 u|| + ||u||$ on $D(H_0)$ are equivalent. Therefore, since H_0 is the closure of its restriction to the Schwartz space $\mathcal{S}(R^l)$ of rapidly decreasing functions, the same is true for H.

Purely as an illustration—and not for later use—we state conditions on $V(x)$ which are more transparent than (1) and imply (1)[4]: Let $V(x) = \sum_{k=1}^{n} V_k(x)$ and suppose that for each k there exists an affine transformation $(x_1 \cdots x_l) \to (y_1 \cdots y_l)$ in R^l such that V_k depends only on $y_1 \cdots y_m$, $m \leq l$: $V_k(x_1 \cdots x_l) = U_k(y_1 \cdots y_m)$. If each U_k can be written as

$$U_k = U_k^2 + U_k^\infty, \tag{2}$$

* Supported in part by the U. S. Air Force through the Air Force Office of Scientific Research.
† Present address: Seminar für theoretische Physik der ETH, Zürich, Switzerland.
[1] W. Hunziker, J. Math. Phys. 6, 6 (1965).
[2] K. Hepp, Helv. Phys. Acta 37, 659 (1964), and J. Math. Phys. 6, 1762 (1965).
[3] T. Kato, Trans. Am. Math. Soc. 70, 195 (1951).
[4] E. Nelson, J. Math. Phys. 5, 332 (1964).

with $U_k^p \in L^p(R^m)$, $U_k^\infty \in L^\infty(R^m)$, where

$$m/2 < p \tag{3}$$

and

$$2 \leq p, \tag{4}$$

then (1) holds, and a can even be chosen arbitrarily small (at the expense of increasing b). If the system consists of a (finite) number of particles in R^3 interacting via two-body forces, then $m = 3$ for all k. Equation (4) then requires that each two-body potential, as a function of the three relative coordinates of the two particles, belongs to $L^2(R^3) + L^\infty(R^3)$, in the sense of (2). For many-body forces one has $m \geq 6$ and sufficient conditions follow from (3).

3. RESULTS

Definitions: $x^n = x_1^{n_1} x_2^{n_2} \cdots x_l^{n_l}$, n being the multi-index $(n_1 \cdots n_l)$ with n_j integer ≥ 0; $|n| = \sum_j n_j$; $n = 0$ means $|n| = 0$; $k \leq n$ means $k_j \leq n_j$ for all j; $k < n$ means $k \leq n$ and $|k| < |n|$.

For any multi-index n, x^n also denotes the operator of multiplication with the function x^n, defined on all $u \in L^2(R^l)$ for which this product is again in $L^2(R^l)$. So defined, x^n is self-adjoint (and different, in general, from the monomial x^n of the *operators* $x_1 \cdots x_l$).

$x^n{}_{,j}$ and $x^n{}_{,ij}$ denote first- and second-order partial derivatives of x^n with respect to x_j. They are again monomials and thus defined as operators on $L^2(R^l)$.

For any multi-index n we define a linear subset D_n of $L^2(R^l)$ and a norm $||\ ||_n$ on D_n by

$$D_n = \bigcap_{\substack{k \leq n \\ m \leq |n|-|k|}} D(x^k H^m),$$

$$||u||_n = \sup_{\substack{k \leq n \\ m \leq |n|-|k|}} ||x^k H^m u||,$$

m integer ≥ 0.

Theorem 1. Under the assumptions stated in Sec. 2, the following holds for any multi-index n:

(a) D_n is invariant under the unitary group e^{-iHt}.

(b) For any $u \in D_n$, $e^{-iHt}u$ is continuous in t in the sense of the norm $||\ ||_n$, and there exists a constant c_n such that

$$||e^{-iHt}u||_n \leq c_n(1 + |t|)^{|n|}\ ||u||_n.$$

(c) For any $u \in D_n$,

$$x^n e^{-iHt}u = e^{-iHt}x^n u + i\int_0^t e^{-iH(t-\tau)}[H, x^n]e^{-iH\tau}u\, d\tau,$$

where the commutator is *defined* as

$$[H, x^n] \equiv \sum_{j=1}^{l} (-2ip_j x^n{}_{,j} + x^n{}_{,jj}).$$

In the L^2-norm, the integrand is continuous in τ and bounded by const $(1 + |\tau|)^{|n|-1}\ ||u||_n$.

Without further assumptions on the potential we are not able to exclude the possibility that D_n (for $|n|$ sufficiently large) contains only the vector 0. If $V(x)$ is a C^∞-function in some open set O of R^l, then any C^∞-function of compact support contained in O belongs to $D_\infty \equiv \bigcap_n D_n$. In the cases usually considered in physics, such as a system of particles interacting via Coulomb forces, for example, it is clear that D_∞ is dense in $L^2(R^l)$. If, in particular, $V(x)$ is a bounded C^∞-function on R^l with bounded derivatives (the bounds depending on the derivatives), it follows that

$$D_n = \bigcap_{\substack{k \leq n \\ m \leq |n|-|k|}} D(x^k H_0^m)$$

and that the norm $||\ ||_n$ is equivalent to the norm

$$||u||_n' \equiv \sup_{\substack{k \leq n \\ m \leq |n|-|k|}} ||x^k H_0^m u||.$$

D_∞ is then simply $\mathcal{S}(R^l)$, and since the system of norms $||\ ||_n'$ generates the usual topology on $\mathcal{S}(R^l)$, we obtain as a consequence of Theorem 1:

Theorem 2. If $V(x)$ is a bounded C^∞-function on R^l with bounded derivatives, then $\mathcal{S}(R^l)$ is invariant under the unitary group e^{-iHt} and the mapping $(u, t) \to e^{-iHt}u$ of $\mathcal{S}(R^l) \times R$ onto $\mathcal{S}(R^l)$ is continuous [in the sense of the conventional topology on $\mathcal{S}(R^l)$]

4. PROOFS

Lemma 1. For any multi-index n we define four linear subsets M_n^i of $L^2(R^l)$, $i = 1 \cdots 4$, and a norm $||\ ||_n^i$ on each M_n^i by

$$M_n^1 = \bigcap_{k \leq n} D(H_0 x^k),$$

$$||u||_n^1 = \sup_{k \leq n} (||x^k u|| + ||H_0 x^k u||),$$

$$M_n^2 = \bigcap_{k \leq n} D(x^k) \cap D(x^k H_0),$$

$$||u||_n^2 = \sup_{k \leq n} (||x^k u|| + ||x^k H_0 u||),$$

$$M_n^3 = \bigcap_{k \leq n} D(H x^k),$$

$$||u||_n^3 = \sup_{k \leq n} (||x^k u|| + ||H x^k u||),$$

$$M_n^4 = \bigcap_{k \le n} D(x^k) \cap D(x^k H),$$

$$||u||_n^4 = \sup_{k \le n} (||x^k u|| + ||x^k H u||).$$

Then:

(d) For fixed n, the 4 spaces M_n^i coincide:

$$M_n^i = M_n^j \tag{5}$$

for all i, j; and on this single space, which we call M_n, the four norms $||\ ||_n^i$ are equivalent:

$$||\ ||_n^i \cong ||\ ||_n^j \tag{6}$$

for all i, j.

(e) $\mathcal{S}(R^l)$ is dense in M_n in the sense of any of the norms $||\ ||_n^i$.

(f) On M_n, the formal commutation rules hold between p_j and any monomial x^k with $k \le n$.

Proof: For $i, j = 1, 3$, (5) follows from $D(H) = D(H_0)$ and (6) from the equivalence of the norms $||Hu|| + ||u||$ and $||H_0 u|| + ||u||$ on $D(H)$. For the same reason, (d) holds for $n = 0$. Hence it suffices to prove (5) and (6) for $i, j = 3, 4$, assuming that (d) holds for any multi-index $< n$ (this also covers the case $i, j = 1, 2$, which is obtained by setting $V = 0$). We only show that

$$M_n^4 \subset M_n^3 \quad \text{and, for all} \quad u \in M_n^4, \qquad ||u||_n^3 \le \text{const}\ ||u||_n^4, \tag{7}$$

since the proof of (7) with 3 and 4 interchanged is completely analogous (and in fact not needed for the demonstration of theorem 1). Let $u \in M_n^4$. For any $k < n$, $M_n^4 \subset M_k^4$, so that by induction hypothesis $u \in M_k^3$ and

$$||u||_k^3 \le \text{const}\ ||u||_k^4 \le \text{const}\ ||u||_n^4.$$

It remains to prove that $u \in D(Hx^n)$ and that $||Hx^n u|| \le \text{const}\ ||u||_n^4$. For any $v \in \mathcal{S}(R^l)$ we have

$$(u, x^n H v) = (u, H x^n v) + \sum_{j=1}^{l} [-2i(u, x^n{}_{,j} p_j v) + (u, x^n{}_{,jj} v)]. \tag{8}$$

By induction hypothesis, $u \in M_k^1$ and $||u||_k^1 \le \text{const}\ ||u||_k^4 \le \text{const}\ ||u||_n^4$ for any $k < n$, in particular for $k = (n_1 \cdots, n_i - 1, \cdots n)$. Therefore $u \in D(p_j x^n{}_{,j})$ and

$$||p_j x^n{}_{,j} u|| \le ||x^n{}_{,j} u|| + ||H_0 x^n{}_{,j} u|| \le \text{const}\ ||u||_n^4, \tag{9}$$

so that we can write (8) in the form

$$(x^n u, Hv) = (x^n H u, v) + \sum_{j=1}^{l} [-2i(p_j x^n{}_{,j} u, v) + (x^n{}_{,jj} u, v)]. \tag{10}$$

Since H is the closure of its restriction to $\mathcal{S}(R^l)$, this extends by continuity to all $v \in D(H)$. Therefore, by (9) and because $u \in M_n^4$,

$$|(x^n u, Hv)| \le \text{const}\ ||u||_n^4\ ||v||$$

for all $v \in D(H)$. Hence $x^n u \in D(H)$ and $||Hx^n u|| \le \text{const}\ ||u||_n^4$, which is the desired result.

(e) is proved by showing that $\mathcal{S}(R^l)$ is dense in M_n in the norm $||\ ||_n^1$. This is true for $n = 0$, so again we proceed by induction, assuming (e) for all multi-indices $< n$. Let $u \in M_n$ and suppose that $n_1 \neq 0$. Then $(i + x_1)u \in M_{n'}$, $n' = (n_1 - 1, n_2, \cdots, n_l)$. By induction hypothesis there is a sequence v_j, $v_j \in \mathcal{S}(R^l)$, $||v_j - (i + x_1)u||_{n'}^1 \to 0$ as $j \to \infty$. Defining $u_j = (i + x_1)^{-1} v_j$ we find that $u_j \in \mathcal{S}(R^l)$, $u_j \to u$, and

$$x^k (i + x_1) u_j \to x^k (i + x_1) u, \tag{11}$$

$$H_0 x^k (i + x_1) u_j \to H_0 x^k (i + x_1) u, \tag{12}$$

for all $k \le n'$, the arrow indicating convergence in the L^2-norm. Multiplying (11) by $(i + x_1)^{-1}$ we see that $x^k u_j \to x^k u$ for all $k \le n'$ and therefore also for $k = n$, by (11). Suppose now that $H_0 x^k u_j$ is a Cauchy sequence for all $k \le n'$. Since H_0 is closed, this implies $H_0 x^k u_j \to H_0 x^k u$ for all $k \le n'$ and therefore also for $k = n$, by (12), or altogether $||u_j - u||_n^1 \to 0$. It remains to prove that $H_0 x^k u_j$ is a Cauchy sequence for any $k \le n'$. For $w \in \mathcal{S}(R^l)$ one has $||w|| \le ||(i + x_1)w||$ and $||p_j w|| \le ||H_0 w|| + ||w||$. Using this, one easily verifies that

$$||H_0 x^k w|| \le 3\ ||H_0 x^k (i + x_1) w|| + 4\ ||x^k (i + x_1) w||$$

for $w \in \mathcal{S}(R^l)$. Hence (11) and (12) imply that $H_0 x^k u_j$ is a Cauchy sequence.

The proof of (f) is now obvious: the formal commutation rules between p_j and any monomial x^k hold on $\mathcal{S}(R^l)$ and, for $k \le n$, extend to M_n by continuity.

Proof of Theorem 1. The theorem holds for $n = 0$. For $n > 0$, we assume that it holds for all multi-indices $< n$ and prove it for n.

Let $u \in D_n$. Then, for any $k < n$ and $m \le |n| - |k|$, $H^m u \in D_k$ and $||H^m u||_k \le ||u||_n$. By induction hypothesis, $x^k H^m e^{-iHt} u$ exists and is continuous in t and bounded by const $(1 + |t|)^{|k|} ||u||_n$.[5] The same follows from (c) for the remaining case $k, m = n, 0$, hence it suffices to prove (c).

First we show that the integrand in (c) is well defined, continuous in τ and bounded by const $(1 + |\tau|)^{|n|-1} ||u||_n$. By induction hypothesis, this

[5] Unless stated differently, the terms "continuous" and "bounded", applied to L^2-valued functions, are understood in the sense of the L^2-norm.

clearly holds for the contribution of the terms $x^n_{,ii}$ in the commutator. Also by induction hypothesis, in the terminology of Lemma 1,

$$e^{-iH\tau}u \in M^4_k \tag{13}$$

for all τ if $k < n$, and, in the sense of $||\ ||^4_k$, $e^{-iH\tau}u$ is continuous in τ and bounded by const $(1 + |\tau|)^{|k|}||u||_n$. In particular, this applies to $k = n' = (n_1 \cdots, n_i - 1, \cdots n_l)$. By Lemma 1, we have for any $v \in M_{n'}$

$$||p_i x^n_{,i} v|| \leq ||x^n_{,i} v|| + ||H_0 x^n_{,i} v|| \leq ||v||^1_{n'} \leq \text{const}\ ||v||^4_{n'},$$

hence we conclude that $p_i x^n_{,i} e^{-iH\tau}u$ exists, is continuous in τ and bounded by const $(1 + |\tau|)^{|n|-1}||u||_n$.

It remains to be proved that the lhs (left-hand side) of (c) exists and is equal to the rhs. Let A be a bounded operator on $L^2(R^l)$ mapping $D(H)$ into itself continuously, in the sense of the norm $||Hu|| + ||u||$ on $D(H)$. Then, for any $u, v \in D(H)$,

$$(d/d\tau)(v, e^{iH\tau}Ae^{-iH\tau}u) = (v, ie^{iH\tau}[H, A]e^{-iH\tau}u).$$

Integration over $0 \leq \tau \leq t$ yields

$$(v, e^{iHt}Ae^{-iHt}u - Au) = \left(v, i\int_0^t e^{iH\tau}[H, A]e^{-iH\tau}u\, d\tau\right), \tag{14}$$

where we have used that the integrand in (14) is continuous in τ, so that the integration can be carried out inside the scalar product. Since $D(H)$ is dense in $L^2(R^l)$, we conclude that for any $u \in D(H)$

$$Ae^{-iHt}u = e^{-iHt}Au + i\int_0^t e^{-iH(t-\tau)}[H, A]e^{-iH\tau}u\, d\tau.$$

An example of such an operator A is the operator of multiplication by a function $A(x)$ of $\mathcal{S}(R^l)$. Choosing $A(x) = x^n e^{-|\alpha x|^2}$, α real $\neq 0$, we obtain

$$x^n e^{-|\alpha x|^2}e^{-iHt}u = e^{-iHt}x^n e^{-|\alpha x|^2}u + i\int_0^t e^{-iH(t-\tau)}[H, x^n e^{-|\alpha x|^2}]e^{-iH\tau}u\, d\tau. \tag{15}$$

To discuss the limit as $\alpha \to 0$, we apply the following Lemma:

Lemma 2. Let $t \to u(t)$ be a continuous mapping $R \to L^2(R^l)$. Then, for any multi-index n,

$$\lim_{\alpha\to 0} (\alpha x)^n e^{-|\alpha x|^2}u(t) = \begin{cases} u(t) & \text{if } n = 0 \\ 0 & \text{if } n > 0 \end{cases}$$

in the sense of the L^2-norm, uniformly in finite t-intervals.

Proof: On $L^2(R^l)$, let P_r be the projection defined by

$$P_r u(x) = \begin{cases} u(x) & \text{if } |x| \leq r \\ 0 & \text{if } |x| > r. \end{cases}$$

$||(1 - P_r)u(t)||$ is continuous in t and, for fixed t, converges to 0 monotonically as $r \to \infty$. By Dini's lemma, the convergence to 0 is therefore uniform in finite t-intervals. The norm of the operator $(\alpha x)^n e^{-|\alpha x|^2}$ is independent of α and, for fixed r,

$$\lim_{\alpha\to 0} (\alpha x)^n e^{-|\alpha x|^2}P_r = \begin{cases} P_r & \text{if } n = 0 \\ 0 & \text{if } n > 0, \end{cases}$$

in the sense of the operator norm. The rest of the proof is clear.

Now let $u \in D_n$. As we shall see below, the rhs of (15) converges in the L^2-norm to the rhs of (c) for any t as $\alpha \to 0$. Also, by Lemma 2,

$$e^{-|\alpha x|^2}e^{-iHt}u \to e^{-iHt}u.$$

Since x^n is closed we conclude that $e^{-iHt}u \in D(x^n)$ and that $x^n e^{-iHt}u$ is given by (c).

To complete the proof we show now that, as $\alpha \to 0$, the integrand in (15) converges in norm to the integrand in (c), uniformly in $0 \leq \tau \leq t$. By (13) and (f) we can write the integrand in (c) in the form

$$-e^{-iH(t-\tau)} \sum_{i=1}^{l} (2ix^n_{,i}p_i e^{-iH\tau}u + x^n_{,ii}e^{-iH\tau}u), \tag{16}$$

each term in the sum being continuous in τ. On the other hand, the integrand in (15) is, explicitly,

$$-e^{-iH(t-\tau)} \sum_{i=1}^{l} \left(\left[1 - \frac{2}{n_i}(\alpha x_i)^2\right]e^{-|\alpha x|^2}2ix^n_{,i}p_i e^{-iH\tau}u + \left[1 - \frac{4n_i + 2}{n_i(n_i - 1)}(\alpha x_i)^2 + \frac{4}{n_i(n_i - 1)}(\alpha x_i)^4\right]e^{-|\alpha x|^2}x^n_{,ii}e^{-iH\tau}u\right). \tag{17}$$

By Lemma 2, (17) converges in norm to (16) as $\alpha \to 0$, uniformly in $0 \leq \tau \leq t$.

5. CONCLUDING REMARKS

In classical mechanics, the velocity of a particle can be estimated in terms of the total energy if $V(x)$ is bounded from below. This suggests the following question: let $V(x)$ be such that $D(V) \cap D(H_0)$ is dense in $L^2(R^l)$ and that $H_0 + V$ is bounded from below, and let H be the Friedrichs extension of $H_0 + V$. Does Theorem 1 still hold in this case?

If $V(x)$ is the potential of an N-particle system

($l = 3N$) which is not affected by external forces, then $V(x)$ is invariant under simultaneous translations and rotations of the N particles, and there is a unitary/antiunitary representation (up to a factor) $U(G)$ of the Galilei group (including time inversion). For any integer $p \geq 0$, let $D_p = \bigcap_{|n| \leq p} D_n$ and let $\sum_{|n| \leq p} || \, ||_n$ be the norm on D_p. Then $U(G)$ is a bounded operator mapping D_p onto itself and depending continuously on the Galilei transformation G, in the strong sense on D_p.

A final remark applies to Ref. 1: Suppose that the N particles interact via two-body forces of short range, in the sense that $\int d^3x \, |x|^n \, |V(x)|^2 < \infty$ for all $n \geq 0$ and any of the two-body potentials $V(x)$. Using Theorem 1 we can show that

$$T(-a_1 \cdots -a_n)e^{-iHt}T(a_1 \cdots a_n),$$

applied to a certain class of states, converges to the cluster limit faster than any inverse power of a. However, we have not been able to obtain a similar result for the wave operators $\Omega_\alpha^\pm$. The difficulty is the following: we do not know sufficient conditions on the channel state f_α such that $\Omega_\alpha^\pm f_\alpha \in D_\infty$—except for the very special case where Theorem 2 applies.

ACKNOWLEDGMENT

The author gratefully acknowledges the partial financial support obtained from the Swiss National Fund.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 2 FEBRUARY 1966

New Approach to the Ising Problem

C. A. HURST
University of Adelaide, Adelaide, South Australia
(Received 26 July 1965)

The partition function for the Ising model on a two-dimensional rectangular lattice is cast into a form which closely resembles the vacuum expectation value of the S-matrix in quantum field theory. The standard Onsager expression is obtained very simply in this formalism. It is further shown how the nonsoluble models can be expressed as field theories with quartic interactions, thereby resembling the standard many-body fermion theories.

1. INTRODUCTION

ALTHOUGH the Pfaffian method[1] hitherto used for solving many examples of the Ising model was originally expressed in an operator formalism, which implied some connection with quantum field theory, the analogy was rather slight, and the method of solution depended little on the vast armory of methods of field theory. This was unfortunate because the sophisticated techniques developed for handling the deep problems of elementary particles have proved very powerful in other contexts such as many-body theory and statistical mechanics.[2] In particular the development of approximation methods for handling the unsolved problems would be accelerated if such techniques were available.

This paper is concerned with two questions. The first is to show how the solution of the Ising model can be reduced to the problem of calculating the vacuum expectation value of an expression which can be regarded as a "time"-ordered product of exponentials and which closely resembles the standard expressions employed in the functional equation approach to field theory[3]. For the soluble models the exponent is a quadratic expression in fermion-type field operators, and for this case the problem of evaluation is almost trivial. For the well-known unsolved problems, such as the rectangular next-nearest-neighbor problem and the cubic lattice, it is shown how the exponent has an additional quartic term. Such an expression is of the same type as that encountered in the treatment of many-fermion problems such as in the electron gas, the nuclear model, and in superconductivity.

The second question is the one already referred to, namely the derivation of the Onsager expression for the partition function for the rectangular lattice with nearest-neighbor interactions. The calculation of partition functions for other soluble lattices, such as the triangular lattice, presents no further difficulties.

The first question is discussed in Sec. 2 and the second question in Sec. 3 of this paper.

2. TRANSFORMATION OF THE PARTITION FUNCTION

The starting point is an abstract algebraic definition of a Pfaffian.[4] An algebra E, generated by forming all possible sums and products, with numerical coefficients, from a set of elements $x_1, \cdots, x_n$ is called an *exterior algebra* if the relation

$$x_i x_j + x_j x_i = 0 \tag{1}$$

holds between all pairs of generators. In particular we note the relation

$$x_i^2 = 0. \tag{2}$$

Both the relations (1) and (2) hold not only for the generators themselves but also for all elements of the linear space spanned by these generators.

Because of these defining relations the algebra E will have dimensions 2^n. The linear subspace E_m spanned by all products $x_{i_1} \cdots x_{i_m}$ of degree m is called homogeneous of degree m, and every element $x \in E$ can be uniquely written as a sum

$$x = \sum x_m \quad \text{with} \quad x_m \in E_m. \tag{3}$$

x_m is called the *homogeneous m-component* of x. In particular if $m = 2$, we have a subspace E_2 all of whose elements commute with each other. In this case the exponential addition law is valid, namely

$$(\exp a)(\exp b) = \exp (a + b) \tag{4}$$

[1] H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964). This book will be referred to as G.

[2] D. Pines, *The Many Body Problem* (W. A. Benjamin, Inc., New York, 1962).

[3] Quantum field theory methods have been applied by other authors to the Ising model, but their approaches are quite different from that adopted here. For example, F. H. Stillinger, Jr., Phys. Rev. 135, A1646 (1964).

[4] C. Chevally, *The Construction and Study of Certain Important Algebras* (The Mathematical Society of Japan, Tokyo, 1955).

for $a, b \in E_2$ (more generally if $a, b \in E_m$ for m even). Now suppose

$$\Gamma = \sum_{i<j} \alpha_{ij} x_i x_j \tag{5}$$

is an arbitrary element contained in E_2 and $n = 2p$ is even. Then the homogeneous $2p$-component of $\exp \Gamma$ is a multiple of $x_1 \cdots x_{2p}$,

$$(\exp \Gamma)_{2p} = P_\Gamma(x_1 \cdots x_{2p}) \tag{6}$$

and P_Γ, *a number, is called the Pfaffian of* $\Gamma \in E_2$. As an example consider the case $p = 2$, and write

$$\Gamma = \alpha_{12}x_1x_2 + \alpha_{13}x_1x_3 + \alpha_{14}x_1x_4 + \alpha_{23}x_2x_3 + \alpha_{24}x_2x_4 + \alpha_{34}x_3x_4.$$

Then, from Eqs. (1) and (2) we find

$$\Gamma^2 = (\alpha_{12}x_1x_2 + \alpha_{13}x_1x_3 + \alpha_{14}x_1x_4 + \alpha_{23}x_2x_3 + \alpha_{24}x_2x_4 + \alpha_{34}x_3x_4)^2 = 2(\alpha_{12}\alpha_{34} - \alpha_{13}\alpha_{24} + \alpha_{14}\alpha_{23})x_1x_2x_3x_4$$

and

$$P_\Gamma = \alpha_{12}\alpha_{34} - \alpha_{13}\alpha_{24} + \alpha_{14}\alpha_{23}, \tag{7}$$

which is the usual definition. Incidentally it is simple to prove from the definition (4) that $P_\Gamma^2 = D$, where D is the antisymmetric determinant formed from the coefficients α_{ij} of Γ. Conversely, if $\Delta \in E$, for $p = 2$, is an arbitrary element of even order, with

$$\Delta = \alpha_0 + \alpha_{12}x_1x_2 + \alpha_{13}x_1x_3 + \alpha_{14}x_1x_4 + \alpha_{23}x_2x_3 + \alpha_{24}x_2x_4 + \alpha_{34}x_3x_4 + \alpha_{1234}x_1x_2x_3x_4, \tag{8}$$

$\alpha_0 \neq 0$,

then the necessary and sufficient condition for $\Delta = \exp \Gamma$ for some Γ is that

$$\alpha_0\alpha_{1234} = \alpha_{12}\alpha_{34} - \alpha_{13}\alpha_{24} + \alpha_{14}\alpha_{23}, \tag{9}$$

which is *just the consistency condition for factorizability* introduced earlier[5] to justify the application of the Pfaffian method to the solution of the Ising problem for the rectangular lattice. For $p > 2$, additional relations must be satisfied if the representation $\Delta = \exp \Gamma$ is to be possible, and the whole set of such relations are identical with the consistency conditions derived in C for the more general case. So the condition for solubility of an Ising model by the Pfaffian (and hence by the Onsager method) appears to be equivalent to the condition that an even element of an exterior algebra be expressible as an exponential. We notice that if the condition (9) is not satisfied, then we must write instead

$$\Delta = \alpha_0 \exp\Big(\sum_{i>j} \alpha'_{ij}x_ix_j + \beta_{1234}x_1x_2x_3x_4\Big) \tag{10}$$

with

$$\alpha'_{ij} = \alpha_{ij}/\alpha_0 \quad \text{and} \quad \alpha_0^2\beta_{1234} = \alpha_0\alpha_{1234} - (\alpha_{12}\alpha_{34} - \alpha_{13}\alpha_{24} + \alpha_{14}\alpha_{23}).$$

In order to show the connection with the Ising model, consider first of all the case of the rectangular two-dimensional lattice. The problem is to evaluate the expression

$$Z_1 = \Big(\Omega, \prod_{j=1}^{N} (1 + a_{j-m}^{(2)}a_{j-1}^{(1)} + xa_j^{(1)\dagger}a_{j-1}^{(1)} + ya_j^{(2)\dagger}a_{j-1}^{(1)} + xa_j^{(1)\dagger}a_{j-m}^{(2)} + ya_j^{(2)\dagger}a_{j-m}^{(2)} + xya_j^{(2)\dagger}a_j^{(1)\dagger} + xya_j^{(2)\dagger}a_j^{(1)\dagger}a_{j-m}^{(2)}a_{j-1}^{(1)})\Omega\Big), \tag{11}$$

where $a_j^{(i)}$, $a_j^{(i)\dagger}$ are a set of $2N$ fermion annihilation and creation operators satisfying the anticommutation relations

$$[a_j^{(i)}, a_{j'}^{(i')}]_+ = [a_j^{(i)\dagger}, a_{j'}^{(i')\dagger}]_+ = 0, \qquad [a_j^{(i)}, a_{j'}^{(i')\dagger}]_+ \equiv a_j^{(i)}a_{j'}^{(i')\dagger} + a_{j'}^{(i')\dagger}a_j^{(i)} = \delta_{ii'}\delta_{jj'}. \tag{12}$$

The indices (1) and (2) refer to horizontal and vertical bonds respectively. The product in equation (11) is to be read from right to left in order of increasing j and the special considerations required to account for edge conditions are neglected. Ω is the vacuum state defined by the simultaneous equations

$$a_j^{(1)}\Omega = a_j^{(2)}\Omega = 0, \quad \text{all } j. \tag{13}$$

For a particular j the four operators $a_j^{(2)\dagger}$, $a_j^{(1)\dagger}$, $a_{j-m}^{(2)}$ and $a_{j-1}^{(1)}$ generate an exterior algebra of dimensions 2^4, and the condition for the bracket expression in Z_1 to be written as an exponential is satisfied. Hence Z_1 can be written as

$$Z_1 = (\Omega, \prod_j \exp[a_{j-m}^{(2)}a_{j-1}^{(1)} + xa_j^{(1)\dagger}a_{j-1}^{(1)} + ya_j^{(2)\dagger}a_{j-1}^{(1)} + xa_j^{(1)\dagger}a_{j-m}^{(2)} + ya_j^{(2)\dagger}a_{j-m}^{(2)} + xya_j^{(2)}a_j^{(1)\dagger}]\Omega). \tag{14}$$

We now define the following operators

$$A^{(1)}(j) = a_{j-1}^{(1)}, \quad A^{(3)}(j) = xa_j^{(1)\dagger}, \quad A^{(2)}(j) = a_{j-m}^{(2)}, \quad A^{(4)}(j) = ya_j^{(2)\dagger}, \tag{15}$$

and then the expression in square brackets in Eq. (14) can be written as

[5] C. A. Hurst, J. Math. Phys. 5, 90 (1964), to be referred to as C.

$$\frac{1}{2}\sum_{p=1}^{4}\sum_{q=1}^{4} k_{pq}A^{(p)}(j)A^{(q)}(j), \tag{16}$$

where the matrix (k_{pq}) is given by

$$K = (k_{pq}) = \begin{vmatrix} 0 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 0 \end{vmatrix} = (-k_{qp}) = -\tilde{K}. \tag{17}$$

In terms of this notation Eq. (14) can now be written in the very suggestive form

$$Z_1 = (\Omega, T[\exp \tfrac{1}{2}\sum_{j}\sum_{p}\sum_{q} k_{pq}A^{(p)}(j)A^{(q)}(j)]\Omega), \tag{18}$$

where T denotes an ordering operator which requires the terms obtained on expanding the exponential to be written from right to left in order of increasing j. Such an operator is the discrete analog of an operator very well known in quantum field theory—the Dyson chronological operator.[6] In this formalism Z_1 is the analog of the vacuum expectation value of the S-matrix with an interaction Lagrangian (or Hamiltonian) which is a quadratic function of the field operators. Such a theory is known as a pair field theory and was first introduced by Wentzel,[7] who obtained exact solutions by diagonalization. In this context therefore, it is not at all surprising that the rectangular lattice can be solved exactly. Also, from the remarks made earlier about consistency conditions and exponential representation it is clear that all soluble models lead to an expression for Z_1 which contains an exponent which is quadratic in fermion operators, and so all are types of pair field theories.

The way in which the unsolved models, containing crossed bonds, are represented can best be explained by considering the two-dimensional next-nearest-neighbor problem. An abortive attempt to solve this problem was described in Chap. 7 of G. At each point of intersection of the diagonal bonds a new lattice point was introduced, to which a cross-over condition is applied thereby permitting bonds only to go straight through and not to turn through 90°. The corresponding new factor in the operator product in Z_1 is given by

$$\Delta = (1 + ua_{j'}^{(3)\dagger}a_j^{(3)} + va_{j'}^{(4)\dagger}a_{j+1}^{(4)} + uva_{j'}^{(4)\dagger}a_{j'}^{(3)\dagger}a_{j+1}^{(4)}a_j^{(3)}), \tag{19}$$

where $a^{(3)}$, $a^{(3)\dagger}$ are operators annihilating and creating bonds in the diagonal j to $j + m + 1$ and u is the corresponding weight and $a^{(4)}$, $a^{(4)\dagger}$, v are the corresponding quantities for the diagonal $j + 1$ to $j + m$. The index j' refers to the new lattice point inserted at the intersection of the diagonals. Comparing Eq. (19) with Eq. (8) we see that

$$\alpha_0 = 1, \alpha_{12} = \alpha_{14} = \alpha_{23} = \alpha_{34} = 0,$$

$$\alpha_{13} = u, \alpha_{24} = v, \alpha_{1234} = uv$$

and hence

$$\Delta = \exp(ua_{j'}^{(3)\dagger}a_j^{(3)} + va_{j'}^{(4)\dagger}a_{j+1}^{(4)} + 2uva_{j'}^{(4)\dagger}a_{j'}^{(3)\dagger}a_{j+1}^{(4)}a_j^{(3)}). \tag{20}$$

So the next-nearest-neighbor problem is analogous to a quantum field theory (or a many-fermion problem) with a quartic interaction. The same argument may be applied to any lattice with crossed bonds, so that their solution will depend on the development of a successful method for handling such problems. These sorts of problems are well known in many-body theory, and the question is now what sort of approximate methods would be the most suitable. It is not claimed that all problems with crossed bonds are now equivalent, for it could well be that the cubic lattice, for example, has a far more singular "interaction" than the next-nearest-neighbor problem.

Also arising from the representation (18) is the question of the meaning of the Green's functions which may be obtained by considering vacuum expectation values of products of operators $A^{(p)}$. For example we can define a function

$$G_2^{(rs)}(j', j'') = (\Omega, T[\exp \tfrac{1}{2}\sum_{j}\sum_{p}\sum_{q} k_{pq}A^{(p)}(j) \times A^{(q)}(j), A^{(r)}(j'), A^{(s)}(j'')]\Omega), \tag{21}$$

the two-point Green's function. Such a function will have some connection with correlations, although, because of the anticommutation properties of the operators $A^{(r)}$, $A^{(s)}$, the connection is certainly not as direct as might be expected. Similarly a function G_4 could be defined which will be relevant to the discussion of susceptibilities. All these questions are still under investigation. In order to demonstrate not only the greater perspicuity that Eq. (18) affords but also the much greater simplicity of

[6] N. N. Bogoliubov and D. V. Shirokov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959).

[7] G. Wentzel, Helv. Phys. Acta 15, 111 (1942).

calculation, the next section will be devoted to its evaluation.

3. THE EVALUATION OF THE PARTITION FUNCTION

In order to evaluate Z_1 the exponential can be expanded out and the operator products in the resulting infinite series arranged in normal order according to Wick's theorem.[1] In the process of reordering, there appear time-ordered contractions of pairs of operators $A^{(p)}\cdot A^{(q)}\cdot$ which are defined as follows:

$$T(A^{(p)}(j), A^{(q)}(j')) = N(A^{(p)}(j), A^{(q)}(j')) + A^{(p)}(j)\cdot A^{(q)}(j')\cdot, \qquad (22)$$

where

$$T(A^{(p)}(j), A^{(q)}(j')) = \theta(j - j')A^{(p)}(j)A^{(q)}(j') - \theta(j' - j)A^{(q)}(j')A^{(p)}(j),$$

$$\theta(j) = 1 \quad \text{for} \quad j \geqq 0$$
$$= 0 \quad \text{for} \quad j < 0.$$

The normal product $N(\cdots)$ is defined as the product of the operator arguments written with annihilation operators to the right, creation operators to the left, together with a negative sign if there is a reordering of fermion operators involving an odd permutation. So, for example, we have

$$T(A^{(1)}(j), A^{(3)}(j')) = x\theta(j - j')a^{(1)}_{i-1}a^{(3)\dagger}_{i'} - x\theta(j' - j)a^{(3)\dagger}_{i'}a'^{(1)}_{i-1}$$
$$= -xa^{(3)\dagger}_{i'}a^{(1)}_{j-1} + x\theta(j - j')\delta_{i',i-1}$$

and therefore

$$A^{(1)}(j)\cdot A^{(3)}(j')\cdot = x\delta_{i-1,i'}. \qquad (23)$$

If $A(j, j')$ denotes the matrix of time-ordered contractions, we find that

$$A(j, j') = \begin{bmatrix} 0 & 0 & x\delta_{i-1,i'} & 0 \\ 0 & 0 & 0 & y\delta_{i-m,i'} \\ -x\delta_{i+1,i'} & 0 & 0 & 0 \\ 0 & -y\delta_{i+m,i'} & 0 & 0 \end{bmatrix}$$
$$= -A^{\sim}(j', j). \qquad (24)$$

With these preliminaries we can now consider the evaluation of Z_1. The idea will be to represent this as a linked cluster expansion. We can write the expression arising from the expansion of the exponential in Z_1 as

$$Z_1 = \sum_{n=0}^{\infty} Z_1^{(n)}, \qquad (25)$$

where

$$Z_1^{(n)} = (2^n n!)^{-1}(\Omega, T[(\sum_i \sum_p \sum_q k_{pq}A^{(p)}(j \times A^{(q)}(j))^n]\Omega), \qquad (26)$$

and an application of Wick's theorem enables this to be written as

$$Z_1^{(n)} = (2^n n!)^{-1} \sum_{\text{contractions}} \sum_{j_1} \cdots \sum_{j_n} \times \sum_{p_1} \cdots \sum_{p_n} \sum_{q_1} \cdots \sum_{q_n} (-1)^P k_{p_1q_1} \cdots k_{p_nq_n} \times A^{(p)}(j)\cdot A^{(p')}(j')\cdot \cdots, \qquad (27)$$

where the symbol $\sum_{\text{contractions}}$ means that the sum is taken over all possible pairs of contractions of the operators. No terms containing normal products survive because of the conditions on the vacuum state given by Eq. (13) and its Hermitian conjugate. The factor $(-1)^P$ is the sign factor depending on the parity of the permutation of the fermion operators required to bring paired operators together. The factor 2^{-n} in (27) can be removed because for every j there are two operators $A^{(p)}(j)$ and $A^{(q)}(j)$ available to form contractions. In detail we have, corresponding to every contribution

$$\tfrac{1}{2}k_{pq}A^{(p)}(j)\cdot A^{(q')}(j')\cdot A^{(q)}(j)\cdot\cdot A^{(p')}(j')\cdot\cdot,$$

a further contribution

$$-\tfrac{1}{2}k_{pq}A^{(q)}(j)\cdot A^{(q')}(j')\cdot A^{(p)}(j)\cdot\cdot A^{(p')}(j')\cdot\cdot$$
$$= -\tfrac{1}{2}k_{pq}A^{(p)}(j)\cdot A^{(q')}(j')\cdot A^{(q)}(j)\cdot\cdot A^{(p')}(j')\cdot\cdot$$
$$= \tfrac{1}{2}k_{pq}A^{(p)}(j)\cdot A^{(q')}(j')\cdot A^{(q)}(j)\cdot\cdot A^{(p')}(j')\cdot\cdot,$$

where the minus sign in the first line arises from the interchange of order of the operators $A^{(p)}(j)$ and $A^{(q)}(j)$, and the compensating minus sign in the third line from the antisymmetry of the matrix K. Adding these terms compensates for the factor $\frac{1}{2}$, and this can be done for every j.

The expression $Z_1^{(n)}$ can be given an interpretation in terms of Feynman graphs in the usual way. The labels $j_1, \cdots, j_n$ are represented as n vertices of a graph and a contraction $A(j)\cdot A(j')\cdot$ by a line joining the vertices labeled j and j'. Following the remarks of the previous paragraph we can assign a direction to these lines such that the operator $A^{(p)}(j)$ is associated with an *outgoing* line and the operator $A^{(q)}(j)$ with an *ingoing* line. As there are only two operators for each j there will be only two lines incident on each vertex and the sense of the arrow specifying the direction of these two lines will be continuous through a vertex. This means that each product of pairs of contractions in $Z_1^{(n)}$ will be represented by a graph which consists of

a set of nonoverlapping closed loops, and the sum $\sum_{\text{contractions}}$ will be represented by the set of all graphs which contain loops of all sizes and numbers restricted only by the condition that the total number of vertices in each graph is n. The permutation P referred to in Eq. (27) can be effected in two steps. The first step is to rearrange the order so that all operators appearing in a single loop are brought together, taking care not to alter the relative order of operators within a single loop. This permutation must be an even permutation because each vertex will contribute two operators to a loop. The next step is to rearrange the operators in a loop so as to put them in the order corresponding to the associated graph. In order to see the effect of this second rearrangement, which effects all the loops independently of each other, consider a loop with l operators. This will contribute a factor

$$\sum_{j_1}\cdots\sum_{j_l}\sum_{p_1}\cdots\sum_{p_l}\sum_{q_1}\cdots\sum_{q_l}k_{p_1q_1}\cdots k_{p_lq_l}\times A^{(p_1)}(j_1)A^{(q_1)}(j_1)\cdots A^{(p_l)}(j_l))A^{(q_l)}(j_l),$$

for which a possible contraction will be with the indices in the order written:

$$\sum_{j_1}\cdots\sum_{j_l}\sum_{p_1}\cdots\sum_{p_l}\sum_{q_1}\cdots\sum_{q_l}k_{p_1q_1}\cdots k_{p_lq_l}\times A^{(p_1)}(j_1)\cdot A^{(p_2)}(j_2)\cdot\cdots A^{(p_l)}(j_l)\cdot\cdot A^{(q_l)}(j_l)\cdot\cdot.$$

This can be written in matrix notation as

$$\sum_{j_1}\cdots\sum_{j_l}\mathrm{Tr}\,(KA(j_1,j_2)KA(j_2,j_3)\cdots KA^{-}(j_1,j_l)) = -\sum_{j_1}\cdots\sum_{j_l}\mathrm{Tr}\,(KA(j_1,j_2)KA(j_2,j_3)\times\cdots KA(j_l,j_1)), \tag{28}$$

A reordering of the indices $j_1,\cdots,j_l$ in Eq. (28) will not change the sign of this expression because such a reordering will entail the shifting of a pair of operators at a time. Hence there will be $\frac{1}{2}(l-1)!$ equal contributions to the term (28). The factor $(l-1)!$ arises because all cyclic permutations are counted only once, as the loop can be regarded as starting from any vertex. The factor $\frac{1}{2}$ is included because a permutation which differs from another permutation by describing the same set of vertices in the opposite order can be produced by interchanging $A^{(p)}$ and $A^{(q)}$ at every vertex and such an interchange has already been allowed for.

If we denote the expression (28) by $-L^{(l)}$ the expression (27) can be written as

$$Z_1^{(n)} = (n!)^{-1}\sum_{\nu_1+2\nu_2+\cdots=n}C(\nu_1,\nu_2,\cdots)\times\left(-\frac{L^{(1)}}{2}\right)^{\nu_1}\left(-\frac{L^{(2)}}{2}\right)^{\nu_2}\cdots, \tag{29}$$

where $C(\nu_1,\nu_2\cdots)$ is a combinatorial factor, which is given by

$$C(\nu_1,\nu_2\cdots) = \left(\frac{n!}{1^{\nu_1}(2!)^{\nu_2}(3!)^{\nu_3}\cdots\nu_1!\,\nu_2!\cdots}\right)\times(\nu_1!\,\nu_2!\cdots)((2!)^{\nu_2}(3!)^{\nu_3}\cdots). \tag{30}$$

The first factor counts the number of ways of assigning the n vertices to the set of ν_1 loops with a single point, ν_2 loops with two points and so on. The second factor counts the number of ways of permuting the points belonging to loops of the same size among themselves, while the third factor counts permutations of points within a loop as already described. Collecting (25), (29), and (30) together, Z_1, can be written as

$$Z_1 = \exp(-\tfrac{1}{2}L), \tag{31}$$

where $L=\sum_{l=1}^{\infty}L^{(l)}$, and $L^{(l)}$ is given by Eq. (28).

By analogy with the methods used for evaluation associated with Feynman graphs, we make a Fourier transformation, using the identity

$$\delta_{ii'} = \frac{1}{N}\sum_{r=0}^{N-1}\omega^{r(i-i')} = \frac{1}{N}\sum_{r=0}^{N-1}\omega^{-r(i-i')}, \tag{32}$$

with $\omega=\exp(2\pi i/N)$. Then $A(j,j')$ can be written

$$A(j,j') = \frac{1}{N}\sum_{r=0}^{N-1}\omega^{r(i-i')}\times\begin{bmatrix}0&0&x\omega^{-r}&0\\0&0&0&y\omega^{-mr}\\-x\omega^{r}&0&0&0\\0&-y\omega^{mr}&0&0\end{bmatrix} = \frac{1}{N}\sum_{r=0}^{N-1}\omega^{r(i-i')}A(r,\omega). \tag{33}$$

Then $L^{(l)}$ becomes

$$\begin{aligned}L^{(l)} &= N^{-1}\sum_{j_1}\cdots\sum_{j_l}\sum_{r_1=0}^{N-1}\cdots\sum_{r_l=0}^{N-1}\omega^{r_1(j_1-j_2)+r_2(j_2-j_3)+\cdots+r_l(j_l-j_1)}\times\mathrm{Tr}\,[KA(r_1,\omega)KA(r_2,\omega)\cdots KA(r_l,\omega)]\\ &= \sum_{r_1=0}^{N-1}\cdots\sum_{r_l=0}^{N-1}\delta_{r_1r_2}\delta_{r_2r_3}\cdots\delta_{r_lr_1}\times\mathrm{Tr}\,(KA(r_1,\omega)\cdots KA(r_l,\omega))\\ &= N\sum_{r=0}^{N-1}\mathrm{Tr}\,[(KA(r,\omega))^{l}].\end{aligned} \tag{34}$$

Hence

$$N^{-1}\log Z_1 = \frac{1}{2}\sum_{r=0}^{N-1}\mathrm{Tr}\log[I-KA(r,\omega)], \tag{35}$$

where I is the unit 4×4 matrix. Now for any matrix D with determinant $|D|$,

$$\mathrm{Tr} \log D = \log |D|,$$

and, using $|K| = 1$, we have

$$N^{-1} \log Z_1 = \frac{1}{2} \sum_{r=0}^{N-1} \log |K^{-1} - A| \qquad (36)$$

where

$$K^{-1} - A = \begin{bmatrix} 0 & 1 & -1-x\omega^{-r} & 1 \\ -1 & 0 & 1 & -1-y\omega^{-mr} \\ 1+x\omega^{r} & -1 & 0 & 1 \\ -1 & 1+y\omega^{mr} & -1 & 0 \end{bmatrix}, \qquad (37)$$

which is the usual result. The advantage of this method is that it replaces the problem of summing over all closed polygons drawn on a lattice with the possible replications and overlappings that this entails by the much simpler problem of summing over single closed polygons without any special topological structure.

The results for the triangular lattice can be written down immediately if we put

$$K = \begin{bmatrix} 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}. \qquad (38)$$

$$A(j, j') = \begin{bmatrix} 0 & 0 & 0 & x\delta_{j-1,j'} & 0 & 0 \\ 0 & 0 & 0 & 0 & y\delta_{j-m,j'} & 0 \\ 0 & 0 & 0 & 0 & 0 & z\delta_{j-m+1,j'} \\ -x\delta_{j+1,j'} & 0 & 0 & 0 & 0 & 0 \\ 0 & -y\delta_{j+m,j'} & 0 & 0 & 0 & 0 \\ 0 & 0 & -z\delta_{j+m-1,j'} & 0 & 0 & 0 \end{bmatrix}$$

Summarizing, the rules for evaluating the partition function for any soluble Ising lattice are:

(1) Transform the operator polynomial for a lattice point into an exponential, and calculate the matrix K from the structure of the exponent.

(2) Calculate the matrices $A(j, j')$ and $A(r, \omega)$.

(3) Draw all possible closed loops with an arbitrary number of vertices, and connected with directed lines so that the sense of the arrow is preserved through a vertex and only two lines are incident on a vertex.

(4) With each vertex associate on index j and a matrix K, and with each directed line leading from j' to j associate a matrix $A(j, j')$.

(5) Form the matrix product in the order of the directed lines, take the trace of this product, and sum over all j from 1 to N. Finally sum over all contributions from all possible loops.

(6) Instead of labels j attached to vertices, a label r may be attached to lines and the label r must be the same along a line, i.e., "conserved" at each vertex. The label r is analogous to the momentum in field theory.

(7) Green's functions are obtained by allowing vertices to have only a single line terminating on them. If there are two such vertices, the graphs correspond to the two-point function, four vertices give the four-point function, and so on.

There is no need to assume that all vertices are of the same type, and so one may consider problems in which there are several sorts of vertices corresponding to lattices with different connections at different lattice points. Also one may allow for several bonds joining pairs of lattice points, so long as they run in parallel.

When crossed bonds are present, the additional quartic term in the exponent will lead to loops which are no longer simple but have self crossings. Each such self-crossing would be marked by a vertex point, and there will then appear many more topologically distinct such loops. It would be interesting to see whether these graphs can be classified and whether at least partial summations can be made, in order to form an initial approximation.

ACKNOWLEDGMENT

I would like to thank R. W. Gibberd for some useful discussions, particularly on the presentation of Sec. 3.

Numerical Solution of Non-Fredholm (Singular) Integral Equations by Matrix Inversion*

C. EDWARD JONES AND GEORGE TIKTOPOULOS
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 22 June 1965)

It is shown that the simple matrix-inversion techniques often used in numerically solving linear integral equations with a Fredholm–Schmidt (i.e., square-integrable) kernel can also be employed for a wide class of non-Fredholm ("singular") equations. This class includes equations the kernel of which is the sum of a Fredholm–Schmidt kernel and a kernel whose norm (in the operator sense) is less than one. In particular, the integral equations of the so-called "new strip approximation" in particle dynamics belong to this class.

I. INTRODUCTION

THE use of electronic computers has made possible the rapid and accurate numerical solution of linear integral equations of the Fredholm type. For the purposes of numerical solution, a given integral equation of the standard form

$$\psi(x) = \varphi(x) + \int_a^b dx' K(x, x')\psi(x') \qquad (1)$$

is usually replaced by a system of linear algebraic equations involving as unknown quantities the values of $\psi(x)$ at a finite number of values of x ("mesh" points). This is accomplished by approximating the integral by a "weighted" sum according to some rule of numerical integration. The linear system is then solved by the ordinary algebraic method which amounts essentially to the inversion of a finite-dimensional numerical matrix. The matrix inversion is carried out efficiently by electronic computers. The resulting approximate values of $\psi(x)$ at the "mesh" points can then be interpolated to produce an approximation to $\psi(x)$. The convergence of this procedure as the number of mesh points increases is easily established if $\varphi(x)$ and $K(x, x')$ satisfy (besides certain continuity properties) the conditions

$$\int_a^b |\varphi(x)|^2 \, dx < \infty, \qquad (2)$$

$$\int_a^b \int_a^b dx \, dx' \, |K(x, x')|^2 < \infty. \qquad (3)$$

In the language of functional analysis,[1] condition (2) means that $\varphi(x)$ belongs to the Hilbert space of $L^2(a, b)$ functions and condition (3) means that $K(x, x')$ is the kernel of a Schmidt operator on $L^2(a, b)$. An enormous literature exists on these Fredholm or Schmidt equations.[2] Quite often, however, in physical problems one encounters integral equations whose kernel $K(x, x')$ is not of the Schmidt type, although it represents a *bounded*[3] operator on $L^2(a, b)$. To this category belong certain types of "singular" integral equations such as integral equations of the Wiener–Hopf[4] type. We mention here an example which, in fact, motivated the present work. In the framework of the S-matrix approach to particle dynamics one encounters the following equation[5–12]:

$$N(x) = B(x) + \frac{1}{\pi}\int_{x_0}^{x_1} dx' \frac{B(x') - B(x)}{x' - x}\left[\frac{x' - 4}{x'}\right]^{\frac{1}{2}} N(x'), \qquad (4)$$

where $B(x)$ is a continuous smooth function having a logarithmic singularity at $x = x_1$ [i.e. $B(x) \sim \log(x_1 - x)$ near $x = x_1$]. This equation embodies the so-called "New Strip Approximation,"[6] the solutions to which are used to compute Regge trajectories in the relativistic scattering problem.

Due to the singularity of $B(x)$ at x_1, Eq. (4) is not of the Schmidt type and special methods have

* Work supported by the U. S. Air Force Office of Scientific Research and Development Command.

[1] R. Riesz and B. Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955).

[2] A concise account of the theory may be found, for example, in F. G. Tricomi, *Integral Equations* (Interscience Publishers, New York, London, 1957). See also F. Smithies, *Integral Equations* (Cambridge University Press, Cambridge, England, 1958).

[3] A *bounded* operator K is an operator of finite *norm*. The norm of K is defined as the sup $|K\phi|$ for $|\phi| = 1$. A practical method for obtaining upper bounds on the norm of integral operators is given in G. Tiktopoulos, J. Math. Phys. **6**, 573 (1965).

[4] See B. Noble, *The Wiener-Hopf Technique* (Pergamon Press, New York, 1958).

[5] G. F. Chew, Phys. Rev. **129**, 2363 (1963).

[6] G. F. Chew and C. E. Jones, Phys. Rev. **135**, B208 (1964).

[7] C. E. Jones, Phys. Rev. **135**, B214 (1964).

[8] V. L. Teplitz, Phys. Rev. **137**, B136 (1965).

[9] G. F. Chew and V. L. Teplitz, Phys. Rev. **137**, B139 (1965).

[10] D. C. Teplitz and V. L. Teplitz, Phys. Rev. **137**, B142 (1965).

[11] G. F. Chew, Phys. Rev. **130**, 1264 (1963).

[12] C. E. Jones, "N/D Equations with a Finite Strip," (to be published).

been used for its numerical solution. Chew[11] observed that the kernel of Eq. (4) can be written as the sum of a kernel $W(x, x')$ of the Wiener–Hopf type and a Schmidt kernel $C(x, x')$. He then showed that Eq. (4) can be solved by first obtaining the resolvent $R(x, x')$ of $W(x, x')$ by the Weiner–Hopf technique and then reducing Eq. (4) to a Fredholm problem with kernel $\int dx''R(x, x'')C(x'', x')$ and inhomogeneous term $\int dx'R(x, x')\varphi(x')$. A numerical solution is possible in this way. However, the labor and computer time involved is considerable.[8,10]

In this paper we show that *a wide class of non-Fredholm integral equations* [to which (4) belongs] can be solved numerically by simple matrix inversion. The actual numerical procedure does not differ from that used for the ordinary Fredholm equations except for some care needed in assigning values to the kernel at mesh points where it is discontinuous or singular.

Our class of equations can be briefly (but not completely) described by saying that $\varphi(x)$ is in $L^2(a, b)$ and that

$$K(x, x') = W(x, x') + C(x, x'),$$

where the norm[3] of $W(x, x')$ as an operator on $L^2(a, b)$ is less than one and $C(x, x')$ is a Schmidt kernel.[13] The relevant theorems are stated and proved in Sec. II. A discussion of the numerical procedures and examples are given in Sec. III.

All integrals occurring in our discussion are understood in the Lebesgue sense. Also only (bounded) *integral* operators are considered. These are defined by a function $K(x, x')$ on $(a, b) \times (a, b)$ (called the "kernel") such that for every $f(x)$ in $L^2(a, b)$ the integral $\int_a^b K(x, x')f(x')\,dx'$ exists almost everywhere and defines a function in $L^2(a, b)$.

II. THE THEOREMS

Theorem 1. The integral equation

$$\psi(x) = \varphi(x) + \int_a^b dx'K(x, x')\psi(x') \qquad (5)$$

is given where $\varphi(x)$ is in $L^2(a, b)$, i.e.,

$$\int_a^b |\varphi(x)|^2\,dx = |\varphi|^2 < \infty.$$

The integral operator K is defined on $L^2(a, b)$ by the kernel $K(x, x')$ and its resolvent $(1 - K)^{-1}$ exists. Consider a sequence of approximations φ_n and K_n $(n = 1, 2, \cdots)$ to φ and K such that

(i) $\lim_{n\to\infty} \varphi_n(x) = \varphi(x)$ a.e.,

$\lim_{n\to\infty} K_n(x, x') = K(x, x')$ a.e.,

where a.e. means almost everywhere;

(ii) for $n > n_0$, $|(I - K_n)^{-1}| < a$;

(iii) for $n > n_0$, $|\varphi_n(x)| < \bar{\varphi}(x)$, where $\bar{\varphi}$ is in L^2 $|K_n(x, x')| < \bar{K}(x, x')$, where $\bar{K}(x, x')$ is the kernel of a *bounded* operator in L^2.

Then the solution $\psi = (1 - K)^{-1}\varphi$ of (1) is approximated in the mean:

$$\lim_{n\to\infty} |\psi - (I - K_n)^{-1}\varphi_n| = 0.$$

Proof: We have

$$\begin{aligned}
&|(I - K)^{-1}\varphi - (I - K_n)^{-1}\varphi_n| \\
&= |(I - K_n)^{-1}(K - K_n)(I - K)^{-1}\varphi \\
&\qquad + (I - K_n)^{-1}(\varphi - \varphi_n)| \\
&\leq |(1 - K_n)^{-1}|\,\{|(K - K_n)(1 - K)^{-1}\varphi| + |\varphi - \varphi_n|\} \\
&\leq a\,|(K - K_n)(1 - K)^{-1}\varphi| + a\,|\varphi - \varphi_n|
\end{aligned}$$

and

$$\lim_{n\to\infty} |\varphi - \varphi_n| = \lim_{n\to\infty} \left\{\int_a^b |\varphi(x) - \varphi_n(x)|^2\,dx\right\} = 0 \qquad (7)$$

because (i) and (iii) ensure the validity of Lebesgue's "dominated convergence" theorem.[14]

The existence of $(1 - K)^{-1}$ implies that $v = (1 - K)^{-1}\varphi$ exists as a square-integrable function. Because of (i) and (iii), we can apply twice the Lebesgue theorem to show that

$$\lim_{n\to\infty} \int_a^b K_n(x, x')v(x')\,dx' = \int_a^b K(x, x')v(x')\,dx'$$

and that

$$\lim_{n\to\infty} \int_a^b dx \left|\int_a^b dx'[K_n(x, x') - K(x, x')]v(x')\right|^2 = 0.$$

[13] It should be made clear that the numerical procedure will approximate the solution of Eq. (1) which is unique in $L^2(a, b)$. In general, equations of our class may have additional solutions outside the $L^2(a, b)$ space. This happens frequently, for instance, in Wiener–Hopf problems. (See, in this regard, Ref. 12.)

[14] Lebesgue's "dominated convergence" theorem, one of the most important results of the integral calculus, can be stated as follows: "Let $f_1(x), f_2(x), \cdots$ be a sequence of functions which are (Lebesgue) integrable over some measurable set E of the real line. If $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere in E, and $|f_n(x)| \leq g(x)$, where $g(x)$ is a fixed (i.e., independent of n) function integrable over E, then $f(x)$ is integrable over E and

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx.$$

See, for example, M. E. Munroe, *Measure and Integration* (Addison-Wesley Publishing Company, Inc., Cambridge, Massachusetts, 1953).

This means that

$$\lim_{n\to\infty} |(K - K_n)(I - K)^{-1}\varphi| = 0 ,$$

and from (6), (7), and (8) we deduce $|(I - K)^{-1}\varphi - (I - K_n)^{-1}\varphi_n| \to 0$ Q.E.D.

In practice, condition (ii) is the most difficult to verify. If K is a Schmidt operator, namely if $\int dxdx' \, |K(x, x')|^2 < \infty$, then (i) and (iii) imply $|K - K_n| \to 0$ (assuming $\bar{K}$ is also a Schmidt operator) so that condition (ii) is satisfied automatically because

$$|(I - K_n)^{-1}| \leq \frac{|(I - K)^{-1}|}{1 - |(I - K)^{-1}| \cdot |K - K_n|}$$

$$\text{provided} \quad |K - K_n| < |(I - K)^{-1}|^{-1}.$$

Another case in which condition (ii) can be established rather easily is when $K(x, x') \geq 0$ and represents an operator of norm $|K|$ less than one. We can then simply consider approximations in which

$$|K_n(x, x')| < [(a - 1)/a] \cdot K(x, x')/|K|,$$

where a is a fixed positive number greater than one.

The most important case, however, for applications is $K = W + C$ where $|W| < 1$ and C is a Schmidt kernel. In this case, condition (ii) can be replaced by more manageable ones and the theorem can be reformulated as follows.

Theorem 2. The integral equation

$$\psi(x) = \varphi(x) + \int_a^b dx' K(x, x')\psi(x')$$

is given where $\varphi(x)$ is in $L^2(a, b)$, i.e., $\int_a^b dx \, |\varphi(x)|^2 < \infty$ and the resolvent $(1 - K)^{-1}$ of the integral operator K exists. Let $K = W + C$ where $|W| < 1$ and C is a Schmidt operator (i.e., $\int |C(x, x')|^2 \, dxdx' < \infty$).

Consider a sequence of approximations φ_n, W_n, and C_n $(n = 1, 2 \cdots)$ to φ, W, and C such that

(i) $\lim_{n\to\infty} \varphi_n(x) = \varphi(x)$ a.e.,

$\lim_{n\to\infty} W_n(x, x') = W(x, x')$ a.e.,

$\lim_{n\to\infty} C_n(x, x') = C(x, x')$, a.e.;

(ii) $|\varphi_n(x)| < |\bar{\varphi}(x)|$

where $\int_a^b |\bar{\varphi}(x)|^2 \, dx < \infty$,

$|C_n(x, x')| < |\bar{C}(x, x')|$

where $\int_a^b \int_a^b dx \, dx' \, |\bar{C}(x, x')|^2 < \infty$,

$|W_n(x, x')| < |\bar{W}(x, x')|$ where $\bar{W}(x, x')$ is the kernel of a *bounded* operator, and

$|W_n| < 1 - \epsilon$ where ϵ is a fixed positive number.

Then the solution $\psi = (1 - K)^{-1}\varphi$ of the integral equation is approximated in the mean:

$$\lim_{n\to\infty} |\psi - (I - K_n)^{-1}\varphi_n| = 0.$$

Note: Conditions (ii) for W_n can be ensured if, for example, $W(x, x') \geq 0$ and

$$|W_n(x, x') < (1 - \epsilon)W(x, x')/|W|.$$

Proof. Since $|W| < 1$, we may write

$$\psi = [I - (1 - W)^{-1}C]^{-1}(1 - W)^{-1}\varphi.$$

We note that

$$\lim_{n\to\infty} |(1 - W)^{-1}\varphi - (1 - W_n)^{-1}\varphi_n| = 0$$

because W, φ and their approximations satisfy the conditions of Theorem 1. Thus we only have to show that

$$|(I - W)^{-1}C - (I - W_n)^{-1}C_n| \to 0.$$

We write

$$\begin{aligned} &|(I - W)^{-1}C - (I - W_n)^{-1}C_n| \\ &= |(I - W_n)^{-1}(W - W_n)(I - W)^{-1}C \\ &\qquad + (I - W_n)^{-1}(C - C_n)| \\ &\leq (1/\epsilon) \, |(W - W_n)(I - W)^{-1}C| + (1/\epsilon) \, |C - C_n|. \end{aligned}$$

Now under the assumed properties for C and C_n the Lebesgue "dominated convergence" theorem[14] implies that $|C - C_n| \to 0$. In order to treat the $|(W - W_n)(1 - W)^{-1}C|$ term we introduce

$$B_n(x, x') = \int_a^b dy[W(x, y) - W_n(x, y)]A(y, x'),$$

where $A(y, x')$ is the kernel of $(1 - W)^{-1}C$. Since $|W(x, y) - W_n(x, y)|$ is bounded by $|W(x, y)| + |\bar{W}(x, y)|$, the Lebesgue theorem[14] can be used to show first that $B_n(x, x') \to 0$, a.e., and then that $\int_a^b dxdx' \, |B_n(x, x')|^2 \to 0$. Here we have used the fact that B_n is a Schmidt operator because it is the product of a bounded operator and a Schmidt operator. From $(1 - W_n)^{-1}\varphi_n \to (1 - W)^{-1}\varphi$ and $(1 - W_n)^{-1}C_n \to (1 - W)^{-1}C$ it easily follows that

$$\begin{aligned} &[I - (1 - W_n)^{-1}C_n]^{-1}(1 - W_n)^{-1}\varphi_n \\ &\qquad \to [I - (1 - W)^{-1}C]^{-1}(1 - W)\varphi \end{aligned}$$

or $(I - K_n)^{-1}\varphi_n \to \psi$ in the mean.

III. NUMERICAL APPROXIMATIONS

In numerical applications, one replaces the given equation

$$\psi(x) = \varphi(x) + \int_a^b dx' K(x, x')\psi(x')$$

by a system of linear equations

$$\tilde{\psi}(x_i) = \varphi(x_i) + \sum_j C_j K(x_i, x_j)\tilde{\psi}(x_j); \qquad i, j = 0, 1, 2, \cdots, n,$$

where the values of φ and K at only a finite number of "mesh" points are used. The "weighted" sum has replaced the integral according to some convenient rule of numerical integration, e.g., Simpson's rule. This presupposes that $\varphi(x)$ and $K(x, x')$ have certain continuity properties as functions of x and x'. They may be continuous with a finite number of discontinuities: for example $\phi(x) = \log x$ is discontinuous at $x = 0$; $\log (x'/x)\cdot(x' - x)^{-1}$ is discontinuous if either x or $x' = 0$. If one or more of the points x_i of subdivision of the basic interval (a, b) happens to lie at points of discontinuity of φ or K, then $\varphi(x_i)$ or $K(x_i, x_j)$ can be assigned an arbitrary value (e.g., zero) at these points.

In order to apply our results the sums must first be interpreted as integrals over appropriately defined "step" functions. For concreteness, let us consider a specific simple recipe for numerical integration: To evaluate $\int_a^b g(x)\,dx$ divide the interval into n equal parts by the division points

$$a, \quad a + \frac{b-a}{n}, \quad a + 2\frac{b-a}{n}, \cdots, \quad a + (n-1)\frac{b-a}{n}, \quad b.$$

An approximation to the integral is then provided by the sum

$$\sum_{m=1}^{n} \frac{b-a}{n}\, g\left(a + \left(m - \frac{1}{2}\right)\frac{b-a}{n}\right).$$

The use of this rule for numerical integration in our integral equation amounts to considering the following "approximate" integral equation

$$\tilde{\psi}_n(x) = \varphi_n(x) + \int_a^b dx' K_n(x, x')\tilde{\psi}_n(x'),$$

where $\varphi_n(x)$ and $K_n(x, x')$ are step functions defined as follows:

$$\varphi_n(x) = \varphi\left[x = a + \frac{m - \frac{1}{2}}{n}(b-a)\right]$$

$$\text{for} \quad a + \frac{m-1}{n}(b-a) \le x < a + \frac{m}{n}(b-a),$$

$$K_n(x, x') = K\left[a + \frac{m - \frac{1}{2}}{n}(b-a),\ a + \frac{m' - \frac{1}{2}}{n}(b-a)\right],$$

$$\text{for} \quad a + \frac{m-1}{n}(b-a) \le x < a + \frac{m}{n}(b-a)$$

$$\text{and} \quad a + \frac{m'-1}{n}(b-a) \le x' < a + \frac{m'}{n}(b-a).$$

Such step functions are capable of approximating $\varphi(x)$ and $K(x, x')$ even if these latter have a finite number of discontinuities (where they may become infinite). As $n \to \infty$ we have $\varphi_n(x) \to \varphi(x)$ and $K_n(x, x') \to K(x, x')$ except at these points where $\varphi(x)$ and $K(x, x')$ are discontinuous. Thus convergence almost everywhere (as required in the theorems) is ensured.

At this point we would like to indicate by an example how one can choose $W_n(x, x')$ to fulfill the condition described in the note following Theorem 2. We consider the case

$$W(x, x') = \frac{\lambda}{\pi^2}\cdot\left(\log\frac{x'}{x}\right)(x' - x)^{-1}; \qquad a = 0, \quad b = 1, \quad \lambda > 0. \tag{9}$$

It can be shown that the norm of this operator[8] is $|W| = \lambda$. If we now use the described step functions for $W_n(x, x')$, the condition

$$|W_n(x, x')| < (1 - \epsilon)W(x, x')/|W| \tag{10}$$

applied at $x = x' = 1/n$ demands $2\lambda < (1 - \epsilon)$ which excludes values of λ in the interval $\frac{1}{2} < \lambda = |W| < 1$. However, the situation can be easily remedied by a slight modification in $W_n(x, x')$. We take $W_n(x, x')$ to be zero whenever x or $x' \le l/n$ where l is a fixed integer depending on λ (for instance if $\lambda < \frac{3}{4}$, it suffices to take $l = 1$).

Better results are obtained in practice through more sophisticated integration procedures like for instance Simpson's rule, which in our case amounts to replacing the given integral equation by the following set of linear equations:

$$\tilde{\psi}(x_i) = \varphi(x_i) + \sum_{j=0}^{n-1} h\{\tfrac{1}{3}K(x_i, x_{2j})\tilde{\psi}(x_{2j}) + \tfrac{4}{3}K(x_i, x_{2j+1})\tilde{\psi}(x_{2j+1}) + \tfrac{1}{3}K(x_i, x_{2j+2})\tilde{\psi}(x_{2j+2})\}, \tag{11}$$

where

$$x_m = a + mh, \qquad h = (b-a)/2n, \qquad m = 0, 1, 2, \cdots, 2n.$$

These approximate equations are *equivalent* to the integral equation

$$\bar{\psi}_n(x) = \varphi_n(x) + \int_a^b K_n(x, x')\bar{\psi}_n(x')\,dx',$$

where $\varphi_n(x)$ and $K_n(x, x')$ are step functions defined as follows:

$$\begin{aligned}
\varphi_n(x) &= \varphi(x_{2i}) \quad \text{for} \quad |x - x_{2i}| < \tfrac{1}{3}h,\\
&= \varphi(x_{2i+1}) \qquad\quad |x - x_{2i+1}| < \tfrac{2}{3}h,\\
K_n(x, x') &= K(x_{2i}, x_{2k}) \qquad |x - x_{2i}| < \tfrac{1}{3}h, \qquad |x' - x_{2k}| < \tfrac{1}{3}h,\\
&= K(x_{2i}, x_{2k+1}) \qquad |x - x_{2i}| < \tfrac{1}{3}h, \qquad |x' - x_{2k+1}| < \tfrac{2}{3}h,\\
&= K(x_{2i+1}, x_{2k}) \qquad |x - x_{2i+1}| < \tfrac{2}{3}h, \qquad |x' - x_{2k}| < \tfrac{1}{3}h,\\
&= K(x_{2i+1}, x_{2k+1}) \qquad |x - x_{2i+1}| < \tfrac{2}{3}h, \qquad |x' - x_{2k+1}| < \tfrac{2}{3}h.
\end{aligned}$$

These definitions are modified at points of discontinuity as mentioned above. As an example we take $K(x, x') = W(x, x')$ as given in Eq. (9). Since the kernel becomes infinite at $x = 0$ or $x' = 0$, we set $K_n(x, x') = 0$ for $x < \frac{1}{3}h$ or $x' < \frac{1}{3}h$. It may also be verified that with this modification condition (10) is satisfied for $\lambda < \frac{3}{5}$. [For larger values of $\lambda < 1$, one would have to modify $K_n(x, x')$ from the strict Simpson rule values for $x < \frac{1}{3}lh$ and $x' < \frac{1}{3}lh$ where l is some fixed integer depending on λ.] The theorems presented in this paper show that the step function $\bar{\psi}_n(x)$ defined as

$$\begin{aligned}
\bar{\psi}_n(x) &= \bar{\psi}(x_{2i}), \qquad |x - x_{2i}| < \tfrac{1}{3}h\\
&= \bar{\psi}(x_{2i+1}), \qquad |x - x_{2i+1}| < \tfrac{2}{3}h
\end{aligned} \tag{12}$$

converges in the mean to $\psi(x) = (I - K)^{-1}\varphi(x)$ as $n \to \infty$.

In practice one deals only with the set of linear equations (11). [Care need be taken only for the values of φ and K at points of discontinuity: in our example, we take $K(0, x) = K(x, 0) = 0$.] One solves this system by *numerical matrix inversion* to obtain a set of $2n + 1$ numbers for $\bar{\psi}(x_0)$, $\bar{\psi}(x_1)$, $\cdots$, $\bar{\psi}(x_{2n})$ which represent a step function $\bar{\psi}_n(x)$ according to Eqs. (12). *By increasing the number of "mesh" points, $\bar{\psi}_n(x)$ thus obtained by matrix inversion converges in the mean to the exact solution of the given integral equation.*

In order to test this procedure we applied it to the equation[15]

$$\psi(x) = 1 + \frac{\lambda}{\pi^2}\int_0^1 \frac{\log(x'/x)}{x' - x}\,\psi(x')\,dx' \tag{13}$$

for $\lambda = \frac{1}{2}$. We used the Simpson rule with 32 and 64 mesh points and compared the results of the matrix inversion with the exact solution obtained by the Wiener–Hopf method. Table I gives a sample of the numerical results.

TABLE I. Comparison of the matrix inversion solution to Eq. (13) for mesh sizes $2n = 32$ and $2n = 64$ with the analytical solution.

	$\psi(x)$		
x	$2n = 32$	$2n = 64$	exact
0.0625	0.024750	0.025160	0.025574
0.1875	0.016357	0.016433	0.016513
0.3125	0.013981	0.014015	0.014052
0.4375	0.012834	0.012855	0.012877
0.5625	0.012161	0.012175	0.012191
0.6875	0.011722	0.011732	0.011743
0.8125	0.011414	0.011421	0.011430
0.9375	0.011188	0.011194	0.011200

ACKNOWLEDGMENT

We gladly acknowledge a useful conversation with S. M. Roy.

[15] A change of variables $x' = y'^2$ was employed for convenience in the numerical work.

Metrical Lattice and the Problem of Electricity

C. LANCZOS
Dublin Institute for Advanced Studies, Dublin, Ireland and Scientific Laboratory, Ford Motor Company, Dearborn, Michigan
(Received 4 June 1965)

A reinterpretation of Einstein's tetrad geometry leads to new results if we abandon the oversimplified picture of the vacuum as an almost empty Minkowskian manifold. The basic tetrad is identified with the four principal axes of the matter tensor which belongs to a strongly curved fourfold periodic Riemannian world. The macroscopic perturbation of the metrical lattice is investigated, corresponding to a mere rotation of the principal axes and assuming the quadratic action principle of general relativity. The perturbation Lagrangian yields the scalar $E^2 - H^2$, (and thus the Maxwellian equations), although the basic manifold is strictly Riemannian, with a positive-definite line element.

1. INTRODUCTION

EVER since the successful completion of general relativity, Einstein was looking for an all-comprehensive geometrical principle which would include electromagnetism and gravitation in a unified world picture.[1] Since the equivalence principle has shown that all forms of energy must influence the geometry of the universe, it appeared imperative to look for a basically geometrical interpretation of all physical action. However, the difficulty existed that Riemannian geometry seemed to be void of antisymmetric elements. This suggested that we must generalize the basic geometrical structure, although none of the attempted generalizations could compete in simplicity and naturalness with the original Gauss–Riemannian concepts.

The postulate of "cosmic wisdom," which asserts the admissibility of a speculative (against the purely empirical) approach to the fundamental problems of the physical universe,[2] dominated Einstein's endeavors during the last thirty years of his life. Although contemporary physics denies the possibility of such a program, Einstein's thought constructions never lost their inherent magic. A revision of his work may reveal the point, at which we have to depart from his assumptions, in order to make further progress in this field. In the author's opinion the fundamental departure must occur in the theoretical evaluation of the vacuum. The wave-mechanical phenomena of vacuum polarization and zero-point energy clearly indicate that the vacuum cannot be considered as something almost empty, namely a small deviation from the flat Minkowskian universe $g_{ik} = \eta_{ik}(= 0$ if $i \neq k$ and $-1, -1, -1, +1$, if $i = k)$ but as something strongly agitated. Is such a possibility reconcilable with the apparent validity of the Lorentz transformations? The answer is yes, if we admit the possibility of a metrical substructure of crystalline (fourfold periodic) character, with a lattice constant of submicroscopic smallness; ($\mu = 10^{-32}$cm, obtained by putting the fundamental constants c, $\hbar$, and $8\pi\kappa$ equal to 1).[3] The objection that such a lattice would establish a preferential frame of reference is macroscopically invalid, as the example of crystals of cubic symmetry demonstrates.[4] Such crystals are in all macroscopic relations entirely isotropic, although they have three well-defined mutually perpendicular axes. These axes are macroscopically equivalent, with the result that their privileged position is macroscopically unobservable. Something similar may hold in relation to the privileged axes of the metrical lattice.

It was in Einstein's theory of "distant parallelism" that four mutually perpendicular axes came in evidence. Einstein postulated these axes, in order to enrich Riemannian geometry by new elements, namely the notion of "distant parallelism," which exists in Euclidean but ordinarily not in Riemannian geometry.[5] For our present purposes the "local tetrads" introduced by Einstein are eminently useful, although not in the Einsteinian context. They appear as the principal axes of the matter tensor; (Ricci's "principal directions"[6]). Hence they are entirely within the scope of Riemannian geometry. It so happens that Einstein's tetrad geometry is singularly well suited to the discussion of the quadratic action principle in its relation to the problem of electromagnetism.

[1] A. Einstein, Sitzber. Preuss. Akad Wiss. **1925**, 414 (1925).
[2] A. Einstein, *Festschrift A. Stodola* (Füssli, Zürich, 1929), p. 126–132.
[3] C. Lanczos, J. Math. Phys. **4**, 951 (1963); Phys. Rev. **134** B476 (1964).
[4] Cf. M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Inc., New York, 1959), p. 675.
[5] A. Einstein, Math. Ann. **102**, 685 (1930).
[6] Cf. L. P. Eisenhart, *Riemannian geometry* (Princeton University Press, Princeton, New Jersey, 1925), p. 114.

While previously the author focused his attention on the scalar wave equation of a positive-definite Riemannian space and endeavored to show that the existence of high metrical ridges make signal propagation possible by acting as wave guides, the present paper pursues a different line. The macroscopic perturbation of the basic lattice is investigated, which must come in evidence as a small deformation of the basic tetrads. Such a deformation can be of a twofold type; it can be "elastic," influencing the metric of the manifold, or purely "rotational," without metrical change. The latter type of deformation demands an antisymmetric tensor for its description, which can be correlated to the electromagnetic field strength. It is this type of deformation which is investigated in the present paper, on the basis of an action principle which is quadratic in the curvature components and thus in harmony with the demand of gauge invariance.[7] The result of this investigation is that we obtain the Lagrangian $E^2 - H^2$ of the electromagnetic field without any artifices, on the basis of the general properties of a (genuinely Riemannian) metrical field whose field equations are governed by a quadratic action principle.

2. EINSTEIN'S TETRAD GEOMETRY

In his celebrated investigation of curved surfaces, C. F. Gauss introduced two differential forms for the intrinsic and extrinsic characterization of the geometry of the surface. The coefficients of these two forms were not independent of each other, but interconnected by a number of differential relations, the "Gauss–Codazzi equations".[8] A similar situation is encountered in the application of Riemannian geometry to the problems of physics. The first fundamental form is once more the line element of Gauss–Riemann:

$$ds^2 = g_{ik}\, dx^i\, dx^k, \tag{2.1}$$

but it was Einstein's great discovery that the physical manifestation of matter, expressed in the form of the energy–momentum tensor T_{ik}, which gives rise to the differential form

$$dw^2 = T_{ik}\, dx^i\, dx^k \tag{2.2}$$

must be equated to a purely metrical quantity, obtained with the help of the contracted curvature tensor R_{ik}, in the sense of the equation

$$T_{ik} = R_{ik} - \tfrac{1}{2}Rg_{ik}. \tag{2.3}$$

It is true that if the g_{ik} are given as functions of the coordinates, the entire geometry of the manifold is already determined. The matter tensor is obtainable with the help of the first and second derivatives of the g_{ik}. Yet the matter tensor has a significance of its own and is in a similar relation to the metric as the electric current is to the vector potential. Given the vector potential φ_i, the current vector ρ_i is obtainable by applying the wave operator on φ_i. But in physical problems the current vector is the primary source of the field, from which the vector potential is obtainable with the help of an integral operation. Similarly the matter tensor can be considered as the primary source of the field, which is physically of greater importance than the metrical field, in spite of the geometrical primacy of the latter.

Einstein's tetrad geometry provides us with a mathematical tool for the simultaneous representation of *both* g_{ik} and T_{ik}. We consider the algebraic problem of finding the four principal axes of the tensor T_{ik}. This problem is meaningful in a genuine Riemannian geometry (of positive-definite signature), while in a Minkowskian world the eigenvalue problem will generally not possess real solutions, although it is entirely possible that in a given physical situation—such as Maxwell's phenomenological matter tensor—a real Lorentz transformation can be found, by which the matter tensor can be diagonalized. In that case both eigenvalues and eigenvectors become real. We assume the positive-definiteness of the lattice geometry on the basis that a genuine Riemannian geometry is the most natural generalization of our Euclidean concepts, while an indefinite metric violates the minimum property of the distance and does not satisfy the conditions which we could reasonably demand of a rational metric.[9] (The strongly nonlinear nature of the field equations does not exclude the possibility that the macroscopic geometry of the physically observable phenomena may appear Minkowskian, as we will see later.)

We now obtain four real eigenvalues and the corresponding eigenvectors, which we assume as uniquely determined, due to the distinctness of the four eigenvalues. Of course, the gravitational equations $R_{ik} = 0$ would make the eigenvalue problem meaningless, but the existence of a highly agitated metrical lattice means that the matter tensor is strong at all points of the universe and thus the principal axes at all points well defined.

[7] Cf. C. Lanczos, Rev. Mod. Phys. 29, 337 (1957).

[8] Cf., e.g., L. P. Eisenhart, *Introduction to Differential Geometry*, (Princeton University Press, Princeton, New Jersey, 1947), p. 219.

[9] Cf. The second paper quoted in Ref. 3, Eqs. (1) and (2).

With a slight modification we can replace the matter tensor T_{ik} by the somewhat simpler contracted curvature tensor R_{ik}:

$$R_{ik} = T_{ik} - \tfrac{1}{2}Tg_{ik}. \tag{2.4}$$

It is clear that T_{ik} has the same principal axes as R_{ik}, while the eigenvalues are shifted by a mere constant. Hence we will consider our eigenvalue problem in the form

$$R_{ik}h^k = \sigma g_{ik}h^k \tag{2.5}$$

to be solved at a given point P of the manifold. We assume that the four (real) eigenvalues σ_i are distinct:

$$\sigma = \sigma_1, \sigma_2, \sigma_3, \sigma_4 \tag{2.6}$$

and denote the corresponding four solutions by

$$h^i = h^{i1}, h^{i2}, h^{i3}, h^{i4}. \tag{2.7}$$

Hence in the notation h^{ia} the first superscript i represents a genuine contravariant index, while the second superscript a, the "list index," refers to a mere enumeration of our four vectors. The pulling up and down of covariant and contravariant indices occurs in the usual fashion, with the understanding that this operation is meaningful only with respect to the *first* index:

$$h_{ia} = g_{ik}h^{ka}. \tag{2.8}$$

(For the sake of convenience we write the list index next to the component index, irrespective of what type of component is meant.)

In matrix notation our eigenvalue problem can be written in the form

$$RH = GH\Lambda \tag{2.9}$$

where the matrices R and G are symmetric, while Λ is diagonal (with the diagonal elements $\sigma_1, \cdots, \sigma_4$). According to the rules of matrix algebra we obtain the orthogonality relation

$$\tilde{H}GH = I \tag{2.10}$$

which leads to

$$G = \tilde{H}^{-1}H^{-1} \tag{2.11}$$

and

$$R = \tilde{H}^{-1}\Lambda H^{-1}. \tag{2.12}$$

Written out in components, the equations (2.10) and (2.11) contain the fundamental operational rules to which the four vectors h^{ia} (which form the matrix H) and the four vectors h_{ia} (which form the matrix $\tilde{H}^{-1}$) are subjected:

$$g_{ik}h^{ia}h^{kb} = \delta_a^b, \tag{2.13}$$

$$h^{ib}h_{ia} = \delta_a^b, \tag{2.14}$$

$$h^{ia}h_{ka} = \delta_k^i, \tag{2.15}$$

$$g_{ik} = h_{ia}h_{ka}. \tag{2.16}$$

All these relations are contained in Einstein's papers on distant parallelism. To this has to be added another fundamental relation which is not contained in Einstein's work, since he postulated the h_{ia} quantities without any relation to an eigenvalue problem. Equation (2.12) yields

$$R_{ik} = \sigma_a h_{ia}h_{ka}. \tag{2.17}$$

This formula shows the peculiarity that the summation index occurs in *three*, instead of *two*, factors. In order to avoid misunderstandings, we will discard the eigenvalue σ_a in the counting of indices. Accordingly in a term of the form $\sigma_a h_{ia}$ the index a shall *not* refer to summation, but denote a single term, at variance with a term of the form $h_{ia}h_{ka}$ where a is a genuine sum index.

The 20 quantities thus obtained ($4\sigma_a$ and $16h_{ia}$) are characteristic for that particular metrical field and are uniquely determined, except for an arbitrary transformation of the coordinates, which will modify the four vectors h_{ia} in the sense of the transformation

$$h'_{ia} = (\partial f^\mu/\partial x'_i)h_{\mu a}, \tag{2.18}$$

where the four functions $f^\mu(x_i)$ are defined by the coordinate transformation

$$x_i = f^i(x'_1, \cdots, x'_4). \tag{2.19}$$

The eigenvalues σ_i remain unaffected by this transformation. (We assume, of course, that the metrical field g_{ik} is twice differentiable and that the determinant g does not vanish anywhere in the domain.)

3. THE QUADRATIC ACTION PRINCIPLE

The operation with the four vectors h_{ia} has some definite advantages compared with the g_{ik}, as it was pointed out by Einstein. While the volume element of a Riemannian manifold demands that we shall take the square root of the determinant of the g_{ik}:

$$d\tau = g^{\frac{1}{2}}\, dx^1 \cdots dx^4, \tag{3.1}$$

the corresponding quantity now becomes

$$d\tau = h\, dx^1 \cdots dx^4, \tag{3.2}$$

where h is the determinant of the matrix h_{ia}. Furthermore, the relation

$$g_{ik} = h_{ia}h_{ka} \tag{3.3}$$

can be interpreted as the operation of taking the square root of the metrical tensor. Considering the relation of Dirac's equation to the wave equation, this can easily be of fundamental significance. And in fact, a possible connection between Einstein's distant parallelism and Dirac's equation was suggested by Wigner[10] and by Weyl,[11] soon after Einstein's first publications in this field.

In our present investigation the h_{ia} will play a dominant role, not on account of Einstein's motivation, but because they seem to provide an exceptionally adequate tool for the discussion of the action principle of general relativity. Einstein's Lagrangian, which leads to the gravitational equations, makes use of the scalar curvature R as the basic invariant. This now appears in the form

$$L = (\sigma_1 + \cdots + \sigma_4)h = \sigma h \tag{3.4}$$

if we denote

$$\sigma = \sigma_1 + \cdots + \sigma_4. \tag{3.5}$$

The much more complicated Lagrangian of the quadratic action principle is reducible to the two invariants[12] $R_{ik}R^{ik}$ and R^2 and can now be written in the simple form

$$L = \tfrac{1}{2}[(\sigma_1^2 + \cdots \sigma_4^2) - C\sigma^2]h$$

where C is an *a priori* undetermined constant. In both cases we have to add the auxiliary condition

$$\sigma_a h_{ia}h_{ka} - R_{ik} = 0, \tag{3.7}$$

where the contracted curvature tensor R_{ik} is defined in the usual way by the differential operator[13]

$$R_{ik} \equiv \frac{\partial^2 \log (g^{\frac{1}{2}})}{\partial x_i\, \partial x_k} - \frac{\partial g^{\frac{1}{2}}\, \Gamma^m_{ik}}{g^{\frac{1}{2}}\, \partial x_m} + \Gamma^m_{in}\Gamma^n_{km} \tag{3.8}$$

and g_{ik} is to be replaced by (3.3). However, instead of this substitution we will consider the g_{ik} as added action variables, making use of the Lagrangian multiplier method. Our final Lagrangian thus becomes

$$L' = [\tfrac{1}{2}(\sigma_i^2 - C\sigma^2) - p^{ik}(\sigma_a h_{ia}h_{ka} - R_{ik}) - w^{ik}(h_{ia}h_{ka} - g_{ik})]h \tag{3.9}$$

(the tensors p^{ik} and w^{ik} are symmetric). The action variables are the $16h_{ik}$, the $4\sigma_i$, the $10g_{ik}$, the $10p^{ik}$ and the $10w^{ik}$, altogether 50 quantities. The variation with respect to p^{ik} and w^{ik} yields, of course, the auxiliary conditions (3.7) and (3.3). The variation with respect to g_{ik} determines the Lagrangian factor w^{ik}, while the variation with respect to the h_{ia} and σ_a yields 20 equations, 10 of which determine the Lagrangian factor p^{ik}, while the other 10 characterize that particular Riemannian geometry, which obeys the quadratic action principle.

Let us observe that L' is purely *algebraic*, except for R_{ik} which is a differential operator of second order. But the fortunate circumstance holds that the second derivatives appear only *linearly* and thus by integrating by parts we can change L' to a first-order operator. The term involving R_{ik} can in this case be written in the following form:

$$\left[\frac{\partial p^{ik}}{\partial x_m}\Gamma^m_{ik} - \frac{\partial p^{ik}}{\partial x_k}\Gamma^m_{im} + p^{ik}(\Gamma^n_{im}\Gamma^m_{kn} - \Gamma^m_{im}\Gamma^n_{kn})\right]h. \tag{3.10}$$

We will also need the adjoint operator of R_{ik}, obtained by varying the g_{ik} and integrating by parts. Denoting covariant differentiation by ; we obtain

$$\delta(R_{ik}p^{ik}h) = B(p^{ik})\,\delta g_{ik}h = h\delta g_{ik}\tfrac{1}{2}[p^{ik}g^{mn} + p^{mn}g^{ik} - p^{in}g^{mk} - p^{kn}g^{mi}]_{;mn} \tag{3.11}$$

and thus

$$B(p^{ik}) = \tfrac{1}{2}[p^{ik}g^{mn} + p^{mn}g^{ik} - p^{in}g^{mk} - p^{kn}g^{mi}]_{;mn}. \tag{3.12}$$

Variation with respect to g_{ik} yields

$$w^{ik} = -B(p^{ik}). \tag{3.13}$$

Variation with respect to σ_a yields

$$\sigma_a - C\sigma = p^{ik}h_{ia}h_{ka} \tag{3.14}$$

(*no* summation over a), while variation with respect to h_{ia} yields

$$\tfrac{1}{2}(\sigma_\mu^2 - C\sigma^2)h^{ia} - 2\sigma_a p^{ik}h_{ka} - 2w^{ik}h_{ka} = 0. \tag{3.15}$$

If the last equation is multiplied by h^{ma} (summing over a) and we make use of the operational rules displayed in Sec. 2, we obtain the relation

$$\tfrac{1}{4}(\sigma_\mu^2 - C\sigma^2)g^{im} - p^{ik}R_k{}^m - w^{im} = 0. \tag{3.16}$$

This equation shows the peculiarity that the first and the third terms are symmetric in i, m but not the second. In consequence we obtain

$$p^{ik}R_k{}^m - p^{mk}R_k{}^i = 0. \tag{3.17}$$

Let us write down the equations (3.14) and (3.17) in the local reference system of the principal axes.

[10] E. Wigner, Z. f. Phys. **53**, 592 (1929).
[11] H. Weyl, Z. f. Phys. **56**, 330 (1929).
[12] Cf. Ref. 7, Eq. (5.1).
[13] Cf. J. L. Synge, *Relativity, the General Theory*, (North-Holland Publishing Company, Amsterdam, 1960), p. 17.

Here

$$h_{ia} = \delta_{ia}, \qquad R_{ik} = \sigma_i \delta_{ik}, \tag{3.18}$$

and (3.14) becomes

$$\sigma_a - C\sigma = p^{aa}, \tag{3.19}$$

while (3.17) yields

$$(\sigma_m - \sigma_i)p^{im} = 0. \tag{3.20}$$

But then (considering the distinctness of the σ_i), we obtain

$$p^{ik} = 0 \qquad (i \neq k). \tag{3.21}$$

Now (3.19) and (3.21) can be combined to the single equation

$$p^{ik} = R^{ik} - CRg^{ik}, \tag{3.22}$$

which is first established in the reference system of the principal axes only, but then, being a tensor equation, must hold generally. The Lagrangian factor p^{ik} is thus determined.

Our equations permit us to deduce two important consequences of a quadratic action principle. One

$$R_{ik} = \lambda g_{ik} \tag{3.23}$$

(λ = const) is an exact solution of the field equations.[14] In this case

$$p^{ik} = (1 - 4C)\lambda g^{ik}, \qquad \sigma_1 = \cdots \sigma_4 = \lambda. \tag{3.24}$$

Substitution in (3.11) gives $w^{ik} = 0$, while Eq. (3.15) becomes

$$\tfrac{1}{4}(4\lambda^2 - 16C\lambda^2)g^{im} - (1 - 4C)\lambda^2 g^{im} = 0, \tag{3.25}$$

which is identically satisfied.

A second conclusion is obtained if (3.12) is multiplied by g_{ik}. The expression (3.11) shows that

$$B(p^{ik}g_{ik}) = \tfrac{1}{2}[(p^{ik}g_{ik})_{;mn}g^{mn} + 2p^{mn}{}_{;mn}], \tag{3.26}$$

while multiplication of (3.15) by g_{im} yields for the scalar $w^{im}g_{im}$ in the principal axis system (and thus also generally)

$$w^{im}g_{im} = (\sigma_\mu^2 - C\sigma^2) - (\sigma_i - C\sigma)\sigma_i = 0. \tag{3.27}$$

Hence we obtain, according to (3.12) and (3.26),

$$\Delta_4(p^{ik}g_{ik}) - p^{mn}{}_{;mn} = 0, \tag{3.28}$$

where Δ_4 is the four-dimensional scalar potential operator. On the other hand, the conservation law of the matter tensor yields, in view of (3.22),

$$p^{mn}{}_{;m} = (\tfrac{1}{2} - C)R_{,m}g^{mn}, \tag{3.29}$$

and thus (3.28) becomes

$$(1 - 4C)\Delta_4 R + 2(\tfrac{1}{2} - C)\Delta_4 R = 2(1 - 3C)\Delta_4 R = 0 \tag{3.30}$$

in harmony with the usual result.[15]

4. ELECTROMAGNETIC FIELDS

We do not attempt here to demonstrate that the field equations demanded by the quadratic action principle allow solutions of fourfold periodicity. We postulate the existence of these solutions, which establish a metrical lattice with a lattice constant of the order of magnitude $\mu = 10^{-32}$cm. Our aim will be to investigate the *perturbations* of this lattice, comparable to the bending of a crystal. The perturbation shall be weak in comparison to the existing field of the lattice and we assume it to be macroscopic, i.e., the deformation of the lattice shall be such that it shall extend over many lattice cells with practically constant amplitude.

A perturbation of this kind must come into evidence in the form of a deformation of the local tetrads, i.e., the magnitude and the orientation of the local tetrads must change slightly in consequence of the superimposed perturbation field. Generally such a deformation can be interpreted as a mere rotation of the axes, plus a change of their mutual orientation, i.e., in physical terms an *elastic* deformation. The latter type of deformation modifies the metrical tensor of the manifold. Einstein pointed out repeatedly[16] that in first approximation electric and gravitational fields must be considered as practically independent of each other. If we accept the energy–momentum tensor of Maxwell as describing the macroscopic metrical effect of an electromagnetic field, then the fact that this tensor is quadratic in the field strength indicates that the metrical effect of a weak electromagnetic field must become negligibly small. In that case an electromagnetic type of deformation of the basic tetrad can only become a mere *rotation* of the axes, without any change of the g_{ik}. This harmonizes well with the antisymmetric nature of the electromagnetic field tensor F_{ik}, whose six components exactly imitate the six coefficients of an infinitesimal Lorentz transformation. But what significance can we give to such an infinitesimal rotation, if there is no basic frame which is to be rotated? In the present theory the basic frame exists in the form of the principal axes which establish a natural frame of reference of the space–

[14] Cf. Ref. 7, Eq. (5.16).

[15] Cf. Ref. 7, Eq. (5.22).

[16] E.g., Ref. 5, p. 696.

time world. Einstein in his theory had the 16 components h_{ia} freely at his disposal by considering them as the basic parameters of world geometry, for which the proper field equations had to be found. He could thus separate the symmetric and the antisymmetric parts of h_{ia} and correlate the former to gravitational, the latter to electromagnetic effects. In our case such freedom does not exist since the h_{ia} are defined in terms of a definite Riemannian geometry. If the g_{ik} remain unperturbed, then also the curvature tensor R_{ik} remains unchanged, and thus the principal axes cannot change their orientation. A mere rotation of the axes without elastic deformation is not possible.

There exists, however, an exceptional case, in which the metrical deformation can recede arbitrarily strongly in comparison to the rotation of the axes. We have seen in Sec. 3 that the so-called "cosmological equations" (3.23) are *exact* (although by no means the *only*) solutions of the basic field equations. If we accept this solution, our principal axis problem loses its significance, since all four principal axes collapse into one and *any* four orthogonal unit vectors can be chosen as principal axes. An arbitrarily small perturbation can then cause a finite rotation of the axes. We will assume that the actual solution of the field equations, which belongs to the metrical lattice, is *near* to the solution (3.23), i.e., that our eigenvalues σ_i can be put equal to

$$\sigma_i = \lambda + \epsilon_i, \tag{4.1}$$

where the ϵ_i are small in comparison to the large constant λ. Then we have distinct eigenvalues and the principal axes of the matter tensor are well defined. But we are still near to the case of degeneracy and thus the tendency for a rotation of the axes with practically no elastic changes will be strong.

We will consider the weak perturbation of h_{ia} in the form

$$\delta h_{ia} = h_{ka}\varphi^k{}_i, \tag{4.2}$$

where $\varphi^k{}_i$ is an infinitesimal tensor of second order. This is always possible, since multiplication by h^{ma} yields

$$\varphi^m{}_i = h^{ma}\delta h_{ia}. \tag{4.3}$$

The advantage of operating with $\varphi^k{}_i$ is that it is a genuine tensor of second order, in contrast to the four vectors δh_{ia}. (In Einstein's work this tensor did not come in appearance since he considered perturbations of the flat field $h_{ia} = \delta_{ia}$, in which case δh_{ia} and φ_{ai} coincide.) We can equally lower the index k and write (4.2) in the form

$$\delta h_{ia} = h^{ka}\varphi_{ki}. \tag{4.4}$$

The tensor φ_{ki} need not satisfy any symmetry conditions and has generally 16 components. However, the auxiliary condition (3.7) imposes a restriction on the permissible variations δh_{ia}. Only such variations are permitted which satisfy the condition

$$\delta(\sigma_a h_{ia}h_{ka}) - \delta R_{ik} = 0. \tag{4.5}$$

We will first consider a variation, in which $\varphi^k{}_i$ is chosen as the gradient of a continuous vector field φ^k:

$$\varphi^k{}_i = \varphi^k{}_{;i} \tag{4.6}$$

(the eigenvalues σ_a shall not be varied). Such a variational field is certainly permissible, since it is caused by the infinitesimal coordinate transformation

$$dx^{i\prime} = dx^i + d\varphi^i, \tag{4.7}$$

where $\varphi^i(x_j)$ is an arbitrary vector field. The corresponding variation of g_{ik} becomes

$$\begin{aligned}\delta g_{ik} &= g_{im}\varphi^m{}_{;k} + g_{km}\varphi^m{}_{;i} \\ &= \varphi_{i;k} + \varphi_{k;i}.\end{aligned} \tag{4.8}$$

Similarly,

$$\begin{aligned}\delta R_{ik} &= R_{im}\varphi^m{}_{;k} + R_{km}\varphi^m{}_{;i} \\ &= R_i{}^m\varphi_{m;k} + R_k{}^m\varphi_{m;i}.\end{aligned} \tag{4.9}$$

On the other hand, we consider the perturbation field

$$\delta h_{ia} = h^{ma}\varphi_{i;m}, \tag{4.10}$$

in which case the variation of g_{ik} and of $\sigma_a h_{ia}h_{ka}$ becomes

$$\delta g_{ik} = \varphi_{k;i} + \varphi_{i;k}, \tag{4.11}$$

$$\sigma_a\delta(h_{ia}h_{ka}) = R_i{}^m\varphi_{k;m} + R_k{}^m\varphi_{i;m}. \tag{4.12}$$

The variation of g_{ik} is the same in both fields and thus the difference of the two fields, that is

$$\varphi_{ik} = F_{ik} = \varphi_{i,k} - \varphi_{k,i} \tag{4.13}$$

is free of any metrical change. This, however, cannot hold without proper correction, since the second field does not satisfy the auxiliary condition (4.5) of the variation. The difference between the two terms on the left side of (4.5) becomes

$$\begin{aligned}&R_i{}^m(\varphi_{k,m} - \varphi_{m,k}) + R_k{}^m(\varphi_{i,m} - \varphi_{m,i}) \\ &\qquad = R_i{}^mF_{km} + R_k{}^mF_{im}.\end{aligned} \tag{4.14}$$

This imbalance has to be corrected by a proper metrical change. However, if we write down (4.14) in the reference system of the principal axes, considering our condition (4.1), we find that the error of our equation becomes

$$\epsilon_i F_{ki} + \epsilon_k F_{ik} = (\epsilon_k - \epsilon_i)F_{ik}. \tag{4.15}$$

The metrical correction needed is of the order of magnitude

$$(\epsilon_m/\lambda)F_{ik}, \tag{4.16}$$

which for our present purposes can be considered as negligible. We have thus obtained a perturbation field which in close approximation realizes the non-metrical field of Einstein, represented in his case by the antisymmetric part of h_{ia}, in our case by the antisymmetric tensor F_{ik}.

Before we proceed to the construction of the perturbation Lagrangian, we make one more observation. We have seen in (3.30) that one of the important consequences of a quadratic action principle is that the scalar curvature R satisfies the four-dimensional potential equation

$$\Delta_4 R = 0 \tag{4.17}$$

(we exclude the singular value $C = \frac{1}{3}$), which in a space of positive-definite signature allows no regular solutions, except

$$R = \text{const.} \tag{4.18}$$

This is an *exact local first integral* of the quadratic action principle. It has the consequence that at every point of the manifold

$$\epsilon_1 + \cdots \epsilon_4 = \text{const} = \epsilon. \tag{4.19}$$

Now we will show that without loss of generality this constant can be equated to zero. Instead of the original λ in the cosmological equations (4.1), we can use a slightly different $\lambda' = \lambda + \frac{1}{4}\epsilon$ with the consequence that now the new $\epsilon_i' = \epsilon_i - \frac{1}{4}\epsilon$ satisfy the condition $\sum \epsilon_i' = 0$. Omitting the primes we can immediately submit the ϵ_i to the condition

$$\epsilon = \epsilon_1 + \cdots + \epsilon_4 = 0. \tag{4.20}$$

This is of crucial importance for our later conclusions.

5. THE PERTURBATION LAGRANGIAN

We have assumed that the lattice structure of our basic metrical field represents a regular solution of the field equations of the quadratic action principle. If now we want to investigate a weak *perturbation* of the basic field, we are in a similar situation as in the problem of "small vibrations around a state of equilibrium," encountered in the mechanics of solids. We have to obtain the perturbation Lagrangian up to quantities of *second* order, since the first-order perturbation vanishes, considering that we have started with a genuine solution of the field equations. The variation of this quadratic Lagrangian will result in linear differential equations, expressing the perturbation (in our case macroscopic perturbation) of the basic lattice vibrations. We will not consider the problem in its full generality but restrict ourselves to the perturbation studied in the previous chapter, in which the variation of the g_{ik} becomes negligible and the deformation of the lattice consists essentially in a mere *rotation* of the principal axes. Moreover, we are not interested in the local variations of the lattice field but purely in the *macroscopic* change which extends over many lattice cells. For this reason we will integrate the second variation of L' over the entire lattice cell and minimize the resultant action, instead of obtaining local field equations by considering the variational problem in its totality.

We consider the Lagrangian (3.9). The variation of p^{ik} and w^{ik} give the following contribution to the second variation of L':

$$-\delta p^{ik}[\delta(\sigma_a h_{ia}h_{ka}) - \delta R_{ik}]h - \delta w^{ik}[\delta g_{ik} - \delta h_{ia}h_{ka}]h. \tag{5.1}$$

Since, however, the factors of δp^{ik} and δw^{ik} must vanish, this part of $\delta^2 L'$ can be omitted. What remains, is the following contribution, caused by the variation of the h_{ia}; (in the perturbation field here considered the σ_a are not varied):

$$\delta^2 L' = \tfrac{1}{2}[(\lambda + \epsilon_i)^2 - C(4\lambda + \epsilon)^2]\delta^2 h - hp^{ik}\sigma_a \delta h_{ia}\delta h_{ka} - hw^{ik}\delta h_{ia}\delta h_{ka}. \tag{5.2}$$

The factor of $\delta^2 h$ becomes greatly simplified, in view of the condition (4.20). We will consistently neglect second-order quantities of the form $\epsilon_i\epsilon_k$. Then the first term of (5.2) becomes

$$2(1 - 4C)\lambda^2\delta^2 h, \tag{5.3}$$

and our first task will be to obtain the second variation of the determinant of the matrix h_{ik}. Now the modified value of h_{ia}, given by (4.4), may be written in the following form:

$$h'_{ia} = h_{ia} + h_{ka}\varphi^k{}_i = h_{ka}(\delta^k_i + \varphi^k{}_i), \tag{5.4}$$

and by the determinant theorem of the product of two matrices we obtain

$$||h'_{ik}|| = h\,||\delta^k_i + \varphi^k{}_i|| = h[1 + \varphi^k{}_k + \tfrac{1}{2}(\varphi^k{}_k)^2 - \tfrac{1}{2}\varphi^k{}_i\varphi^i{}_k]. \tag{5.5}$$

Hence

$$2\delta^2 h = h[(\varphi^k{}_k)^2 - \varphi^k{}_i\varphi^i{}_k]. \tag{5.6}$$

In our problem $\varphi_{ik} = F_{ik} = -F_{ki}$ is antisymmetric. Hence $\varphi^k{}_k = 0$ and we obtain

$$2\delta^2 h = -hF^{ki}F_{ik} = hF^{ik}F_{ik}. \tag{5.7}$$

We now come to the investigation of the second term on the right side of (5.2). Here we obtain

$$\sigma_a \delta h_{ia}\delta h_{ka} = R_{jm}\varphi^j{}_i\varphi^m{}_k = R^{jm}\varphi_{ji}\varphi_{mk} \tag{5.8}$$

and the complete second term becomes, if we substitute for p^{ik} its value (3.22):

$$-h[R^{ik} - 4\lambda C g^{ik}]R^{jm}F_{ji}F_{mk}. \tag{5.9}$$

Let us now introduce the reference system of the principal axes as a local reference system. Here the term (5.3) becomes

$$(1 - 4C)\lambda^2(F_{ik})^2, \tag{5.10}$$

while the term (5.9) becomes

$$-\sigma_i\sigma_k(F_{ik})^2 + 4\lambda C\sigma_i(F_{ik})^2, \tag{5.11}$$

but in the last term we can symmetrize σ_i and replace it by $(\frac{1}{2})(\sigma_i + \sigma_k)$. If we now make use of our assumption (4.1) and neglect second-order terms in ϵ_i, we obtain for (5.11)

$$-(1 - 4C)\lambda^2(F_{ik})^2 - (\epsilon_i + \epsilon_k)\lambda(1 - 2C)(F_{ik})^2. \tag{5.12}$$

In the sum of (5.10) and (5.12), the term with λ^2 drops out and thus the contribution of the first two terms of $\delta^2 L'$ becomes

$$-\lambda(1 - 2C)(\epsilon_i + \epsilon_k)(F_{ik})^2. \tag{5.13}$$

We now come to the investigation of the last term on the right side of (5.2):

$$-hw^{ik}\delta h_{ia}\delta h_{ka} = -hw^{ik}g_{jm}\varphi^j{}_i\varphi^m{}_k = -hw^{ik}F_{mi}F^m{}_k = hB(p^{ik})F_{mi}F^m{}_k. \tag{5.14}$$

First of all we observe from the expression (3.12) that if we put

$$p^{ik} = (1 - 4C)\lambda g^{ik} + \bar{p}^{ik}, \tag{5.15}$$

the first term gives identically zero. Hence in the evaluation of (5.14) we can replace p^{ik} by $\bar{p}^{ik}$, with the consequence that we know in advance that $B(\bar{p}^{ik})$ will be of the order ϵ_i. Furthermore, let us write R_{ik} in the following form:

$$R_{ik} = R_{ijkm}g^{jm} = \tfrac{1}{2}[g_{ik,jm} + g_{jm,ik} - g_{im,jk} - g_{jk,im} + (\Gamma_{ik,a}\Gamma_{jm,b} - \Gamma_{im,a}\Gamma_{jk,b})g^{ab}]g^{jm}. \tag{5.16}$$

We notice that varying the g_{ik} and integrating by parts we obtain a large number of terms which are not more than *derivatives*. But integrating over a whole lattice cell all these quantities must disappear, on account of the periodic nature of our lattice. The only part which contributes to our action integral is obtained by varying g^{ab} and g^{jm}. The resulting expression is of the general form

$$\bar{p}_{jn}A^{jink}, \tag{5.17}$$

where A^{jink} has the symmetry properties of the Riemann tensor. Since $\bar{p}_{jn}$ is already of the order ϵ_i, it suffices to know the tensor A^{jink} in first approximation only, neglecting quantities of the order ϵ_i. We assume that it is permissible to put with sufficient accuracy

$$A^{jink} = \alpha\lambda(g^{jn}g^{ik} - g^{jk}g^{in}), \tag{5.18}$$

where α is a numerical constant. Then, if we utilize once more the reference system of the principal axes (which is permissible since the quantity to be evaluated is a mere scalar), we obtain

$$-\alpha\lambda\epsilon_j\delta_{jn}\delta_{jk}\delta_{in}F_{mi}F_{mk} = -\alpha\lambda\epsilon_k(F_{mk})^2 = -\tfrac{1}{2}\alpha\lambda(\epsilon_i + \epsilon_k)(F_{ik})^2.$$

This term is once more of the general form (5.13).

In summary we can say that the resulting perturbation Lagrangian is proportional to the following scalar:

$$\tfrac{1}{2}(\epsilon_i + \epsilon_k)(F_{ik})^2. \tag{5.20}$$

The six terms of this sum show a remarkable property. Considering the condition (4.20) between the eigenvalues ϵ_i, the six terms split into the sum of only three terms

$$(\epsilon_1 + \epsilon_2)(F_{12}^2 - F_{34}^2) + (\epsilon_2 + \epsilon_3)(F_{23}^2 - F_{14}^2) + (\epsilon_3 + \epsilon_1)(F_{31}^2 - F_{24}^2), \tag{5.21}$$

and if we assume that the average values of ϵ_1, ϵ_2, and ϵ_3, integrated over the entire lattice cell, are equal to each other, we obtain (in customary notation) the Lagrangian of the superimposed perturbation field as proportional to

$$H^2 - E^2. \tag{5.22}$$

We have thus obtained the customary Lagrangian of the electromagnetic field—and thus the Maxwellian equations—in spite of the fact that the basis of our consideration was a genuine Riemannian metric, with a positive-definite line element.

This result remains unaltered by the following modification. In evaluating the perturbation Lagrangian we had to know the perturbation field in

second approximation. This brought in the products $\delta h_{ia}\delta h_{ka}$ while $\delta^2 h_{ia}$ was considered as zero. But if in the representation (4.2) we consider φ_i^k as a *constant* tensor, then the second variation of h_{ia} is not zero but becomes

$$\delta^2 h_{ia} = h_{ma}\varphi^m{}_k\varphi^k{}_i. \tag{5.23}$$

The addition of these terms does not change, however, the resulting expression (5.21) of the perturbation Lagrangian—it only modifies the constant factor by which it is multiplied.

The weakness of the present development is that we do not know enough about the structure of the basic lattice field and had to make assumptions concerning its behavior, which must be corroborated by future research. We have, furthermore, focused our attention on the electromagnetic field, neglecting the gravitational effects and the possible interaction between the two types of fields.

However, our results permit us to draw the following general conclusion. If we erect the perturbation field on an empty (flat) universe, then the perturbation Lagrangian must be interpretable as a mere scalar of a certain (constant) metrical field. But if the perturbation field is superimposed on a metrical crystalline lattice, then this lattice will impose its own structure on the perturbation Lagrangian, although we know in advance that the high-frequency periodicity of this lattice can only give rise to *constant* tensors (comparable to the dielectric tensor of a crystal). In that case the positive-definite nature of a Riemannian line element need not come in direct evidence in the perturbation Lagrangian and the Minkowskian geometry might merely be an attempt to interpret the superposition Lagrangian as an invariant of a properly chosen metric, although in fact—as we have seen it in the calculations concerning the electromagnetic field—this Lagrangian is an invariant of a much more complicated type, vitally influenced by the (highly agitated) metrical substructure. The Riemannian line element and the apparently Minkowskian nature of the superposition field are thus in no contradiction to each other.

ACKNOWLEDGMENTS

The principal ideas developed in this paper germinated during the author's memorable six months' stay at the Mathematical and Theoretical Sciences Department of the Scientific Laboratory of the Ford Motor Company, Dearborn, Michigan, by invitation of Dr. J. E. Goldman, Director. The author acknowledges with thanks the stimulating discussions with the members of the weekly Relativity Seminars, in particular, Dr. A. W. Overhauser, Dr. R. Penney, and Professor G. Y. Rainich.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 2 FEBRUARY 1966

Methods of Quadrature for Euler Transform Integrals*

PETER M. LIVINGSTON
Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin
(Received 22 February 1965)

Several methods of treating Euler transform integrals exist. One such method follows from the expression of the Euler transform kernel as a bilinear series of independent solutions to the Jacobi equation valid for the integration variable in the real interval -1 to 1 and the transform variable outside. The transform function then is expressed as a series of solutions of the second kind to the Jacobi equation whose coefficients are the expansion coefficients of the function to be transformed in the complete set of Jacobi polynomials, provided the latter exist. Such a series is absolutely convergent for the transform variable not on the real interval cited above. Another method, due to MacRobert, permits quadrature of the Euler transform integral directly for certain integrands. Finally, the expansion of the Euler kernel in a bilinear series of Bessel functions and Neumann polynomials valid for the integration variable on the finite interval, 0 to a, is mentioned, and applied to several integrals. Examples of all three methods are given.

INTRODUCTION

In certain applications of applied mathematics such as the study of potential problems in quantum mechanics or plasma physics, integrals of the *Euler transform* type occasionally arise;

$$g(z) = \int_{-1}^{1} (z - x)^{-\mu} f(x)\, dx, \qquad (1)$$

where $f(x)$ is sufficiently well behaved to be uniformly approximated on the interval $[-1, 1]$ by a complete set of polynomials, and μ is any number such that $\mathrm{Re}\ \mu > 0$. If the order of the transform μ is an integer, the restriction that z not lie on the real axis segment, $[-1, 1]$ may be relaxed by taking the principal part of the integral. As it stands, $g(z)$ is an analytic function of z for all neighborhoods not overlapping the real axis cut as given above, and therefore the kernel may be expanded in a Taylor series such that the integration may be carried out term by term. There is, however, another expansion of the kernel in a bilinear series of functions which are solutions of the first and second kind of a hypergeometric equation. This representation of the kernel is also absolutely convergent for all z restricted as above; therefore term by term integration in Eq. (1) is also justified. However, solutions of the first kind with integer indices are polynomials that form closed sets on the interval, $[-1, 1]$ with respect to specified weight functions. Thus $g(z)$ may be expressed as a series in solutions of the second kind with coefficients that are the expansion coefficients of the arbitrary function in the polynomial set. An obvious advantage of this kernel representation follows if $f(x)$ is orthogonal to all but one of the polynomials, in which case $g(z)$ is proportional to a solution of the second kind. From a numerical point of view, expression of $g(z)$ in a series of solutions permits the use of recursion relations to "build up" the series as might be done in a computer evaluation of $g(z)$. Finally, the establishment of a bilinear expansion of the Euler kernel permits the extension of integral tables to cover integrals of the type shown in Eq. (1) if the appropriate expansion coefficients are already evaluated. Examples of a few of these integrals are given in Appendix A.

The following three sections summarize the development of several bilinear expansions, and treat the Euler transform of a special integrand. The first of these reviews some basic properties of the Jacobi function system necessary to subsequent development.

JACOBI FUNCTION SYSTEM

The hypergeometric equation with three regular singular points located at ± 1 and ∞ is known as the Jacobi equation. Its two independent solutions are characterized by three parameters α, β, and n; if the latter is integer, one of the two solutions is a polynomial of order n. The second solution is regular everywhere in the complex plane but has branch points at ± 1, with a branch cut joining these two singularities to make it single-valued. The Jacobi polynomials, or solutions of the first kind, form a complete set on the closed interval -1 to 1 with respect to the integer index n and a weight function given in the table below. At infinity, these polynomials have a simple pole of order n. Both solutions satisfy well-known recursion formulas

* This research was carried out under grant NsG-275-62 from the National Aeronautics and Space Administration.

TABLE I.

(α, β)	Name	Symbol	Weight function, $W^{(\alpha,\beta)}$	Normalization, $N_n^{(\alpha,\beta)}$
	Jacobi	$P_n^{(\alpha,\beta)}$	$(1+x)^\beta(1-x)^\alpha$	$\dfrac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!(2n+\alpha+\beta+1)\Gamma(\alpha+\beta+n+1)}$
$\alpha=\beta=\lambda-\frac{1}{2}$	Gegenbauer	$C_n^\lambda = \dfrac{\Gamma(2\lambda+n)\Gamma(\lambda+\frac{1}{2})}{\Gamma(2\lambda)\Gamma(\lambda+\frac{1}{2}+n)} P_n^{(\lambda-\frac{1}{2},\lambda-\frac{1}{2})}$	$(1-x^2)^{\lambda-\frac{1}{2}}$	$\dfrac{\pi 2^{1-2\lambda}\Gamma(2\lambda+n)}{n!(n+\lambda)\{\Gamma(\lambda)\}^2}$
$\alpha=\beta=-\frac{1}{2}$	Tchebycheff 1st kind	$T_n = \dfrac{n!\Gamma(\frac{1}{2})}{2\Gamma(n+\frac{1}{2})} P_n^{(-\frac{1}{2},-\frac{1}{2})}$	$(1-x^2)^{-\frac{1}{2}}$	$\frac{1}{2}\pi,\ n\neq 0;\quad \pi,\ n=0$
$\alpha=\beta=\frac{1}{2}$	Tchebycheff 2nd kind	$U_n = \dfrac{(n+1)!\Gamma(\frac{3}{2})}{2\Gamma(n+\frac{3}{2})} P_n^{(\frac{1}{2},\frac{1}{2})}$	$(1-x^2)^{\frac{1}{2}}$	$\frac{1}{2}\pi,\ n\neq 0;\quad \pi,\ n=0$
$\alpha=\beta=0$	Legendre	$P_n = P_n^{(0,0)}$	1	$\dfrac{2}{2n+1}$

given elsewhere.[1] For certain specified values of the parameters α, and β, the Jacobi polynomials are proportional to the Gegenbauer, Tchebycheff, and Legendre polynomials as shown in Table I.

The parameters α, β, and λ have real parts greater than -1. Thus we see that any result that holds for the Jacobi polynomial system is also true for any of the systems listed in the table. For the purpose of conveniently expressing later results, we introduce a general polynomial/function system in the next section. These are the Jacobi solutions for arbitrary α and β but with an additional multiplicative factor to account for the various interrelations among the polynomial sets given by special values of α, β as listed above.

As shown in Appendix B, the Euler transform method solution of the Jacobi equation leads to a convenient integral representation of its solutions. The choice of the contour determines which of the two independent solutions is represented; and, for the real interval $[-1, 1]$, we obtain the integral form of the Jacobi function of the second kind:

$$Q_n^{(\alpha,\beta)}(z) = 2^{-n-1}\int_{-1}^{1}(z-t)^{-n-\alpha-\beta-1} \times (1+t)^{n+\alpha}(1-t)^{n+\beta}\,dt. \quad (2)$$

Here $z \notin [-1, 1]$. This function satisfies the same recursion formulas as $P_n^{(\alpha,\beta)}$, except for $n = 0$, and if Re $(\alpha, \beta) > -1$ and is analytic everywhere except for the branch cut between -1 and 1. Its value on the cut is defined to be

$$Q_n^{(\alpha,\beta)}(x) \equiv \tfrac{1}{2}[Q_n^{(\alpha,\beta)}(x+i0) + Q_n^{(\alpha,\beta)}(x-i0)]$$
$$= \frac{-1}{2\pi\sin\alpha\pi}P_n^{(\alpha,\beta)}(x) + \frac{2^{\alpha+\beta-1}\cos\alpha\pi\,\Gamma(\alpha)\Gamma(n+\beta+1)(1-x)^{-\alpha}(1+x)^{-\beta}}{\Gamma(n+\alpha+\beta+1)} \times F(n+1, -n-\alpha-\beta; 1-\alpha; \tfrac{1}{2}-\tfrac{1}{2}x); \quad -1 < x < 1. \quad (3)$$

The latter equality follows from considering the contour integral about the branch cut and taking the limit Im $z = 0$; and F is the hypergeometric function ${}_2F_1$. Additional properties of the solutions of the second kind associated with the Jacobi, Gegenbauer, Tchebycheff, and Legendre polynomials are found in standard texts.[1-3]

It is known from the theory of the hypergeometric function that these quantities map into themselves under the fractional linear substitution:

$$t' = (At+B)/(Ct+D); \qquad AD - BC \neq 0. \quad (4)$$

Therefore it is possible to obtain several equivalent integral representations of the same solution. One very useful form is derived from Eq. (2) by the substitution

$$t = \left\{\frac{z}{1+z}(1+u) - 1\right\} \Big/ \left\{\frac{1+u}{1+z} - 1\right\}. \quad (5)$$

Equation (2) then reads

[1] Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II.

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

[3] E. Hobson, *Spherical and Ellipsoidal Harmonics* (Chelsea Publishing Company, New York, 1955).

$$Q_n^{(\alpha,\beta)}(z) = 2^{-n-1}(z-1)^{-\alpha}(z+1)^{-\beta} \times \int_{-1}^{1} du\,(z-u)^{-n-1}(1+u)^{n+\beta}(1-u)^{n+\alpha}; \quad z \notin [-1, 1]. \tag{6}$$

In a similar manner, the integral representations of the polynomial solution may be developed:

$$P_n^{(\alpha,\beta)}(z) = 2^{-n-1}(\pi i)^{-1}(1-z)^{-\alpha}(1+z)^{-\beta} \times \oint_C (z-u)^{-n-1}(1+u)^{n+\beta}(1-u)^{n+\alpha}\,du; \quad -1 \le z \le 1. \tag{7}$$

Here the phase of the integrand has been chosen such that $P_n^{(\alpha,\beta)}$ is real for z on the positive real axis. The selected contour encloses the points z and 1 where the complex plane has been cut from -1 to $-\infty$. These two relations may be combined to give the key equation upon which this paper is based. First we see that the integrand of Eq. (7) is analytic everywhere in C except at z, and hence its residue is $(n!)^{-1}(-1)^n$ times the nth derivative of $(1+z)^{n+\beta}(1-z)^{n+\alpha}$. Therefore the polynomial $P_n^{(\alpha,\beta)}$ is expressed by Rodrigues' formula

$$P_n^{(\alpha,\beta)}(x) = \frac{(1-x)^{-\alpha}(1+x)^{-\beta}(-1)^n}{2^n n!}\frac{d^n}{dx^n} \times (1+x)^{n+\beta}(1-x)^{n+\alpha}. \tag{8}$$

Now, since the integrand of Eq. (6) vanishes at $u = \pm 1$, the integrated terms vanish upon integration by parts to give

$$Q_n^{(\alpha,\beta)}(z) = \frac{(z-1)^{-\alpha}(z+1)^{-\beta}(-1)^n}{2^{n+1}n!} \times \int_{-1}^{1} \frac{1}{(z-t)}\frac{d^n}{dt^n}(1-t)^{n+\alpha}(1+t)^{n+\beta}\,dt. \tag{9}$$

Rodrigues' formula may be substituted for the integrand in the above equation to give the fundamental relationship between the polynomials and solutions of the second kind, often called the Neumann integral,

$$Q_n^{(\alpha,\beta)}(z) = 2^{-1}(z-1)^{-\alpha}(z+1)^{-\beta} \times \int_{-1}^{1} (z-t)^{-1}(1+t)^{\beta}(1-t)^{\alpha}P_n^{(\alpha,\beta)}(t)\,dt. \tag{10}$$

Here $z \notin [-1, 1]$. The equation may be extended by definition to values of $|\mathrm{Re}\ z| < 1$, $\mathrm{Im}\ z = 0$, if the principal part of the above integral is taken. This Euler transform relationship holds for all of the specializations of the Jacobi polynomials and solutions of the second kind as listed in Table I.

BILINEAR EXPANSIONS

The Jacobi polynomial system and its specializations form closed sets on the interval $[-1, 1]$. In order to express this property as well as subsequent expansion formulas, it is convenient to introduce a general polynomial of the first kind $\psi_n^{(\alpha,\beta)}$, and a general function of the second kind $\varphi_n^{(\alpha,\beta)}$, which, for special values of their indices are proportional to various of the polynomial systems listed in Table I. Thus, for α and β as given in the table below, we have

$$\begin{Bmatrix}\psi_n^{(\alpha,\beta)}\\ \varphi_n^{(\alpha,\beta)}\end{Bmatrix} = B_n^{(\alpha,\beta)} \times \begin{Bmatrix}\text{any of the Jacobi polynomial/function}\\ \text{systems given in Table I}\end{Bmatrix}.$$

The constant of proportionality, $B_n^{(\alpha,\beta)}$ has the values given in Table II.

The general polynomials are orthogonal on the closed interval from -1 to 1 with respect to the weight function $w(x)$ for different integer indices:

$$\int_{-1}^{1} w(x)\psi_n^{(\alpha,\beta)}(x)\psi_m^{(\alpha,\beta)}(x)\,dx = \mathfrak{N}_n^{(\alpha,\beta)}\delta_{nm}.$$

TABLE II.

α, β	Name	Polynomial/function		$B_n^{(\alpha,\beta)}$
α, β	Jacobi	$P_n^{(\alpha,\beta)}(x)$;	$Q_n^{(\alpha,\beta)}(x)$	1
$\lambda - \frac{1}{2}, \lambda - \frac{1}{2}$	Gegenbauer	$C_n^{\lambda}(x)$;	$D_n^{\lambda}(x)$	$\Gamma(2\lambda)\Gamma(\lambda + \frac{1}{2} + n)/\{\Gamma(2\lambda + n)\Gamma(\lambda + \frac{1}{2})\}$
$-\frac{1}{2}, -\frac{1}{2}$	Tchebycheff 1st kind	$T_n(x)$;	$R_n(x)$	$\Gamma(n + \frac{1}{2})/\{\Gamma(\frac{1}{2})\Gamma(n+1)\}$
$\frac{1}{2}, \frac{1}{2}$	Tchebycheff 2nd kind	$U_n(x)$;	$S_n(x)$	$\Gamma(n + \frac{3}{2})/\{\Gamma(\frac{1}{2})\Gamma(n+2)\}$
0, 0	Legendre	$P_n(x)$;	$Q_n(x)$	1

The normalization constant for the general polynomials is related to those given in Table I for the indicated choice of parameters by

$$\mathfrak{N}_n^{(\alpha,\beta)} = [B_n^{(\alpha,\beta)}]^2 N_n^{(\alpha,\beta)}.$$

The closure relation for the general case is

$$w(x) \sum_{n=0}^{\infty} [\mathfrak{N}_n^{(\alpha,\beta)}]^{-1} \psi_n^{(\alpha,\beta)}(x) \times \psi_n^{(\alpha,\beta)}(x') = \delta(x - x'). \quad (11)$$

It follows immediately from this result and the Euler transform relation between the polynomials of the first kind and functions of the second kind, Eq. (10), that $(z - x)^{-1}$ has the expansion

$$(z - x)^{-1} = 2(-1)^{\alpha} w(z) \times \sum_{n=0}^{\infty} \{\mathfrak{N}_n^{(\alpha,\beta)}\}^{-1} \psi_n^{(\alpha,\beta)}(x) \varphi_n^{(\alpha,\beta)}(z). \quad (12)$$

This equation is the familiar Christoffel–Darboux identity as the upper limit on the index n passes to infinity. It may be derived from the recursion relations of three different indices satisfied by all of the Jacobi family for any upper limit, as shown in Bateman.[1] For fixed z, the region of absolute convergence for x is any point on the interior of an ellipse, passing through z in the complex plane with foci at ± 1. The quadrature formulas now become straightforward in terms of this expansion.

A generalization of Eq. (12) to integral powers of $(z - x)^{-1}$ follows from the definition of the associated polynomial and function of the second kind. Let these quantities be defined by

$$\psi_n^{(\alpha,\beta;m)}(x) = \frac{(x^2 - 1)^{\frac{1}{2}m}}{w(x)} \frac{d^m}{dx^m} \{w(x) \psi_n^{(\alpha,\beta)}(x)\} \quad (13)$$

and

$$\varphi_n^{(\alpha,\beta;m)}(z) = \frac{(z^2 - 1)^{m/2}}{w(z)} \frac{d^m}{dz^m} \{w(z) \varphi_n^{(\alpha,\beta)}(z)\}, \quad (14)$$

where the index m is integer and positive. It may easily be shown that these functions satisfy the associated equation.

$$(1 - x^2)y'' + [\beta - \alpha - (\alpha + \beta + 2)x]y' + (n - m)(n + m + \alpha + \beta + 1)y = 0, \quad (15)$$

where y represents either type of solution. Recursion relations for these functions follow from the above equation and the previous definitions. Integral representations follow from the definitions and the integral representations of the nonassociated quantities previously discussed. Thus differentiation of Eq. (12) with respect to z leads to the required generalization of that formula in terms of the associated functions of the second kind:

$$(z - x)^{-m-1} = (m!)^{-1} 2(-1)^{\alpha+m} w(z)(z^2 - 1)^{-\frac{1}{2}m} \times \sum_{n=0}^{\infty} \{\mathfrak{N}_n^{(\alpha,\beta)}\}^{-1} \psi_n^{(\alpha,\beta)}(x) \varphi_n^{(\alpha,\beta;m)}(z). \quad (16)$$

If, in the integrand of Eq. (1), $h(x)$ is defined by $f(x) = w(x)h(x)$ and $\mu = m + 1$ then that integral becomes, in view of the expansion developed above,

$$g(z) = \int_{-1}^{1} \frac{w(x)h(x)\,dx}{(z - x)^{m+1}} = \frac{(-1)^{\alpha+m} w(z)(z^2 - 1)^{\frac{1}{2}m}}{m!} \sum_{n=0}^{\infty} A_n^{(\alpha,\beta)} \varphi_n^{(\alpha,\beta;m)}(z), \quad (17)$$

where the coefficients of $\varphi_n^{(\alpha,\beta;m)}$ are

$$A_n^{(\alpha,\beta)} = 2\{\mathfrak{N}_n^{(\alpha,\beta)}\}^{-1} \int_{-1}^{1} w(x) \psi_n^{(\alpha,\beta)}(x) h(x)\,dx. \quad (18)$$

By this means we have expressed $g(z)$ in an absolutely convergent series for all $z \notin [-1, 1]$, and have reduced the integral to the problem of determining the coefficients given in the above equation. In cases where these are listed in integral tables, or may be easily determined, the tables then can be expanded to include generalized Euler transform integrals of the type given in Eq. (17). As an example, consider $h(x) = \cos \alpha x$; $w(x) = 1$; $\psi_n^{(\alpha,\beta)} = P_n(x)$; and $\varphi_n^{(\alpha,\beta)} = Q_n(z)$; then we have

$$\int_{-1}^{1} \cos \alpha x (z - x)^{-1}\,dx = \sum_{n=0}^{\infty} A_n Q_n(z) = \left(\frac{2\pi}{\alpha}\right)^{\frac{1}{2}} \sum_{\substack{n=0 \\ \text{even}}}^{\infty} (-1)^{\frac{1}{2}n}(2n + 1) J_{n+\frac{1}{2}}(\alpha) Q_n(z), \quad (19)$$

where

$$A_n = \begin{cases} \left(\frac{2\pi}{\alpha}\right)^{\frac{1}{2}} (2n + 1)(-1)^{\frac{1}{2}n} J_{n+\frac{1}{2}}(\alpha); & n = \text{even}, \\ 0; & n = \text{odd}. \end{cases} \quad (20)$$

Other examples are listed in Appendix A.

A simple generalization of the integral shown in Eq. (1) follows by replacing the denominator by a polynomial of finite order whose roots do not lie on the real axis segment, $[-1, 1]$. By means of an improper fraction expansion, the integral may be reduced to a sum over distinct roots of integrals of the form shown in Eq. (17). Here, the integer $m + 1$ is the order of the degenerate root if $m > 0$. Integrals with polynomial denominators that do have a finite number of roots on the real axis

segment may be handled in like manner, but with the real axis segment definition of functions of the second kind, Eq. (3), and replacement of the integrals with their principal values.

A third method of expansion of an Euler kernel depends on the fact that the integral representation of the solutions of the second kind, Eq. (6), is valid for any n whose real part is greater than minus one. Thus we show that a bilinear expansion exists for the quantity $(z-x)^{-\gamma-1}$, for $\mathrm{Re}\,\gamma > -1$. For clarity, we explicitly indicate the parameter dependence of the weight function as superscripts. Now from the choice of the parameter B given in Table II it follows that Eqs. (8) and (6) may be taken to be definitions of the general polynomial and function of the second kind, respectively. If, in the latter equation, n is set equal to $l+\gamma$, where l is a positive integer, Eq. (6) becomes

$$\varphi_{l+\gamma}^{(\alpha,\beta)}(z) = \frac{(-1)^{\alpha}}{2^{l+\gamma+1}w^{(\alpha,\beta)}(z)} \times \int_{-1}^{1} \frac{dt\, w^{(\alpha,\beta)}(t)(1-t^2)^{l+\gamma}}{(z-t)^{l+\gamma+1}}. \quad (21)$$

But

$$w^{(\alpha,\beta)}(t)(1-t^2)^{l+\gamma} = w^{(\alpha+\gamma,\beta+\gamma)}(t)(1-t^2)^{l},$$

and therefore, after integrating Eq. (21) l times by parts, we have

$$\varphi_{l+\gamma}^{(\alpha,\beta)}(z) = \frac{(-1)^{\alpha+l}\Gamma(\gamma+1)}{2^{l+\gamma+1}w^{(\alpha,\beta)}(z)\Gamma(l+\gamma+1)} \times \int_{-1}^{1} \frac{dt}{(z-t)^{\gamma+1}} \frac{d^l}{dt^l}\{w^{(\alpha+\gamma,\beta+\gamma)}(t)(1-t^2)^{l}\}. \quad (22)$$

As in the previous cases, the integrand vanishes at the end points, and thus the integrated terms are zero. Rodrigues' formula may be used to replace the derivative appearing in the integrand with the appropriate general polynomial to give the following integral representation:

$$\varphi_{l+\gamma}^{(\alpha,\beta)}(z) = \frac{\Gamma(1+\gamma)\Gamma(l+1)(-1)^{\alpha}}{w^{(\alpha,\beta)}(z)\Gamma(l+\gamma+1)2^{\gamma+1}} \times \int_{-1}^{1} \frac{dt}{(z-t)^{\gamma+1}} w^{(\alpha+\gamma,\beta+\gamma)}(t)\psi_l^{(\alpha+\gamma,\beta+\gamma)}(t). \quad (23)$$

The closure formula, Eq. (11), may now be applied to extract the denominator; thus

$$(z-x)^{-\gamma-1} = (-1)^{\alpha}w^{(\alpha,\beta)}(z) \sum_{l=0}^{\infty} \{\mathfrak{N}_l^{(\alpha+\gamma,\beta+\gamma)}\}^{-1} \times \psi_l^{(\alpha+\gamma,\beta+\gamma)}(x)\varphi_{l+\gamma}^{(\alpha,\beta)}(z) \frac{\Gamma(l+\gamma+1)2^{\gamma+1}}{\Gamma(\gamma+1)l!}. \quad (24)$$

In this formula, as in others, z is assumed not to lie on the cut. Several interesting cases arise from particular choices of the parameters α, β, and γ. For example, if $\alpha = \beta = 0$ and $\gamma = -\frac{1}{2}$, the above equation reduces to

$$(z-x)^{-\frac{1}{2}} = \frac{2^{\frac{1}{2}}}{\pi} Q_{-\frac{1}{2}}(z) + \frac{2\sqrt{2}}{\pi} \sum_{l=1}^{\infty} Q_{l-\frac{1}{2}}(z)T_n(x), \quad (25)$$

which, if the interval is changed to 0, π becomes the Fourier cosine expansion of the square root of $z - \cos\theta$. The functions $Q_{l-\frac{1}{2}}$ are analytic everywhere in the complex plane with logarithmic singularities at ± 1, and discontinuous along the cut from -1 to 1. These functions arise in the theory of toroidal harmonics, and are discussed by Hobson.[3] This same quantity $(z-x)^{-\frac{1}{2}}$ may be written as a bilinear series in Legendre polynomials by setting $\alpha = \beta = \frac{1}{2}$, $\gamma = -\frac{1}{2}$. The result is

$$(z-x)^{-\frac{1}{2}} = (z^2-1)^{\frac{1}{2}}2^{\frac{1}{2}} \sum_{l=1}^{\infty} lP_l(x)S_l(z), \quad (26)$$

where $S_l(z)$ is a Tchebycheff solution of the second kind defined by

$$S_l(z) = 2^{-1}(z^2-1)^{-\frac{1}{2}} \times \int_{-1}^{1} (z-t)^{-1}(1-t^2)^{\frac{1}{2}}U_l(t)\,dt. \quad (27)$$

Quadrature formulas applying these expansions are straightforward, and examples are given in Appendix A.

As a concluding remark on this section on bilinear expansions, we briefly comment on one other bilinear form for the Euler kernel. As shown in Bateman,[1] for x in the finite interval, $[0, a]$, and for $|z| > |x|$, an expansion of the Euler kernel may be written

$$(z-x)^{-1} = \sum_{n=0}^{\infty} \epsilon_n O_n(z)J_n(x); \qquad \epsilon_0 = 1, \quad n = 0; \quad \epsilon_n = 2, \quad n \geq 1.$$

The coefficients of the Bessel functions in the above equation are (Neumann) polynomials in z^{-1} of degree one greater than the order and are bounded for large z by an exponential form in z^2. Therefore the expansion shown above is absolutely convergent whenever $|x| < |z|$. These polynomials, however, do not satisfy Bessel's equation for arbitrary index n and therefore do not possess the same relationship to the Bessel functions as the Jacobi polynomials do to the Jacobi solutions of the second kind. Recursion relations, and integral properties are found in the above reference to Bateman. Several

examples of this type of integral are included in Appendix A.

A SPECIAL INTEGRAND

Special methods for carrying out the integration of Eq. (1) for certain integrands exist. One, due to MacRobert,[4] is given below. Consider the quantity $\zeta_q^n(z)$ to be defined as

$$\zeta_n^q(z) = z^q\varphi_n^{(\alpha,\beta)}(z) - \frac{(-1)^\alpha}{2w^{(\alpha,\beta)}(z)} \times \int_{-1}^{1} (z-t)^{-1}t^q\psi_n^{(\alpha,\beta)}(t)w^{(\alpha,\beta)}(t)\,dt. \quad (28)$$

We now may show that for $q \leq n$, $\zeta_q^n(z) = 0$. From the Euler transform relation between solutions of the first and second kind as given in Eq. (10), we may substitute for $\varphi_n^{(\alpha,\beta)}$ in the above integral to give

$$\zeta_q^n(z) = \frac{(-1)^\alpha}{2w^{(\alpha,\beta)}(z)} \int_{-1}^{1} (z-t)^{-1}(z^q - t^q) \times \psi_n^{(\alpha,\beta)}(t)w^{(\alpha,\beta)}(t)\,dt. \quad (29)$$

However, for $q \leq n$, $(z^q - t^q)/(z-t)$ is a polynomial of $q-1$ order, and hence is orthogonal to the general polynomial $\psi_n^{(\alpha,\beta)}$, in Eq. (29). Thus $\zeta_q^n = 0$, from which it follows that

$$z^q\varphi_n^{(\alpha,\beta)}(z) = \frac{(-1)^\alpha}{2w^{(\alpha,\beta)}(z)} \int_{-1}^{1} (z-t)^{-1}t^q \times \psi_n^{(\alpha,\beta)}(t)w^{(\alpha,\beta)}(t)\,dt. \quad (30)$$

Or, by taking the appropriate linear combinations of this expression, it also follows that

$$\psi_q^{(\alpha,\beta)}(z)\varphi_n^{(\alpha,\beta)}(z) = \frac{(-1)^\alpha}{2w^{(\alpha,\beta)}(z)} \times \int_{-1}^{1} (z-t)^{-1}\psi_q^{(\alpha,\beta)}(t)\psi_n^{(\alpha,\beta)}(t)w^{(\alpha,\beta)}(t)\,dt. \quad (31)$$

This result may be generalized somewhat by setting $q = n+1$. Then, by the arguments above we have

$$\zeta_{n+1}^n(z) = (-1)^\alpha 2^{-1}[w^{(\alpha,\beta)}(z)]^{-1} \times \int_{-1}^{1} t^n\psi_n^{(\alpha,\beta)}(t)w^{(\alpha,\beta)}(t)\,dt. \quad (32)$$

To evaluate this integral, we replace the polynomial by Rodrigues' formula and integrate by parts n times. We then obtain

$$\zeta_{n+1}^n(z) = (-1)^\alpha 2^{-n-1}[w^{(\alpha,\beta)}(z)]^{-1} \times \int_{-1}^{1} (1-t^2)^n w^{(\alpha,\beta)}(t)\,dt. \quad (33)$$

This integral is just the normalization integral for the Jacobi polynomials with n set equal to zero, and α replaced by $n+\alpha$ and β replaced by $n+\beta$. Therefore

$$\zeta_{n+1}^n(z) = 2^{n+\alpha+\beta}\Gamma(n+\alpha+1) \times \Gamma(n+\beta+1)/\{w^{(\alpha,\beta)}(z)\Gamma(2n+\alpha+\beta+2)\}, \quad (34)$$

and finally giving the result

$$z^{n+1}\varphi_n^{(\alpha,\beta)}(z) - \frac{(-1)^\alpha}{2w^{(\alpha,\beta)}(z)} \int_{-1}^{1} \frac{t^n\psi_n^{(\alpha,\beta)}(t)w^{(\alpha,\beta)}(t)\,dt}{(z-t)} = \frac{(-1)^\alpha 2^{n+\alpha+\beta}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{w^{(\alpha,\beta)}(z)\Gamma(2n+\alpha+\beta+2)}. \quad (35)$$

Again, after taking the appropriate linear combinations of the above equation, it may be shown that

$$\psi_{n+1}^{(\alpha,\beta)}(z)\varphi_n^{(\alpha,\beta)}(z) = (-1)^\alpha 2^{-1}[w^{(\alpha,\beta)}(z)]^{-1} \times \int_{-1}^{1} (z-t)^{-1}\psi_n^{(\alpha,\beta)}\psi_{n+1}^{(\alpha,\beta)}w^{(\alpha,\beta)}\,dt + \frac{2^{\alpha+\beta-1}(2n+\alpha+\beta+2)\Gamma(\alpha+n+1)\Gamma(\beta+n+1)(-1)^\alpha}{n!\,\Gamma(2n+\alpha+\beta+2)w^{(\alpha,\beta)}(z)}. \quad (36)$$

The convergence of these integrals for large z is shown in the case of the Legendre functions for all real indices in the reference cited above.

COMMENTS

The expansions of the Euler transform kernel and its generalizations given above represent a general technique for the reduction of the generalized Euler transform integrals as shown in Eq. (1). In each case, the quadrature is expressed as a series, convergent for all values of z not on the branch cut, of functions that form the second solution to the variation of the Jacobi equation as listed in Table I. The well-known recursion relations and asymptotic behavior of these functions are valuable aids in the numerical computation and analytical study of integrals of the Euler transform type.

ACKNOWLEDGMENTS

The author would like to thank Professor W. Byers Brown for his suggestions, and Professor C. F. Curtiss for his comments and critical reading of the manuscript.

[4] T. M. MacRobert, Proc. Glasgow Math. Assoc. 1, 10–12 (1948).

APPENDIX A

A miscellaneous collection of integrals evaluated by the procedures discussed in the text is listed below. The appropriate expansion coefficients are taken from Refs. 5 and 6. Integrals involving the general polynomial/function system defined in the text are valid for all of the special cases of the Jacobi polynomial/function set. Roman indices represent positive integers or zero; Greek indices represent numbers restricted by the requirement that their real parts be greater than -1 unless otherwise noted. The variable z is an arbitrary complex number not lying on the real axis segment -1, 1 unless otherwise noted. Integrals with a slash are principal parts.

$$\int_{-1}^{1} \frac{\psi_n^{(\alpha,\beta)}(x)w^{(\alpha,\beta)}(x)\,dx}{(a+bx^2)} = \frac{iw^{(\alpha,\beta)}}{(ab)^{\frac{1}{2}}}\left[i\left(\frac{a}{b}\right)^{\frac{1}{2}}\right]\left\{\varphi_n^{(\alpha,\beta)}\left[i\left(\frac{a}{b}\right)^{\frac{1}{2}}\right] - (-1)^{n+1}\varphi_n^{(\beta,\alpha)}\left[i\left(\frac{a}{b}\right)^{\frac{1}{2}}\right]\right\}, \tag{A1}$$

$$\frac{1}{2}\int_{-1}^{1} \frac{\psi_n^{(\alpha,\beta)}(x)\psi_{n+1}^{(\alpha,\beta)}(x)w^{(\alpha,\beta)}(x)\,dx}{(z-x)} = w^{(\alpha,\beta)}(z)(-1)^{\alpha}\psi_{n+1}^{(\alpha,\beta)}(z)\varphi_n^{(\alpha,\beta)}(z) - \frac{2^{\alpha+\beta-1}(2n+\alpha+\beta+2)\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{n!\,\Gamma(2n+\alpha+\beta+2)}, \tag{A2}$$

$$\frac{1}{2}\int_{-1}^{1} \frac{x^n\psi_n^{(\alpha,\beta)}(x)w^{(\alpha,\beta)}(x)\,dx}{(z-x)} = z^{n+1}w^{(\alpha,\beta)}(z)(-1)^{\alpha}\varphi_n^{(\alpha,\beta)}(z) - \frac{2^{n+\alpha+\beta}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(2n+\alpha+\beta+2)}, \tag{A3}$$

$$\frac{1}{2}\int_{-1}^{1} \frac{\psi_m^{(\alpha,\beta)}\psi_n^{(\alpha,\beta)}w^{(\alpha,\beta)}\,dx}{(z-x)} = (-1)^{\alpha}w^{(\alpha,\beta)}(z)\psi_m^{(\alpha,\beta)}(z)\varphi_n^{(\alpha,\beta)}(z); \quad m \le n, \tag{A4}$$

$$\int_{-1}^{1} \frac{\psi_n^{(\alpha+\gamma,\beta+\gamma)}(x)w^{(\alpha+\gamma,\beta+\gamma)}(x)\,dx}{(z-x)^{\gamma+1}} = \frac{(-1)^{\alpha}w^{(\alpha,\beta)}(z)2^{\gamma+1}\Gamma(n+\gamma+1)}{\Gamma(1+\gamma)n!}\varphi_{n+\gamma}^{(\alpha,\beta)}(z), \quad \operatorname{Re}\gamma > -1, \tag{A5}$$

$$\int_{-1}^{1} \frac{\psi_n^{(\alpha,\beta)}(x)w^{(\alpha,\beta)}(x)\,dx}{(z-x)^{m+1}} = 2(-1)^{\alpha+m}[m!]^{-1}w^{(\alpha,\beta)}(z)(z^2-1)^{-m/2}\varphi_n^{(\alpha,\beta;m)}(z), \tag{A6}$$

$$\int_{-1}^{1} \frac{e^{iax}(1-x^2)^{\gamma+\lambda-\frac{1}{2}}\,dx}{(z-x)^{\gamma+1}} = (z^2-1)^{\gamma+\lambda-\frac{1}{2}}2^{1+\lambda+2\gamma}a^{-\lambda-\gamma}\Gamma(\lambda+\gamma)/\Gamma(\gamma+1) \times \sum_{l=0}^{\infty} i^l\frac{(l+\lambda+\gamma)}{l!}\Gamma(l+\gamma+1)D_{l+\gamma}^{\lambda}(z)J_{\lambda+\gamma+l}(a), \quad \operatorname{Re}\gamma > -1; \quad \operatorname{Re}(\gamma+\lambda) > -\tfrac{1}{2}, \tag{A7}$$

$$\int_{-1}^{1} \frac{e^{iax}\,dx}{(z-x)} = \left(\frac{2\pi}{a}\right)^{\frac{1}{2}}\sum_{l=0}^{\infty} i^l(2l+1)Q_l(z)J_{l+\frac{1}{2}}(a), \tag{A8}$$

$$\int_{-1}^{1} \frac{\sin(\lambda R)}{(z-x)R}\,dx = \frac{\pi}{(ab)^{\frac{1}{2}}}\sum_{n=0}^{\infty}(2n+1)Q_n(z)J_{n+\frac{1}{2}}(a\lambda)J_{n+\frac{1}{2}}(b\lambda), \quad \text{where} \quad R = a^2+b^2-2abx, \tag{A9}$$

$$\int_{0}^{a} \frac{\cos(\lambda S)}{S(z-x)}\,dx = \tfrac{1}{2}\pi\sum_{n=0}^{\infty}\epsilon_n O_n(z)J_{\frac{1}{2}n} \times \{\tfrac{1}{2}a[\lambda+(1+\lambda^2)^{\frac{1}{2}}]\}J_{\frac{1}{2}n}\{\tfrac{1}{2}a[\lambda+(1+\lambda^2)^{\frac{1}{2}}]^{-1}\}; \quad \epsilon_n = 1, \ n = 0, \quad \epsilon_n = 2, \ n \ge 1; \quad |z| > |a|; \quad S = (a^2-x^2)^{\frac{1}{2}}. \tag{A10}$$

Properties of the Neumann polynomials are given in Ref. 1.

$$\int_{0}^{a} (a-x)^{-\frac{1}{2}}(z-x)^{-1}\,dx = \pi(\tfrac{1}{2}a)^{\frac{1}{2}}\sum_{n=0}^{\infty}\epsilon_n O_n(z)J_{\frac{1}{2}n+\frac{1}{4}}(\tfrac{1}{2}a)J_{\frac{1}{2}n-\frac{1}{4}}(\tfrac{1}{2}a) = 2(z-a)^{-\frac{1}{2}}\tan^{-1}\left[\frac{a^{\frac{1}{2}}}{(z-a)^{\frac{1}{2}}}\right], \tag{A11}$$

$$\int_{-1}^{1} (z-x)^{-1}Q_k(x)\,dx = \sum_{n=0}^{\infty}(2n+1)Q_n(z) \times [1-(-1)^{n+k}]/\{(n-k)(n+k+1)\}; \quad Q_k(x) \text{ is defined by Eq. (3)}, \tag{A12}$$

$$\int_{-1}^{1} (1-x^2)^{-\frac{1}{2}}(z-x)^{-1}(1+a^2-2ax)^{-1}\,dx = -\pi\sum_{n=0}^{\infty}\epsilon_n U_{n-1}(x)\begin{cases}\dfrac{a^n}{1-a^2}; & |a|<1\\[2ex] \dfrac{1}{a^n(a^2-1)}; & |a|>1\end{cases}. \tag{A13}$$

[5] Bateman Project Staff, *Tables of Integral Transforms* (McGraw–Hill Book Company, Inc., New York, 1953), Vol. II.

[6] W. Gröbner and N. Hofreiter, *Integraltafel*, Zweiter Teil, (Springer-Verlag, Berlin, 1961), 2nd ed.

APPENDIX B

The Jacobi equation is derived from the hypergeometric equation with three regular singular points at 0, 1, and ∞ by an appropriate coordinate change and linear combination of parameters such that the new equation has its singular points at ± 1 and ∞. Written in standard from, the Jacobi equation is

$$\left\{(1-x^2)\frac{d^2}{dx^2} + [\beta - \alpha - (\alpha+\beta+2)x]\frac{d}{dx} + n(n+\alpha+\beta+1)\right\}y(x) = 0, \quad \text{(B1)}$$

where the parameters α and β have real parts > -1, and the index n is taken as integer. In this case, one of the two independent solutions of the above equation is a polynomial, regular at ± 1 and having a simple pole of order n at infinity; the other solution is a function, regular at infinity and single-valued if a branch cut is made on the Riemann sheet between the branch points -1 and 1. An integral representation of the general solution of the above equation may be had by means of a generalized Euler transform

$$y(x) = \oint_C (x-t)^{+\mu} v(t)\, dt, \quad \text{(B2)}$$

where μ is a parameter to be fixed, and the contour C will determine what linear combination of the two independent solutions $y(x)$ represents. Let L_x stand for the operator in Eq. (B1), then there exists an operator A such that

$$L_x(x-t)^{+\mu} = A_t(x-t)^{+\mu}, \quad \text{(B3)}$$

where A operates on the variable t and is a linear operator with the same type of coefficients for its second and first derivatives as appear in the operator L. These coefficients may readily be determined by a Taylor expansion about t of the correspondent coefficient of the operator L in the above equation. We find that

$$A_t = (1-t^2)\frac{d^2}{dt^2} + [\alpha - \beta + (\alpha+\beta+2\mu)t]\frac{d}{dt} + n(n+\alpha+\beta+1) - \mu(\mu+\alpha+\beta+1). \quad \text{(B4)}$$

The adjoint operator $\tilde{A}$ is determined from the above by Green's theorem and turns out to be

$$\tilde{A}_t = \frac{d^2}{dt^2}(1-t^2) - \frac{d}{dt}[\alpha - \beta + (\alpha+\beta+2\mu)t] + n(n+\alpha+\beta+1) - \mu(\mu+\alpha+\beta+1). \quad \text{(B5)}$$

Now all of the above equations can be combined to give the following sequence of results:

$$\begin{aligned} 0 = L_x y &= \oint_C L_x(x-t)^{+\mu} v(t)\, dt \\ &= \oint_C A_t(x-t)^{+\mu} v(t)\, dt \qquad \text{(B6)} \\ &= \oint_C (x-t)^{+\mu}\tilde{A}_t v(t)\, dt + \oint_C \frac{d}{dt}\left\{(1-t^2)\right. \\ &\quad \left.\times\left[v(t)\frac{d}{dt}(x-t)^{-\mu} - (x-t)^{+\mu}\frac{d}{dt}v(t)\right]\right\} dt. \end{aligned}$$

Now any contour that forms a complete circuit in the Riemann sheet or any open contour that begins and ends in a zero of the integrand of the second term (the bilinear concomitant) will cause that term to vanish. Since $v(t)$ is as yet arbitrary, the satisfaction of the resulting equation demands that

$$\begin{aligned} \tilde{A}_t v(t) &= \frac{d^2}{dt^2}[(1-t^2)v(t)] \\ &\quad - \frac{d}{dt}\{[\alpha - \beta - (\alpha+\beta+2\mu)t]v(t)\} \\ &\quad + \{n(n+\alpha+\beta+1) \\ &\quad - \mu(\mu+\alpha+\beta+1)\}v(t) = 0. \qquad \text{(B7)} \end{aligned}$$

We have, as yet, not chosen the parameter μ. Quadrature of the above equation follows immediately if the last term vanishes; therefore it follows that

$$\mu = \begin{Bmatrix} n \\ -n-\alpha-\beta-1 \end{Bmatrix}.$$

If $v(t)$ is not to have an essential singularity at ∞, the constant of the first integration must be taken to be zero, therefore, the second integration can be readily performed to give

$$v(t) = A(1+t)^{n+\alpha}(1-t)^{n+\beta}. \quad \text{(B8)}$$

Hence, the general solution to the Jacobi equation may be written

$$y(x) = A\oint_C (x-t)^{-n-\alpha-\beta-1} \times (1+t)^{n+\alpha}(1-t)^{n+\beta}\, dt, \quad \text{(B9)}$$

where we have made the choice $\mu = -n - \alpha -$

$\beta - 1$, and A is an integration constant. The combination of fundamental solutions that $y(x)$ represents is determined by the choice of the contour in the above representation. It may be shown that if a simple loop enclosing the point x on the line segment $[-1, 1]$ and 1 is selected, with the branch cut made from -1 to $-\infty$, $y(x)$ is proportional to the Jacobi polynomial, $P_n^{(\alpha,\beta)}(x)$; and on the other hand, if the contour is chosen to be the real axis segment from -1 to 1, the second solution, $Q_n^{(\alpha,\beta)}(x)$ is represented if x does not lie on the real axis segment. The proportionality constant in both cases is the integration constant A, and turns out to be 2^{-n-1} as determined by comparison with the hypergeometric series solution of the same equation.

Orbits in a Magnetic Universe

M. A. MELVIN* AND J. S. WALLINGFORD
Physics Department, Florida State University, Tallahassee, Florida
(Received 22 June 1965)

A cylindrically symmetric parallel bundle of magnetic lines of force, in equilibrium under their mutual gravitational attraction ("magnetic universe"), has recently received attention. While a Newtonian analysis suggests that the equilibrium is unstable, the complete general relativity analysis shows that the equilibrium is stable. This discrepancy may have to do with the unusually slow falloff of the gravitational field at large distances in this geometry. In order to understand the gravitational field of the static magnetic universe somewhat better, we have studied its timelike and lightlike geodesics (i.e., the orbits in it of electromagnetically neutral test particles with unit or zero rest mass). Since the density of magnetic flux—and energy and stress and, therefore, "gravitating mass"—is approximately uniform in the vicinity of the axis, the motion of test particles there is like that in a Newtonian simple harmonic oscillator field. "Vicinity" here means within a small fraction ρ of the range radius $\bar{a} = (6.96/B_0) \times 10^{24}$ cm (B_0 is the magnetic field on the axis measured in gauss). As is to be expected from the universality of the angular frequency ω_0 in the harmonic oscillator field and the relation: orbital velocity $\cong \omega_0\rho$, no motion can get too far from the axis. Otherwise the physical orbital velocity would exceed the speed of light. It is in this way that the strength of the attractive field, though it does not remain strictly of the harmonic oscillator type as one proceeds outward, implies that there is a critical straddling radius $\rho = 1/\sqrt{3}$. Circular or circular helical light tracks occur only at the critical radius, and with $B_0 = 10^5$ G, the time required for light to circumnavigate the critical circle is about 200 years. The cylinder marked out by this radius plays a unique limiting role: All particles, whether of zero or nonzero mass, and no matter what their initial positions and velocities (except in the one singular subcase of light tracks parallel to the cylindrical axis), must have their orbits lying wholly or partially within the cylindrical region $\rho < 1/\sqrt{3}$; hence the use of the adjective "straddling." Constants of motion which correspond closely to ζ-component linear momentum, angular momentum, and energy in Newtonian mechanics are defined. Bounds are placed on these dynamical constants and on the apsidal radii by the requirement that the range of motion be real. Finally, the magnetic universe is complete in the sense that "no news can enter or leave"—all orbits are of infinite duration.

I. INTRODUCTION, MOTIVATION, AND SUMMARY

GRAVITATIONAL collapse has recently received increased attention both as a process posing fundamental issues of principle[1] and as a mechanism conceivably primarily responsible for the $\sim 10^{61}$ ergs of energy output of quasistellar sources.[2] In connection with the study of collapse, it is important to distinguish between energy-carrying systems which are unstable against gravitational infall and those which are stable.

Among configurations which are in static equilibrium under their own gravitational attraction, one of the simplest and most interesting is a parallel bundle comprising some specified number of magnetic lines of force, Φ.[3,4] The density B of lines of

* This work was performed in part at the Oak Ridge National Laboratory.

[1] See, e.g., J. A. Wheeler, "Geometrodynamics and the Issue of the Final State," a chapter in *Relativity, Groups and Topology*, edited by C. DeWitt and B. S. DeWitt, (Gordon and Breach, New York, 1964); and *Gravitation Theory and Gravitational Collapse* (Proceedings of the December 1963 Dallas International Conference on Gravitational Collapse) (University of Chicago Press, Chicago, 1965), Vol. II.

[2] See Vol. I of the Proceedings cited in Ref. 1.

[3] M. A. Melvin, Phys. Letters **8**, 65 (1964). See also W. B. Bonnor, Proc. Phys. Soc. (London) **67A**, 225 (1953); M. Misra and L. Radhakrishna, Proc. Natl. Inst. Sci. India **28A**, 632 (1962).

[4] M. A. Melvin, "Dynamics of Cylindrical Electromagnetic Universes," ORNL-3758 (April 1965); Phys. Rev. **139**, B225 (1965).

$\beta - 1$, and A is an integration constant. The combination of fundamental solutions that $y(x)$ represents is determined by the choice of the contour in the above representation. It may be shown that if a simple loop enclosing the point x on the line segment $[-1, 1]$ and 1 is selected, with the branch cut made from -1 to $-\infty$, $y(x)$ is proportional to the Jacobi polynomial, $P_n^{(\alpha,\beta)}(x)$; and on the other hand, if the contour is chosen to be the real axis segment from -1 to 1, the second solution, $Q_n^{(\alpha,\beta)}(x)$ is represented if x does not lie on the real axis segment. The proportionality constant in both cases is the integration constant A, and turns out to be 2^{-n-1} as determined by comparison with the hypergeometric series solution of the same equation.

Orbits in a Magnetic Universe

M. A. MELVIN* AND J. S. WALLINGFORD
Physics Department, Florida State University, Tallahassee, Florida
(Received 22 June 1965)

A cylindrically symmetric parallel bundle of magnetic lines of force, in equilibrium under their mutual gravitational attraction ("magnetic universe"), has recently received attention. While a Newtonian analysis suggests that the equilibrium is unstable, the complete general relativity analysis shows that the equilibrium is stable. This discrepancy may have to do with the unusually slow falloff of the gravitational field at large distances in this geometry. In order to understand the gravitational field of the static magnetic universe somewhat better, we have studied its timelike and lightlike geodesics (i.e., the orbits in it of electromagnetically neutral test particles with unit or zero rest mass). Since the density of magnetic flux—and energy and stress and, therefore, "gravitating mass"—is approximately uniform in the vicinity of the axis, the motion of test particles there is like that in a Newtonian simple harmonic oscillator field. "Vicinity" here means within a small fraction ρ of the range radius $\bar{a} = (6.96/B_0) \times 10^{24}$ cm (B_0 is the magnetic field on the axis measured in gauss). As is to be expected from the universality of the angular frequency ω_0 in the harmonic oscillator field and the relation: orbital velocity $\cong \omega_0\rho$, no motion can get too far from the axis. Otherwise the physical orbital velocity would exceed the speed of light. It is in this way that the strength of the attractive field, though it does not remain strictly of the harmonic oscillator type as one proceeds outward, implies that there is a critical straddling radius $\rho = 1/\sqrt{3}$. Circular or circular helical light tracks occur only at the critical radius, and with $B_0 = 10^5$ G, the time required for light to circumnavigate the critical circle is about 200 years. The cylinder marked out by this radius plays a unique limiting role: All particles, whether of zero or nonzero mass, and no matter what their initial positions and velocities (except in the one singular subcase of light tracks parallel to the cylindrical axis), must have their orbits lying wholly or partially within the cylindrical region $\rho < 1/\sqrt{3}$; hence the use of the adjective "straddling." Constants of motion which correspond closely to ζ-component linear momentum, angular momentum, and energy in Newtonian mechanics are defined. Bounds are placed on these dynamical constants and on the apsidal radii by the requirement that the range of motion be real. Finally, the magnetic universe is complete in the sense that "no news can enter or leave"—all orbits are of infinite duration.

I. INTRODUCTION, MOTIVATION, AND SUMMARY

GRAVITATIONAL collapse has recently received increased attention both as a process posing fundamental issues of principle[1] and as a mechanism conceivably primarily responsible for the $\sim 10^{61}$ ergs of energy output of quasistellar sources.[2] In connection with the study of collapse, it is important to distinguish between energy-carrying systems which are unstable against gravitational infall and those which are stable.

Among configurations which are in static equilibrium under their own gravitational attraction, one of the simplest and most interesting is a parallel bundle comprising some specified number of magnetic lines of force, Φ.[3,4] The density B of lines of

* This work was performed in part at the Oak Ridge National Laboratory.

[1] See, e.g., J. A. Wheeler, "Geometrodynamics and the Issue of the Final State," a chapter in *Relativity, Groups and Topology*, edited by C. DeWitt and B. S. DeWitt, (Gordon and Breach, New York, 1964); and *Gravitation Theory and Gravitational Collapse* (Proceedings of the December 1963 Dallas International Conference on Gravitational Collapse) (University of Chicago Press, Chicago, 1965), Vol. II.

[2] See Vol. I of the Proceedings cited in Ref. 1.

[3] M. A. Melvin, Phys. Letters **8**, 65 (1964). See also W. B. Bonnor, Proc. Phys. Soc. (London) **67A**, 225 (1953); M. Misra and L. Radhakrishna, Proc. Natl. Inst. Sci. India **28A**, 632 (1962).

[4] M. A. Melvin, "Dynamics of Cylindrical Electromagnetic Universes," ORNL-3758 (April 1965); Phys. Rev. **139**, B225 (1965).

force is finite at the center of the configuration and falls off with radial coordinate r (axis of rotational symmetry at $r = 0$!) according to the law

$$B = B_0[1 + (r/\bar{a})^2]^{-2}.$$

Here B_0, the magnetic flux density at the axis, and $\bar{a}$, the "range radius," are related to each other by an equation which results, in the process of finding the equilibrium solution, from identifying the energy and stress density as a gravitational source density ("gravitating mass")[5]:

$$B_0\bar{a} = 2c^2/G^{\frac{1}{2}} = 6.96 \times 10^{24}(\text{G-cm}).$$

The physical significance of $\bar{a}$ is discussed later in this paper (Sec. II). For a magnetic flux $B_0 \sim 10^5$ G, of the order of the polar values observed in the most strongly magnetic stars, $\bar{a}$ is about one million times the diameter of the earth's orbit about the sun. The time unit for the static magnetic universe is

$$\bar{a}/c = 2.32 \times 10^{14}(\text{G}/B_0) \text{ sec}.$$

For an axis magnetic flux $B_0 \sim 10^5$ G, $\bar{a}/c$ is about 75 years.

Associated with the equilibrium magnetic field distribution is a well-determined geometry, whose symmetry is also that of a whole cylinder. The field distribution together with its associated geometry is denoted here for brevity as a "magnetic universe." The magnetic field configuration has the following remarkable property: Although a suitable Newtonian analysis and that given by general relativity yield identically the same equilibrium configuration, *the conclusions as to stability are opposite.* According to a Newtonian analysis, the equilibrium is *unstable* against gravitational collapse[6]; according to the full general relativity analysis by various methods, it is stable.[4,7] Though this striking difference is almost certainly connected with the non-asymptotically-flat geometry associated with the magnetic universe according to general relativity,[4] the situation cannot be said to have been fully clarified. Therefore it is of interest to examine the gravitational field associated with the bundle of magnetic field energy as a step toward what one can hope will eventually be a full illumination of the reason for the difference. In principle, the geodesics, or tracks of electrically neutral test particles moving in the given gravitational field, tell all that one can know about the geometry associated with this field. To analyze these geodesics is the object of this paper.

The results of the investigation are easily summarized. In broad survey, one can say that *there is no escape from the gravitational attraction of the parallel bundle of magnetic field energy.* The gravitational attraction here falls off more slowly with distance than in the case of a spherically symmetric center of attraction (Schwarzschild solution). One is reminded of how the electrical attraction of an infinite line charge falls off with distance to one power lower than does the electrical attraction of a point charge. As a consequence of this reach of the gravitational field to great distances, a particle with an angular momentum of any nonzero value whatsoever is necessarily bound to orbit periodically around the axis of symmetry an infinite number of times. This behavior is in contrast to that of a particle of a given velocity passing a *spherically symmetric* center of attraction at a distance b. In that case, when b exceeds a critical limit, dependent upon velocity, the particle escapes on a trajectory which asymptotically approaches a straight line. It does not execute periodic turns. But *in the case of the magnetic universe, a particle with a nonzero angular momentum is never able to get away on an asymptotically straight trajectory.*

The slow falloff of the gravitational field with distance is manifest also in another simple and interesting consequence. Consider a particle traveling in a circular orbit about the axis of symmetry. When the orbit is small, it lies in the region where the effective potential of the attraction follows the harmonic oscillator law. The circular motion is characterized by a certain circular frequency ω. This frequency is fully specified by the axis flux density (or total flux, or range radius). It is $\sqrt{2}$ in magnetic universe units or in standard units

$$\omega = (\tfrac{1}{2}G)^{\frac{1}{2}}B_0/c = 6.09 \times 10^{-15}(B_0/\text{G}) \text{ rad/sec}.$$

The associated period, $2\pi/\omega = \sqrt{2}\pi$ in magnetic universe units, equals 1.032×10^{15} (G/B_0) sec in ordinary units. (When $B_0 \sim 10^5$ G, this is about 327 years.) The velocity v required for motion in an orbit of coordinate radius r increases for small r according to the law $v = \omega r$. For large orbits, the velocity needed to balance the attraction increases even more rapidly with r, and reaches at

$$\rho = \rho_{\text{straddle}} = 1/\sqrt{3}$$

$$r = r_{\text{straddle}} = 4.02 \times 10^{24}(\text{G}/B_0) \text{ cm}$$

a limit which cannot be surpassed: the speed of light.

[5] Reference 3, Eq. (8).

[6] This was demonstrated by an argument of J. A. Wheeler in *Relativity, Groups and Topology* (Ref. 1) following some early discussion with one of us concerning the magnetic universe.

[7] K. S. Thorne, Phys. Rev. **139**, B244 (1965).

There is no possibility for a neutral test particle to move in a circular orbit of radius r greater than r_{straddle}. *Only zero-rest-mass particles may orbit circularly or helically at the radius* $r = r_{\text{straddle}}$.

The period of circular motion depends only on the radius and decreases steadily as one takes circles closer and closer to the critical circle (i.e., to the orbit which is allowed only for light). The physical time period required by light to traverse this critical circular orbit is $\sqrt{3}\pi/2$ in magnetic universe units or

$$\tfrac{1}{2}\sqrt{3}\pi\,\frac{\bar{a}}{c} = \tfrac{1}{2}\sqrt{3}\pi\,\frac{6.96\times10^{24}}{3\times10^{10}B_0} = 6.31\times10^{14}\left(\frac{\text{G}}{B_0}\right)\text{ sec.}$$

With $B_0 = 10^5$ G, of the order of the polar values observed in the magnetically strongest stars, the time for light to circumnavigate the magnetic universe at the critical radius comes out nearly exactly 200 years.

The quantity r_{straddle} measures the pulling power of the long-range gravitational attraction in a more general sense: Every orbit which is *not* circular (or helical with $r = r_{\text{straddle}}$) and which is not a straight line parallel to the z axis must lie at least partially within r_{straddle}.

In addition to displaying so clearly the slow falloff of the gravitational field with distance, the geodesics also establish the "completeness," in the mathematical sense, of the geometry associated with this gravitational field. Modern developments in differential geometry have raised interest in the question of "completeness" for any given Riemannian manifold. Kundt,[8] following earlier more abstract discussions of Avez[9] and Lichnerowicz,[10] has suggested four types of completeness relevant to physics in the sense of determining whether or not "news can enter or leave" the space–time. The most restrictive type, g completeness, requires that all geodesics (considered as one-dimensional submanifolds) are either closed or of infinite length in both directions (doubly infinite). The remaining types—g_t, g_l, and g_s completeness—are less restrictive classes, referring to space–times with doubly infinite lightlike, timelike, and spacelike geodesics, respectively; the lightlike geodesics are described with the help of an affine parameter such as the coordinate time t.

To determine g_t or g_l completeness, we need only investigate geodesics lying in a meridian plane, in short "meridian geodesics," in that these are the cases in which the largest range is attained. In these cases the angular momentum about the symmetry axis is zero. The explicit integration yields, in the case of timelike orbits of this type, the time as a first-kind elliptic integral of the radial variable. Just as with the simple pendulum, the time goes to infinity as the radial variable oscillates back and forth progressively. For meridianal lightlike orbits the situation is even simpler. The increment in coordinate time is equal always to the absolute value of the increment in the radial variable (in our units, in which the light velocity equals unity). Thus the duration of all orbits is infinite—no news can enter or leave the magnetic universe—or the space–time is complete.

The details of the analysis of equations and constants of motion appear in the following section. The equations of motion are obtained via the Hamilton–Jacobi formalism. The formalism shows particularly simply how the invariances of the metric with respect to shifts in time t, in the coordinate z parallel to the symmetry axis, and in the azimuthal angle ϕ are responsible for the existence of a conserved energy, z-component linear momentum, and angular momentum, respectively. In the Appendix, the complete integrals of the equations of motion are obtained in the form of elliptic integrals.

II. GENERAL MATHEMATICAL ANALYSIS

II.1 Geometry and Magnetic Properties of the Static Magnetic Universe

The geometry of the cylindrical magnetic universe is given by

$$d\sigma^2 = (1+\rho^2)^2(d\tau^2 - d\rho^2 - d\zeta^2) - (1+\rho^2)^{-2}\rho^2\,d\phi^2. \qquad (1)$$

Here the coordinates have been made dimensionless through the introduction of an appropriate unit of length for which we choose the range radius $\bar{a}$ of the magnetic universe[3]; thus,

$$\text{coordinate for}\begin{Bmatrix}(c\cdot\text{time})\\(\text{translation parallel to symmetry axis})\\(\text{radius})\end{Bmatrix} = \bar{a}\begin{Bmatrix}\tau\\\zeta\\\rho\end{Bmatrix}.$$

The associated Newtonian gravitational potential for the metric is

$$\psi = \tfrac{1}{2}\ln g_{44} = \ln(1+\rho^2)$$

and the range radius $r = \bar{a}$ or $\rho = 1$ corresponds to the place where the "acceleration of gravity," $d\psi/d\rho = 2\rho/(1+\rho^2)$, is a maximum. (There is an analogy here with the gravitational definition of the radius of the earth.)

[8] W. Kundt, Z. Physik **172**, 488 (1963).
[9] A. Avez, Compt. Rend. **240**, 485 (1955).
[10] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme*, p. 126, (Masson, Paris, 1955).

That the length $\bar{a}$ plays a role as an "effective radius" in several other respects is seen as follows: We have,[11] in terms of the central value of the magnetic field B_0,

$$B_0\bar{a} = (f/\pi G)^{\frac{1}{2}}c^2 = f^{\frac{1}{2}} \times 1.963 \times 10^{24} \text{ (flux density units} \times \text{cm)}$$

($f = 1$ for rationalized units, $f = 4\pi$ for unrationalized units) or

$$\bar{a} = 2c^2/G^{\frac{1}{2}}B_0 = (6.96 \times 10^{24}/B_0) \text{ cm} \qquad (B_0 \text{ in G}).$$

The integral of the magnetic mass density over the equilibrium distribution[12] endows the equation for $B_0\bar{a}$ with a simple physical meaning. Realizing that the energy or stress density at the axis equals $B_0^2/2f$, we have

$$\text{(axis energy density)} \times \frac{\text{area of range radius circle}}{\text{light velocity squared}}$$

$$= \pi\bar{a}^2B_0^2/2fc^2 = c^2/2G = \text{total magnetic mass}$$

$$= 6.738 \times 10^{27} \text{ g/cm}.$$

Thus $\bar{a}$ plays the role of an *effective radius for the magnetic-mass distribution.*

A second equation, relating the three characterizing quantities B_0, $\bar{a}$, and the total flux Φ, is obtained by integrating the flux density over the entire space. One finds[13] $\Phi = \pi\bar{a}^2B_0$. Here again we see $\bar{a}$ playing the role of an "effective radius," this time for the flux distribution. One can then express any two of the three quantities in terms of the third with only universal physical constants appearing. For instance, we have $B_0\Phi = fc^4/G$ or

$$\Phi = 1.522 \times 10^{50}(\text{G}/B_0) \text{ Mx}.$$

The length unit is connected with the total flux Φ through the relation

$$\bar{a} = \Phi G^{\frac{1}{2}}/2\pi c^2 = 4.56 \times 10^{-26}\,\Phi \text{ cm} \qquad (\Phi \text{ in Gaussian units} \equiv \text{maxwells}).$$

II.2 Constants of Motion and the Radial First Integral by Hamilton–Jacobi Theory

We originally determined the geodesics and constants of motion for the geometry [Eq. (1)] and related results by solving the usual geodesic equations of motion. However, J. A. Wheeler pointed out that one can arrive at the same results more elegantly by the Hamilton–Jacobi method.[14] More descriptively perhaps, this may be called the method of "constructive interference of ideal wave trains." Each "wave" is characterized by a phase proportional to $S = S(\tau, \rho, \zeta, \phi; E, P, L)$.

This phase depends upon position and time and in addition upon constants of integration (E, P, L), which, as we shall see, can be variously interpreted as energy, z-component linear momentum, and angular momentum or as frequency, z-component wave-number, and azimuthal index number. [An additive constant $\Delta(E, P, L)$ in S is always allowed because the equation for the phase—the Hamilton–Jacobi equation—involves only the derivatives of the phase with respect to the coordinates and not the phase itself.]

To determine the motion explicitly involves two stages. Stage 1: We have to determine S as a function of space, time, and the constants of motion. Stage 2: We have to determine the explicit functional forms of the first integrals (i.e., relations between the constants of motion and the coordinates and their first time derivatives) so that by quadratures alone ρ, ζ, and ϕ may be determined as functions of τ (the "orbit"), or ρ and ζ as functions of ϕ (the "trajectory").

In this section we concern ourselves with the first stage. In Sec. A of the Appendix, we shall see how the second stage is accomplished by an interesting equivalent of the Jacobi procedure to be found in Wheeler and Power.[14]

We turn then to the determination of the function S. As behooves a function which is, after all, the phase in a wave motion in the limit of short wavelength, S satisfies the Hamilton–Jacobi equation for a particle of rest mass M,

$$(\text{momentum})^2 - (\text{energy}/c)^2 + (Mc)^2 = 0; \qquad (2)$$

or

$$-g^{\alpha\beta}(\partial S/\partial x^\alpha)(\partial S/\partial x^\beta) + (Mc)^2 = 0.$$

In the present case we use such units, and a test particle of such a mass, that in our coordinates the constant Mc can be replaced by unity or zero; thus

$$(1+\rho^2)^{-2}\left[-\left(\frac{\partial S}{\partial \tau}\right)^2 + \left(\frac{\partial S}{\partial \rho}\right)^2 + \left(\frac{\partial S}{\partial \zeta}\right)^2\right] + (1+\rho^2)^2\rho^{-2}\left(\frac{\partial S}{\partial \phi}\right)^2 + \epsilon = 0, \qquad (3)$$

[11] Reference 3, Eq. (8).

[12] Reference 4, Sec. 1.3, last equation; if Ref. 3, Eq. (19), is used instead, the extra factor $\frac{1}{2}$ occurs because the integrated local Lorentz-frame energy densities have been used instead of the integrated "conserved" magnetic energy density.

[13] Reference 3, Eq. (18).

[14] L. D. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1962) 2d ed., p. 285; E. Power and J. A. Wheeler, Rev. Mod. Phys. **29**, 480 (1957), Appendix.

where $\epsilon = +1, 0$ for timelike and lightlike orbits, respectively.

We look for a solution by the method of separation of variables. Here S is expressed as the sum of four terms, each depending upon only one of the four coordinates. The absence of any explicit dependence upon τ, ζ, and ϕ in Eq. (3) then says that $\partial S/\partial\tau$, $\partial S/\partial\zeta$, and $\partial S/\partial\phi$ are all constants. Thus one has

$$S = \int R(\rho)\, d\rho + P\zeta + L\phi - E\tau. \qquad (3')$$

Here (see Appendix, Sec. A) P is a measure of the linear momentum; L is a measure of the angular momentum, in each case parallel to the symmetry axis; and E is a measure of the energy. In terms of these constants of integration, the Hamilton–Jacobi equation (3) yields for the measure of radial momentum

$$R(\rho) = [E^2 - P^2 - (1 + \rho^2)^4 L^2/\rho^2 - \epsilon(1 + \rho^2)^2]^{\frac{1}{2}}. \qquad (4)$$

II.3 Bounds on the Radial Motion

Most of the significant features of the motion follow at once from Eq. (4):

First, the radial motion depends upon the energy factor E and the momentum factor P only in the combination $U^2 = E^2 - P^2$. Therefore it is appropriate to call U the "transverse energy factor" and write

$$R(\rho) = [U^2 - (1 + \rho^2)^4 L^2/\rho^2 - \epsilon(1 + \rho^2)^2]^{\frac{1}{2}}. \qquad (4')$$

We have $U^2 = 0$ only for a photon traveling parallel to the ζ axis, and $U^2 \geq 1$ for every nonzero rest mass particle, where $U^2 = 1$ only when the particle is moving parallel to the ζ axis (actually, as we shall see, *on* the ζ axis). For circular or helical motion about the symmetry axis, the radius ρ is a constant determined by setting $R(\rho) = 0$. Motion in a meridian plane is given by setting $L = 0$. These two special classes of orbits—helical and meridian plane—intersect in the limiting case of motion parallel to the ζ axis; this can occur only if the particle happens to be initially directed parallel to the ζ axis at the distance

$$\rho_0 = [(U/\epsilon^{\frac{1}{2}}) - 1]^{\frac{1}{2}}.$$

The only possible value of ρ_0 for a nonzero rest mass particle is zero. For a zero rest mass particle $U/\epsilon^{\frac{1}{2}}$ is indeterminate and parallel motion at any distance from the axis can occur.

Second, in a general orbit, all values of the radius vector ρ satisfy the inequality

$$\rho \leq [(U/\epsilon^{\frac{1}{2}}) - 1]^{\frac{1}{2}},$$

where the upper limit for ρ is attained only in a meridian plane orbit ($L = 0$) at the turning point ρ_0. Thus when there is no angular momentum ($L = 0$), and we are not dealing with a zero rest mass particle, the radial coordinate oscillates through zero out to the value

$$\rho_0 = (U - 1)^{\frac{1}{2}}.$$

For a zero rest mass particle traveling in a meridian plane there are no turning points. All such tracks are "straight" as viewed in the canonical coordinate system. This is a peculiar simplification in the description of motions due to the use of canonical coordinates. In this connection Thorne[7], who has independently discussed certain special cases of the geodesics which we have here considered quite generally, has made the following interesting remarks:

"The only geodesic of constant (ρ, ϕ) is the null geodesic $\zeta = \pm\tau$.* It is strange that, although a photon moving in the plane perpendicular to the axis of symmetry (ζ = constant) is strongly deflected by the mass inside its orbit, a photon moving parallel to the axis of symmetry (ρ, ϕ constant) is not deflected at all."

To make one feel more comfortable about this apparent paradox Thorne then argues essentially as follows: The static magnetic universe is indeed invariant under translations by a given $\Delta\zeta$ along the ζ direction. But such invariant translations mean translating the universe at every value of ρ by $(1 + \rho^2)$ times the proper distance of the translation on the symmetry axis [because $g_{\zeta\zeta} = (1 + \rho^2)^2$]. Hence an invariant translation of the static magnetic universe "is more like a rotation of Euclidean space than like a translation of it; and ζ is more like an angular coordinate of Euclidean space than like a rectilinear coordinate. Just as it requires mass to "deflect a photon into circular motion in Euclidean space, so it takes mass to 'deflect' one into motion along a ζ-coordinate line in Melvin's universe. 'Undeflected motion' would correspond to some path other than ζ = constant."

Third, when there is angular momentum of given amount L, the minimum allowable value for U^2 equals the minimum value of the positive-definite quantity

$$V^2(\rho^2) = (1 + \rho^2)^4 L^2/\rho^2 + \epsilon(1 + \rho^2)^2.$$

* Here Thorne's statement is not quite literally correct. Possible non-null geodesics parallel to the symmetry axis are of course, those on the axis itself—with arbitrary energy and momentum ($U = 1$).

When U^2 has its minimum value, $R(\rho) = 0$ and there is no radial oscillation, the particle moves in a circular orbit of a certain radius, $\rho = \rho_L$. Here ρ_L is that value of ρ which minimizes $V^2(\rho^2)$. Calculating $dV^2/d\rho^2$, and annulling this derivative, one finds an equation which is cubic in ρ_L^2:

$$3\rho_L^6 + 5\rho_L^4 + \rho_L^2 - 1 + 2\epsilon L^{-2}\rho_L^4 = 0. \quad (5)$$

Rather than try to solve for ρ_L^2 for a specified L, it is easier to calculate L for a given ρ_L; thus we have

$$L^2 = \frac{2\epsilon\rho_L^4}{(1 + \rho_L^2)^2(1 - 3\rho_L^2)}, \qquad U_{\min}^2 = \frac{\epsilon(1 + \rho_L^2)^2(1 - \rho_L^2)}{1 - 3\rho_L^2}. \quad (6)$$

Clearly, physically acceptable values of ρ_L lie only between 0 and $1/\sqrt{3}$, the latter value occurring only for a zero rest mass particle.

When U^2 is larger than $U_{\min}^2 = V_{\min}^2$, the particle has turning points ρ_1 and ρ_2 which *straddle* ρ_L. That this is so may be seen simply from the following argument: Writing $\rho^2 \equiv x$, we have

$$U_{\min}^2 = [P(x)/x]_{\min} \equiv P(x_L)/x_L$$
$$[P(x) \equiv (1 + x)^4 L^2 + \epsilon x(1 + x)^2]. \quad (7)$$

We have only to verify that x_L lies between the roots of $P(x)/x = U^2$ for any value of U^2 larger than $U_{\min}^2$. The polynomial $P(x)$ and all its derivatives are positive-definite in the physically acceptable domain $\rho > 0$; the graph of $P(x)$ vs x is then necessarily concave upward and increasing with increasing x. The intersections of $P(x)$ with the straight line U^2x for $U^2 > U_{\min}^2$ necessarily lie on both sides of the point of tangency of $P(x)$ with U^2x; but this point of tangency occurs for $x = x_L$, $U^2 = U_{\min}^2$, as is evident from comparing the minimum condition

$$x_L P'(x_L) - P(x_L) = 0$$

and the tangency condition

$$P(x) = U^2 x = P'(x)x.$$

The *largest permissible radius* for a circular orbit, expressed in the present units, is $\rho_L = 1/\sqrt{3}$, and this is the critical "straddle" radius for all orbits. We see this by referring back to Eq. (5) and noting that the largest value of ρ_L that can occur for various values of the parameter $2\epsilon/L^2$ is an upper bound for the inner apse of any orbit. (The smallest value of ρ_L is a lower bound for the outer apse.) The largest value of ρ_L occurs for minimum $2\epsilon/L^2$; ϵ/L^2 has its minimum value zero for a light track. We then find for the roots of Eq. (5), $1/\sqrt{3}$, -1, -1, of which the only physically meaningful root is the first: $\rho_{L\,\max} = 1/\sqrt{3}$. *Thus we have the remarkable result that the circular helical geodesics for light, which occur at the critical radius* $\rho = 1/\sqrt{3}$ ($\psi = \ln \frac{4}{3}$), *play a unique limiting role: No matter what the angular momentum* L *or* ζ *component of linear momentum* $(E^2 - U^2)^{\frac{1}{2}}$ *of any particle—whether of zero or nonzero rest mass—this particle must "fall" in its orbit to within the critical radius (except when* $L = U = 0$).

II.4 Limiting Values for Orbital Velocity and Period of Motion

For a *circular orbit of radius* ρ *small compared with* $1/\sqrt{3}$, the denominator in Eq. (6) approaches unity. The angular momentum approaches

$$L \cong 2^{\frac{1}{2}}\rho_L^2 \text{ (particle of nonzero rest mass!)}.$$

From Eq. (4′) we see that the energy squared then approaches $1 + 2\rho_L^2$ ($\epsilon = 1$). Thus the orbital velocity approaches $\sqrt{2}\rho_L$. The angular velocity, given by the quotient of orbital velocity by radius, approaches a constant value, as is to be expected in a harmonic oscillator field. This value, $\omega = \sqrt{2}$, translated to physical units, gives the frequency listed in Sec. I of this report.

We now consider the case of a *circular orbit of radius* ρ *not small compared with* $1/\sqrt{3}$. Upon substituting the expression for L^2 in Eq. (4′) we find under the conditions for a circular orbit ($R = 0$, $P = 0$)

$$U_{\min}^2 = E_{\min}^2 = \epsilon[(1 - \rho_L^2)/(1 - 3\rho_L^2)](1 + \rho_L^2)^2,$$
$$\phi_r = (1 + \rho^2)^2[2/(1 - \rho^2)]^{\frac{1}{2}}.$$

We see that though the angular momentum and the energy required to maintain a nonzero rest mass particle in a circular orbit go to infinity as the radius approaches the critical radius, the frequency or period, which depends on the *ratio* of angular momentum to energy, remains finite. The period depends only on the radius and decreases steadily as one takes circles closer and closer to the critical circle. There the period has its least value

P_{least} :

$$2\pi \frac{[(1 - \rho^2)/2]^{\frac{1}{2}}}{(1 + \rho^2)^2}\bigg|_{\rho=1/\sqrt{3}} = \tfrac{8}{9}\sqrt{3}\pi \text{ (coordinate time)},$$

$$2\pi \frac{[(1 - \rho^2)/2]^{\frac{1}{2}}}{(1 + \rho^2)}\bigg|_{\rho=1/\sqrt{3}} = \tfrac{1}{2}\sqrt{3}\pi \text{ (physical time)}.$$

The last value corresponds of course to the critical circle circumference $2\pi\rho/(1 + \rho^2)\,|_{\rho=1/\sqrt{3}}$ being traversed with the unit velocity of light.

The results concerning orbits going to infinity,

mentioned in the discussion of "completeness" in Sec. I, also follow from inspection of Eq. (4′). The fact that the final quadrature, giving the time or proper time as a function of radius vector, leads to a simple elliptic integral, or a linear combination of simple elliptic integrals (Appendix), enables one to conclude that the magnetic universe is "complete" with respect to its timelike and lightlike geodesics. All orbits go on for an infinite timelike interval (for mass points) or time (for light tracks), and "no news can enter or leave the universe."

ACKNOWLEDGMENT

We wish to express our appreciation to K. Fox, R. Roskies, and K. S. Thorne for a helpful reading and discussion. We are very grateful to J. A. Wheeler for an intensive review of this paper and detailed valuable suggestions for formulating our results most simply.

APPENDIX: DETAILS OF ORBITS

A. First Integrals of Motion by Method of Constructive Interference of Ideal Wave Trains (Hamilton–Jacobi–Wheeler)

When the orbit is not circular or "straight" in the canonical coordinates we are using, it is desirable to spell out the equations for orbit or trajectory. The motion of the particle is given by the condition of "constructive interference" of wave trains with values of the constants of integration which lie in a small range

$$E \text{ to } E + dE, \qquad P \text{ to } P + dP, \qquad L \text{ to } L + dL.$$

This condition—that waves with nearly identical values of E, P, and L will at all times have the same value of the phase S—requires then

$$\begin{aligned}
&\frac{\partial S}{\partial E} = 0: \quad \frac{d}{d\tau}(\partial S/\partial E) = 0,\\
\text{or}\quad &S_{E\tau} + S_{E\rho}\rho_\tau + S_{E\zeta}\zeta_\tau + S_{E\phi}\phi_\tau = 0,\\
&\frac{\partial S}{\partial P} = 0: \quad \frac{d}{d\tau}(\partial S/\partial P) = 0,\\
\text{or}\quad &S_{P\tau} + S_{P\rho}\rho_\tau + S_{P\zeta}\zeta_\tau + S_{P\phi}\phi_\tau = 0,\\
&\frac{\partial S}{\partial L} = 0: \quad \frac{d}{d\tau}(\partial S/\partial L) = 0,\\
\text{or}\quad &S_{L\tau} + S_{L\rho}\rho_\tau + S_{L\zeta}\zeta_\tau + S_{L\phi}\phi_\tau = 0,
\end{aligned} \tag{8}$$

where the subscripts in the equations on the right indicate partial derivatives with respect to the subscript quantities. These three linear equations give ρ_τ, ζ_τ, and ϕ_τ directly; ultimately they determine ρ, ζ, and ϕ as functions of τ (the "orbit"); the last two of them give ρ and ζ as functions of ϕ (the "trajectory"). Equations (8) contain the equivalent of the first integrals of the geodesic equations in the usual pedestrian procedure. It is characteristic of this elegant Jacobi–Wheeler approach that it should replace the finding of first integrals of geodesic differential equations (Lagrange's equations) by simple differentiation and solution of a system of linear equations once the general form of the solution of the Hamilton–Jacobi equation has been found. We obtain, from Eqs. (3′) and (4) substituted in Eq. (8),

$$\rho_\tau = -S_{E\tau}/S_{E\rho} = R(\rho)/E$$
$$= [U^2 - (1+\rho^2)^4 L^2/\rho^2 - \epsilon(1+\rho^2)^2]^{\frac{1}{2}} E^{-1}, \tag{8a}$$

$$\zeta_\tau = P/E, \tag{8b}$$

$$\phi_\tau = LE^{-1}(1+\rho^2)^4/\rho^2, \tag{8c}$$

which specify the functional forms of the three first integrals E, P, and L.

Equation (8b) considered as a conservation law identifies P as a sort of "ζ-component linear momentum," with E representing an effective Newtonian mass. Again Eq. (8c) expresses a conservation law for a "ζ-component angular momentum L" to whose Newtonian form for a mass E it approximates for small ρ (weak gravitational field). Finally Eq. (8a) may be rewritten exactly

$$E = (\epsilon e^{2\psi} + E^2\rho_\tau^2 + P^2 + e^{4\psi}L^2/\rho^2)^{\frac{1}{2}}. \tag{9}$$

The last three terms in the parentheses represent the square of the quantity: [E multiplied by the physical orbital velocity ($\equiv v_{\text{phys}}$)] as is readily verified from Eqs. (8b) and (8c) with the help of the line element, Eq. (1). Thus Eq. (9) gives

$$Ee^{-\psi} = [\epsilon + (Ee^{-\psi}v_{\text{phys}})^2]^{\frac{1}{2}}$$
$$\equiv [(\text{rest mass})^2 + (\text{momentum})^2]^{\frac{1}{2}} \tag{9a}$$

which is exactly the special relativity expression if $Ee^{-\psi}$ represents the kinetic energy. For a unit rest mass particle

$$E(1-\psi) \simeq \{1 + [E(1-\psi)v_{\text{phys}}]^2\}^{\frac{1}{2}},$$
$$Ee^{-\psi} \cong 1 + \tfrac{1}{2}[Ee^{-\psi}v_{\text{phys}}]^2$$
$$\equiv \text{rest mass} + \tfrac{1}{2}[\text{momentum}]^2/\text{rest mass},$$

where the first approximation is the special relativistic approximation for weak fields, and the second one is the Newtonian approximation involving the limitation to small velocities. We note that for a zero rest mass particle Eq. (9a) verifies the fundamental principle: $v_{\text{phys}} = 1$.

The three equations (8a)–(8c) are entirely equivalent to those obtained by the standard Jacobi integration method for a system of which every coordinate but one is cyclic[15]:

$$\tau = \int R_E(\rho)\, d\rho = E \int R^{-1}(\rho)\, d\rho, \tag{10a}$$

$$\zeta = \int R_P(\rho)\, d\rho = P \int R^{-1}(\rho)\, d\rho, \tag{10b}$$

$$\phi = \int R_L(\rho)\, d\rho = L \int \frac{(1+\rho^2)^4}{\rho^2} R^{-1}(\rho)\, d\rho. \tag{10c}$$

The standard Jacobi procedure lacks, however, the intuitive picture that went into the derivation of Eqs. (8)—i.e., of a particle represented by a wave packet formed by "constructive interference of ideal wave trains."

B. Classification of Types of Orbits. The Meridian Plane Cases

It is clarifying to analyze all possible orbits in two classes: (I) those for which the angular momentum about the symmetry axis is zero ($L = 0$), which we call "meridian-plane orbits," and (II) the general nonmeridian-plane orbits for which $L \neq 0$.

For meridian-plane light tracks, Eq. (8a) yields

$$\rho = \dot{\rho}_0 \tau + \rho_0 \qquad (\dot{\rho}_0 \equiv U/E,\ \rho_0 \text{ constants}).$$

This together with Eq. (8b) shows that all such tracks are "straight" as viewed in the canonical coordinate system. Projected either forward or backward in time, any such geodesic is bound to intersect the ζ axis, except in the one subcase where $\dot{\rho}_0 = 0$. In this instance the geodesic is forever parallel to the ζ axis. *It is only in this subcase that the orbit does not straddle the critical straddling radius* $\rho = 1/\sqrt{3}$.

Meridian-plane orbits of particles with nonzero rest mass satisfy the simplified equation

$$\rho_\tau^2 = [U^2 - (1+\rho^2)^2]/E^2. \tag{11}$$

Any such orbit certainly intersects the ζ axis and has a turning point given by the positive root

$$(U-1)^{\frac{1}{2}} \equiv \rho_0.$$

Integrated a second time, Eq. (11) gives ρ as a Jacobi elliptic function of the time

$$\rho = \rho_0 \operatorname{cn} \alpha\tau \quad \alpha^2 \equiv 2U/E^2 = 2(1+\rho_0^2)/E^2 \tag{12}$$

with the modulus

$$k \equiv [(U-1)/2U]^{\frac{1}{2}} = \rho_0[2(1+\rho_0^2)]^{-\frac{1}{2}}$$

and period $4K(k)$, where $K(k)$ is the complete elliptic integral of the first kind.

C. Solutions for General Orbits and Trajectories

The orbit and trajectory of a particle from Eqs. (10a) and (10c) is given by

$$\tau = \frac{E}{2L} \int_{x_0}^{\rho^2} \frac{dx}{[(x_0 - x)(x^3 + Ax^2 + Bx + C)]^{\frac{1}{2}}} \qquad (x_0 \equiv \rho_0^2), \tag{13}$$

$$\begin{aligned}\phi &= \frac{L}{E} \int_0^\tau \frac{(1+\rho^2)^4}{\rho^2}\, d\tau \\ &= \frac{1}{2} \int_{\rho_0^2}^{\rho^2} \frac{(1+x)^4\, dx}{x[(x_0 - x)(x^3 + Ax^2 + Bx + C)]^{\frac{1}{2}}} \\ &= \frac{1}{2} \int_{\rho_0^2}^{\rho^2} \frac{(1+x)^4\, dx}{x[-(1+x)^4 - (\epsilon/L^2)(1+x)^2 x + (U^2/L^2)x]^{\frac{1}{2}}},\end{aligned} \tag{14}$$

$$A \equiv \frac{U^2}{L^2(1+x_0)^2} + 2 - \frac{1}{x_0},$$

$$B \equiv \frac{U^2(2+x_0)}{L^2(1+x_0)^2} - \frac{2-x_0}{x_0},$$

$$C \equiv -\frac{1}{x_0},$$

where the initial conditions at $\tau = 0$, $x = x_0 = \rho_0^2$ (an apse) have been inserted.

If $L = 0$, Eq. (13) reduces to the simple elliptic integral integrated in Part B of this Appendix.

For $L \neq 0$, $\tau = \tau(\rho^2)$ is still an elliptic integral of the first kind, whereas $\phi(\rho^2)$ is a linear combination of first- and third-kind elliptic integrals. We omit here a detailed analysis which brings out once again the existence of the maximum straddling circle.

[15] See, e.g., Max Born, *The Mechanics of the Atom* (G. Bell and Sons, London, 1927 or Frederick Ungar Publishing Company, New York, 1960), p. 41.

Correlation Functions for Eigenvalues of Real Quaternian Matrices

M. L. MEHTA AND P. K. SRIVASTAVA
Department of Physics and Astrophysics, University of Delhi, Delhi, India
(Received 14 September 1965)

The eigenvalue density, the two-, and the three-point correlation functions for the ensemble of real quaternion matrices are calculated. The forms suggest a generalization for the n-point correlation function. The probability that no eigenvalues lie inside a circle of radius r around the origin is also calculated for the ensemble of real quaternion matrices as well as for that of complex matrices. Upper and lower bounding functions for the last probability density are given.

MATRIX ensembles have been studied by several authors in order to describe large systems with unknown Hamiltonians. The aim is to obtain a knowledge of the types of restrictions which arise from the most general invariance of the physical problem. Besides these, the matrix elements are considered random variables. The Hamiltonians of physical systems are Hermitian, and depending on the time-reversal properties, can be divided into three classes, i.e., complex Hermitian, real symmetric, and self-dual quaternion matrices. These have been studied by many authors.[1] A more general problem, that of non-Hermitian matrices, has been studied by Ginibre,[2] who has been able to obtain the joint probability density of the eigenvalues of $N \times N$ matrices, in each of the three cases.

In case of complex matrices, correlation functions for the eigenvalues have been extracted[2] from the joint probability density, and the limit $N \to \infty$ taken. In this communication we have done the same for the case of quaternion matrices. The correlation functions up to the third order are explicitly obtained, and the forms suggest an expression for the general nth-order correlation function. One can also obtain an expression for the probability of there being no eigenvalues in a circle of radius r with the origin as center. A set of lower and upper bounds for this expression is found.

I

Every $N \times N$ matrix formed out of real quaternians can be considered as a $2N \times 2N$ matrix of complex numbers.[3] The eigenvalues of such matrices occur in complex conjugate pairs (z_i, z_i^*). The joint probability density of the eigenvalues $z_1, z_2, \cdots, z_N$ is given by[2]

$$P_N(z_1, \cdots, z_N) \prod_i dx_i\, dy_i = K \exp\left(-\sum_{i=1}^{N} |z_i|^2\right) \prod_{1\le i<j\le N} |z_i - z_j|^2 \cdot |z_i - z_j^*|^2 \cdot \prod_{i=1}^{N} |z_i - z_i^*|^2\, dx_i\, dy_i, \tag{1.1}$$

where

$$K^{-1} = N! \prod_{j=1}^{N} \{2\pi \cdot (2j-1)!\}. \tag{1.2}$$

In order to obtain the correlation functions we first obtain a functional integral,

$$\rho_N(1 + a) = \int \cdots \int P_N(z_1, \cdots, z_N) \times \prod_{i=1}^{N} (1 + a(z_i))\, dx_i\, dy_i. \tag{1.3}$$

The correlation functions can then be obtained by taking the functional derivatives of $\rho_N(1 + a)$ around $a(z) = 0$.[4]

We make use of the identity

$$\prod_{i=1}^{N} (z_i^* - z_i) \prod_{1\le i<j\le N} |z_i - z_j|^2 \cdot |z_i - z_j^*|^2 = \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 \\ z_1 & z_1^* & \cdots & z_N & z_N^* \\ z_1^2 & z_1^{*2} & \cdots & z_N^2 & z_N^{*2} \\ \vdots & \vdots & & \vdots & \vdots \\ z_1^{2N-1} & z_1^{*2N-1} & \cdots & z_N^{2N-1} & z_N^{*2N-1} \end{vmatrix} \cdot \tag{1.4}$$

The integral in Eq. (1.3) consists of the above determinant multiplied by the product

$$\prod_{i=1}^{N} e^{-|z_i|^2} (z_i - z_i^*)[1 + a(z_i)].$$

Noting that one variable occurs in two columns

[1] The relevant articles have been collected in the reprint volume "Statistical Properties of Spectra: Fluctuations" edited by C. E. Porter (Academic Press Inc., New York, to be published).
[2] J. Ginibre, J. Math. Phys. 6, 440 (1965).
[3] C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), pp. 16–18.
[4] F. J. Dyson, J. Math. Phys. 3, 166 (1962); also included in Ref. 1.

only; we can use standard methods[5] to express the integral as a Pfaffian. Before doing this one replaces, in the determinant,

$$z_i^{2l+1} \quad \text{by} \quad b_{2l+1}(z_i) = (2\pi)^{-\frac{1}{2}} \frac{z_i^{2l+1}}{(2l+1)!} \tag{1.5}$$

and

$$z_i^{2l} \quad \text{by} \quad b_{2l}(z_i) = (2\pi)^{-\frac{1}{2}} \sum_{k=0}^{l} \frac{2^l l!}{2^k k!} z_i^{2k}. \tag{1.6}$$

This changes the value of the determinant by

$$\prod_{j=1}^{N} \{2\pi \cdot (2j-1)!\}^{-1}$$

and we get

$$\rho_N(1+a) = \frac{1}{N!} \int \cdots \int \left\{ \prod_{i=1}^{N} \{[1+a(z_i)] \times e^{-|z_i|^2}(z_i - z_i^*)\, dx_i\, dy_i\} \times \det [b_k(z_i)\ b_k(z_i^*)]_{k=0,\cdots,2N-1}^{i=1,\cdots,N} \right\} \tag{1.7}$$

$$= \{\det [f_{ij}]^{i,j=0,1,\cdots,2N-1}\}^{\frac{1}{2}}, \tag{1.8}$$

where

$$f_{ij} = \iint e^{-|z|^2}(z - z^*)[1 + a(z)] \times [b_i(z)b_j(z^*) - b_j(z)b_i(z^*)]\, dx\, dy. \tag{1.9}$$

We notice that

$$f_{2i,2k}(a=0) = f_{2i+1,2k+1}(a=0) = 0 \tag{1.10}$$

and

$$f_{2i,2k+1}(a=0) = -f_{2k+1,2i}(a=0) = \delta_{ik}. \tag{1.11}$$

Thus

$$\rho_N(1) = \{\det [f_{ik}(a=0)]\}^{\frac{1}{2}} = 1. \tag{1.12}$$

The functional $\rho_N(1+a)$ can now be written as a Pfaffian which can be expanded into smaller Pfaffians,

$$\rho_N(1+a) = \left\{ \det \begin{bmatrix} \lambda_{ij} & \delta_{ij} + \nu_{ij} \\ -\delta_{ij} - \nu_{ji} & \mu_{ij} \end{bmatrix}_{i,j=0,\cdots,N-1} \right\}^{\frac{1}{2}}, \tag{1.13}$$

where λ, μ, and ν are small terms:

$$\lambda_{ij} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-|z|^2}(z - z^*)a(z) \sum_{k=0}^{i} \sum_{l=0}^{j} \frac{2^i i!\, 2^j j!}{2^k k!\, 2^l l!} \times (z^{2k} z^{*2l} - \text{c.c.})\, dx\, dy, \tag{1.14}$$

$$\nu_{ij} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-|z|^2}(z - z^*)a(z) \sum_{k=0}^{i} \frac{2^i i!}{2^k k!} \frac{1}{(2j+1)!} \times (z^{2k} z^{*2j+1} - \text{c.c.})\, dx\, dy, \tag{1.15}$$

$$\mu_{ij} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-|z|^2}(z - z^*)a(z) \times \frac{1}{(2i+1)!\,(2j+1)!} (z^{2i+1} z^{*2j+1} - \text{c.c.})\, dx\, dy. \tag{1.16}$$

We can now expand $\rho_N(1+a)$ (see Appendix A),

$$\rho_N(1+a) = 1 + \sum_{i=0}^{N-1} \begin{vmatrix} 0 & \nu_{ii} \\ -\nu_{ii} & 0 \end{vmatrix}^{\frac{1}{2}} + \frac{1}{2!} \sum_{i_1 i_2 = 0}^{N-1} \begin{vmatrix} 0 & \nu_{i_1 i_1} & \lambda_{i_1 i_2} & \nu_{i_1 i_2} \\ -\nu_{i_1 i_1} & 0 & -\nu_{i_2 i_1} & \mu_{i_1 i_2} \\ \lambda_{i_2 i_1} & \nu_{i_2 i_1} & 0 & \nu_{i_2 i_2} \\ -\nu_{i_1 i_2} & \mu_{i_2 i_1} & -\nu_{i_2 i_2} & 0 \end{vmatrix}^{\frac{1}{2}} + \cdots + \frac{1}{n!} \sum_{i_1 i_2 \cdots i_n = 0}^{N-1} \left\{ \det \begin{bmatrix} \lambda_{i_\alpha i_\beta} & \nu_{i_\alpha i_\beta} \\ -\nu_{i_\beta i_\alpha} & \mu_{i_\alpha i_\beta} \end{bmatrix}_{\alpha,\beta=1,\cdots,n} \right\}^{\frac{1}{2}} + \cdots. \tag{1.17}$$

II.

The one-level correlation function (i.e., the eigenvalue density) can be obtained by differentiating $\rho_N(1+a)$ functionally with respect to $a(z)$,

$$R_1(z) = \left[\frac{\delta}{\delta a(z)} \rho_N(1+a) \right]_{a=0} \tag{2.1}$$

$$= \frac{\delta}{\delta a(z)} \sum_{i=0}^{N-1} \begin{vmatrix} 0 & \nu_{ii} \\ -\nu_{ii} & 0 \end{vmatrix}^{\frac{1}{2}}. \tag{2.2}$$

Taking the limit $N \to \infty$,

$$R_1(z) = \frac{1}{2\pi} \sum_{i=0}^{\infty} \sum_{k=0}^{i} \frac{2^i i!}{2^k k!} \frac{1}{(2i+1)!} \times (z^{2k} z^{*2i+1} - \text{c.c.}) e^{-|z|^2}(z - z^*) \tag{2.3}$$

$$= \frac{1}{2\pi} e^{-|z|^2}(z - z^*) \sum_{i=0}^{\infty} \sum_{k=0}^{i} \sqrt{\frac{\pi}{2}} \frac{1}{k!\, \Gamma(i + \frac{3}{2})} \times (z^{2k} z^{*2i+1} - \text{c.c.}) \tag{2.4}$$

$$= \frac{1}{2\pi} e^{-|z|^2}(z - z^*)\phi(z, z^*) \tag{2.5}$$

$$= \frac{1}{4\pi} |z - z^*|^2 \cdot \int_0^1 e^{\frac{1}{4}(z - z^*)^2 \alpha} \frac{d\alpha}{(1-\alpha)^{\frac{1}{2}}}. \tag{2.6}$$

For the last step see Appendix B.

[5] M. L. Mehta, Nucl. Phys. 18, 395 (1960), Appendix I; also included in Ref. 1.

The two-level correlation function can be calculated from $\rho(1+a) = \lim_{N\to\infty} \rho_N(1+a)$ by differentiating twice,

$$R_2(z_1, z_2) = \left[\frac{\delta^2}{\delta a(z_1)\,\delta a(z_2)}\,\rho(1+a)\right]_{a(z)=0} \tag{2.7}$$

$$= \frac{\delta^2}{\delta a(z_1)\,\delta a(z_2)} \sum_{i_1, i_2=0}^{\infty} \begin{vmatrix} 0 & \nu_{i_1 i_1} & \lambda_{i_1 i_2} & \nu_{i_1 i_2} \\ -\nu_{i_1 i_1} & 0 & -\nu_{i_2 i_1} & \nu_{i_2 i_2} \\ \lambda_{i_2 i_1} & \nu_{i_2 i_1} & 0 & \nu_{i_2 i_2} \\ -\nu_{i_1 i_2} & \mu_{i_2 i_1} & -\nu_{i_2 i_2} & 0 \end{vmatrix}^{\frac{1}{2}}. \tag{2.8}$$

Carrying out the indicated functional differentiation and the summation we get

$$R_2(z_1, z_2) = (2\pi)^{-2} e^{-|z_1|^2 - |z_2|^2} (z_1 - z_1^*)(z_2 - z_2^*) \cdot \begin{vmatrix} 0 & \phi(z_1, z_1^*) & \phi(z_1, z_2) & \phi(z_1, z_2^*) \\ \phi(z_1^*, z_1) & 0 & \phi(z_1^*, z_2) & \phi(z_1^*, z_2^*) \\ \phi(z_2, z_1) & \phi(z_2, z_1^*) & 0 & \phi(z_2, z_2^*) \\ \phi(z_2^*, z_1) & \phi(z_2^*, z_1^*) & \phi(z_2^*, z_2) & 0 \end{vmatrix}^{\frac{1}{2}}. \tag{2.9}$$

Higher correlation functions are believed to be of the form

$$R_n(z_1, \cdots, z_n) = \prod_{i=1}^{N} \{(2\pi)^{-1} e^{-|z_i|^2} (z_i - z_i^*)\} \cdot \left\{\det \left[\begin{matrix} \phi(z_i, z_j) & \phi(z_i, z_j^*) \\ \phi(z_i^*, z_j) & \phi(z_i^*, z_j^*) \end{matrix}\right]_{i,j=1,\cdots,n}\right\}^{\frac{1}{2}}. \tag{2.10}$$

We have verified the form for $n = 3$ also.

III.

In this section we calculate the probability of all the eigenvalues lying outside a circle of radius r. We put

$$1 + a(z) = 0, \quad \text{for} \quad |z| < r$$
$$= 1, \quad \text{for} \quad |z| > r,$$

and express the result again as a pfaffian,

$$E_N(r^2) = \left\{(2\pi)^N \prod_{i=1}^{N} (2i-1)!\right\}^{-1} \times \{\det [g_{ij}]_{i,j=0,\cdots,2N-1}\}^{\frac{1}{2}}, \tag{3.1}$$

where

$$g_{ij} = \iint_{|z|>r} e^{-|z|^2} (z - z^*)(z^i z^{*j} - \text{c.c.})\, dx\, dy \tag{3.2}$$

$$= 2\pi\{\Gamma(j+1, r^2)\delta_{i+1,j} - \Gamma(i+1, r^2)\delta_{j+1,i}\}, \tag{3.3}$$

where $\Gamma(n, r^2)$ is an incomplete gamma function. Thus

$$E_N(r^2) = \prod_{i=1}^{N} f_{2i-1}(r^2), \tag{3.4}$$

where

$$f_i(x) = e^{-x} e_i(x) \tag{3.5}$$

and

$$e_i(x) = \sum_{k=0}^{i} \frac{x^k}{k!}. \tag{3.6}$$

As $0 < f_i(n) \leq 1$, we get the inequalities

$$E_\infty(r^2) \leq \cdots \leq E_N(r^2) \leq E_{N-1}(r^2) \leq \cdots \leq 1. \tag{3.7}$$

Also from

$$\prod_{i=N+1}^{\infty} f_{2i-1}(r^2) \geq \exp\left\{-\sum_{i=N+1}^{\infty} \int_0^{r^2} \frac{x^{2i-1}\,dx}{(2i-1)!\, e_{2N+1}(x)}\right\} = F_N(r^2), \tag{3.8}$$

we get

$$E_\infty(r^2) \geq \cdots \geq E_N(r^2)F_N(r^2) \geq E_{N-1}(r^2)F_{N-1}(r^2) \geq \cdots. \tag{3.9}$$

A similar result can also be obtained for eigenvalue of complex matrices. For the probability $E_\infty^{(o)}(r^2)$ of obtaining no eigenvalues in a circle of radius r, we have

$$E_\infty^{(o)}(r^2) = \lim_{N\to\infty} \prod_{i=1}^{N} e^{-r^2} e_i(r^2). \tag{3.10}$$

Bounding functions similar to (3.7) and (3.9) can be written down.

ACKNOWLEDGMENT

We are thankful to Professor R. C. Majumdar for his hospitality at the Physics Department.

APPENDIX A

A Pfaffian is the square root of an even-order antisymmetric determinant

$$\{\det [\phi_{ij}]_{i,j=1,\cdots,2n}\}^{\frac{1}{2}} = \frac{1}{n!} \sum \pm \phi_{i_1 i_2} \phi_{i_3 i_4} \cdots \phi_{i_{2n-1} i_{2n}}, \tag{A1}$$

where the summation is extended over all permutations $i_1, i_2, \cdots, i_{2n}$ of $1, 2, \cdots, 2n$ with the

restrictions $i_1 < i_2$, $i_3 < i_4$, $\cdots$, $i_{2n-1} < i_{2n}$ and the sign is plus or minus according as whether the permutation

$$\begin{pmatrix} i_1 & i_2 & \cdots & i_{2n} \\ 1 & 2 & \cdots & 2n \end{pmatrix}$$

is even or odd.

The expansion of a determinant with large diagonal and small off-diagonal elements is very well known. The coefficient of ε^k in the power series expansion of

$$\det [\delta_{ij} + \varepsilon\alpha_{ij}]_{i,j=1,\cdots,n}$$

is the sum of all possible $k \times k$ principal subdeterminants of $[\alpha_{ij}]$ obtained by suppressing symmetrically $(n - k)$ rows and columns of $[\alpha_{ij}]$. There is an analogous, but not so widely known, expansion of a Pfaffian. To get the expansion of

$$\left\{\det \begin{bmatrix} \varepsilon\lambda_{ij} & \delta_{ij} + \varepsilon\nu_{ij} \\ -\delta_{ij} - \varepsilon\nu_{ji} & \varepsilon\mu_{ij} \end{bmatrix}_{i,j=1,2,\cdots,n}\right\}^{\frac{1}{2}},$$

with

$$\lambda_{ji} = -\lambda_{ij}, \qquad \mu_{ji} = -\mu_{ij} \tag{A2}$$

in powers of ε we proceed in a similar manner. To get ε^k one must have $(n - k)$ factors in Eq. (A1) equal to unity, while the remaining k factors containing ε can be regrouped into Pfaffians. Thus

$$\begin{aligned} &\left\{\det \begin{bmatrix} \varepsilon\lambda_{ij} & \delta_{ij} + \varepsilon\nu_{ij} \\ -\delta_{ij} - \varepsilon\nu_{ji} & \varepsilon\mu_{ij} \end{bmatrix}\right\}^{\frac{1}{2}} \\ &= 1 + \varepsilon \sum_i \begin{vmatrix} 0 & \nu_{ii} \\ -\nu_{ii} & 0 \end{vmatrix}^{\frac{1}{2}} \\ &\quad + \cdots + \varepsilon^k \sum_{i_1<i_2<\cdots<i_k} \left\{\begin{vmatrix} \lambda_{ij} & \nu_{ij} \\ -\nu_{ji} & \mu_{ij} \end{vmatrix}_{i,j=i_1,\cdots,i_k}\right\}^{\frac{1}{2}} \\ &\quad + \cdots . \end{aligned} \tag{A3}$$

Whenever any two indices are equal, the corresponding Pfaffian is identically zero. Hence one can sum independently over the indices and replace ε^k by $\varepsilon^k/k!$.

APPENDIX B

In this paper we come across the function $\phi(a, b)$ of two complex variables, defined by

$$\begin{aligned} \phi(a, b) = &\left(\frac{\pi}{2}\right)^{\frac{1}{2}} \sum_{i=0}^{\infty} \sum_{k=0}^{i} \frac{1}{k!\,\Gamma(i + \frac{3}{2})} \\ &\times \left\{\left(\frac{a^2}{2}\right)^k \left(\frac{b^2}{2}\right)^{i+\frac{1}{2}} - \left(\frac{b^2}{2}\right)^k \left(\frac{a^2}{2}\right)^{i+\frac{1}{2}}\right\}. \end{aligned} \tag{B1}$$

Changing the summation index, and re-expressing as the function of $x = ab$ and $y = b/a$, we have

$$\phi(a, b) = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \sum_{k=0}^{\infty} \{y^{k+\frac{1}{2}} - y^{-(k-\frac{1}{2})}\} I_{k+\frac{1}{2}}(x), \tag{B2}$$

where $I_{k+\frac{1}{2}}(x)$ is the modified Bessel function. Using the expression

$$2I'_{k+\frac{1}{2}}(x) = I_{k-\frac{1}{2}}(x) + I_{k+}(x),$$

We get

$$\begin{aligned} d\phi(a, b)/dx = &\tfrac{1}{2}(y^{\frac{1}{2}} - y^{-\frac{1}{2}})\{I_{\frac{1}{2}}(x) + I_{-\frac{1}{2}}(x)\} \\ &+ \tfrac{1}{2}(y + y^{-1})\phi(a, b). \end{aligned} \tag{B3}$$

This, coupled with the fact that

$$\phi(a, b) \to 0 \quad \text{as} \quad x \to 0, \tag{B4}$$

gives us

$$\phi(a, b) = \tfrac{1}{2}e^{ab}(b - a) \int_0^1 e^{\frac{1}{2}(a-b)^2\alpha} \frac{d\alpha}{(1 - \alpha)^{\frac{1}{2}}}. \tag{B5}$$

As $b \to a$, we have

$$\phi(a, b) \to e^{ab}(b - a)\{1 + \tfrac{2}{3}(b - a)^2 + \cdots\}. \tag{B6}$$

As $(b - a) \to i\infty$,

$$\phi(a, b) \to e^{ab}/(a - b), \tag{B7}$$

and as $(b - a) \to \infty$,

$$\phi(a, b) \to (\tfrac{1}{2}\pi)^{\frac{1}{2}} e^{(a^2+b^2)/2} \operatorname{sign}(b - a). \tag{B8}$$

First-Order Solutions of the One-Speed Transport Equation in Plane Multiregion Geometries*

M. R. MENDELSON
Knolls Atomic Power Laboratory, Schenectady, New York†
(Received 1 March 1965)

Case's method is applied to multiregion problems in plane geometry. Half-range and "two-media" orthogonality relations are used to obtain a system of coupled integral equations for the appropriate expansion coefficients. The system is explicitly solved in a first-order approximation, which is valid for wide regions. Numerical analysis and comparison with transport computer codes indicates that such solutions are essentially exact for region thicknesses of five mean free paths or more.

I. INTRODUCTION

RECENTLY Case's method has been applied to the solution of several problems defined in two adjacent half-spaces.[1] This paper is an extension of that work to the analysis of multiregion problems. In particular, three problems are considered: (1) The slab with infinite reflectors; (2) the slab with finite reflectors; and (3) the uniform, infinite lattice.

These problems (and indeed all multiregion problems) do not in general have exact solutions in closed form. However, by making use of the orthogonality relations recently developed for the one and two half-space problems,[2] a solution can be obtained implicitly in the form of a system of integral equations for the discrete and continuum coefficients. From this system an explicit first-order solution, valid for "wide regions," is easily obtained. This technique was first employed by Case in connection with the bare multiplying slab.[3] Such solutions are generally of much better accuracy than would be implied from the assumptions invoked to obtain them. Numerical analysis, in fact, indicates that they are nearly exact in regions which are more than five mean free paths thick.

The multiregion problem is the most general problem that arises in the context of the one-speed, one-dimensional theory and therefore represents, in a sense, a culmination of the earlier work on one- and two-region problems. Aside from this academic motivation, there are several additional reasons for considering these problems:

1. A similar set of multiregion problems has been treated by Kuzell,[4] also using Case's method. His approach was based upon the use of full range integral relations (not orthogonality relations) between the normal modes of different regions. This led to extremely complicated solutions in terms of multiple singular integrals which were not readily amenable to approximation in explicit form. The formalism presented here does not suffer from that disadvantage.

2. Although some selected numerical information has been obtained by Mitsis regarding the accuracy of the low-order approximate solutions in a bare slab problem,[5] there has not been a detailed investigation of the range of parameters for which these approximate representations of the neutron distribution function are valid. In the present analysis the accuracy of the first- (lowest-) order solution for the neutron distribution is examined by comparison with a numerical integration of the transport equation.

3. A practical motivation exists in that these analytical methods may lead to an accurate calculation of certain reactor parameters. For example, quantities, such as generalized escape probabilities, flux depression factors, disadvantage factors, and the like can obviously be obtained in explicit form from the solution of these problems. Furthermore, this analysis can be used to advantage in treating large homogeneous regions for which numerical integration methods break down because of mesh spacing limitations. Finally, a possible application exists in constructing an improved asymptotic diffusion theory, in the manner of Selengut[6] and Pomraning and Clark,[7] by using only the discrete

* Based in part upon a Ph.D. thesis at the University of Michigan.
† Operated by the General Electric Company for the U. S. Atomic Energy Commission, Contract No. W-31-109-Eng-52.
[1] M. R. Mendelson and G. C. Summerfield, J. Math. Phys. 5, 668 (1964).
[2] I. Kuscer, M. J. McCormick, and G. C. Summerfield, Ann. Phys. (N. Y.) 30, 411 (1964).
[3] K. M. Case, *Recent Developments in Neutron Transport Theory*, Michigan Memorial Phoenix Project Report, The University of Michigan (1961).
[4] A. Kuszell, *Critical Problems for Multilayer Slab Systems*, Polish Academy of Sciences, Institute of Nuclear Research, Report No. 206/IX, Warsaw (1961).
[5] G. J. Mitsis, Nucl. Sci. Eng. 17, 55 (1963).
[6] D. S. Selengut, Trans. Am. Nucl. Soc. 5, 40 (1962).
[7] G. C. Pomraning and M. Clark, Jr., Nucl. Sci. Eng. 17, 227 (1963).

portion of the solution in the wide-region approximation.

II. ELEMENTARY SOLUTIONS AND ORTHOGONALITY RELATIONS

The homogeneous one-speed neutron transport equation in plane geometry for isotropic scattering may be written as

$$\mu \frac{\partial \psi(x, \mu)}{\partial x} + \psi(x, \mu) = \frac{C}{2} \int_{-1}^{1} \psi(x, \mu')\, d\mu' \qquad (1)$$

where $\psi(x, \mu)$ is the angular density measured in units of mean free paths, C is the mean number of secondaries per collision and is assumed to be constant, and the arguments have their usual meaning. The general solution of this equation is the following[8]:

$$\psi(x, \mu) = a_{0+}\phi_{0+}(\mu)e^{-x/\nu_0} + a_{0-}\phi_{0-}(\mu)e^{+x/\nu_0} + \int_{-1}^{1} A(\nu)\phi_\nu(\mu)e^{-x/\nu}\, d\nu \qquad (2)$$

where $a_{0\pm}$ and $A(\nu)$ are arbitrary expansion coefficients and $\phi_{0\pm}(\mu)$ and $\phi_\nu(\mu)$ are angular eigenfunctions given by

$$\phi_{0\pm}(\mu) = C\nu_0/2(\nu_0 \mp \mu), \qquad (3)$$

$$\phi_\nu(\mu) = (C\nu/2)[P/(\nu - \mu)] + \lambda(\nu)\, \delta(\nu - \mu). \qquad (4)$$

The eigenvalue ν_0 is the positive root of the equation

$$\Lambda(\nu) = 1 - \frac{C\nu}{2} \int_{-1}^{1} \frac{d\mu}{\nu - \mu} = 0 \qquad (5)$$

and the function $\lambda(\nu)$ is defined as

$$\lambda(\nu) = 1 - \frac{C\nu}{2} P \int_{-1}^{1} \frac{d\mu}{\nu - \mu} = 1 - C\nu \tanh^{-1} \nu \qquad (6)$$

where P denotes the Cauchy principal value.

The functions $\phi_{0+}(\mu)$ and $\phi_\nu(\mu)$ are orthogonal with respect to the weight function $w(\mu)$ on the interval $0 \leq \mu \leq 1$, where

$$w(\mu) = \gamma(\mu)(\nu_0 - \mu), \qquad (7)$$

$$\gamma(\mu) = C\mu/[2(1 - C)(\nu_0^2 - \mu^2)X(-\mu)], \qquad (8)$$

and

$$X(z) = \frac{1}{1 - z} \exp\left[\frac{1}{\pi} \int_0^1 \frac{d\mu}{\mu - z} \tan^{-1}\left\{\frac{c\mu\pi/2}{\lambda(\mu)}\right\}\right]. \qquad (9)$$

The following orthogonality and normalization relations have been derived by Kuscer *et al.*[2]:

$$\int_0^1 \phi_\nu(\mu)\phi_{\nu'}(\mu)w(\mu)\, d\mu = \frac{w(\nu)\, \delta(\nu - \nu')}{g(C, \nu)}, \qquad (10)$$

$$\int_0^1 \phi_{0+}(\mu)\phi_\nu(\mu)w(\mu)\, d\mu = 0, \qquad (11)$$

$$\int_0^1 \phi_{0-}(\mu)\phi_\nu(\mu)w(\mu)\, d\mu = C\nu\, \nu_0 X(-\nu_0)\phi_{0-}(\nu), \qquad (12)$$

$$\int_0^1 \phi_{0\pm}(\mu)\phi_{0+}(\mu)w(\mu)\, d\mu = \mp(\tfrac{1}{2}C\nu_0)^2 X(\pm\nu_0), \qquad (13)$$

$$\int_0^1 \phi_{-\nu}(\mu)\phi_{0+}(\mu)w(\mu)\, d\mu = \frac{C^2\nu\, \nu_0}{4} X(-\nu), \qquad (14)$$

$$\int_0^1 \phi_{-\nu}(\mu)\phi_{\nu'}(\mu)w(\mu)\, d\mu = \frac{C\nu'}{2} \phi_{-\nu}(\nu')(\nu_0 + \nu)X(-\nu). \qquad (15)$$

In Eq. (10) the function $g(C, \nu)$ refers to the function defined by Case, DeHoffmann, and Placzek[9]:

$$g(C, \nu) = [(1 - C\nu \tanh^{-1} \nu)^2 + (\pi C\nu/2)^2]^{-1}.$$

In the case of two adjacent media, where $C = C_1$ on the right-hand side and $C = C_2$ on the left, one distinguishes the angular eigenfunctions and other quantities by the appropriate subscripts, e.g., $\phi_{10\pm}(\mu)$, $\phi_{1\nu}(\mu)$, $\phi_{20\pm}(\mu)$, $\phi_{2\nu}(\mu)$, according to their respective values of C. In addition to the half-range orthogonality relations indicated above, it has been shown by Kuscer *et al.* that the functions $\phi_{10+}(\mu)$, $\phi_{20-}(\mu)$, $\phi_{1\nu}(\mu)$, $\nu > 0$, and $\phi_{2\nu}(\mu)$, $\nu < 0$, are orthogonal with respect to the weight function $W(\mu)$ on the interval $-1 \leq \mu \leq 1$ where

$$W(\mu) = \Gamma(\mu)(\nu_{01} - \mu)(\nu_{02} + \mu) \qquad (16)$$

and where

$$\Gamma(\mu) = \begin{cases} \gamma_1(\mu)X_2(-\mu), & \mu > 0, \\ \gamma_2(\mu)X_1(\mu), & \mu < 0. \end{cases}$$

If we define

$$C(\nu) = \begin{cases} C_1, & \nu > 0 \\ C_2, & \nu < 0 \end{cases} \quad \text{and} \quad \Phi_\nu(\mu) = \begin{cases} \phi_{1\nu}(\mu), & \nu > 0 \\ \phi_{2\nu}(\mu), & \nu < 0, \end{cases}$$

then the following "two-media" orthogonality relations and some related integrals are obtained[2]:

$$\int_{-1}^{1} \Phi_\nu(\mu)\Phi_{\nu'}(\mu)W(\mu)\, d\mu = \frac{W(\nu)\, \delta(\nu - \nu')}{g(C(\nu), \nu)}, \qquad (18)$$

$$\int_{-1}^{1} \phi_{10+}(\mu)\Phi_\nu(\mu)W(\mu)\, d\mu = 0, \qquad (19)$$

$$\int_{-1}^{1} \phi_{20-}(\mu)\Phi_\nu(\mu)W(\mu)\, d\mu = 0, \qquad (20)$$

[8] K. M. Case, Ann. Phys. 9, 1 (1960).

[9] K. M. Case, F. DeHoffmann, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (U. S. Government Printing Office, 1963), Vol. 1.

$$\int_{-1}^{1} \phi_{10+}(\mu)\phi_{20-}(\mu)W(\mu)\,d\mu = 0, \tag{21}$$

$$\int_{-1}^{1} \phi_{10-}(\mu)\Phi_{\nu}(\mu)W(\mu)\,d\mu = \tfrac{1}{2}C_1\nu C(\nu)\nu_{01}^2 \times \frac{(\nu_{02}-\nu_{01})}{(\nu_{02}+\nu)}\chi(-\nu_{01}), \tag{22}$$

$$\int_{-1}^{1} \phi_{20+}(\mu)\Phi_{\nu}(\mu)W(\mu)\,d\mu = \tfrac{1}{2}C_2\nu C(\nu)\nu_{02}^2 \times \frac{(\nu_{01}-\nu_{02})}{(\nu_{02}-\nu)}\chi(\nu_{02}), \tag{23}$$

$$\int_{-1}^{1} \phi_{10\pm}(\mu)\phi_{10+}(\mu)W(\mu)\,d\mu = -(\tfrac{1}{2}C_1\nu_{01})^2 \times (\nu_{01}\pm\nu_{02})\chi(\pm\nu_{01}), \tag{24}$$

$$\int_{-1}^{1} \phi_{20\pm}(\mu)\phi_{20-}(\mu)W(\mu)\,d\mu = (\tfrac{1}{2}C_2\nu_{02})^2 \times (\nu_{02}^2\mp\nu_{01})\chi(\pm\nu_{02}), \tag{25}$$

$$\int_{-1}^{1} \phi_{20+}(\mu)\phi_{10+}(\mu)W(\mu)\,d\mu = -\tfrac{1}{2}C_1C_2\nu_{01}\nu_{02}^2\chi(\nu_{02}), \tag{26}$$

$$\int_{-1}^{1} \phi_{10-}(\mu)\phi_{20-}(\mu)W(\mu)\,d\mu = \tfrac{1}{2}C_1C_2\nu_{01}^2\nu_{02}\chi(-\nu_{02}), \tag{27}$$

$$\int_{-1}^{1} \phi_{20-}(\mu)W(\mu)\,d\mu = -\frac{C_2\nu_{02}}{2}, \tag{28}$$

$$\int_{-1}^{1} \phi_{10+}(\mu)W(\mu)\,d\mu = \frac{C_1\nu_{01}}{2}, \tag{29}$$

$$\int_{-1}^{1} \Phi_{\nu}(\mu)W(\mu)\,d\mu = \frac{\nu C(\nu)}{2}, \tag{30}$$

where the χ-function appearing in the above equations is defined by

$$\chi(z) = X_1(z)X_2(-z). \tag{31}$$

III. SLAB WITH INFINITE REFLECTORS

Consider the problem of a slab with a uniformly distributed source adjoined by two symmetric infinite reflectors. Let $C = C_2$ in the central region and $C = C_1$ in the outer regions, where $C_1, C_2 < 1$, and, in general, $C_1 \neq C_2$.

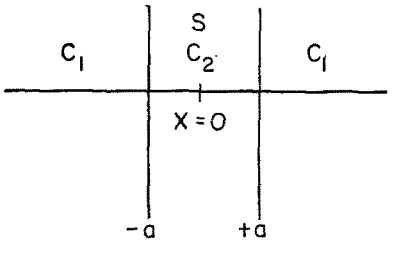

FIG. 1. Coordinate system for the slab with infinite reflectors.

The angular density satisfies the following equations:

$$\mu\frac{\partial\psi_1(x,\mu)}{\partial x} + \psi_1(x,\mu) = \frac{C_1}{2}\int_{-1}^{1}\psi_1(x,\mu')\,d\mu', \qquad x > a, \tag{32}$$

$$\mu\frac{\partial\psi_2(x,\mu)}{\partial x} + \psi_2(x,\mu) = \frac{C_2}{2}\int_{-1}^{1}\psi_2(x,\mu')\,d\mu' + S, \quad 0 < x < a, \tag{33}$$

where

(1) $\psi_1(0,\mu) = \psi_1(0,-\mu)$ (reflection symmetry about the $x = 0$ plane),

(2) $\psi_1(a,\mu) = \psi_2(a,\mu)$ (interface continuity),

(3) $\lim_{x\to\infty}\psi_1(x,\mu) = 0$ (boundedness).

In the usual manner[3,8] we express the solution of the problem in terms of "elementary solutions" (normal modes). A solution which obeys boundary condition (3) is

$$\psi_1(x,\mu) = a_0\phi_{10+}(\mu)e^{-x/\nu_{01}} + \int_0^1 A(\nu)\phi_{1\nu}(\mu)e^{-x/\nu}\,d\nu, \tag{34}$$

$$\psi_2(x,\mu) = b_{0+}\phi_{20+}(\mu)e^{-x/\nu_{02}} + b_{0-}\phi_{20-}(\mu)e^{x/\nu_{02}} - \int_{-1}^{1}B(\nu)\phi_{2\nu}(\mu)e^{-x/\nu}\,d\nu + \frac{S}{1-C_2}, \tag{35}$$

where $S/(1 - C_2)$ is a particular solution. Upon noting that $\phi_{0\pm}(-\mu) = \phi_{0\mp}(\mu)$ and $\phi_{\nu}(-\mu) = \phi_{-\nu}(\mu)$, it follows from the symmetry condition that

$$b_{0+} = b_{0-} \triangleq b_0, \tag{36a}$$

$$B(\nu) = B(-\nu). \tag{36b}$$

Hence application of boundary condition (2) results in the continuity equation

$$\psi(\mu) = \int_{-1}^{0} B(\nu)e^{-a/\nu}\phi_{2\nu}(\mu)\,d\nu + \int_0^1 A(\nu)e^{-a/\nu}\phi_{1\nu}(\mu)\,d\nu \tag{37}$$

where

$$\psi(\mu) = b_0[\phi_{20+}(\mu)e^{-a/\nu_{02}} + \phi_{20-}(\mu)e^{a/\nu_{02}}] + \frac{S}{1-C_2} - a_0\phi_{10+}(\mu)e^{-a/\nu_{01}} - \int_0^1 B(\nu)e^{-a/\nu}\phi_{2\nu}(\mu)\,d\nu. \tag{38}$$

This is a "two-media" full-range expansion of the function $\psi(\mu)$, the existence of which has been proven in Ref. 1.

In order to more conveniently apply the orthogonality relations given in Sec. II we invoke the following identity[10]:

$$\phi_{2\nu}(\mu) = (C_2/C_1)\phi_{1\nu}(\mu) + [(C_1 - C_2)/C_1]\,\delta(\nu - \mu). \tag{39}$$

The continuity equation may then be written

$$b_0[\phi_{20+}(\mu)e^{-a/\nu_{02}} + \phi_{20-}(\mu)e^{a/\nu_{02}}] + \frac{S}{1 - C_2} - \frac{C_1 - C_2}{C_1}B(\mu)e^{-a/\mu}H(\mu) - a_0\phi_{10+}(\mu)e^{-a/\nu_{01}} = \int_{-1}^{0} B(\nu)e^{-a/\nu}\phi_{2\nu}(\mu)\,d\nu + \int_0^1 \left\{A(\nu) + \frac{C_2}{C_1}B(\nu)\right\}e^{-a/\nu}\phi_{1\nu}(\mu)\,d\nu \tag{40}$$

where

$$H(\mu) \quad \begin{cases} 1, & \mu > 0 \\ 0, & \mu < 0. \end{cases}$$

If we now multiply Eq. (40) by $\phi_{20-}(\mu)W(\mu)$ and integrate over all μ we obtain, using Eqs. (20), (21), (25), and (28),

$$b_0 = \frac{1}{C_2\nu_{02}\Delta}\left\{\frac{2S}{1 - C_2} + \frac{2(C_1 - C_2)}{C_1} \times \int_0^1 B(\mu)e^{-a/\nu}(\nu_{01} - \mu)\gamma_1(\mu)\,d\mu\right\} \tag{41}$$

where

$$\Delta = [(\nu_{02} - \nu_{01})\chi(\nu_{02})e^{-a/\nu_{02}} + (\nu_{02} + \nu_{01})\chi(-\nu_{02})e^{a/\nu_{02}}]. \tag{42}$$

Similarly, upon multiplying Eq. (40) by $\phi_{10+}(\mu)W(\mu)$ and integrating over all μ we obtain, using Eqs. (19), (21), (24), and (29),

$$a_0 = \frac{1}{e^{-a/\nu_{01}}C_1\nu_{01}(\nu_{01} + \nu_{02})\chi(\nu_{01})} \times \left\{2b_0e^{-a/\nu_{02}}C_2\nu_{02}^2\chi(\nu_{02}) - \frac{2S}{1 - C_2} + \frac{2(C_1 - C_2)}{C_1} \times \int_0^1 B(\mu)e^{-a/\mu}(\nu_{02} + \mu)\gamma_1(\mu)\,d\mu\right\}. \tag{43}$$

Upon substitution of b_0 from Eq. (41), and after much algebraic manipulation, a_0 may be written in the form

$$a_0 = \left\{e^{-a/\nu_{02}}\chi(\nu_{02})\left[\frac{2(C_1 - C_2)}{C_1} \times \int_0^1 B(\mu)e^{-a/\mu}(\nu_{02} - \mu)\gamma_1(\mu)\,d\mu + \frac{2S}{1 - C_2}\right] + e^{a/\nu_{02}}\chi(-\nu_{02})\left[\frac{2(C_1 - C_2)}{C_1} \times \int_0^1 B(\mu)e^{-a/\mu}(\nu_{02} + \mu)\gamma_1(\mu)\,d\mu - \frac{2S}{1 - C_2}\right]\right\} \times \frac{1}{C_1\nu_{01}e^{-a/\nu_{01}}\chi(\nu_{01})\,\Delta}. \tag{44}$$

To obtain expressions for the continuum coefficients, we first multiply Eq. (40) by $\phi_{1\nu}(\mu)W(\mu)$, $\nu > 0$, and integrate over all μ. Using Eqs. (23), (20), (30), (19), and (18) we find

$$A(\nu) = \frac{e^{a/\nu}g(C_1, \nu)}{W(\nu)}\left\{b_0e^{-a/\nu_{02}}\frac{C_2\nu C_1}{2}\nu_{02}^2\frac{(\nu_{01} - \nu_{02})}{(\nu_{02} - \nu)}\chi(\nu_{02}) + \frac{S\nu C_1}{2(1 - C_2)} - \frac{C_1 - C_2}{C_1}\int_0^1 B(\mu)e^{-a/\mu}\phi_{1\nu}(\mu)W(\mu)\,d\mu\right\} - \frac{C_2}{C_1}B(\nu), \qquad \nu > 0 \tag{45}$$

or, equivalently,

$$A(\nu) = \frac{e^{a/\nu}(1 - C_1)(\nu_{01}^2 - \nu^2)X_1(-\nu)g(C_1, \nu)}{X_2(-\nu)} \times \left\{\begin{bmatrix} b_0\phi_{20+}(\nu)\chi(\nu_{02})e^{-a/\nu_{02}} \\ +b_0\phi_{20-}(\nu)\chi(-\nu_{02})e^{+a/\nu_{02}} \\ -a_0\phi_{10+}(\nu)\chi(\nu_{01})e^{-a/\nu_{01}} \end{bmatrix} - \frac{2(C_1 - C_2)}{\nu C_1^2}\int_0^1 B(\mu)e^{-a/\mu}\gamma_1(\mu)\phi_{1\nu}(\mu)\,d\mu\right\} - \frac{C_2}{C_1}B(\nu), \qquad \nu > 0. \tag{46}$$

Finally, multiplying Eq. (40) by $\phi_{2\nu}(\mu)W(\mu)$, $\nu < 0$, and integrating over all μ we obtain, with the aid of the appropriate orthogonality relations,

$$B(\nu) = \frac{e^{a/\nu}g(C_2, \nu)}{W(\nu)}\left\{b_0e^{-a/\nu_{02}}\frac{C_2^2\nu\,\nu_{02}^2}{2}\frac{(\nu_{01} - \nu_{02})}{(\nu_{02} - \nu)}\chi(\nu_{02}) + \frac{S\nu C_2}{2(1 - C_2)} - \frac{(C_1 - C_2)}{C_1}\int_0^1 B(\mu)e^{-a/\mu}\phi_{2\nu}(\mu)W(\mu)\,d\mu\right\}, \qquad \nu < 0 \tag{47}$$

[10] M. R. Mendelson, Ph.D. Thesis, The University of Michigan (1964).

or alternately,

$$B(\nu) = \frac{e^{+a/\nu}(1 - C_2)(\nu_{02}^2 - \nu^2)X_2(\nu)g(C_2, \nu)}{X_1(\nu)} \times \left\{ \begin{bmatrix} b_0\phi_{20+}(\nu)\chi(\nu_{02})e^{-a/\nu_{02}} \\ +b_0\phi_{20-}(\nu)\chi(-\nu_{02})e^{a/\nu_{02}} \\ -a_0\phi_{10+}(\nu)\chi(\nu_{01})e^{-a/\nu_{01}} \end{bmatrix} - \frac{C_2 - C_1}{C_1}\int_0^1 \frac{B(\mu)e^{-a/\mu}\gamma_1(\mu)\,d\mu}{\nu - \mu} \right\}, \quad \nu < 0. \tag{48}$$

Equations (36b), (41), (44), (46), and (48) form a system of coupled equations for the coefficients a_0, b_0, $A(\nu)$, $\nu > 0$, $B(\nu)$, $\nu > 0$, and $B(\nu)$, $\nu < 0$. In particular, Eq. (48) is a Fredholm equation for $B(\nu)$, $\nu < 0$, which can be solved by Neumann iteration if the resultant series is convergent, and which then leads to the solution of the system for the remaining coefficients in corresponding orders of approximation.

It is obvious, however, that the solutions will become enormously complicated as the order of approximation increases. We shall therefore obtain only "first" order solutions in this paper, which correspond to the retention of the first (integral-free) term in the Neumann series for $B(\nu)$, $\nu < 0$. Indeed we go one step further in that we shall interpret this approximation as corresponding physically to a wide slab and therefore justify the neglect of the integral term in Eq. (48) on the grounds that it makes a small contribution for large a. In the same sense we may also neglect the integral contributions in Eqs. (41) and (44), and the contributions of the last two terms in Eq. (46). This decouples the system and gives the following physically consistent first-order approximation:

$$b_0^{(1)} = \frac{2S}{(1 - C_2)C_2\nu_{02}\,\Delta}, \tag{49}$$

$$a_0^{(1)} = \frac{-2S[\chi(-\nu_{02})e^{a/\nu_{02}} - \chi(\nu_{02})e^{-a/\nu_{02}}]}{(1 - C_2)C_1\nu_{01}\chi(\nu_{01})e^{-a/\nu_{01}}\,\Delta}, \tag{50}$$

where Δ is given by Eq. (42);

$$A^{(1)}(\nu) = \frac{(1 - C_1)(\nu_{01}^2 - \nu^2)X_1(-\nu)g(C_1, \nu)f(\nu)e^{a/\nu}}{X_2(-\nu)}, \quad \nu > 0, \tag{51}$$

$$B^{(1)}(\nu) = \frac{(1 - C_2)(\nu_{02}^2 - \nu^2)X_2(\nu)g(C_2, \nu)f(\nu)e^{a/\nu}}{X_1(\nu)}, \quad \nu < 0, \tag{52}$$

where

$$f(\nu) = \begin{bmatrix} b_0\phi_{20+}(\nu)\chi(\nu_{02})e^{-a/\nu_{02}} \\ +b_0\phi_{20-}(\nu)\chi(-\nu_{02})e^{a/\nu_{02}} \\ -a_0\phi_{10+}(\nu)\chi(\nu_{01})e^{-a/\nu_{01}} \end{bmatrix}. \tag{53}$$

It should be pointed out that it is not necessary to first obtain the exact solution in order to arrive at the first-order approximation given above. The "wide region" approximation can be invoked much earlier, in the statement of the continuity equation. In fact, the procedure followed here is completely equivalent to ignoring the integral term in $\psi(\mu)$, Eq. (38), and then applying the orthogonality relations directly to the simplified continuity condition. This is an advantage of the "wide region" interpretation of the first order approximation, as opposed to other types of first-order approximations, in that it greatly simplifies the manipulation. This simpler method of attack will be followed in the next two problems.

There are two interesting characteristics of the first-order solution that should be pointed out:

(1) The first-order solution, or indeed any finite-order solution, will not be continuous at the interfaces since the equation resulting from the continuity condition was only approximately solved. The magnitude of this discontinuity, which will be small for "wide regions," is investigated in Sec. VI.

(2) When $C_1 = C_2$, one sees that the discrete coefficients and $B^{(1)}(\nu)$, $\nu < 0$ as given by the first-order solution are exact, while $A^{(1)}(\nu)$ is too large by the quantity $B^{(1)}(\nu)$. Thus the solution is exact in the inside region. This behavior is quite useful as a check on the accuracy of the numerical analysis.

IV. SLAB WITH FINITE REFLECTORS

Consider the case of a slab with a uniformly distributed source (Fig. 2), adjoined by two finite symmetric reflectors. As before, $C = C_2$ in the central region, $C = C_1$ in the outer regions, and C_1, $C_2 < 1$. The angular density in the region $0 \leq x \leq b$ satisfies the equations

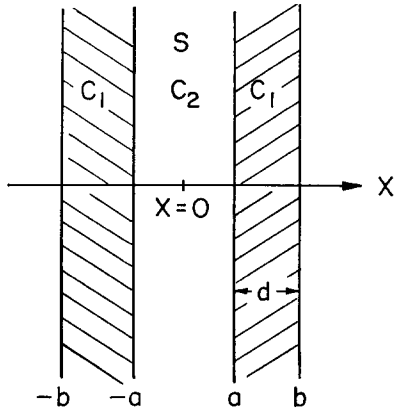

FIG. 2. Coordinate system for the slab with finite reflectors.

$$\mu \frac{\partial \psi_1(x,\mu)}{\partial x} + \psi_1(x,\mu) = \frac{C_1}{2}\int_{-1}^{1} \psi_1(x,\mu')\,d\mu', \qquad a < x < b, \tag{54}$$

$$\mu \frac{\partial \psi_2(x,\mu)}{\partial x} + \psi_2(x,\mu) = \frac{C_2}{2}\int_{-1}^{1} \psi_2(x,\mu')\,d\mu' + S, \qquad 0 < x < a \tag{55}$$

with the following boundary conditions:

(1) $\psi_2(0,\mu) = \psi_2(0,-\mu)$,

(2) $\psi_1(a,\mu) = \psi_2(a,\mu)$,

(3) $\psi_1(b,\mu) = 0, \qquad \mu < 0.$

The general solution is

$$\psi_1(x,\mu) = a_{0+}\phi_{10+}(\mu)e^{-x/\nu_{01}} + a_{0-}\phi_{10-}(\mu)e^{+x/\nu_{01}} + \int_{-1}^{1} A(\nu)\phi_{1\nu}(\mu)e^{-x/\nu}\,d\nu, \qquad a < x < b, \tag{56}$$

$$\psi_2(x,\mu) = b_{0+}\phi_{20+}(\mu)e^{-x/\nu_{02}} + b_{0-}\phi_{20-}(\mu)e^{x/\nu_{02}} - \int_{-1}^{1} B(\nu)\phi_{2\nu}(\mu)e^{-x/\nu}\,d\nu + \frac{S}{1-C_2}, \qquad 0 < x < a.$$

Upon applying boundary condition (3), we obtain

$$\alpha(\mu) = \int_0^1 A(-\nu)e^{b/\nu}\phi_{1\nu}(\mu)\,d\nu, \qquad \mu > 0, \tag{57}$$

where

$$\alpha(\mu) = -a_{0+}\phi_{10-}(\mu)e^{-b/\nu_{01}} - a_{0-}\phi_{10+}(\mu)e^{b/\nu_{01}} - \frac{C_1}{2}\int_0^1 A(\nu)e^{-b/\nu}\frac{\nu}{\nu+\mu}\,d\nu. \tag{58}$$

Equation (57) is a half-range expansion of the function $\alpha(\mu)$. Hence expressions for the coefficients appearing in this equation can be obtained by means of the half-range orthogonality relations given in Sec. II. However, we first impose the wide-region approximation, i.e., we assume that the term

$$\frac{C_1}{2}\int_0^1 A(\nu)e^{-b/\nu}\frac{\nu}{\nu+\mu}\,d\nu$$

makes a negligible contribution to $\alpha(\mu)$ if b is sufficiently large. (This assumption must be checked later for consistency.) Then, upon multiplying Eq. (57) by $\phi_{10+}(\mu)w(\mu)$ and integrating over μ from 0 to 1, we obtain

$$a_{0+} = a_{0-}\frac{X_1(\nu_{01})}{X_1(-\nu_{01})}e^{2b/\nu_{01}} \tag{59}$$

and similarly we find

$$A(-\nu)e^{b/\nu} = -a_{0+}e^{-b/\nu_{01}}X_1(-\nu) \times (1-C_1)C_1\nu_{01}^2 g(C_1,\nu)X_1(-\nu_{01}), \quad \nu > 0. \tag{60}$$

Now consider the solution in region (2). The symmetry condition (1) gives

$$b_{0+} = b_{0-} \triangleq b_0, \qquad B(\nu) = B(-\nu). \tag{61}$$

Boundary condition (2) results in the continuity equation

$$\psi(\mu) = \int_0^1 A(\nu)\phi_{1\nu}(\mu)e^{-a/\nu} + \int_{-1}^{0} B(\nu)\phi_{2\nu}(\mu)e^{-a/\nu}\,d\nu \tag{62}$$

where

$$\psi(\mu) = -a_{0-} \times \left[e^{(2b-a)/\nu_{01}}\frac{X_1(\nu_{01})}{X_1(-\nu_{01})}\phi_{10+}(\mu) + \phi_{10-}(\mu)e^{a/\nu_{01}}\right] + b_0[\phi_{20+}(\mu)e^{-a/\nu_{02}} + \phi_{20-}(\mu)e^{a/\nu_{02}}] - \int_0^1 B(\nu)e^{-a/\nu}\phi_{2\nu}(\mu)\,d\nu - \int_{-1}^{0} A(\nu)e^{-a/\nu}\phi_{1\nu}(\mu)\,d\nu + \frac{S}{1-C_2}, \tag{63}$$

and where we have used Eqs. (59) and (61).

In keeping with the wide-region approximation, we ignore the two integral terms in $\psi(\mu)$. Using the "two-media" full-range orthogonality relations we then easily find that

$$a_{0-} = \frac{2SX_1(-\nu_{01})}{(1-C_2)C_1\nu_{01}X_1(\nu_{01})\chi(\nu_{01})\,\Delta} \times \left[\frac{\chi(\nu_{02})}{\chi(-\nu_{02})}e^{-a/\nu_{02}} - e^{a/\nu_{02}}\right], \tag{64}$$

$$b_0 = \frac{2S}{(1 - C_2)C_2\nu_{02}\chi(-\nu_{02})\,\Delta} \times \left[e^{2(b-a)/\nu_{01}} - \frac{X_1(-\nu_{01})\chi(-\nu_{01})e^{a/\nu_{01}}}{X_1(\nu_{01})\chi(\nu_{01})}\right], \quad (65)$$

where

$$\Delta = \nu_{01}\left[e^{a/\nu_{02}} - \frac{\chi(\nu_{02})}{\chi(-\nu_{02})}e^{-a/\nu_{02}}\right] \times \left[e^{(2b-a)/\nu_{01}} + \frac{X_1(-\nu_{01})\chi(-\nu_{01})}{X_1(\nu_{01})\chi(\nu_{01})}e^{a/\nu_{01}}\right] + \nu_{02}\left[e^{a/\nu_{02}} + \frac{\chi(\nu_{02})}{\chi(-\nu_{02})}e^{-a/\nu_{02}}\right] \times \left[e^{(2b-a)/\nu_{01}} - \frac{X_1(-\nu_{01})\chi(-\nu_{01})}{X_1(\nu_{01})\chi(\nu_{01})}e^{a/\nu_{01}}\right], \quad (66)$$

$$A(\nu) = \frac{e^{a/\nu}(1 - C_1)(\nu_{01}^2 - \nu^2)X_1(-\nu)}{X_2(-\nu)} \times g(C_1, \nu)f(\nu), \qquad \nu > 0, \quad (67)$$

and

$$B(\nu) = e^{a/\nu}(1 - C_2)(\nu_{02}^2 - \nu^2)\frac{X_2(\nu)}{X_1(\nu)} \times g(C_2, \nu)f(\nu), \qquad \nu < 0 \quad (68)$$

where

$$f(\nu) = b_0[\phi_{20+}(\nu)\chi(\nu_{02})e^{-a/\nu_{02}} + \phi_{20-}(\nu)\chi(-\nu_{02})e^{a/\nu_{02}}] - a_{0-}\left[e^{(2b-a)/\nu_{01}}\frac{X_1(\nu_{01})}{X_1(-\nu_{01})}\phi_{10+}(\nu)\chi(\nu_{01}) + \phi_{10-}(\nu)\chi(-\nu_{01})e^{a/\nu_{01}}\right]. \quad (69)$$

Equations (59), (60), (61), (64), (65), (67), and (68) comprise the first-order solution of this problem. One can now readily check that the approximations made in the expressions for $\alpha(\mu)$ and $\psi(\mu)$ were consistent ones, i.e., the integral terms do indeed vanish rapidly for large a and b.

The comments in the previous chapter about the discontinuity of the solution at the interfaces apply equally well here. There does not, however, appear to be any special significance in the first-order solution when $C_1 = C_2$.

V. REPEATING INFINITE LATTICE

The last problem to be considered is that of a repeating infinite lattice composed of alternating source slabs and source-free slabs, with configuration as shown in Fig. 3.

Because of symmetry we may confine our attention to a half-cell of the lattice, i.e., $0 \leq x \leq a + \frac{1}{2}d$.

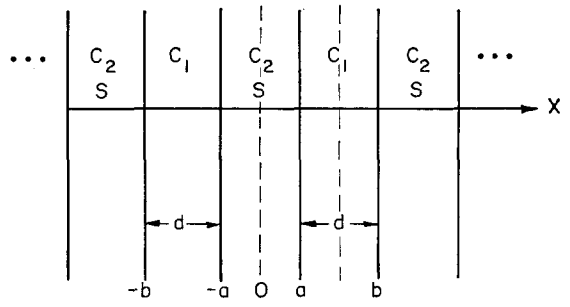

FIG. 3. Coordinate system for a repeating infinite lattice.

The equations to be solved in this region are

$$\mu\frac{\partial\psi_1(x,\mu)}{\partial x} + \psi_1(x,\mu) = \frac{C_1}{2}\int_{-1}^{1}\psi_1(x,\mu')\,d\mu', \quad a < x < \frac{a+b}{2}, \quad (70)$$

$$\mu\frac{\partial\psi_2(x,\mu)}{\partial x} + \psi_2(x,\mu) = \frac{C_2}{2}\int_{-1}^{1}\psi_2(x,\mu')\,d\mu' + S, \quad 0 < x < a, \quad (71)$$

where

(1) $\psi_2(0,\mu) = \psi_2(0,-\mu)$,

(2) $\psi_1[(b+a)/2,\mu] = \psi_1[(b+a)/2,-\mu]$,

(3) $\psi_1(a,\mu) = \psi_2(a,\mu)$.

Solutions are

$$\psi_1(x,\mu) = a_{0+}\phi_{10+}(\mu)e^{-x/\nu_{01}} + a_{0-}\phi_{10-}(\mu)e^{x/\nu_{01}} + \int_{-1}^{1}A(\nu)\phi_{1\nu}(\mu)e^{-x/\nu}\,d\nu, \quad (72)$$

$$\psi_2(x,\mu) = b_{0+}\phi_{20+}(\mu)e^{-x/\nu_{02}} + b_{0-}\phi_{20-}(\mu)e^{x/\nu_{02}} - \int_{-1}^{1}B(\nu)\phi_{2\nu}(\mu)e^{-x/\nu}\,d\nu + \frac{S}{1 - C_2}. \quad (73)$$

Symmetry conditions (1) and (2) give

$$b_{0+} = b_{0-} \triangleq b_0, \quad (74)$$

$$B(\nu) = B(-\nu), \quad (75)$$

$$a_{0+} = a_{0-}e^{(b+a)/\nu_{01}}, \quad (76)$$

$$A(\nu) = A(-\nu)e^{(b+a)/\nu}. \quad (77)$$

Using these relations in Eqs. (72) and (73) and applying the continuity condition (3), we obtain

$$\psi(\mu) = \int_0^1 A(\nu)\phi_{1\nu}(\mu)e^{-a/\nu} + \int_{-1}^{0}B(\nu)\phi_{2\nu}(\mu)e^{-a/\nu}\,d\nu \quad (78)$$

where

$$\psi(\mu) = b_0[\phi_{20+}(\mu)e^{-a/\nu_{02}} + \phi_{20-}(\mu)e^{a/\nu_{02}}] - a_{0+}[\phi_{10+}(\mu)e^{-a/\nu_{01}} + \phi_{10-}(\mu)e^{-b/\nu_{01}}] - \int_{-1}^{0} A(\nu)\phi_{1\nu}(\mu)e^{-a/\nu}\,d\nu - \int_{0}^{1} B(\nu)\phi_{2\nu}(\mu)e^{-a/\nu}\,d\nu + \frac{S}{1-C_2}. \quad (79)$$

As before we ignore the contribution of the two integral terms to $\psi(\mu)$ in Eq. (79) in the case of wide regions. Then, using the "two-media" orthogonality relations, the first-order solution follows straightforwardly:

$$b_0 = \frac{2S}{\Delta(1-C_2)C_2\nu_{02}\chi(-\nu_{02})} \times \left[e^{-a/\nu_{01}} - e^{-b/\nu_{01}}\frac{\chi(-\nu_{01})}{\chi(\nu_{01})}\right], \quad (80)$$

$$a_{0+} = \frac{2S}{\Delta(1-C_2)C_1\nu_{01}\chi(\nu_{01})} \times \left[e^{-a/\nu_{02}}\frac{\chi(\nu_{02})}{\chi(-\nu_{02})} - e^{a/\nu_{02}}\right], \quad (81)$$

where

$$\Delta = \nu_{01}\left[e^{a/\nu_{02}} - \frac{\chi(\nu_{02})}{\chi(-\nu_{02})}e^{-a/\nu_{02}}\right] \times \left[e^{-a/\nu_{01}} + \frac{\chi(-\nu_{01})}{\chi(\nu_{01})}e^{-b/\nu_{01}}\right] + \nu_{02}\left[e^{a/\nu_{02}} + \frac{\chi(\nu_{02})}{\chi(-\nu_{02})}e^{-a/\nu_{02}}\right] \times \left[e^{-a/\nu_{01}} - \frac{\chi(-\nu_{01})}{\chi(\nu_{01})}e^{-b/\nu_{01}}\right], \quad (82)$$

and

$$A(\nu) = e^{a/\nu}(1-C_1)(\nu_{01}^2 - \nu^2)\frac{X_1(-\nu)}{X_2(-\nu)} \times g(C_1,\nu)f(\nu), \qquad \nu > 0, \quad (83)$$

$$B(\nu) = e^{a/\nu}(1-C_2)(\nu_{02}^2 - \nu^2)\frac{X_2(\nu)}{X_1(\nu)} \times g(C_2,\nu)f(\nu), \qquad \nu < 0, \quad (84)$$

where

$$f(\nu) = b_0[\phi_{20+}(\nu)\chi(\nu_{02})e^{-a/\nu_{02}} + \phi_{20-}(\nu)\chi(-\nu_{02})e^{a/\nu_{02}}] - a_{0+}[\phi_{10+}(\nu)\chi(\nu_{01})e^{-a/\nu_{01}} + \phi_{10-}(\nu)\chi(-\nu_{01})e^{-b/\nu_{01}}]. \quad (85)$$

One easily verifies that the first-order approximation is consistent. With the usual comments about continuity at the interfaces, we proceed to the numerical analysis.

VI. NUMERICAL ANALYSIS

For calculational purposes we shall consider not the angular density, but rather the neutron density, obtained by integrating the previously derived solutions over all μ. Making use of the normalization condition of the normal modes,[3] i.e.,

$$\int_{-1}^{1}\phi_\nu(\mu)\,d\mu = 1, \qquad \text{all } \nu, \quad (86)$$

as well as the appropriate symmetry conditions, we find that the solutions of these problems may be put into the following forms:

(1) the slab with infinite reflectors:

$$\rho_1(x) = a_0e^{-x/\nu_{01}} + \int_0^1 A(\nu)e^{-x/\nu}\,d\nu, \quad (87)$$

$$\rho_2(x) = \frac{2S}{1-C_2} + 2b_0\cosh x/\nu_{02} - 2\int_0^1 B(-\nu)\cosh x/\nu\,d\nu; \quad (88)$$

(2) the slab with finite reflectors:

$$\rho_1(x) = a_{0+}e^{-x/\nu_{01}} + a_{0-}e^{x/\nu_{01}} + \int_0^1 A(\nu)e^{-x/\nu}\,d\nu + \int_0^1 A(-\nu)e^{+x/\nu}\,d\nu, \quad (89)$$

$$\rho_2(x) = \frac{2S}{1-C_2} + 2b_0\cosh x/\nu_{02} - 2\int_0^1 B(-\nu)\cosh x/\nu\,d\nu; \quad (90)$$

(3) the uniform, infinite lattice:

$$\rho_1(x) = a_{0+}\{e^{-x/\nu_{01}} + e^{-(b+a-x)/\nu_{01}}\} + \int_0^1 A(\nu)e^{-x/\nu}\,d\nu + \int_0^1 A(\nu)e^{-(b+a-x)/\nu}\,d\nu, \quad (91)$$

$$\rho_2(x) = \frac{2S}{1-C_2} + 2b_0\cosh x/\nu_{02} - 2\int_0^1 B(-\nu)\cosh x/\nu\,d\nu \quad (92)$$

where the appropriate discrete and continuum coefficients are given in the previous three sections.

These solutions have been programmed in the FORTRAN language, originally for the IBM 7090 Computer at the University of Michigan, and later for the Philco 2000 Computer at the Knolls Atomic

Power Laboratory. The various integrals were evaluated by means of Simpson's rule, with 200 intervals for $0 \leq \nu \leq 0.99$ and 200 intervals for $0.99 \leq \nu \leq 1.0$. The very fine quadrature near $\nu = 1$ is necessary to properly represent the function $g(C, \nu)$. This function, while vanishing at $\nu = 1$, becomes nearly singular as ν approaches 1, for values of C near 0.[9] The calculation of the integral terms was facilitated by a table of X-functions, which was constructed from an iterative solution of the integral equation for Chandreskhar's H-function.[10] The parameters used in the numerical analysis are shown in Table I.

TABLE I. Parameters and ETC1 options used in numerical evaluation.

C_1	C_2	a	d (Prob. 2)	$d/2$ (Prob. 3)	Angles ($0 \leq \mu \leq 1$)	Mesh points
0.01	0.99	5.0	5.0	5.0	8	1000
0.01	0.99	1.0	1.0	1.0	12	500
0.01	0.99	0.1	0.1	0.1	16	500
0.99	0.01	5.0	5.0	5.0	6	1000
0.99	0.01	1.0	1.0	1.0	10	500
0.99	0.01	0.1	0.1	0.1	14	500

As a comparative standard, we have used the ETC1 program[11] which performs a numerical integration of the transport equation by the double-Gaussian quadrature method. This program has provision for up to 18 angles on the range $0 \leq \mu \leq 1$ and as many as 2000 spatial mesh points. Only the second two problems, i.e., the slab with finite reflectors and the lattice, were analyzed with the ETC1 program. The first problem, the slab with infinite reflectors, was not analyzed with a transport code because of the great difficulty in obtaining highly accurate solutions in infinite media. The angular and spatial options used were the same for corresponding cases of the two problems considered. These are shown in the last two columns of Table I.

Tables II, III, IV, and V show the resultant values of the neutron density for the two types of solutions as well as percent error, defined by

$$\text{percent error} = \frac{\rho_{\text{CASE}} - \rho_{\text{ETC1}}}{\rho_{\text{ETC1}}} \times 100.$$

Percent errors less than 0.01 are regarded as not meaningful and are therefore not indicated.

It is noted that in the four problems for which "wide-region" comparisons are made, i.e., regions five and ten mean free paths thick, the first-order solutions are essentially exact in comparison with the ETC1 solutions. One exception to this is observed in

TABLE II. Slab with finite reflectors. $C_1 = 0.01$, $C_2 = 0.99$; $a = d = 5.0$, 1.0, 0.1, respectively.

x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error
0	34.4729	34.4730	<0.01	0	3.98055	3.98956	−0.226	0	0.259611	0.37581	−30.912
1.0	33.4940	33.4941	<0.01	0.2	3.91276	3.92179	−0.230	0.02	0.256856	0.373132	−31.162
2.0	30.5255	30.5255	<0.01	0.4	3.70636	3.71546	−0.245	0.04	0.248362	0.364963	−31.949
3.0	25.4623	25.4624	<0.01	0.6	3.34994	3.35915	−0.274	0.06	0.233297	0.350519	−33.442
4.0	18.0635	18.0635	<0.01	0.8	2.81092	2.82045	−0.338	0.08	0.209475	0.328025	−36.141
5.0	6.65515	6.65514	<0.01	1.0	1.88856	1.89886	−0.542	0.10	0.165265	0.286303	−42.276
5.0	6.65497	6.65514	<0.01	1.0	1.90702	1.89886	+0.430	0.10	0.406634	0.286303	+42.029
6.0	1.16268	1.16269	<0.01	1.2	1.12392	1.12006	0.345	0.12	0.351025	0.244212	43.738
7.0	0.304931	0.304941	<0.01	1.4	0.764690	0.762965	0.226	0.14	0.315063	0.219366	43.624
8.0	0.088226	0.088228	<0.01	1.6	0.542456	0.541787	0.123	0.16	0.286321	0.200094	43.093
9.0	0.026871	0.026869	<0.01	1.8	0.394061	0.393860	0.051	0.18	0.262140	0.184321	42.219
10.0	0.008454	0.008426	<0.331	2.0	0.290870	0.290545	0.112	0.20	0.241241	0.170958	41.111

TABLE III. Slab with finite reflectors. $C_1 = 0.99$, $C_2 = 0.01$; $a = d = 5.0$, 1.0, 0.1, respectively.

x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error
0	1.00979	1.00979	<0.01	0	0.920314	0.918188	0.232	0	0.413014	0.308241	33.991
1.0	1.00956	1.00956	<0.01	0.2	0.916288	0.913981	0.252	0.02	0.412007	0.306604	34.378
2.0	1.00848	1.00848	<0.01	0.4	0.903491	0.900590	0.322	0.04	0.408910	0.301561	35.598
3.0	1.00457	1.00457	<0.01	0.6	0.879376	0.875293	0.466	0.06	0.403451	0.292667	37.853
4.0	0.989178	0.989178	<0.01	0.8	0.837577	0.831273	0.758	0.08	0.394935	0.278779	41.666
5.0	0.892898	0.892892	<0.01	1.0	0.750440	0.738626	1.599	0.10	0.379904	0.252762	50.301
5.0	0.892889	0.892892	<0.01	1.0	0.757451	0.738626	2.549	0.10	0.570479	0.252762	125.698
6.0	0.680254	0.680253	<0.01	1.2	0.639334	0.618874	3.306	0.12	0.541144	0.225076	140.427
7.0	0.513208	0.513209	<0.01	1.4	0.542475	0.523157	3.693	0.14	0.515982	0.206919	149.364
8.0	0.363058	0.363057	<0.01	1.6	0.449114	0.431811	4.007	0.16	0.491499	0.191307	156.916
9.0	0.223110	0.223109	<0.01	1.8	0.354155	0.339423	4.340	0.18	0.466438	0.176893	163.684
10.0	0.075405	0.075398	<0.01	2.0	0.241172	0.229787	4.955	0.20	0.437390	0.161731	170.443

[11] Written by Charles Dawson, David Taylor Model Basin, Washington, D. C.

TABLE IV. Infinite lattice. $C_1 = 0.01$, $C_2 = 0.99$; $a = d/2 = 5.0, 1.0, 0.1$, respectively.

x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error
0	34.4729	34.4731	<0.01	0	3.98055	4.24116	6.145	0	0.25961	1.20937	78.533
1.0	33.4940	33.4942	<0.01	0.2	3.91276	4.17237	6.223	0.02	0.25686	1.20591	78.700
2.0	30.5255	30.5256	<0.01	0.4	3.70636	3.96269	6.469	0.04	0.24836	1.19518	79.220
3.0	25.4623	25.4625	<0.01	0.6	3.34994	3.59967	7.938	0.06	0.23330	1.17597	80.161
4.0	18.0635	18.0636	<0.01	0.8	2.81092	3.04810	7.781	0.08	0.20947	1.14545	81.713
5.0	6.65515	6.65522	<0.01	1.0	1.88856	2.09249	9.744	0.10	0.16526	1.08560	84.777
5.0	6.65500	6.65522	<0.01	1.0	1.97984	2.09249	5.384	0.10	0.57317	1.08560	47.202
6.0	1.16278	1.16280	<0.01	1.2	1.21880	1.29005	5.523	0.12	0.52961	1.02678	48.420
7.0	0.305226	0.305239	<0.01	1.4	0.888952	0.942894	5.721	0.14	0.506912	0.997607	49.187
8.0	0.089116	0.089119	<0.01	1.6	0.706157	0.750422	5.899	0.16	0.492843	0.979537	49.686
9.0	0.029591	0.029593	<0.01	1.8	0.611288	0.650505	6.029	0.18	0.485008	0.969542	49.976
10.0	0.016907	0.016908	<0.01	2.0	0.581740	0.619378	6.077	0.20	0.482482	0.966336	50.074

TABLE V. Infinite lattice. $C_1 = 0.99$, $C_2 = 0.01$; $a = d/2 = 5.0, 1.0, 0.1$, respectively.

x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error	x	ρ_{CASE}	ρ_{ETC1}	% Error
0	1.00993	1.00993	<0.01	0	1.00422	1.00384	0.038	0	1.00523	1.00034	0.489
1.0	1.00980	1.00980	<0.01	0.2	1.00393	1.00353	0.040	0.02	1.00520	1.00030	0.490
2.0	1.00920	1.00920	<0.01	0.4	1.00297	1.00252	0.045	0.04	1.00512	1.00020	0.492
3.0	1.00702	1.00702	<0.01	0.6	1.00112	1.00057	0.055	0.06	1.00498	1.00002	0.496
4.0	0.998356	0.998356	<0.01	0.8	0.997790	0.997068	0.072	0.08	1.00475	0.999728	0.502
5.0	0.942879	0.942879	<0.01	1.0	0.990103	0.988961	0.115	0.10	1.00431	0.999131	0.518
5.0	0.942878	0.942879	<0.01	1.0	0.991025	0.088961	0.209	0.10	1.00843	0.999131	0.931
6.0	0.827641	0.827644	<0.01	1.2	0.981708	0.979308	0.245	0.12	1.00799	0.998523	0.948
7.0	0.752906	0.752912	<0.01	1.4	0.976263	0.973736	0.260	0.14	1.00774	0.998218	0.954
8.0	0.701763	0.701770	<0.01	1.6	0.972663	0.970069	0.267	0.16	1.00759	0.998025	0.958
9.0	0.671777	0.671786	<0.01	1.8	0.970578	0.967950	0.272	0.18	1.00751	0.997916	0.961
10.0	0.661890	0.661899	<0.01	2.0	0.969893	0.967256	0.273	0.20	1.00748	0.997881	0.962

the case of the slab with finite reflectors, with $C_1 = 0.01$, $C_2 = 0.99$. Here it is noted that a discrepancy arises very near to the right-hand boundary, of approximately 0.3% in magnitude. This is probably explained by the fact that, in this particular problem, the first-order continuum coefficient $A(-\nu)$, $\nu > 0$ vanishes as $C \to 0$. A closer analysis reveals, however, that the higher-order contributions to this term do not vanish in this limiting case. Since essentially the entire solution is contained in the continuum when $C \to 0$, the neglect of the term $\int_0^1 A(-\nu)e^{x/\nu}\,d\nu$ can contribute an error near the right-hand boundary. This difficulty does not arise in the lattice problem.

The very thin region solutions, i.e., those for regions 0.1 and 0.2 mean free paths thick, appear to break down completely. Solutions may be in error by a factor of 3 to 4, being at times either uniformly high or uniformly low. While difficult to prove analytically, this behavior indicates that the first-order approximation does not conserve neutrons.

The problems which deal with region widths which are one and two mean free paths thick appear to indicate the approximate lower bounds on the of the wide-region approximation. Here the percent errors range from less than 0.1% to almost 10%. Discontinuities at the interface become significant, with observed errors as high as four to five percent. Again the solution may be uniformly high or low.

TABLE VI. Average % error for various values of Σ_s/Σ_t and region thicknesses.

C_2	C_1	Interface coordinate	Right-hand coordinate	⟨% Error⟩ Prob. 2	⟨% Error⟩ Prob. 3
0.01	0.99	5.0	10.0	<0.01	<0.01
0.01	0.99	1.0	2.0	2.173	0.157
0.01	0.99	0.1	0.2	95.037	0.725
0.99	0.01	5.0	10.0	<0.01	<0.01
0.99	0.01	1.0	2.0	0.243	6.405
0.99	0.01	0.1	0.2	38.329	64.715

These results are summarized in Table VI, which shows the behavior of the average error, defined by

$$\langle \text{error} \rangle = \frac{1}{L}\int_0^L |\text{error}|\,dx$$

for the various combinations of parameters. L is the distance from $x = 0$ to the right-hand boundary or symmetry plane.

VII. CONCLUSIONS

It has been demonstrated that Case's method can be applied to the solution of multiregion problems in

slab geometry. First-order solutions have been obtained with relative ease, based upon a wide-region approximation. Numerical analysis and comparison with transport computer codes indicates that such solutions are essentially exact for region thicknesses of five mean free paths or more. Errors of 10% or less are observed for region widths of two mean free paths, indicating an approximate dividing line for the validity of the wide region approximation. Such solutions appear to be greatly advantageous in treating problems with large homogeneous regions, for which numerical transport methods begin to break down because of mesh spacing limitations.

Furthermore, an interesting parallel with the work of Selengut[6] and Pomraning and Clark[7] is observed. These authors have attempted to improve P_1 diffusion theory by using the transport diffusion length and allowing the current and/or the flux to be discontinuous at an interface. This procedure, while predicting the correct exponential behavior, does not necessarily give the magnitudes of the exponential terms correctly. Alternatively, one could use the discrete coefficients obtained by Case's method to form the asymptotic solutions. Although the boundary conditions are more difficult to apply, the exact asymptotic solutions are obtained if the regions are sufficiently wide.

ACKNOWLEDGMENTS

The use of the Computing Center at the University of Michigan, where much of the early numerical analysis was performed, is gratefully acknowledged. I am also indebted to Professor P. F. Zweifel, Professor G. C. Summerfield and Dr. N. C. Francis, for many stimulating discussions, and to Charles Dawson of the David Taylor Model Basin for his aid in performing the ETC1 calculations.

$V\theta$-Bound State and Uniqueness in the Three-Particle Sector of the Lee Model*

ANTONIO PAGNAMENTA†
Department of Physics and Astronomy, University of Maryland, College Park, Maryland
(Received 13 May 1965)

If the coupling constant is large enough, but still smaller than its critical value, there exists a $V\theta$-bound state in the Lee model. We find the bound-state wavefunction via the solution of the homogeneous Källen–Pauli equation. The condition for the existence of a $V\theta$-bound state is seen to give a statement about the uniqueness of the scattering solutions of both the $V\theta$- and $N2\theta$-integral equations. There exists no $N2\theta$-bound state if the V particle is unstable.

I. INTRODUCTION

BOUND-state problems in the Lee model have been discussed at several occasions in the literature. They can be grouped into three classes and have been solved in the order of their complexity. The VN-bound state has been investigated first[1] and most completely.[2] The V particle can be considered as $N\theta$-bound state either for part of the time[3] ($Z \neq 0$) or entirely ($Z = 0$).[4] The $V\theta/N2\theta$-bound state represents the problem which has the most realistic structure due to the fact that the V particle is the only one which undergoes renormalization. The possibility of its existence has been pointed out by investigating the structure of the $V\theta$ amplitude[5,6] which was derived by Amado.[7]

Earlier[8,9] the complete solution of the three-particle sector in the Lee model has been given under the condition that the coupling constant be small enough so that no bound state exists. It was pointed out in I that the presence of a bound state will not affect the scattering solutions of this sector. The scattering states will, however, no longer form a complete set. Also the question of uniqueness was left open.

In this note we address ourselves to the bound-state problem in the $V\theta/N2\theta$ sector. If the coupling constant is large enough but still less than its critical value, which would lead to a non-Hermitian Hamiltonian, there exists a $V\theta/N2\theta$-bound state. We show this by an analysis of the common denominator function of the three-particle amplitudes, using the form we have derived earlier.[9] We find the bound-state wavefunction via the solution of the homogeneous Källen–Pauli equation. To do this we give a simple direct method to solve this type of integral equation. This technique can be applied to more general cases. By relating inhomogeneous and homogeneous equations we show that all the solutions of this sector are unique.

In Sec. II we give precise conditions for which a $V\theta$-bound state exists. In Sec. III we derive its state vector for convenience in the realm of the Tamm–Dancoff method. The solution of the integral equation is given in Sec. IV together with a proof of uniqueness for both scattering and bound-state solutions. The $N2\theta$ sector, i.e., scattering amplitude and the existence of an $N2\theta$-bound state

* Work supported in part by the U. S. Air Force under contract AFOSR 500-64.

† Present address: Rutgers, The State University, New Brunswick, New Jersey.

[1] S. Weinberg, Phys. Rev. 102, 285 (1956).

[2] N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) 25, 803 (1961).

[3] G. Källén and W. Pauli, Kgl. Danske Videnskab. Selskab Mat.-Fys. Medd. 30, No. 7 (1955).

[4] J. C. Houard and B. Jouvet, Nuovo Cimento 18, 466 (1960).

[5] M. T. Vaughn, R. Aaron, and R. D. Amado, Phys. Rev. 124, 1258 (1961). H. Ezawa, K. Kikkawa, and H. Umezawa, Nuovo Cimento 23, 751 (1962). H. Chew, Phys. Rev. 132, 2756 (1963). This last author derives, but does not solve, the $V\theta$ equations for a spin-dependent interaction. We have extended our method of solution to this case and hope to report it in another context.

[6] T. Muta, Progr. Theoret. Phys. (Kyoto) 33, 666 (1965), has independently obtained the wavefunction for the $V\theta$ bound state. His method is very interesting because he relates the expansion coefficients to matrix elements of the V-current for which one can, following Amado, obtain a Muskhelisvili equation. Our result, although given in quite different form, agrees with Muta's. To derive the bound-state condition he uses however the full $V\theta$ amplitude instead of only its denominator. The numerator of the $V\theta$ amplitude [Eq. (6) in I] has a zero for $Z < \frac{1}{2}$. This explains why the value $Z = \frac{1}{2}$ plays a special role in Muta's work. We may emphasize that the value of the coupling constant for which $Z = \frac{1}{2}$ and the value for which $D(\mu) \leq 0$, the bound-state condition, are independent.

[7] R. D. Amado, Phys. Rev. 122, 697 (1961).

[8] R. D. Amado and R. P. Kenschaft, J. Math. Phys. 5, 1340 (1964).

[9] A. Pagnamenta, J. Math. Phys. 6, 955 (1965), hereafter referred to as I. We are using the same notation and Hamiltonian as in I. Also: Im $G^+(\omega) = 4\pi^2 f^2(\omega)k(\omega)$ and $f^2(\omega) = [g^2/(2\pi)^2][u(\omega)/(2\omega)^{\frac{1}{2}}]$, where $u(\omega)$ is the ordinary cutoff function normalized to $u(k^2 = 0) = u(\omega = \mu) = 1$.

in a modified Lee model where the V particle is unstable, is discussed in Sec. V. In Appendix A we give a graphical method to derive the relevant equation. In Appendix B we show that the bound state has a nonvanishing norm. We verify the solution of the $N2\theta$-integral equation for the unstable V in Appendix C. That the spatial dependence of the bound-state wavefunction shows exponential decrease at large distance is shown in Appendix D.

II. EXISTENCE OF A $V\theta$-BOUND STATE

It has been observed some time ago[5] that the $V\theta$ elastic amplitude can develop a dynamical pole below the physical threshold thus indicating a bound state[10]. In I we have shown that all the amplitudes of the $V\theta/N2\theta$ sector of the Lee models contain the denominator function[9]

$$D^+(\omega) = 1 - H^+(\omega)A^+(\omega), \tag{1}$$

where[3]

$$H(z) = zG(z), \tag{2}$$

$$G(z) = 1 + \frac{z}{\pi}\int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')}{\omega'(\omega' - z)}\,d\omega', \tag{3}$$

and

$$A^+(\omega) = \frac{1}{\pi}\int_\mu^\infty \frac{d\omega'}{H^-(\omega - \omega')} \operatorname{Im}\frac{1}{H^+(\omega')}. \tag{4}$$

Here we have indicated the boundary values of our analytic functions as $z \to \omega \pm i\epsilon$ by superscripts $(\pm)$. For example, $G^+(\omega) = G(\omega + i\epsilon)$; hence $G^-(\omega_0 - \omega) = G(\omega_0 - \omega + i\epsilon)$. From (3) we read off: $G(0) = 1$ and find the wavefunction renormalization constant Z by

$$\lim_{\omega\to\infty} G(\omega) = Z. \tag{5}$$

Let g_c be the critical value of the coupling constant for which $Z = 0$. Beyond the ordinary ghost state appears[3] as a zero of $G(\omega)$. Here we assume that the coupling constant be subcritical, $g^2 < g_c^2$, so that $0 < Z \leq 1$.

Both $H(\omega)$ and $A(\omega)$ are real for $\omega < \mu$. $H(0) = 0$ and therefore $D(0) = 1$. Neither $H(\omega)$ nor $A(\omega)$ is singular in the interval $0 \leq \omega \leq \mu$. From (3) follows that $H(\mu) = \mu G(\mu) > 0$. Writing (4) out as

$$A(\omega) = \frac{1}{\pi}\int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{|H(\omega')|^2}\frac{d\omega'}{(\omega' - \omega)G(\omega - \omega')}$$

we see that $H(\mu) > 0$. Hence the product

[10] That a zero on the real axis below threshold in the denominator of the scattering amplitude indicates a bound state was pointed out by R. Jost, Helv. Phys. Acta. **30**, 409 (1957).

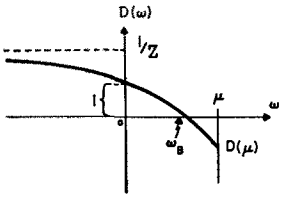

FIG. 1. $D(\omega)$ for g^2 large enough so that $D(\mu) < 0$, but $g^2 < g_c^2$ so that $Z > 0$. For $Z^{-1} < 0$ there is no intersection.

$H(\mu)A(\mu) > 0$. The bound-state condition

$$D(\omega_B) = 0, \qquad 0 < \omega_B < \mu \tag{6}$$

will be fulfilled if $D(\mu) < 0$ (Fig. 1) or equivalently if

$$H(\mu)A(\mu) > 1. \tag{6'}$$

Since $A(\omega)$ contains a factor g^2 in $\operatorname{Im} H^{-1}(\omega)$[8] we expect for a given cutoff function (6') to be fulfilled for large g^2. This will give a $V\theta$-bound state and not a ghost if still $g^2 < g_c^2$. For $\omega < \mu$ the product $H(\omega)A(\omega)$ is monotonically increasing and using (5) in (4) one easily finds (Fig. 1)

$$1 - H(-\infty)A(-\infty) = Z^{-1}.$$

That any finite value of $D(\omega) < 1$, hence of $D(\omega) < 0$, can be reached for $g^2 < g_c^2$ is seen as follows. $H(\mu)$ is finite for all finite $g^2 < g_c^2$. Clearly for $g^2 < g_c^2$, $A(\mu)$ is finite. For $g^2 \geq g_c^2$ the zero of $G(\omega)$ which corresponds to the ghost makes the integral of $A(\omega)$ divergent because the integrand contains the real denominator $G(\omega - \omega')$ and the pole at ω' given by $\omega - \omega' = \lambda < 0$ is in the range of integration. Therefore for $g^2 > g_c^2$, $A(\omega)$ for $\omega \leq \mu$ is divergent and so is the product $H(\mu)A(\mu)$. Hence it can obtain any finite value already for $g^2 < g_c^2$.

This shows also that our discussion is only valid for $Z \geq 0$. The expressions (1) have been obtained by explicitly assuming the no ghost condition which entered the solutions of the integral equations through the statement that $G(\omega)$ had no zero. This explains why there is apparently no intersection in Fig. 1 if $Z < 0$.

In principle one can invert the unfortunately transcendent equation (6) to find the binding energy $E_B = \mu - \omega_B$ as a function of g^2. For $Z \to 0$ we observe that $\omega_B \to 0$ and the $V\theta$-bound state moves into the V particle. If the coupling constant becomes too weak to cause a bound state it may still cause $\operatorname{Re} D(\omega) = 0$ for $\omega > \mu$. If the parameters of the model are favorable this shows up as a resonance in $V\theta$ scattering.

In Sec. V we show that there is no bound state in the Lee model that has only one unstable V particle.[11] We do, however, expect a bound state in the model modified to contain a stable V_1 and

[11] V. Glaser and G. Källén, Nucl. Phys. **2**, 706 (1956).

an unstable V_2.[12] Depending on ratio and size of the two coupling constants we expect none, one, or two bound states in the Lee model with two stable V particles.[13]

III. THE BOUND-STATE WAVEFUNCTION

The $V\theta$-bound state which exists under condition (6) will be an eigenstate of the total Hamiltonian. It can be written as a linear combination of the bare states $|V\theta\rangle$ and $|N\theta_1\theta_2\rangle$ which we take to be symmetrized and to have norm one,

$$|B\rangle = \int d^3k'\, Z^{\frac{1}{2}}\bar{\alpha}(k')\, |V\theta_{k'}\rangle + \int d^3k'\, d^3k''\, \beta(k', k'')\, |N\theta_{k'}\theta_{k''}\rangle. \quad (7)$$

The eigenvalue equation $(\mathfrak{H} - m_B)\, |B\rangle = 0$, where $\mathfrak{H}$ is the Lee model Hamiltonian given in I, shows that the coefficients $\alpha(k) = Z^{-\frac{1}{2}}\langle B \mid V\theta_k\rangle$ and $\beta(k_1, k_2) = \langle B \mid N\theta_1\theta_2\rangle$ obey the relations

$$H^-(\omega_B - \omega)\alpha(k) = -f(\omega)\int d^3k'\, \frac{f(\omega')\alpha(k')}{\omega' - \omega_B + \omega} \quad (8)$$

$$(\omega_1 + \omega_2 - \omega_B)\beta(k_1, k_2) = \tfrac{1}{2}[f(\omega_1)\alpha(k_2) + f(\omega_2)\alpha(k_1)]. \quad (9)$$

m_B is the mass of the bound state and $\omega_B = m_B - m$, where $m = m_N = m_V$. Equation (8) is the homogeneous Källén–Pauli equation. Once we have solved it, $\beta(k_1,k_2)$ follows from (9). A convenient but heuristic method to derive Eqs. (8) and (9) using graphs is given in Appendix A.

In the next section we decide the question of uniqueness for the entire three-particle sector and at the same time solve the homogeneous equation. There we find [Eqs. (36) and (39)]

$$\alpha(\omega) = N[f(\omega)/(\omega - \omega_B)]J(\omega), \quad (10)$$

where

$$J(\omega) = \frac{1}{\pi}\int_\mu^\infty \frac{d\omega'}{\omega' - \omega_B + \omega}\,\frac{1}{H^-(\omega_B - \omega')}\,\mathrm{Im}\,\frac{1}{G^+(\omega')}. \quad (11)$$

As expected this has a pole at $\omega = \omega_B$ and solves Eq. (8) if (6) is fulfilled. From (9) we get

$$\beta(\omega_1, \omega_2) = \frac{Nf(\omega_1)f(\omega_2)}{2(\omega_1 + \omega_2 - \omega_B)} \times \left[\frac{1}{\omega_2 - \omega_B}\,J(\omega_2) + \frac{1}{\omega_1 - \omega_B}\,J(\omega_1)\right]. \quad (12)$$

[12] P. K. Srivastava, Phys. Rev. 131, 461 (1963).
[13] P. K. Srivastava, Phys. Rev. 128, 2906 (1962).

It is interesting to verify that $|B\rangle$ is normalizable. The normalization constant N is determined by

$$\langle B \mid B\rangle = 1. \quad (13)$$

We show in Appendix B that $N^2 \neq 0$ and positive. Therefore we can take $N > 0$.

Note

$$\langle B \mid B'\rangle = 0 \quad (14)$$

if $m_B' \neq m_B$ because then $|B'\rangle \equiv 0$.

The Fourier transforms of $\alpha(\omega)$ and $\beta(\omega', \omega'')$ in x space are the bound-state wavefunctions. They show exponential decrease at large distances, which we verify in Appendix D for $\alpha(\omega)$.

IV. SOLUTIONS OF THE EQUATIONS AND UNIQUENESS

In this section we solve the homogeneous $V\theta$ equation (8) and decide the question of uniqueness for all the solutions in the $V\theta/N2\theta$ vector. To this we have to relate homogeneous and inhomogeneous equations. We therefore start by giving a simple strictly deductive method for solving the inhomogeneous $V\theta$ equation.[14] This method is easily generalized to apply to the other integral equations in this sector as well as to the one obtained in the case of the unstable V particle.

The inhomogeneous integral equation for $V\theta$ scattering[15] can be brought into the form

$$G^-(\omega_0 - \omega)M(\omega) = 1 + \frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}\, G^+(\omega')}{\omega' - \omega_0 - i\epsilon}\cdot\frac{M(\omega')}{\omega' + \omega - \omega_0 - i\epsilon}\,d\omega'. \quad (15)$$

Due to the displacement $\omega \to \omega_0 - \omega$ in the Cauchy denominator this is not a simple Muskhelisvili[16] equation. It can, however, be reduced to a multiple Hilbert problem and we show that in spite of the fact that (15) has not in general a completely continuous kernel and is therefore a singular integral equation a uniqueness theorem similar to the Fredholm alternative is valid. We want the integral in (15) and in the following to exist without subtractions. To this it is sufficient to assume that both

[14] Different methods to obtain solutions of this singular type of integral equations with a displacement in the denominator have been given by G. S. Litvincuk, Izv. Akad. Nauk. SSSR. Ser. Math. 25, 871 (1961); also in Refs 7 and 8. E. Kazes, Pennsylvania State University preprint. Ch. Sommerfield, Yale University preprint. Except in the first of the above references in which, however, a circular contour is considered the homogeneous equation is not discussed. In our discussion we follow to some extent Litvincuk and Kazes.
[15] Equation (62) in Ref. 3 or Eq. (13) in I.
[16] N. J. Muskhelisvili, *Singular Integral Equations* (P. Nordhoff Ltd., Groningen, The Netherlands, 1953).

Im $G(\omega)$ which contains the cutoff and $M(\omega)$ satisfy a Hölder[17] condition. This includes a square cutoff. In special cases this condition can be relaxed.

We can extend the definition of the function $M(\omega)$ in (15) into the complex z plane by writing

$$G(\omega_0 - z)M(z) = 1 + \frac{z}{\pi}\int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')}{\omega' - \omega_0 - i\epsilon}\frac{M^-(\omega')}{\omega' - \omega_0 + z}\,d\omega'. \quad (16)$$

Comparing this with Eq. (15) we see that we recover the sought-for function $M(\omega)$ in the limit $z \to \omega - i\epsilon$ or $M(\omega) \to M^-(\omega) \equiv M(\omega - i\epsilon)$. We have anticipated this in (16) by writing $M^-(\omega)$ under the integral sign.

From (16) we see that $M(z)$ is analytic (regular and bounded) in the entire z plane except for a branch point at $\omega = \omega_0 - \mu$ to which we attach a cut along the negative real axis to $\omega = -\infty$. Further $M(\infty) = \text{const}$.

We now write Eq. (16) twice, once each for the points $z = \omega \pm i\epsilon$ and subtract. This gives

$$G^+(\omega_0 - \omega)[M^+(\omega) - M^-(\omega)] = 2i \operatorname{Im} G^-(\omega_0 - \omega)[M^-(\omega) + M^+(\omega_0 - \omega)]\theta(\omega_0 - \omega - \mu). \quad (17)$$

Equation (17) is usually solved by finding a relation between $M(\omega)$ and $M(\omega_0 - \omega)$.[14] We introduce

$$N(z) = M(z) + M(\omega_0 - z); \quad (18)$$

then for $\omega < \omega_0 - \mu$, Eq. (3) becomes

$$N^+(\omega) - N^-(\omega) = 2i\frac{\operatorname{Im} G^-(\omega_0 - \omega)}{G^+(\omega_0 - \omega)}N^-(\omega)\theta(\omega_0 - \omega - \mu). \quad (19)$$

Due to (22) $N(z)$ has both left- and right-hand cuts. Therefore we replace in (17) ω by $\omega_0 - \omega$. Now we can evaluate it for $\omega > \mu$. In this interval $M^+(\omega) = M^-(\omega)$ and using (18) we obtain

$$N^+(\omega) - N^-(\omega) = -2i\frac{\operatorname{Im} G^+(\omega)}{G^-(\omega)}N^+(\omega)\theta(\omega - \mu). \quad (20)$$

Whether or not there is any space left between $\omega_0 - \mu$ and μ does not matter for this approach; one simply takes the discontinuity across each cut separately. Equations (19) and (20) define a homogeneous Hilbert problem, (two different discontinuities) the solution of which is standard).[16] We find

$$\log N(z) = \frac{1}{2\pi i}\int_{-\infty}^{\omega_0 - \mu}\frac{\log G^+(\omega_0 - \omega') - \log G^-(\omega_0 - \omega')}{\omega' - z}\,d\omega' + \frac{1}{2\pi i}\int_\mu^\infty \frac{\log G^+(\omega') - \log G^-(\omega')}{\omega' - z}\,d\omega' + \log F_1(z). \quad (21)$$

The integrals can be done by contour integration using the reality property of the logarithm and (5) to cancel the contributions of the infinite circles. $F_1(z)$ is an entire analytic function which behaves like a constant at infinity and therefore is just a constant. Thus

$$N(z) = \frac{C_1'}{G(z)G(\omega_0 - z)}. \quad (22)$$

Equation (17) can now be written

$$M^+(\omega) - M^-(\omega) = 2i\frac{\operatorname{Im} G^-(\omega_0 - \omega)}{G^+(\omega_0 - \omega)}N^-(\omega)\theta(\omega_0 - \omega - \mu). \quad (23)$$

$M(z)$ has only the left-hand cut; therefore this Hilbert problem has the solution

$$M(z) = F_2(z) - \frac{1}{\pi}\int_\mu^\infty \frac{d\omega'}{\omega' - \omega_0 + z}\frac{\operatorname{Im} G^+(\omega')}{G^-(\omega')}N^+(\omega_0 - \omega'). \quad (24)$$

For the same reason as above, the entire analytic function $F_2(z)$ is a constant. Collecting we find

$$M(z) = C_2' + \frac{C_1'}{\pi}\int_\mu^\infty \frac{d\omega'}{\omega' - \omega_0 + z} \times \frac{1}{G^-(\omega_0 - \omega')}\operatorname{Im}\frac{1}{G^+(\omega')}. \quad (25)$$

It is convenient to subtract from the integral in (25) its value at $z = 0 - i\epsilon$. Redefining the constants (25) becomes

$$M(z) = C_2 + C_1\frac{z}{\pi} \times \int_\mu^\infty \frac{d\omega'}{(\omega' - \omega_0 + z)H^-(\omega_0 - \omega')}\operatorname{Im}\frac{1}{G^+(\omega')}. \quad (26)$$

Since the inhomogeneous term in (15) had no discontinuity it did not enter the derivation of (26). Evaluating $M(0)$ in (16) and (26) gives

$$M(0) = 1/G^+(\omega_0) = C_2 \quad (27)$$

and the constant C_1 is found easiest evaluating (16) and (26) for $z = \omega_0 - i\epsilon$:

[17] A function $\varphi(x)$ satisfies a Hölder condition of degree k if for any two points x_1, x_2 in the interval L.

$$\varphi(x_1) - \varphi(x_2) < C\,|x_1 - x_2|,$$

where C is an arbitrary constant and $0 \leq k \leq 1$.

$$M^-(\omega_0) = \frac{1}{G^+(\omega_0)} + C_1 \frac{\omega_0}{\pi} \times \int_\mu^\infty \frac{d\omega'}{H^-(\omega_0 - \omega')} \operatorname{Im} \frac{1}{H^+(\omega')} = 1 + \frac{\omega_0}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')}{\omega'(\omega' - \omega_0 - i\epsilon)} M^-(\omega')\, d\omega'. \quad (28)$$

The upper integral is just $A^+(\omega_0)$. The last integral is evaluated by use of (26) and (27).

$$\frac{\omega_0}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')}{\omega'(\omega' - \omega_0 - i\epsilon)} M^-(\omega')\, d\omega' = \frac{\omega_0}{G^+(\omega_0)} \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')\, d\omega'}{\omega'(\omega' - \omega_0 - i\epsilon)} + C_1 \frac{\omega_0}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')\, d\omega'}{\omega' - \omega_0 - i\epsilon} \frac{1}{\pi} \times \int_\mu^\infty \frac{d\omega''}{(\omega'' - \omega_0 + \omega' - i\epsilon) H^-(\omega_0 - \omega'')} \operatorname{Im} \frac{1}{G^+(\omega'')}. \quad (29)$$

Using (3) the first integral becomes $G^+(\omega_0) - 1$. If we interchange orders of integration in the last line, we can do the inner integral and find for the coefficient of C_1: $\omega_0[A^+(\omega_0)G^+(\omega_0) + H^{+-1}(\omega_0) - \omega_0^{-1}]$. Using this in (29) we can solve for C_1 and find

$$C_1 = 2/[1 - H^+(\omega_0)A^+(\omega_0)]. \quad (30)$$

If $1 - H^+(\omega_0)A^+(\omega_0) \neq 0$, the expression

$$M^-(\omega) = \frac{1}{G^+(\omega_0)} - \frac{2\omega}{1 - H^+(\omega_0)A^+(\omega_0)} \times \frac{1}{\pi} \int_\mu^\infty \frac{d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon) H^-(\omega_0 - \omega')} \operatorname{Im} \frac{1}{G^+(\omega')} \quad (31)$$

contains all the analytic properties one can read off from Eq. (16). If this equation has a solution at all it is given uniquely by (31). That (31) actually solves (15) has been shown in I by direct substitution.

If $D(\omega_0) = 0$, (30) does not exist and the inhomogeneous equation (15) has no solution. We show that in this case only the related homogeneous equation has a solution which is unique up to a constant factor.

For a cutoff function $f(\omega)$ which has no zeros for $\omega_0 > \mu$ (long tail) one can show that $D^+(\omega_0)$ has a nonvanishing imaginary part there. Therefore in the region $\omega_0 > \mu$ $D^+(\omega_0) \neq 0$ and the scattering solution to the $V\theta$ equation (15) is unique. The integral equation for $N2\theta$-elastic scattering[18] can also be reduced to a form equivalent to (15) and therefore its solution too is unique.

For $\omega_0 < \mu$, then called ω_B, $D(\omega_B)$ can vanish (Sec. II) and we would like to find the solutions of the homogeneous equation. If we make in (8) the substitution

$$\alpha(k) = [f(\omega)/\omega(\omega_B - \omega)]M_0(\omega)$$

we find for $M_0(\omega)$ the homogeneous part of Eq. (15) with $\omega_0 \to \omega_B$:

$$G^-(\omega_B - \omega)M_0^-(\omega) = \frac{\omega}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')}{\omega' - \omega_B} \frac{M_0^-(\omega')}{\omega' - \omega_B + \omega} d\omega'. \quad (32)$$

Since now $\omega_B < \mu$ the denominators are well defined for ω in the physical domain. For $\omega < \mu$ we consider $M^-(\omega)$ as the analytic continuation which we have indicated in (32). Since we did not use the inhomogeneous term to derive (26) this expression is also valid and gives the most general ansatz for the homogeneous solution. From (26) and (32) we find immediately

$$M_0(0) = C_2 = 0. \quad (33)$$

The solution of a homogeneous equation is only determined up to a factor. Therefore C_1 is free and we expect

$$M_0(\omega) = C_1 \frac{\omega}{\pi} \times \int_\mu^\infty \frac{d\omega'}{(\omega' - \omega_B + \omega) H^-(\omega_B - \omega')} \operatorname{Im} \frac{1}{G^+(\omega')} \quad (34)$$

to solve Eq. (32). We now proceed a little further and show that in general also $C_1 = 0$ and therefore the homogeneous equation has in general no solution; except if the bound-state condition (6) holds. Then it has exactly the solution (34).

Again from (32) we find

$$M_0^-(\omega_B) = \frac{\omega_B}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')}{\omega'(\omega' - \omega_B)} M_0^-(\omega')\, d\omega'.$$

Here we substitute (34) on both sides to find

$$C_1 \omega_B A(\omega_B) = C_1 \frac{\omega_B}{\pi} \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega')\, d\omega'}{\omega' - \omega_B} \frac{1}{\pi} \times \int_\mu^\infty \frac{d\omega''}{(\omega'' - \omega_B + \omega') H^-(\omega_B - \omega'')} \operatorname{Im} \frac{1}{G^+(\omega'')}. \quad (35)$$

Interchanging orders of integration the right-hand

[18] Equation (II) in Ref. 1.

side becomes

$$C_1 \frac{\omega_B}{\pi} \int_\mu^\infty \frac{d\omega'}{H^-(\omega_B - \omega')} \operatorname{Im} \frac{1}{G^+(\omega')} \frac{1}{\pi} \times \int_\mu^\infty \frac{\operatorname{Im} G^+(\omega'')\, d\omega''}{(\omega'' - \omega_B)(\omega'' - \omega_B + \omega')}.$$

The inner integral is $(1/\omega')[G(\omega_B) - G(\omega_B - \omega')]$ and using this in (35) where we take everything to one side gives

$$C_1[\omega_B A(\omega_B) - H(\omega_B)A(\omega_B) - G^{-1}(\omega_B) + 1] = 0$$

or, multiplying by $G(\omega_B)$,

$$C_1 G(\omega_B)[1 - H(\omega_B)A(\omega_B)] = 0. \tag{36}$$

$G(\omega_B) \neq 0$; therefore this can only hold if $C_1 = 0$ unless the square bracket vanishes which is the bound state condition (6). Q.E.D.

V. LEE MODEL WITH AN UNSTABLE V PARTICLE

If we let in the ordinary Lee model the renormalized mass of the V particle becomes larger than the mass of the N plus the mass of a θ meson,

$$m_V > m_N + \mu; \tag{37}$$

the V particle becomes unstable against the decay $V \to N + \theta$. Clearly now $m_V \neq m_N$. The $N\theta$ sector of this model has been solved by Glaser and Källén[11] who find for the T matrix for elastic $N\theta$ scattering

$$T_{N\theta}(\omega) = -f^2(\omega)/\tilde{H}(\omega), \tag{38}$$

where[19]

$$\tilde{H}(\omega) = Z(m_N + \omega - m_V) + Z\delta m + \frac{1}{\pi}\int_\mu^\infty \frac{\operatorname{Im} H(\omega')\, d\omega'}{\omega' - \omega - i\epsilon}. \tag{39}$$

The unstable V shows up as resonance in the $N\theta$ cross section and can be shown to cause zeros of $\tilde{H}(\omega)$ which, however, lay on the second sheet of the Riemann surface of $\tilde{H}(\omega)$.[20] On the first sheet where we perform our integrations $H(\omega)$ has no zero hence $H^{-1}(\omega)$ has no pole. This reflects itself in the solutions for the three-particle sector. Since there is no stable V the V has no in-field and the only amplitude in this sector is the one for elastic $N2\theta$ scattering. To compute it we go out from the state vector

$$|N\theta_{k_1}\theta_{k_2}, \text{in}\rangle = |N\theta_1\theta_2\rangle + Z^{\frac{1}{2}} \int \varphi(k_1k_2; k')\, |V\theta_{k'}\rangle\, d^3k' + \int \psi(k_1k_2; k'k'')\, |N\theta'\theta''\rangle\, d^3k'\, d^3k'', \tag{40}$$

which is an eigenstate of the total Hamiltonian with eigenvalue $m + \omega_1 + \omega_2$. The Schrödinger equation leads to the relations

$$[Z(\omega_1 + \omega_2 - \omega) + Z\delta m]\varphi(k_1, k_2; k) = \tfrac{1}{2}f(\omega_1)\delta(k - k_1) + \tfrac{1}{2}f(\omega_2)\delta(k - k_2) + 2\int \psi(k_1k_2; kk')f(_{m}') \, d^3k', \tag{41}$$

$$(\omega_1 + \omega_2 - \omega' - \omega'')\psi(k_1k_2, k'k'') = \tfrac{1}{2}f(\omega'')\varphi(k_1k_2; k') + \tfrac{1}{2}f(\omega')\varphi(k_1k_2, k''). \tag{42}$$

Eliminating ψ we find for φ the equation

$$\tilde{H}(\omega_1 + \omega_2 - \omega)\varphi(k_1k_2, k) = \tfrac{1}{2}f(\omega_1)\delta(k - k_2) + \tfrac{1}{2}f(\omega_2)\delta(k - k_1) - f(\omega)\int \frac{f(\omega')\varphi(k_1k_2; k')}{\omega' + \omega - \omega_1 - \omega_2 - i\epsilon} d^3k'. \tag{43}$$

Now we may make the substitution

$$\varphi(k_1, k_2; k) = \frac{1}{2\tilde{H}(\omega_1 + \omega_2 - \omega)} [f(\omega_1)\delta(k - k_2) + f(\omega_2)\delta(k - k_1)] - \frac{f(\omega)f(\omega_1)f(\omega_2)}{2\tilde{H}(\omega_1)\tilde{H}(\omega_2)} L(\omega; \omega_1, \omega_2) \tag{44}$$

to get

$$\tilde{H}(\omega_1 + \omega_2 - \omega)L(\omega; \omega_1, \omega_2) = \frac{\tilde{H}(\omega_1)}{\omega - \omega_2 - i\epsilon} + \frac{\tilde{H}(\omega_2)}{\omega - \omega_1 - i\epsilon} - \frac{1}{\pi}\int_\mu^\infty \frac{\operatorname{Im} H^+(\omega')L(\omega', \omega_1, \omega_2)\, d\omega'}{\omega' + \omega - \omega_1 - \omega_2 - i\epsilon}. \tag{45}$$

The solution to this integral equation can be found by a slight generalization of the methods developed for the $V\theta$ equation. Observing that $L(\omega; \omega_1, \omega_2)$ in (45) has two poles with residues $+1$ and a left-hand cut we are led to the ansatz

$$L(\omega; \omega_1, \omega_2) = \frac{1}{\omega - \omega_1 - i\epsilon} + \frac{1}{\omega - \omega_2 - i\epsilon} + C(\omega_1, \omega_2)I^-(\omega_1 + \omega_2 - \omega), \tag{46}$$

where

$$I(z) = \frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde{H}^+(\omega')}\right] \times \frac{1}{\tilde{H}^-(\omega_1 + \omega_2 - \omega')} \frac{d\omega'}{\omega' - z}. \tag{47}$$

The constant C (ω_1 and ω_2 enter only in a parametric way) is determined by direct substitution into Eq.

[19] $\operatorname{Im} \tilde{H}(\omega) = \operatorname{Im} H(\omega)$.
[20] M. Levy, Nuovo Cimento 13, 115 (1959).

(45). We do this in Appendix C. There we find that (46) solves (45) if

$$C(\omega_1, \omega_2) = -2/\tilde{A}(\omega_1 + \omega_2), \tag{48}$$

where

$$\tilde{A}(\omega) = \frac{1}{\pi}\int_\mu^\infty \left[\mathrm{Im}\,\frac{1}{\tilde{H}^+(\omega')}\right]\frac{d\omega'}{\tilde{H}^-(\omega - \omega')}. \tag{49}$$

The T matrix for elastic $N2\theta$ scattering is given by the residue of the pole at $\omega_1 + \omega_2 - \omega' - \omega'' = 0$ in $\psi(\omega_1, \omega_2; \omega', \omega'')$ in (42). Omitting a factor $2\pi i\,\delta(\omega_1 + \omega_2 - \omega' - \omega'')$ we find for the nontrivial part of the T matrix

$$T_{N2\theta}(\omega_1, \omega_2; \omega) = -\frac{f(\omega_1)f(\omega_2)f(\omega)f(\omega_1 + \omega_2 - \omega)}{\tilde{H}^+(\omega_1)\tilde{H}^+(\omega_2)\tilde{H}^-(\omega_1 + \omega_2 - \omega)} \times [I^+(\omega) + I^-(\omega_1 + \omega_2 - \omega)].$$

The last expression can be simplified using relation (C2) in the Appendix. This gives

$$T_{N2\theta}(\omega_1, \omega_2; \omega) = \frac{-1}{\tilde{A}(\omega_1 + \omega_2)}\,\frac{f(\omega_1)f(\omega_2)f(\omega)f(\omega_1 + \omega_2 - \omega)}{\tilde{H}^+(\omega_1)\tilde{H}^+(\omega_2)\tilde{H}^+(\omega)\tilde{H}^-(\omega_1 + \omega_2 - \omega)}.$$

The denominator functions of the second factor represent two-particle interactions, for short called final-state interactions. They are known to be nonvanishing. The only indication for a three-particle bound-state pole could come from a zero of $\tilde{A}(\omega)$. Since $\tilde{H}^{-1}(z)$ has no pole on the first sheet it satisfies the unsubtracted dispersion relation

$$\tilde{H}^{-1}(z) = \frac{1}{\pi}\int_\mu^\infty \left[\mathrm{Im}\,\frac{1}{\tilde{H}(\omega')}\right]\frac{d\omega'}{\omega' - z}; \tag{51}$$

using this in the expression (49) for $\tilde{A}(\omega)$ we find

$$\tilde{A}(\omega) = \frac{1}{\pi}\int_\mu^\infty \left[\mathrm{Im}\,\frac{1}{\tilde{H}^+(\omega')}\right]\frac{d\omega'}{\pi} \times \int_\mu^\infty \left[\mathrm{Im}\,\frac{1}{\tilde{H}^+(\omega'')}\right]\frac{d\omega''}{\omega'' + \omega' - \omega}. \tag{52}$$

For $\omega < 2\mu$ the last denominator never vanishes and therefore in the entire range for a possible bound state $\tilde{A}(\omega) > 0$. Therefore the elastic scattering amplitude shows that there is no $N2\theta$ bound state in this Lee model.

The same conclusion is obtained by deriving the bound-state wavefunction. For an $N2\theta$ bound state we still can make the ansatz (7) where we put again $\sim$ to indicate the unstable case. For the wavefunction $\tilde{\alpha}(\omega)$ we obtain instead of (8) the homogeneous equation

$$\tilde{H}^-(\omega_B - \omega)\tilde{\alpha}(\omega) = -f(\omega)\int \frac{d^3k'\, f(\omega')\tilde{\alpha}(\omega')}{\omega' - \omega_B + \omega}. \tag{53}$$

Since $\tilde{H}(\omega_B - \omega)$ has no zero at $\omega = \omega_B$, $\tilde{\alpha}(\omega)$ has no pole there and cannot be a proper bound-state wavefunction. Indeed a proof similar to the one that led to relation (36) in Sec. IV shows that this homogeneous equation has only the trivial solution $\tilde{\alpha}(\omega) \equiv 0$. This also shows that the solution (46) is unique.

ACKNOWLEDGMENTS

This article grew out of discussions I had with Professor O. W. Greenberg. I am obliged to him for many helpful comments. I am obliged to Dr. Ezawa for pointing out the work of T. Muta and I want to thank Dr. Muta for a helpful discussion. I also want to acknowledge the generous support received from the American–Swiss Foundation for Scientific Exchange.

APPENDIX A

The two equations (8), (9) can be obtained by the use of graphs in the following way. The function $\alpha(k)$ denoted by a square black box in Fig. 2 is

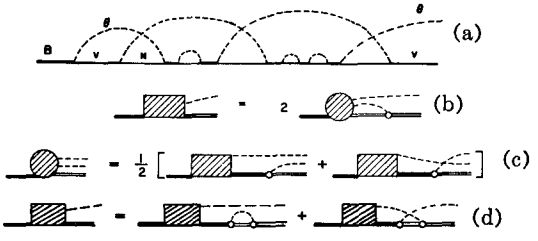

FIG. 2. (a) Sample of a Feynman diagram contributing to the $B \to V\theta$ vertex. (b)–(d) Relations for the vertex functions α (square black box) and β (round black box). Explained in the text.

related to the vertex function describing the decay of the bound state B into $V + \theta$ in that the V has been taken off the mass shell. Similarly, $\beta(k_1, k_2)$, the round black box in Fig. 2, is related to the vertex function for $B \to N2\theta$ in that the N is off-shell. If B is stable these are clearly forbidden processes. If they were allowed, the vertex function for $B \to V\theta$ would be on-shell an infinite sum of Feynman graphs one of which is given in Fig. 2(a). All of them have in common that the last vertex describes the absorption of a θ by an N. Summing all but this last vertex into a round black box we can write the relation of Fig. 2(b), where double lines (off-shell lines) represent propagators. Especially the last double line is to be read as $[Z(E - m) + \delta m]^{-1}$. The N propagator is simply $-[E - m]^{-1}$ and energy conservation gives the value of E. A set of diagrams similar to the one in Fig. 1(a) can be written for $B \to N2\theta$. Their common last vertex is $V \to N\theta$ described in our notation by a factor $f(\omega)$. Taking care of sym-

metrization for the two θ's we find Fig. 2(c), which is immediately recognized as Eq. (9). Substitution of Fig. 2(c) into 2(b) gives Fig. 2(d). The bubble diagram factors and can be taken to the left. Remembering the definition of $H(\omega)$[4]

$$H(\omega) = Z\omega + Z\delta m + \int \frac{f^2(\omega')\, d^3k'}{\omega' - \omega - i\epsilon} \qquad \text{(A1)}$$

one reads off Eq. (8). This derivation also works in the higher sectors of the Lee model. Note that Fig. 2(a) is a Feynman graph, 2(b) – (d) are not because some external lines are off-shell.

APPENDIX B

Here we evaluate the norm of the $V\theta$-bound state. Using (7) in the normalization condition (13) leads to

$$N^{-2} = Z \int \alpha(k')\alpha^*(k')\, d^3k' + 2 \int \beta(k', k'')\beta^*(k', k'')\, d^3k'\, d^3k''.$$

Substituting for α and β and going over to ω' as variable of integration we find

$$N^{-2} = Z \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{(\omega' - \omega_B)^2} J(\omega')J^*(\omega')\, d\omega' + \frac{1}{2\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')\ \operatorname{Im} H(\omega'')}{(\omega' + \omega'' - \omega_B)^2} \times \left[\frac{1}{\omega' - \omega_B} J(\omega') + \frac{1}{\omega'' - \omega_B} J(\omega'')\right] \times \left[\frac{1}{\omega' - \omega_B} J^*(\omega') + \frac{1}{\omega'' - \omega_B} J^*(\omega'')\right] d\omega'\, d\omega''.$$

Observing that the second integrand is symmetric in ω' and ω'' and $J^* = J$ is real since $\omega_B < \mu$ we can replace the product of the two square brackets by

$$\frac{2}{(\omega' - \omega_B)^2} J^2(\omega') + \frac{2}{(\omega' - \omega_B)(\omega'' - \omega_B)} J(\omega')J(\omega'').$$

We then obtain

$$N^{-2} = Z \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{(\omega' - \omega_B)^2} J^2(\omega')\, d\omega' + \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{(\omega' - \omega_B)^2} J^2(\omega') \frac{1}{\pi} \times \int_\mu^\infty \frac{\operatorname{Im} H(\omega'')}{(\omega'' + \omega' - \omega_B)^2} d\omega''\, d\omega' + \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{\omega' - \omega_B} J(\omega') \frac{1}{\pi} \times \int_\mu^\infty \frac{\operatorname{Im} H(\omega'')}{(\omega'' - \omega_B)(\omega'' + \omega' - \omega_B)^2} J(\omega'')\, d\omega''\, d\omega'.$$

The inner integral in the second term can be done by contour integration. Denoting a derivative with a prime, we find

$$\frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega'')}{(\omega'' + \omega' - \omega_B)^2} d\omega'' = -Z + H'(\omega_B - \omega')$$

and we see that the terms with Z cancel. Hence

$$N^{-2} = \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{(\omega' - \omega_B)^2} J^2(\omega')H'(\omega_B - \omega') + \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H(\omega')}{\omega' - \omega_B} J(\omega') \frac{1}{\pi} \times \int_\mu^\infty \frac{\operatorname{Im} H(\omega'')}{(\omega'' - \omega_B)(\omega'' + \omega' - \omega_B)^2} J(\omega'')\, d\omega''\, d\omega'.$$

Here one can verify easily that none of the denominators vanishes. Therefore $N^{-2} > 0$ and, which is important, $N^2 \neq 0$. Q.E.D.

APPENDIX C

Here we determine the constant $C(\omega_1, \omega_2)$ in the ansatz (46) by direct substitution into Eq. (45). Substitution of the ansatz under the integral leads to the following three integrals which we can evaluate ($\omega_0 = \omega_1 + \omega_2$):

$$I_1(\omega; \omega_1, \omega_2) = -\frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H^+(\omega')\, d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon)(\omega' - \omega_1 - i\epsilon)} = \frac{1}{\omega_2 - \omega + i\epsilon} [\tilde{H}^-(\omega_0 - \omega) - \tilde{H}^+(\omega_1)] + Z,$$

$$I_2(\omega; \omega_1, \omega_2) = I_1(\omega; \omega_2, \omega_1),$$

$$I_3(\omega; \omega_0) = -C \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H^+(\omega')}{\omega' - \omega_0 + \omega - i\epsilon} I^-(\omega_0 - \omega')\, d\omega' = -C \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} H^+(\omega')}{\omega' - \omega_0 + \omega - i\epsilon} \frac{d\omega'}{\pi} \int_\mu^\infty \left(\operatorname{Im} \frac{1}{\tilde{H}^+(\omega'')}\right) \times \frac{d\omega''}{\tilde{H}^-(\omega_0 - \omega'')(\omega'' + \omega' - \omega_0 - i\epsilon)}$$

(interchanging orders of integration)

$$= -C \frac{1}{\pi} \int_\mu^\infty \left[\operatorname{Im} \frac{1}{\tilde{H}^+(\omega'')}\right] \frac{d\omega''}{\tilde{H}^-(\omega_0 - \omega'')} \frac{1}{\pi} \times \int_\mu^\infty \frac{\operatorname{Im} H^+(\omega')\, d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon)(\omega' - \omega_0 + \omega'' - i\epsilon)}.$$

The inner integral can be done to give

$$(\omega'' - \omega - i\epsilon)^{-1}[\tilde{H}^-(\omega_0 - \omega) - \tilde{H}^-(\omega_0 - \omega'')] - Z.$$

With this we find

$$I_3(\omega;\omega_0) = -C\,H^-(\omega_0-\omega) \times \frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega'')}\right]\frac{d\omega''}{\tilde H^-(\omega_0-\omega'')(\omega''-\omega-i\epsilon)} + C\frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega'')}\right]\frac{d\omega''}{\omega''-\omega-i\epsilon} + CZ\frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega'')}\right]\frac{d\omega''}{\tilde H^-(\omega_0-\omega'')}.$$

Using the definitions (47), (49), (51), this reads

$$I_3(\omega;\omega_0) = C[\tilde H^{+-1}(\omega) + Z\tilde A^+(\omega_0) - \tilde H^-(\omega_0-\omega)I^+(\omega)]. \qquad \text{(C1)}$$

Before we can proceed to collect all the terms we have to prove the relation

$$I^+(\omega) + I^-(\omega_0-\omega) = 1/\tilde H^+(\omega)\tilde H^-(\omega_0-\omega). \qquad \text{(C2)}$$

This relation is now that $\tilde H(\omega)$ has no zero much simpler than it was in the stable case. Call

$$K(\omega) \equiv [1/\tilde H^+(\omega)\tilde H^-(\omega_0-\omega)] - I^+(\omega)$$

and insert the integral representations (49), (51) for $[\tilde H^+(\omega)]^{-1}$ and $I^+(\omega)$ to get

$$K(\omega) = \frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega')}\right]\frac{d\omega'}{\omega'-\omega-i\epsilon} \times \left[\frac{1}{\tilde H^-(\omega_0-\omega)} - \frac{1}{\tilde H(\omega_0-\omega')}\right]. \qquad \text{(C3)}$$

The last square bracket becomes

$$\frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega'')}\right]d\omega'' \times \left[\frac{1}{\omega''-\omega_0+\omega-i\epsilon} - \frac{1}{\omega''-\omega_0+\omega'-i\epsilon}\right],$$

making a common denominator

$$\frac{\omega'-\omega}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega'')}\right] \times \frac{d\omega''}{(\omega''-\omega_0+\omega-i\epsilon)(\omega''-\omega_0+\omega'-i\epsilon)};$$

we can now insert this into the last expression (C3) for $K(\omega)$. Interchanging orders of integration,

$$K(\omega) = \frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega')}\right]\frac{d\omega'}{\omega'-\omega_0+\omega-i\epsilon} \times \frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega'')}\right]\frac{d\omega''}{\omega''-\omega_0+\omega'-i\epsilon}.$$

The second integral is just $\tilde H^-(\omega_0-\omega)^{-1}$ and K becomes

$$K(\omega) = \frac{1}{\pi}\int_\mu^\infty \left[\operatorname{Im}\frac{1}{\tilde H^+(\omega')}\right] \times \frac{d\omega'}{\tilde H^-(\omega_0-\omega')(\omega'-\omega_0+\omega-i\epsilon)} = I^-(\omega_0-\omega).$$

Q.E.D.

Now we can use identity (C2) in the last expression for I_3 (C1) to write it

$$I_3(\omega,\omega_0) = C\tilde H^-(\omega_0-\omega)I^-(\omega_0-\omega) + ZC\tilde A^+(\omega_0).$$

Collecting $I_1 + I_2 + I_3$ and substituting the ansatz (46) also on the left-hand side of the integral equation (45) we see that most of the terms cancel. The equation reduces to $0 = 2Z + CZ\tilde A^+(\omega_0)$. In general $Z \neq 0$ and we conclude

$$C(\omega_1,\omega_2) = -2/\tilde A^+(\omega_0).$$

APPENDIX D

The decomposition (7) of the bound state can be interpreted that for part of the time B is a $V\theta$-bound state and part of the time an $N2\theta$ system. In the first case $\alpha(\omega)$, or its Fourier transform, is the wavefunction; while in the second interval it is given by $\beta(\omega,\omega')$. We show on the example of $\alpha(\omega)$ that its Fourier transform has the correct spatial dependence for a bound-state wavefunction. Define it by

$$\varphi(\mathbf{x}) = \int e^{i\mathbf{k}\cdot\mathbf{x}}\alpha(\omega)\,d^3k.$$

Substitution of $\alpha(\omega)$ from (10) leads to

$$\varphi(\mathbf{x}) = N\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}\frac{f(\omega)}{\omega-\omega_B}J(\omega).$$

We remember that $f(\omega)$ is, up to factors, the Fourier transform of the source function $U(|\mathbf{x}|)$:

$$f(\omega) = \frac{g}{(2\pi)^{\frac{1}{2}}(2\omega)^{\frac{1}{2}}}\int e^{-i\mathbf{k}\cdot\mathbf{x}'}U(|\mathbf{x}'|)\,d^3\mathbf{x}'.$$

We can insert this into the expression for $\varphi(\mathbf{x})$ to get

$$\varphi(\mathbf{x}) = \frac{gN}{(2\pi)^{\frac{1}{2}}}\int d^3x'\,U(|\mathbf{x}'|)\,d^3k\,\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{(\omega-\omega_B)(2\omega)^{\frac{1}{2}}}J(\omega).$$

In the inner integral we can do the angular integrations. This leads to an expression involving $\sin k\,|\mathbf{x}-\mathbf{x}'|$; $k = |\mathbf{k}|$. We then remove $\sin k\,|\mathbf{x}-\mathbf{x}'|$ by extending the integral over k from $-\infty$ to $+\infty$ where we use the evenness of $\omega(k^2)$. This integral then becomes

$$\frac{1}{|x - x'|} \int_{-\infty}^{\infty} \frac{e^{ik|x-x'|} k\, dk}{[\omega(k^2) - \omega_B] 2\omega(k^2)} J[\omega(k^2)].$$

We want to evaluate the leading term of this integral by contour integration.

The function $J(\omega)$ (11) is regular in the upper k-plane. The function $[\omega(k^2)]^{-\frac{1}{2}}$ has branch points at $k = \pm i\mu$. We attach a cut to the one at $+ i\mu$ up the imaginary axis and to the one at $-i\mu$ down the imaginary axis.

We further expect to get a residue from the vanishing of the denominator $\omega - \omega_B$. Since $k_B^2 = (\omega_B^2 - \mu^2)^{\frac{1}{2}} < 0$, we have also $|k_B| < \mu$. We can write

$$\omega - \omega_B = (k^2 + \mu^2)^{\frac{1}{2}} - (k_B^2 + \mu^2)^{\frac{1}{2}} = \mu\left[\left(1 + \frac{k^2}{\mu^2}\right)^{\frac{1}{2}} - \left(1 + \frac{k_B^2}{\mu^2}\right)^{\frac{1}{2}}\right].$$

Expanding this into a Taylor series around $k = i\,|k_B|$ we find

$$\omega - \omega_B \sim (1/2\mu)(k - i\,|k_B|)(k + i\,|k_B|)$$

and we see that this has zeros at $k = \pm i\,|k_B|$ on the imaginary k axis.

The exponential in the above integrand is damping in the upper k-plane. Therefore we can deform the contour of integration, which is now along the entire real axis, such that it is wrapped around the upper cut on the imaginary axis. Thereby we pick up the residue at $k = i\,|k_B|$ (Fig. 3). Since $|k_B| < \mu$, the contribution from the cut, having larger masses, will be damped faster than the contribution from the pole at $k = i\,|k_B|$. We find for the leading term, taking 2π times the residue,

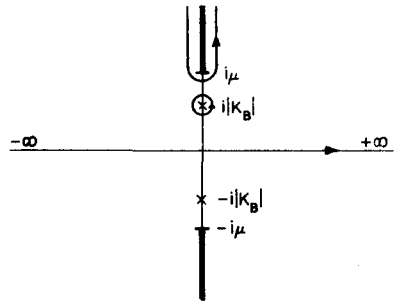

FIG. 3. Contour of integration for the bound-state wavefunction.

$$\frac{\pi}{2\mu} \frac{e^{-|k_B| \cdot |x-x'|}}{[2\omega(i\,|k_B|)]^{\frac{1}{2}}} J[\omega(i\,|k_B|)].$$

Collecting and substituting into the expression for $\varphi(x)$ its leading term becomes, up to a normalization constant N,

$$\varphi(x) \sim N \int d^3x'\, U(|\mathbf{x}'|) \frac{\exp[-|k_B| \cdot |\mathbf{x} - \mathbf{x}'|]}{|\mathbf{x} - \mathbf{x}'|}.$$

Since $U(r)$ is of finite range this behaves for large values of $|\mathbf{x}| = r$ as $r^{-1} \exp[-|k_B| r]$, which is the typical behavior of a bound-state wavefunction.

Canonical Realizations of Lie Symmetry Groups

M. PAURI
Istituto di Fisica dell'Università, Parma, Italy
AND
G. M. PROSPERI
Istituto di Fisica dell' Università, Milan, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Italy
(Received 17 December 1964)

A general theory of the realizations of Lie groups by means of canonical transformations in classical mechanics is given. The problem is the analog to that of the characterization of the projective representations in quantum mechanics considered by Wigner, Bargmann, and others in the case of the Galilei and the Lorentz group. However, no application to particular groups is given in this paper.

It turns out that the generators y_τ of the infinitesimal transformations in a canonical realization of a Lie group $\mathcal{G}$ satisfy relations of the form $\{y_\rho, y_\sigma\} = c_{\rho\sigma}{}^\tau y_\tau + d_{\rho\sigma}$, where $c_{\rho\sigma}{}^\tau$ are the structure constands of $\mathcal{G}$ and $d_{\rho\sigma}$ are constants depending on the particular realization.

It also turns out that any canonical realization of $\mathcal{G}$ can be reduced to a fundamental *typical form* by means of a constant canonical transformation in the phase space of the system. This *typical form* allows one to obtain a complete characterization of all the possible canonical realizations of $\mathcal{G}$. Once a suitable definition of "irreducible" canonical realization is given, a simple classification can be obtained in terms of the values of certain functions of the generators y_τ (*canonical invariants*).

Recalling the correspondence $\{\ \} \to [\ ]$, formal analogy appears to be achieved with the quantum mechanical case. Even more, a parallel development of the outlined theory in quantum mechanics is of interest in the construction of the invariant operators (even in case of non-semisimple groups) and of complete systems of commuting observables acting within the irreducible representations.

INTRODUCTION

IT is well known that the study of the dynamical symmetries of a quantum system is connected with the characterization of the projective unitary representations of symmetry groups, specifically the Galilei and the Lorentz group for nonrelativistic and relativistic quantum mechanics, respectively.[1]

In the present paper, we want to discuss the corresponding problem in classical mechanics. The most natural approach to it consists in the characterization of all possible realizations of the above symmetry groups by means of canonical transformations.

A general axiomatic approach to classical dynamics has been expounded by Dirac[2] who studied in particular the inhomogeneous Lorentz group. However, from our point of view, Dirac's work represents only a preliminary treatment of the problem and, besides, some points require a more detailed and explicit analysis. In this connection, our work may be considered a development of Ref. 2 with a mind nearer to the philosophy of Wigner and Bargmann in their quantum mechanical characterization.

As canonical or contact transformation we mean a transformation in phase space leaving invariant the Poisson brackets among the fundamental dynamical variables, and as canonical realization of a group a class of canonical transformations with the same multiplication rule of the group.

A similar question has been raised by Loinger. Precisely, Loinger and other authors[3] studied, in the case of the rotation, the Galilei and the Lorentz group, the canonical realizations as unitary transformations in the "classical" Hilbert space of Koopman and von Neumann,[4] giving essentially a sort of generalized representation theory for the classical mechanics. We will consider, instead, in a direct way, the canonical realizations rather than their unitary counterparts. Since not all the Hermitian operators in the Koopman–von Neumann space have a direct meaning, it seems more promising

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U. S. **34**, 211 (1948); Iu. M. Shirokov, Soviet Phys.—JETP **6**, 664, 919, 929 (1958); C. Fronsdal, Phys. Rev. **113**, 1367 (1959); D. W. Robinson, Helv. Phys. Acta **35**, 98 (1962); A. S. Wightman in "Rélations de dispersion et particules élémentaires" Proceedings of the Ecole d'été de Physique théorique, Les Houches (Hermann, Paris, 1960). E. Inonü and E. P. Wigner, Nuovo Cimento **9**, 705 (1952); M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts 1962) Sec. 12/5-6-7; M. Hamermesh, Ann. Phys. (N.Y.) **9**, 518 (1960); J. M. Levy-Leblond, J. Math. Phys. **4**, 776 (1963).

[2] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).

[3] A. Loinger, Ann. Phys. (N.Y.) **20**, 132 (1962), **23**, 23 (1963); P. Gulmanelli, Phys. Letters **5**, 320 (1963); G. Lugarini and M. Pauri: "Representation Theory for Classical Mechanics," "Classical Representations of the Inhomogeneous Lorentz Group" (to be published). For a different approach to classical Physics, see also T. F. Jordan and E. C. G. Sudarshan, Rev. Mod. Phys. **33**, 515 (1961).

[4] B. O. Koopman, Proc. Natl. Acad. Sci. U.S. **17**, 315 (1931); J. von Neumann, Ann. Math. **33**, 587 (1932); L. Van Hove, Mem. Acad. Roy. Belg. Cl. Sci. No. 6 **26**, 1951).

from a physical point of view to deal with the canonical generators themselves.

It will be seen that, with a suitable definition of "irreducible" canonical realization, a very close analogy is achieved with the corresponding quantum mechanical problem. Moreover, some features of our treatment will appear to be of direct interest for quantum mechanics itself.

Our characterization of the canonical realizations is limited to the neighborhood of the identity. Thus, the theory directly applies, in the large, only to simply connected groups.[5] For groups which are connected but nonsimply connected, we characterize essentially the canonical realizations of their covering groups. Realizations of the groups themselves can, in principle, be selected paying attention to questions of monodromy.

The present paper contains the general foundation of the theory. In forthcoming papers we shall apply it to investigations of the canonical realizations of the rotation, the Galilei, and the Poincaré groups. We hope to devote a paper to working out the full implications of the present approach in quantum mechanics as well.

In the first section, the concept of canonical realization is introduced in detail. In the second section, a suitable canonical transformation is performed on a general given canonical realization, which allows us to point out the fundamental features of the whole question; we analyze them in the third section. Finally, in the last section, we give the rules for the construction of the most general canonical realization of a Lie group.

1. CANONICAL REALIZATIONS OF A LIE GROUP

Let $\mathcal{G}$ be a given Lie group of order r with parameters $(a_1, \cdots, a_r)$ and let

$$[X_\rho, X_\sigma] = c_{\rho\sigma}^\tau X_\tau; \qquad \rho, \sigma = 1, \cdots, r, \tag{1}$$

be the commutation relations among its infinitesimal operators. Let $q_1, \cdots, q_n, p_1, \cdots, p_n$ be the canonical coordinates for a classical system:

$$\{q_i, q_j\} = \{p_i, p_j\} = 0, \qquad \{q_i, p_j\} = \delta_{ij}; \qquad i, j = 1, \cdots, n, \tag{2}$$

where $\{A, B\}$ denotes the Poisson bracket between A, B. [We think of the Poisson brackets as implicitly defined by their formal properties. Then, from (2), it follows that

$$\{A, B\} = \sum_k \left(\frac{\partial A}{\partial q_k}\frac{\partial B}{\partial p_k} - \frac{\partial A}{\partial p_k}\frac{\partial B}{\partial q_k}\right).$$

Alternatively, this relation can be assumed as a definition.] We define as a contact or canonical realization $\mathfrak{R}$ of the group $\mathcal{G}$, a set of transformations in the canonical variables

$$\begin{aligned} q_i' &= q_i'(q_1, \cdots, q_n, p_1, \cdots, p_n \mid a_1, \cdots, a_r), \\ p_i' &= p_i'(q_1, \cdots, q_n, p_1, \cdots, p_n \mid a_1, \cdots, a_r) \end{aligned} \tag{3}$$

homomorphic to $\mathcal{G}$ and leaving relations (2) invariant. A realization $\mathfrak{R}$ which is not only homomorphic but also isomorphic to $\mathcal{G}$ will be called a faithful realization.

The infinitesimal transformations of $\mathfrak{R}$ can be written as

$$\begin{aligned} q_i' &= q_i + \delta a^\tau \{y_\tau, q_i\}, \\ p_i' &= p_i + \delta a^\tau \{y_\tau, p_i\}. \end{aligned} \tag{4}$$

The y_τ's ($\tau = 1, \cdots, r$) are suitable functions of the $2n$ canonical variables only. They are called the canonical generators of the infinitesimal transformations and are defined up to an additive constant. We have now to satisfy the requirement of homomorphism. Let us consider the infinitesimal transformation $G(\delta\mathbf{a})G(\delta\mathbf{b})G^{-1}(\delta\mathbf{a})G^{-1}(\delta\mathbf{b})$ of the group $\mathcal{G}$, associated with the operator

$$1 + \delta a^\tau\, \delta b^\lambda [X_\tau, X_\lambda].$$

The corresponding transformation in the realization $\mathfrak{R}$ can be written, via the Jacobi identity, in the form

$$\begin{aligned} q_i' &= q_i + \delta a^\tau\, \delta b^\lambda \{\{y_\tau, y_\lambda\}, q_i\}, \\ p_i' &= p_i + \delta a^\tau\, \delta b^\lambda \{\{y_\tau, y_\lambda\}, p_i\}. \end{aligned} \tag{5}$$

Then, from (1) and (5), it follows

$$\{y_\rho, y_\sigma\} = c_{\rho\sigma}^\tau y_\tau + d_{\rho\sigma}, \tag{6}$$

where the $d_{\rho\sigma}$'s are constants which may depend on the particular realization.

Let us look at the operators $\{y_\tau, \cdots\}$: in a faithful realization they are obviously linearly independent. The same is not true in the general case. Since however the nonfaithful realizations are faithful realizations of the related factor groups, in the present paper we shall confine ourselves to this latter class.

Because of the symmetry properties of the Poisson brackets and of the structure constants $c_{\rho\sigma}^\tau$, the conditions

$$\begin{aligned} d_{\rho\sigma} &= -d_{\sigma\rho}, \\ c_{\rho\sigma}^\tau d_{\tau\lambda} + c_{\sigma\lambda}^\tau d_{\tau\rho} + c_{\lambda\rho}^\tau d_{\tau\sigma} &= 0 \end{aligned} \tag{7}$$

must be verified.

[5] See C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), Sec. VI, Theorem 2; see also, for instance: P. M. Cohn, *Lie Groups*, Cambridge Mathematics Tracts (Cambridge University Press, Cambridge, England, 1961), Chap. VII.

We remark that the fundamental relations (6) are formally identical with the commutation relations of the infinitesimal operators of a representation up to a factor of the group $\mathcal{G}$. In particular, by means of a substitution of the form

$$y_\tau \to y_\tau + \gamma_\tau, \tag{8}$$

where the γ_τ's are suitably chosen constants, it is possible to reduce the number of the nonzero $d_{\rho\sigma}$'s to a minimum which is characteristic of the group $\mathcal{G}$. This may be done according to a method given by Bargmann.[6] In particular this minimum number can be easily evaluated by observing that substitution (8) implies the following transformation on the $d_{\rho\sigma}$'s:

$$d'_{\rho\sigma} = d_{\rho\sigma} + c^\tau_{\rho\sigma}\gamma_\tau. \tag{9}$$

Indeed, if q is the number of independent equations of the form

$$c^\tau_{\rho\sigma}\gamma_\tau = 0, \tag{10}$$

we can dispose of q constants to adjust the values of the $d_{\rho\sigma}$'s. If now $m[m \leq \frac{1}{2}r(r-1)]$ is the number of independent $d_{\rho\sigma}$'s, the minimum number of $d_{\rho\sigma}$'s is $s = m - q$. s is, for instance, zero for the rotation group, the Euclidean group in $n > 2$ dimensions, the homogeneous and inhomogeneous Lorentz group, whereas it is one for the Galilei group and the Euclidean group in two dimensions.

2. REDUCTION OF A GENERAL CANONICAL REALIZATION TO A *TYPICAL FORM*

Two given canonical realizations of the group $\mathcal{G}$ in the same number of variables, $\mathfrak{R}$ and $\mathfrak{R}'$ will be said canonically equivalent if a fixed canonical transformation S exists such that symbolically

$$\mathfrak{R}'(a) = S \cdot \mathfrak{R}(a) \cdot S^{-1}. \tag{11}$$

We shall show that, by means of a constant transformation, any canonical realization may be reduced to a *typical form* which allows one to obtain a very significant characterization of all possible canonical realizations of a given group $\mathcal{G}$.

Let us consider a given canonical realization of $\mathcal{G}$ in the $2n$ canonical variables q_i, $p_j (i, j = 1, \cdots, n)$. The r generators of the infinitesimal transformations will satisfy relations (6). It will be seen that a new system of canonical variables can be constructed, in which the generators depend on a number of variables not greater than the order r of the group $\mathcal{G}$. This result is contained in the two following theorems:

Theorem 1: A new system of variables

$$\begin{aligned} \mathfrak{Q}_1 &= \mathfrak{Q}_1(y_1, \cdots, y_r), \cdots, \mathfrak{Q}_h = \mathfrak{Q}_h(y_1, \cdots, y_r), \\ \mathfrak{P}_1 &= \mathfrak{P}_1(y_1, \cdots, y_r), \cdots, \mathfrak{P}_h = \mathfrak{P}_h(y_1, \cdots, y_r), \\ \mathfrak{J}_1 &= \mathfrak{J}_1(y_1, \cdots, y_r), \cdots, \mathfrak{J}_k = \mathfrak{J}_k(y_1, \cdots, y_r), \end{aligned} \tag{12}$$

where $2h + k = r$, can be constructed which are independent functions of the r generators $y_1, \cdots, y_r$ alone and have the following pseudo-canonical Poisson brackets:

$$\begin{aligned} \{\mathfrak{Q}_i, \mathfrak{Q}_j\} &= \{\mathfrak{P}_i, \mathfrak{P}_j\} = \{\mathfrak{Q}_i, \mathfrak{J}_t\} \\ &= \{\mathfrak{P}_i, \mathfrak{J}_t\} = \{\mathfrak{T}_t, \mathfrak{J}_{t'}\} = 0, \\ \{\mathfrak{Q}_i, \mathfrak{P}_j\} &= \delta_{ij}; \qquad i, j = 1 \cdots h; \\ & t, t' = 1 \cdots k, \qquad 2h + k = r. \end{aligned} \tag{13}$$

The numbers h and k and the functional dependence of $\mathfrak{Q}_i$, $\mathfrak{P}_j$, $\mathfrak{J}_t$ on the y_τ depend entirely on the values of the constants $c^\tau_{\rho\sigma}$ and $d_{\rho\sigma}$ and not on other features of the particular realization. On the other hand, the choice of the variables is obviously determined up to transformations leaving relations (13) invariant. Precisely the variables $\mathfrak{J}_1 \cdots \mathfrak{J}_k$ can be replaced by any k independent functions of them, say $\mathfrak{J}'_1 \cdots \mathfrak{J}'_k$, and the variables $\mathfrak{Q}_i$, $\mathfrak{P}_j$ by others obtained through a canonical transformation involving the $\mathfrak{J}_t$ as parameters.

Theorem 2: A set of $2n$ canonical variables

$$Q_i = Q_i(q, p), \quad P_j = P_j(q, p); \quad i, j = 1, \cdots, n$$

can be constructed which are functions of the original variables q_i, p_j such that

$$\begin{aligned} & Q_1 \equiv \mathfrak{Q}_1, \cdots, Q_h \equiv \mathfrak{Q}_h, \; P_1 \equiv \mathfrak{P}_1, \cdots, P_h \equiv \mathfrak{P}_h, \\ & P_{h+1} \equiv \mathfrak{J}'_1, \cdots, P_{h+l} \equiv \mathfrak{J}'_l; \\ & \{Q_i, \mathfrak{J}'_{l+v}\} = \{P_j, \mathfrak{J}'_{l+v}\} = 0; \qquad i, j = 1, \cdots, n; \\ & \qquad v = 1, \cdots, k - l, \qquad l \leq k. \end{aligned} \tag{14}$$

Here the number l depends on the particular realization $\mathfrak{R}$ and the choice of the $\mathfrak{T}'_1 \cdots \mathfrak{T}'_k$ has to be done in a suitable way.

The first theorem is due essentially to S. Lie and has been used in a different context as a part of one of the proofs of the third fundamental theorem.[7] The second one, as far as we know, does not appear in the literature. We will give here the proof of Theorem 1 in view of the necessary modifications suitable to our case and as an introduction to the proof of Theorem 2.

[6] V. Bargmann, Ann Math. 59, 1 (1954); see also Hamermesh's book (Ref. 1), Sec. 12/3-4.

[7] L. Bianchi, *Lezioni sulla teoria dei gruppi continui finiti di trasformazioni* (E. Spoerri, Pisa, Italy, 1918).

Proof of Theorem 1: Let us assume, for instance, that at least one of the expressions $c^r_{\rho 1} y_r + d_{\rho 1}$ is not identically zero. Then we put: $\mathfrak{P}_1 \equiv y_1$. We consider next the differential equation

$$\{\Psi, y_1\} \equiv \sum_{\rho=1}^{r} [c^r_{\rho 1} y_r + d_{\rho 1}] \frac{\partial \Psi}{\partial y_\rho} = 1 \qquad (15)$$

for a function $\Psi = \Psi(y_1 \cdots y_r)$ of the generators. We shall call $\mathfrak{Q}_1$ a particular solution of (15). Then we consider the following system of differential equations:

$$\begin{aligned} \{\mathfrak{P}_1, \Psi\} &= 0, \\ \{\mathfrak{Q}_1, \Psi\} &= 0, \end{aligned} \qquad (16)$$

where Ψ is still a function of the y_r. Since

$$\{\mathfrak{Q}_1, \{\mathfrak{P}_1, \Psi\}\} - \{\mathfrak{P}_1, \{\mathfrak{Q}_1, \Psi\}\} = \{\{\mathfrak{Q}, \mathfrak{P}_1\}, \Psi\} = 0 \qquad (17)$$

holds, the system (16) is a complete Jacobian system and admits $r - 2$ independent solutions $\psi_1, \psi_2, \cdots, \psi_{r-2}$. We note that such solutions are also independent of $\mathfrak{P}_1$ and $\mathfrak{Q}_1$ because if a relation

$$F(\mathfrak{P}_1, \mathfrak{Q}_1, \psi_1, \psi_2, \cdots, \psi_{r-2}) = 0 \qquad (18)$$

existed, we would have from (16)

$$\begin{aligned} \partial F/\partial \mathfrak{Q}_1 &= -\{\mathfrak{P}_1, F\} = 0, \\ \partial F/\partial \mathfrak{P}_1 &= \{\mathfrak{Q}_1, F\} = 0, \end{aligned} \qquad (19)$$

so that relation (18) would involve the $\psi_1, \psi_2, \cdots, \psi_{r-2}$ only. We remark furthermore that, due to the Jacobi identity, the expression $\{\psi_\alpha, \psi_\beta\}$ is also a solution of (16) so we can write

$$\{\psi_\alpha, \psi_\beta\} = \varphi_{\alpha\beta}(\psi_1, \psi_2, \cdots, \psi_{r-2}); \quad \alpha, \beta = 1, \cdots, r-2. \qquad (20)$$

If the $\varphi_{\alpha\beta}$ are not identically zero, we may put, for instance, $\mathfrak{P}_2 \equiv \psi_1$ and look for a function $\mathfrak{Q}_2$ of the variables $\psi_1, \psi_2, \cdots, \psi_{r-2}$ which satisfy the relation:

$$\{\mathfrak{Q}_2, \mathfrak{P}_2\} = 1. \qquad (21)$$

By iterating this procedure we can keep on reducing the number of variables, until we are left with a set of variables having identically zero Poisson brackets. We will denote them by $\mathfrak{I}_1, \mathfrak{I}_2, \cdots, \mathfrak{I}_k$. As to the values of k and $h = \frac{1}{2}(r - k)$, see Sec. 3.

Proof of Theorem 2: We put first of all

$$Q_1 \equiv \mathfrak{Q}_1, \cdots, Q_h \equiv \mathfrak{Q}_h$$
$$P_1 \equiv \mathfrak{P}_1, \cdots, P_h \equiv \mathfrak{P}_h$$

and assume that a definite choice of the variables $\mathfrak{I}_1, \mathfrak{I}_2, \cdots, \mathfrak{I}_k$ has been made. Apart from exceptional values of these variables (we shall return to this point in the following), it must be $n \geq h$. If $n = h$ the theorem is obviously established. Let us assume that $n > h$.

Following a procedure rather more involved but similar to the one used in proving Theorem 1, let us consider the new system of differential equations:

$$\begin{aligned} \{\mathfrak{Q}_i, \varphi\} &= 0 \\ \{\mathfrak{P}_j, \varphi\} &= 0 \end{aligned} \qquad i, j = 1, \cdots, h, \qquad (23)$$

where now φ is a function of the variables q_i, p_i. The system is complete and so it admits $2n - 2h$ independent solutions

$$\varphi_1(q, p), \varphi_2(q, p), \cdots, \varphi_{2n-2h}(q, p).$$

Repeating essentially the same argument given above it is possible to show that such functions are also independent of the variables $\mathfrak{Q}_i$, $\mathfrak{P}_j$. The $\mathfrak{I}_1, \cdots, \mathfrak{I}_k$ are then functions of $\varphi_1, \cdots, \varphi_{2n-2h}$ only and not of the $\mathfrak{Q}_i$'s and $\mathfrak{P}_j$'s.

Let us now consider the system

$$\{\mathfrak{I}_1, \Phi\} = \cdots = \{\mathfrak{I}_k, \Phi\} = 0, \qquad (24)$$

where

$$\Phi = \Phi(\varphi_1, \cdots, \varphi_{2n-2h}).$$

This system is obviously a complete one but the equations are not in general all independent, since the expressions $\mathfrak{I}_1, \cdots, \mathfrak{I}_k$ may be functionally dependent. We remark that indeed, owing to

$$\{\mathfrak{I}_i, \Phi\} = \sum_\alpha \{\mathfrak{I}_i, \varphi_\alpha\} \frac{\partial \Phi}{\partial \varphi_\alpha},$$

the number of independent equations in (24) is given by the rank of the matrix $||\{\mathfrak{I}_i, \varphi_\alpha\}||$. On the other hand, from the relation

$$\{\mathfrak{I}_i, \varphi_\alpha\} = \sum_\beta \frac{\partial \mathfrak{I}_i}{\partial \varphi_\beta} \{\varphi_\beta, \varphi_\alpha\}$$

it follows that the rank of the matrix $||\{\mathfrak{I}_i, \varphi_\alpha\}||$ is identical to that of the matrix $||\partial \mathfrak{I}_i/\partial \varphi_\beta||$,[8] i.e., to the number of independent $\mathfrak{I}_i$. Let $\mathfrak{I}_1, \cdots, \mathfrak{I}_l$ be such independent $\mathfrak{I}_i$. Equations (24) admit $2n - 2h - l$ independent solutions that we can write in the following way:

$$\mathfrak{I}_1, \cdots, \mathfrak{I}_l, w_1, \cdots, w_{2n-2h-2l}.$$

[8] The rank of the matrix formed with the Poisson brackets is invariant under any functional invertible transformation. So, in particular, the determinant of the matrix formed with Poisson brackets of the variables $\mathfrak{Q}_i$, $\mathfrak{P}_j$, φ_α and hence that of the submatrix $||\{\varphi_\alpha, \varphi_\beta\}||$ is different from zero (cf. Sec. 3).

Now let $\omega_1, \cdots, \omega_l$ be l functions of the φ_α, independent of the previous ones.

Then the Poisson brackets $\{\mathfrak{I}_1, \omega_u\}, \cdots, \{\mathfrak{I}_l, \omega_u\}$, $(u = 1, \cdots, l)$ are obviously not all identically zero and it is easy to see that a function Q_{h+1} of the variables $\mathfrak{I}_1, \cdots, \mathfrak{I}_l, \omega_1, \cdots, \omega_l$ exists such that

$$\begin{aligned} \{Q_{h+1}, \mathfrak{I}_1\} &= 1, \\ \{Q_{h+1}, \mathfrak{I}_2\} &= \cdots = \{Q_{h+1}, \mathfrak{I}_l\} = 0. \end{aligned} \tag{25}$$

[The system of equations

$$\{\mathfrak{I}_2, \Theta\} = \cdots = \{\mathfrak{I}_l, \Theta\} = 0,$$

$$\Theta = \Theta(\omega_1, \cdots, \omega_l, \mathfrak{I}_1, \cdots, \mathfrak{I}_l)$$

is complete and admits $l + 1$ solutions. Let them be: $\vartheta, \mathfrak{I}_1, \cdots, \mathfrak{I}_l$. For a function $Q = Q(\vartheta, \mathfrak{I}_1, \cdots, \mathfrak{I}_l)$ we have $\{Q, \mathfrak{I}_1\} = (\partial Q/\partial\vartheta)\{\vartheta, \mathfrak{I}_1\} = f(\vartheta, \mathfrak{I}_1, \cdots \mathfrak{I}_l)$. $\partial Q/\partial\vartheta$. So it is obviously possible to satisfy the first of Eqs. (25) while the others are automatically satisfied.] Let us consider next the system of equations

$$\{Q_{h+1}, \Phi\} = 0, \qquad \{\mathfrak{I}_1, \Phi\} = 0,$$

with $\Phi = \Phi(\mathfrak{I}_1, \cdots, \mathfrak{I}_l, \omega_1, \cdots, \omega_l)$ and let $\mathfrak{I}_2, \cdots, \mathfrak{I}_l, \omega_2', \cdots, \omega_l'$ be the independent solutions. A function Q_{h+2} of the variables $\mathfrak{I}_2, \cdots, \mathfrak{I}_l, \omega_2', \cdots, \omega_l'$ can be constructed such that

$$\begin{aligned} \{Q_{h+2}, \mathfrak{I}_2\} &= 1, \\ \{Q_{h+2}, \mathfrak{I}_3\} &= \cdots = \{Q_{h+2}, \mathfrak{I}_l\} = 0. \end{aligned} \tag{26}$$

The procedure can be iterated until we arrive to the construction of new expressions $Q_{h+1}, \cdots, Q_{h+l}$ such that

$$\begin{aligned} \{Q_{h+u}, Q_{h+u'}\} &= 0 \\ \{Q_{h+u}, \mathfrak{I}_{u'}\} &= \delta_{uu'}; \end{aligned} \qquad u, u' = 1 \cdots l. \tag{27}$$

We put $P_{h+u} \equiv \mathfrak{I}_u$. We are left now with the invariants $\mathfrak{I}_{l+1} \cdots \mathfrak{I}_k$ and the variables w_α.

Let us observe that $\{Q_{h+u}, \mathfrak{I}_{l+v}\}$ $(v = 1, \cdots, k - l;$ $u = 1, \cdots, l)$ is a function of $\mathfrak{I}_1, \cdots, \mathfrak{I}_l$ only.

Then let us consider a function $\Omega(\mathfrak{I}_1, \cdots, \mathfrak{I}_l, \mathfrak{I}_{l+1}, \cdots, \mathfrak{I}_k)$ and look for the solutions of the system

$$\begin{aligned} &\{Q_{h+1}, \Omega\} = 0, \\ &\{Q_{h+2}, \Omega\} = 0, \\ &\cdots\cdots\cdots\cdots \\ &\{Q_{h+l}, \Omega\} = 0. \end{aligned} \tag{28}$$

This system is complete and admits $k - l$ solutions that can be identified with the new expressions $\mathfrak{I}_{l+1}', \mathfrak{I}_{l+2}', \cdots, \mathfrak{I}_k'$. Finally we replace the variables $w_1, \cdots, w_{2n-2h-2l}$ with new ones $w_1', \cdots, w_{2n-2h-2l}'$ which are functions of them and of the $\mathfrak{I}_1, \cdots, \mathfrak{I}_l$ in such a way that their Poisson brackets with the $Q_{h+1}, \cdots, Q_{h+l}$ are zero, and we re-express these last variables w_α' in terms of canonical variables $Q_{h+l+1}, \cdots Q_n, P_{h+l+1}, \cdots, P_n$ closely following the method used in Theorem 1.

3. PROPERTIES OF THE *TYPICAL FORM*: THE IRREDUCIBLE KERNELS OF A CANONICAL REALIZATION

We want to illustrate now some of the results implicitly contained in the last section.

First of all we can invert the functions $\mathfrak{Q}_1 = \mathfrak{Q}_1(y_1, \cdots, y_r), \cdots, \mathfrak{I}_k = \mathfrak{I}_k(y_1, \cdots, y_r)$ of Theorem 1 and obtain

$$y_r = y_r(\mathfrak{Q}_1, \cdots, \mathfrak{Q}_h, \mathfrak{P}_1, \cdots, \mathfrak{P}_h, \mathfrak{I}_1, \cdots, \mathfrak{I}_k). \tag{29}$$

Then, from (13) we have

$$\{y_r, \mathfrak{I}_t\} = 0; \qquad t = 1, \cdots, k \tag{30}$$

so that, in virtue of (4) it is apparent that the $\mathfrak{I}_t$ do not change under $\mathfrak{R}$.

It is also easy to establish (see the following) that the $\mathfrak{I}_1, \mathfrak{I}_2, \cdots, \mathfrak{I}_k$ are the only independent functions of the generators y_r which have zero Poisson brackets with all the y_r themselves. Thus, they are the analogs of the invariant operators of the ordinary representation theory: we shall call them *canonical invariants*. As to their number, we observe that the following property holds true: if $\eta_1, \eta_2, \cdots, \eta_m$ are a set of variables among which certain Poisson brackets are defined and $\xi_1, \xi_2, \cdots, \xi_m$ are obtained from them through some transformations, we have

$$\{\xi_A, \xi_B\} = \frac{\partial\xi_A}{\partial\eta_a}\frac{\partial\xi_B}{\partial\eta_b}\{\eta_a, \eta_b\}; \quad A, B, a, b = 1, \cdots, m \tag{31}$$

so that if $\det \|\partial\xi_A/\partial\eta_a\|$ is different from zero, the rank of the two matrices $\|\{\xi_A, \xi_B\}\|$ and $\|\{\eta_a, \eta_b\}\|$ is the same. Thus, in particular, the two matrices

$$\begin{bmatrix} \{y_1, y_1\} & \{y_1, y_2\} & \cdots & \{y_1, y_r\} \\ \{y_2, y_1\} & \{y_2, y_2\} & \cdots & \\ \cdots & \cdots & \cdots & \\ \{y_r, y_1\} & \cdots & \cdots & \{y_r, y_r\} \end{bmatrix}$$

$$\equiv \begin{bmatrix} 0 & c_{12}^r y_r + d_{12} & \cdots & c_{1r}^r y_r + d_{1r} \\ c_{21}^r y_r + d_{21} & 0 & \cdots & \\ \cdots & \cdots & \cdots & \\ c_{r1}^r y_r + d_{r1} & \cdots & \cdots & 0 \end{bmatrix} \tag{32}$$

and

$$\left|\begin{array}{ccc|ccc} \{\mathfrak{Q}_1,\mathfrak{Q}_1\} & \{\mathfrak{Q}_1,\mathfrak{P}_1\} & \cdots & \{\mathfrak{Q}_1,\mathfrak{I}_1\} & \cdots & \{\mathfrak{Q}_1,\mathfrak{I}_k\} \\ \{\mathfrak{P}_1,\mathfrak{Q}_1\} & \{\mathfrak{P}_1,\mathfrak{P}_1\} & \cdots & \{\mathfrak{P}_1,\mathfrak{I}_1\} & \cdots & \{\mathfrak{P}_1,\mathfrak{I}_k\} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hline \{\mathfrak{I}_1,\mathfrak{Q}_1\} & \{\mathfrak{I}_1,\mathfrak{P}_1\} & \cdots & \{\mathfrak{I}_1,\mathfrak{I}_1\} & \cdots & \{\mathfrak{I}_1,\mathfrak{I}_k\} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \{\mathfrak{I}_k,\mathfrak{Q}_1\} & \{\mathfrak{I}_k,\mathfrak{P}_1\} & \cdots & \{\mathfrak{I}_k,\mathfrak{I}_1\} & \cdots & \{\mathfrak{I}_k,\mathfrak{I}_k\} \end{array}\right| \equiv \left|\begin{array}{ccc|c} 0 & 1 & \cdots & \\ -1 & 0 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \\ \hline & & & \\ & \mathbf{0} & & \mathbf{0} \\ & & & \end{array}\right| \tag{33}$$

have the same rank. Because the rank of the latter is obviously $2h$, we conclude that the number k of invariants is given directly by r minus the generic rank of the matrix (32).

Let us now analyze the properties of the realization $\overline{\mathfrak{R}}$ of $\mathcal{G}$ (*typical form*) induced onto the new system of variables Q_i, P_i [according to our definitions $\overline{\mathfrak{R}}$ is canonically equivalent of $\mathfrak{R}$]. To this aim, let us summarize the properties of the new system of variables in the following scheme:

$$\begin{array}{cc|c|c} \text{I} & \text{II} & \text{III} & \text{IV} \\ P_1 \equiv \mathfrak{P}_1 \cdots P_h \equiv \mathfrak{P}_h & \mathfrak{P}_{h+1} \equiv \mathfrak{I}_1 \cdots \mathfrak{P}_{h+l} \equiv \mathfrak{I}_l & \mathfrak{I}_{l+1} \cdots \mathfrak{I}_k & P_{h+l+1} \cdots P_n \\ Q_1 \equiv \mathfrak{Q}_1 \cdots Q_h \equiv \mathfrak{Q}_h & \mathfrak{Q}_{h+1} \cdots \mathfrak{Q}_{h+l} & & Q_{h+l+1} \cdots Q_n, \end{array} \tag{34}$$

where the Poisson brackets between any two variables is one if they are on the same column and zero otherwise. In particular we have

$$\begin{aligned} &\{Q_i, \mathfrak{I}_{l+v}\} \equiv \partial\mathfrak{I}_{l+v}/\partial P_i = 0; \qquad i, j = 1, \cdots, n, \\ &\{P_j, \mathfrak{I}_{l+v}\} \equiv -\partial\mathfrak{I}_{l+v}/\partial Q_j = 0; \quad v = 1, \cdots, k-l \end{aligned} \tag{35}$$

so that $\mathfrak{I}_{l+1}, \cdots, \mathfrak{I}_k$ are constants.

On the other hand, since they are functions of the generators $y_1, \cdots, y_r$ this means that the generators themselves are not independent functions in the considered realization, but there exist $k - l$ relations among them. (We notice however that no relation of the form $\sum_{r=1}^{r} c_r y_r = \text{const}$ can actually exist among the generators y_r in a faithful realization. In such a case, the infinitesimal operators $\{y_r, \cdots\}$ would indeed be linearly dependent.) Another way to express the above mentioned fact is to observe that, owing to Eqs. (29), it follows that the y_r are functions of the variables $Q_1, \cdots, Q_h$, $P_1, \cdots, P_{h+l}$ only.

Let us examine some consequences of this result. first of all, we have

$$\begin{aligned} &\{y_r, P_{h+\alpha}\} = 0; \qquad \alpha = 1, \cdots, n-h, \\ &\{y_r, Q_{h+l+\beta}\} = 0; \qquad \beta = 1, \cdots, n-h-l, \end{aligned} \tag{36}$$

so that the variables of the set $P_{h+1}, \cdots, P_{h+l}$, $P_{h+l+1}, \cdots, P_n$, $Q_{h+l+1}, \cdots, Q_n$ do not change under the transformation $\overline{\mathfrak{R}}$.

Moreover, it is easy to see that the variables $Q_{h+1}, \cdots, Q_{h+l}$ are transformed, in the neighborhood of the identity, according to the simple law

$$\begin{aligned} Q'_{h+u} = Q_{h+u} \\ - \delta a^r \, \partial y_r(Q_1, \cdots, Q_h, P_1, \cdots, P_{h+l})/\partial P_{h+u}; \\ u = 1, \cdots, l, \end{aligned} \tag{37}$$

which is a sort of translation depending on the values of the other variables. The corresponding finite transformations can be obtained from Eq. (37) by means of a simple quadrature.

The most significant transformation properties are those of the variables of the first set of table (34). We shall call the transformations induced in such variables, for every set of allowable fixed values of the canonical invariants, an *irreducible kernel* of the canonical realization.

We notice that, in particular cases, it may happen to have $n = h + l$, then the fourth set of table (34) is missing; or $l = 0$, then the second set of (34) is missing. A particular interesting situation arises when $n = h$, that is when the realization coincides with an irreducible kernel. In general $2h$ is the minimum number of variables for which we can have a faithful canonical realization of the group $\mathcal{G}$, with the exception of some degenerate cases corresponding to particular values of the invariants for which we are on the boundary of the domain of the variables $\mathfrak{Q}_1, \cdots, \mathfrak{Q}_h$, $\mathfrak{P}_1, \cdots, \mathfrak{P}_h$. We shall meet such degenerate cases in studying the canonical realizations of the Galilei and of the Lorentz group.

We will now make use of the concept of *invariant manifold* in the phase space. Let us consider a manifold of equation

$$\mathfrak{F}(q, p) = 0; \tag{38}$$

it will be said to be invariant if every point in it is transformed into another point of the same manifold by any transformation of the realization $\mathfrak{R}$. A necessary and sufficient condition for that is obviously

$$\{y_\tau, \mathfrak{F}(q, p)\} = 0 \tag{39}$$

or, using the variables of table (34),

$$\{Q_i, \mathfrak{F}(Q, P)\} \equiv \partial\mathfrak{F}/\partial P_i = 0,$$
$$\{P_j, \mathfrak{F}(Q, P)\} \equiv -\partial\mathfrak{F}/\partial Q_j = 0;$$
$$i = 1 \cdots h, \qquad j = 1 \cdots h + l. \tag{40}$$

This implies

$$\mathfrak{F} = \mathfrak{F}(P_{h+1}, \cdots, P_{h+l}; Q_{h+l+1}, \cdots, Q_n, P_{h+l+1}, \cdots, P_n);$$

that is, $\mathfrak{F}$ is simply function of those of the canonical variables that remain unchanged under the realization $\overline{\mathfrak{R}}$. In particular, we observe that the only independent functions of the generators that give rise to invariant manifolds are the functions $\mathfrak{I}_1 \equiv P_{h+1}, \cdots, \mathfrak{I}_l \equiv P_{h+l}$.

Extending a terminology due to Lie, we will call intransitive (or transitive) a canonical realization which admits (or does not admit) invariant manifolds. Thus, apart from the above-mentioned exceptional cases, the transitive faithful realizations, if they exist, are those for which $n = h$ and correspond to definite values of the canonical invariants. As stated above, each of them is canonically equivalent to a single, irreducible kernel. The nontrivial intransitive realizations are those which contain infinitely many irreducible kernels corresponding to different sets of values of the canonical invariants.

It is apparent that the intransitive and transitive canonical realizations are in some way the analogs of the reducible and irreducible representations. It is clear however that a strict analog of the reduction operation, i.e., decomposition in direct sum of irreducible subspaces, does not exist. Finally, as regards the variables of the set IV of (34), it will be apparent in studying the canonical realizations of physical symmetry groups that they correspond to internal variables of isolated systems.

4. CHARACTERIZATION OF THE MOST GENERAL CANONICAL REALIZATION OF A LIE GROUP

We have now to handle the somewhat inverse problem, that is the construction of the most general realization $\mathfrak{R}$ of the group $\mathcal{G}$. In particular we may ask whether a faithful realization can be constructed in which a set of invariants assume certain values prescribed in their accessible domain.

Let us consider, to this purpose, the functions $\mathfrak{Q}_i(y)$, $\mathfrak{P}_i(y)$ and $\mathfrak{I}_i(y)$ where the y_τ have now to be considered as generic independent variables; we have already noted in Sec. 2 that these functions are constructed by using only relations (6).

Let us give fixed values for instance to $\mathfrak{I}_{l+1}, \cdots, \mathfrak{I}_k$. We introduce a set of $2n$ variables Q_i, P_j reproducing a table like (34), with the abstractly defined Poisson brackets relations

$$\{Q_i, Q_j\} = \{P_i, P_j\} = 0,$$
$$\{Q_i, P_j\} = \delta_{ij}; \qquad i, j = 1 \cdots n. \tag{41}$$

The y_τ turn out to be functions of the variables $Q_1, \cdots, Q_h, P_1, \cdots, P_{h+l}$. Such functions will satisfy relations (6), so, for the general theory of transformation groups, they will be generators of a group of canonical transformations homomorphic to $\mathcal{G}$ at least in the neighborhood of the identity. This group will be in addition isomorphic to $\mathcal{G}$ and so a faithful realization, if the infinitesimal operators $\{y_\tau, \cdots\}$ are linearly independent. That amounts to requiring that no relation like

$$\sum_{\tau=1}^{r} \lambda_\tau(\overline{\mathfrak{I}}_{l+1} \cdots \overline{\mathfrak{I}}_k) y_\tau(\mathfrak{Q}_1, \cdots, \mathfrak{Q}_h, \mathfrak{P}_1, \cdots, \mathfrak{P}_h, \cdot \mathfrak{I}_1, \cdots, \mathfrak{I}_l; \overline{\mathfrak{I}}_{l+1}, \cdots, \overline{\mathfrak{I}}_k) = \Omega(\overline{\mathfrak{I}}_{l+1}, \cdots, \overline{\mathfrak{I}}_k) \tag{42}$$

(here the symbols $\overline{\mathfrak{I}}_{l+\tau}$ denote fixed values of the invariants) or, equivalently,

$$\sum_{\tau=1}^{r} \lambda_\tau \frac{\partial y_\tau}{\partial \mathfrak{Q}_i} = 0, \qquad i = 1 \cdots h,$$
$$\sum_{\tau=1}^{r} \lambda_\tau \frac{\partial y_\tau}{\partial \mathfrak{P}_j} = 0, \qquad j = 1 \cdots h, \tag{43}$$
$$\sum_{\tau=1}^{r} \lambda_\tau \frac{\partial y_\tau}{\partial \mathfrak{I}_u} = 0, \qquad u = 1 \cdots l$$

holds true.

Relations (43) must be satisfied for any value of the variables $\mathfrak{Q}_i$, $\mathfrak{P}_i$, $\mathfrak{Q}_u$. Alternatively, we may consider together with (43) the relations obtained with the higher-order derivatives. In this way we get a set of linear equations for λ_τ. As it is well

known, only a finite number of these are independent; precisely in any case, it is sufficient to consider the equations formed with all the derivatives up to order $r - 2h - l + 1$ (see for instance Ref. 7).

Let us then consider the matrix

$$\begin{bmatrix} \frac{\partial y_1}{\partial \Omega_1} \cdots \frac{\partial y_1}{\partial \mathfrak{P}_h} \frac{\partial^2 y_1}{\partial \Omega_1^2} \frac{\partial^2 y_1}{\partial \Omega_1 \partial \mathfrak{P}_1} \cdots \frac{\partial^{k-l+1} y_1}{\partial \mathfrak{I}_l^{k-l+1}} \\ \frac{\partial y_2}{\partial \Omega_1} \cdots \frac{\partial y_2}{\partial \mathfrak{P}_h} \frac{\partial^2 y_2}{\partial \Omega_1^2} \frac{\partial^2 y_2}{\partial \Omega_1 \partial \mathfrak{P}_1} \cdots \frac{\partial^{k-l+1} y_2}{\partial \mathfrak{I}_l^{k-l+1}} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial y_r}{\partial \Omega_1} \cdots \frac{\partial y_r}{\partial \mathfrak{P}_h} \frac{\partial^2 y_r}{\partial \Omega_1^2} \frac{\partial^2 y_r}{\partial \Omega_1 \partial \Omega_1} \cdots \frac{\partial^{k-l+1} y_r}{\partial \mathfrak{I}_l^{k-l+1}} \end{bmatrix}, \qquad (44)$$

Clearly the condition for a relation like (42) to exist is that the rank q_l of (44) be smaller than r. Thus, conversely, the condition under which a faithful realization may be constructed in which the invariants $\mathfrak{I}_{l+1}, \cdots, \mathfrak{I}_k$ have fixed values, is expressed by $q_l = r$.

We observe that, in particular, for $l = k$ (no invariant fixed) the matrix (44) becomes the Jacobian of the transformation $(y_r) \to (\Omega_i, \mathfrak{P}_j, \mathfrak{T}_t)$ which is necessarily different from zero, so that $q_k = r$.

On the other hand, if $q_0 = r$, because obviously $q_0 \leq q_l$, all the realizations constructed by the above procedure are faithful (that is faithful realizations exist which correspond to invariants fixed in an arbitrary number and in an arbitrary way). In particular, only in this latter case all the invariants can be fixed: that gives rise to transitive faithful realizations, or, what is the same, the irreducible kernels give faithful realizations.

As illustration of the above arguments, let us consider a few examples:

(1) Assume the group $\mathfrak{G}$ to have an exceptional subgroup of order m, that is a subgroup the elements of which commute with every element of the group $\mathfrak{G}$ itself. The group $\mathfrak{G}$ is obviously not semi-simple. In this case m invariants exist which are independent linear combinations of the y_r, so, certainly, no more than $k - m$ invariants can be fixed. [One might think that case (1) is the only one in which unfaithful realizations of the kind considered here can actually occur. However, this is not the case. This and other open questions will be dealt with in a succeeding paper.]

(2) Suppose next $\mathfrak{G}$ to be a simple group. In this case only faithful realizations exist and thus all the invariants can be fixed.

(3) Let $\mathfrak{G}$ be now a semisimple group. The same conclusion holds about the invariants as in Case 2. To see this, let us first consider the realization $\mathfrak{R}$ in which no invariant has been fixed. As seen above, $\mathfrak{R}$ is isomorphic to $\mathfrak{G}$ and therefore is itself semisimple. Let us now take into account the realization $\mathfrak{R}'$ (not canonical in our sense!) induced by $\mathfrak{R}$ into an invariant manifold corresponding to certain definite values of all the invariants $\mathfrak{I}_1, \cdots, \mathfrak{I}_k$; we mean here the transformation laws for the variables $P_1, \cdots, P_h$, $Q_1, \cdots, Q_h, Q_{h+1}, \cdots, Q_{h+k}$ corresponding to those values of the $\mathfrak{I}_t$, that is the transformations (37) and the transformations of the irreducible kernel labeled by such fixed values. Moreover, let us consider the realization $\mathfrak{R}_0$ formed only by the irreducible kernel. From general properties of transformation groups theory, we know that $\mathfrak{R}'$ is isomorphic to $\mathfrak{R}$ unless at least one infinitesimal transformation of $\mathfrak{R}$ exists which leaves unchanged every point of the manifold onto which $\mathfrak{R}'$ is induced. This may happen in our case only for exceptional values of the invariants such that the Jacobian determinant $\partial y/\partial(\Omega, \mathfrak{P}, \mathfrak{I})$ be identically zero with respect to Ω_i, $\mathfrak{P}_j$. Apart from these exceptional values, $\mathfrak{R}'$ is isomorphic to $\mathfrak{G}$ and so a semisimple group itself. Now $\mathfrak{R}_0$ is certainly isomorphic to $\mathfrak{R}'$. In fact, had an (invariant) subgroup existed in $\mathfrak{R}'$, corresponding to the identity in $\mathfrak{R}_0$, this would necessarily consist of a set of transformations leaving invariant the variables of the irreducible kernel. On the other hand, due to the particular form of transformations (37), this subgroup would necessarily be an Abelian invariant subgroup. Therefore $\mathfrak{R}_0$ is isomorphic to $\mathfrak{G}$.

The remaining questions to be answered are now the following:

(i) are the realizations considered above all distinct?

(ii) are they inequivalent?

As to the second point, we find that canonical realizations corresponding to a different choice of the invariants to be fixed or to different values of such fixed invariants, if distinct, are inequivalent. This is a consequence of the fact that a canonical transformation in the variables Q_i, P_i does not change the values of the fixed invariants $\mathfrak{I}_{l+1}, \cdots, \mathfrak{I}_k$ and can modify the values of the remaining invariants in a prescribed way. All this can be very easily seen for an infinitesimal canonical transformation, remembering that we have, being $G(Q, P)$ the generating function,

$$\begin{aligned} &\{G(Q,P), \mathfrak{I}_{l+v}\} = 0; \quad v = 1, \cdots, k-l, \\ &\{G(Q,P), \mathfrak{I}_u\} = \partial G/\partial Q_{h+u}; \quad u = 1, \cdots, l. \end{aligned} \qquad (45)$$

A similar conclusion obviously holds true for canonical realizations corresponding to different values of the $d_{\rho\sigma}$'s since a canonical transformation does not change the Poisson brackets. These results are strictly analogous to well-known properties of the unitary ray representations considered in quantum mechanics.

The first point requires a more detailed answer. We assume that the $d_{\rho\sigma}$'s have already been reduced to the minimum number. In order that two realizations with generators $y_\tau(Q, P)$ and $y'_\tau(Q, P)$, in the same number of variables, be nondistinct, we must have

$$\{y'_\tau - y_\tau, Q_i\} = \{y'_\tau - y_\tau, P_i\} = 0; \qquad i, j = 1, \cdots, n; \tag{46}$$

that is $y'_\tau = y_\tau + \alpha_\tau$, where the α_τ's are constants in the given realizations.

Moreover, if the y_τ's and y'_τ's have to satisfy the relations (6) with the same values of the $d_{\rho\sigma}$'s, the conditions

$$c^\tau_{\rho\sigma}\alpha_\tau = 0 \tag{47}$$

must be satisfied.

Let us denote by $\alpha^{(1)}_\tau, \cdots, \alpha^{(p)}_\tau$ the linearly independent solutions of Eqs. (47); obviously $p = r - q$ (cf. Sec. I). A necessary and sufficient condition in order that identical realizations exist corresponding to different fixed values of the invariants $\mathfrak{I}_1, \cdots, \mathfrak{I}_\nu$, is that the generators y_τ assume the form

$$y_\tau = g_\tau(\mathfrak{Q}, \mathfrak{P}, \mathfrak{I}_{\nu+1}, \cdots, \mathfrak{I}_k) + \sum_{\sigma=1}^{p} \alpha^{(\sigma)}_\tau f_\sigma(\mathfrak{I}_1, \cdots, \mathfrak{I}_\nu).$$

More generally, we shall assume that we have

$$y_\tau = g_\tau(\mathfrak{Q}, \mathfrak{P}, \mathfrak{I}_{\nu+1}, \cdots, \mathfrak{I}_k) + \sum_{\sigma=1}^{p} \alpha^{(\sigma)}_\tau \mathfrak{I}'_\sigma(\mathfrak{I}_1, \cdots, \mathfrak{I}_k), \tag{48}$$

where the $\mathfrak{I}'_\sigma$ are ν arbitrary independent functions of all the invariants. We shall also assume that, in Eqs. (48), ν has the maximum possible value.

[For given solutions $\mathfrak{Q}_i(y)$, $\mathfrak{P}_i(y)$, $\mathfrak{I}_i(y)$ of the problem of Theorem 1, let us write the inverse relations (38). Then, consider the expressions

$$y'_\tau = y_\tau(\mathfrak{Q}, \mathfrak{P}, \mathfrak{I}^0_1, \mathfrak{I}^0_2, \cdots, \mathfrak{I}^0_\nu, \mathfrak{I}_{\nu+1}, \cdots \mathfrak{I}_k) + \sum_{\sigma=1}^{p} \alpha^{(\sigma)}_\tau \mathfrak{I}'_\sigma(\mathfrak{I}_1, \cdots, \mathfrak{I}_k), \tag{48'}$$

where $\mathfrak{I}^0_1, \cdots, \mathfrak{I}^0_\nu$ are certain fixed values of the invariants. The y'_τ (48') satisfy the relations (6) with the same values of the $d_{\rho\sigma}$'s. Now, if the ν invariants $\mathfrak{I}'_1, \cdots, \mathfrak{I}'_\nu$ can be chosen in such a way that the Jacobian $\partial(y'_1, \cdots, y'_r)/\partial(\mathfrak{Q}\mathfrak{P}\mathfrak{I})$ in (48') is not identically zero, we can consider the Eqs. (48') as an implicit definition of certain new solutions $\mathfrak{Q}''_i(y_1, \cdots, y_r)$, $\mathfrak{P}''_i(y_1, \cdots, y_r)$, $\mathfrak{I}''_i(y_1, \cdots, y_r)$ of the problem of Theorem 1. For these new solutions, Eqs. (48) evidently hold if we put

$$g_\tau(\mathfrak{Q}'', \mathfrak{P}'', \mathfrak{I}''_{\nu+1}, \cdots, \mathfrak{I}''_k) = y_\tau(\mathfrak{Q}'', \mathfrak{P}'', \mathfrak{I}^0_1, \cdots \mathfrak{I}^0_\nu, \mathfrak{I}''_{\nu+1}, \cdots, \mathfrak{I}''_k).]$$

Equations (48) enable us to recognize easily whether two realizations are distinct or not. First of all we remark that the invariants are divided into two classes, namely $\mathfrak{I}_{\nu+1}, \cdots, \mathfrak{I}_k$ and $\mathfrak{I}'_1, \cdots, \mathfrak{I}'_\nu$. It is clear that two realizations corresponding to different sets of fixed invariants of the first class or to different values of the same set of this same class are certainly distinct. Then, let us restrict ourselves to realizations for which a common set of invariants is fixed in the first class, with the same values. Among these, let us consider a realization in which η relations exist among the invariants of the second class. We can write such relations as

$$\begin{aligned} \mathfrak{I}'_1 - V_1(\mathfrak{I}'_{\eta+1}, \cdots, \mathfrak{I}'_\nu) &= 0 \\ \vdots \qquad\qquad & \vdots \\ \mathfrak{I}'_\eta - V_\eta(\mathfrak{I}'_{\eta+1}, \cdots, \mathfrak{I}'_\nu) &= 0; \qquad \eta \leq \nu. \end{aligned} \tag{49}$$

Then, all other realizations in which the expressions

$$\begin{aligned} U_1 &\equiv \mathfrak{I}'_1 - V_1(\mathfrak{I}'_{\eta+1}, \cdots, \mathfrak{I}'_\nu) \\ \vdots \qquad & \vdots \\ U_\eta &\equiv \mathfrak{I}'_\eta - V_\eta(\mathfrak{I}'_{\eta+1}, \cdots, \mathfrak{I}'_\nu) \end{aligned} \tag{50}$$

have fixed arbitrary values are identical. This, obviously, includes the simple case in which $\mathfrak{I}'_1, \cdots, \mathfrak{I}'_\eta$ have arbitrary fixed values.

An interesting physical example of the preceeding situation is found in studying the canonical realizations of the Galilei group. We have in this group two invariants: one of the first class which represents the square of the intrinsic angular momentum, the other one of the second class which represents the internal energy of the physical system. We shall see in a future paper that the canonical realizations of the Galilei group labeled by fixed values of such latter invariant are the canonical realizations corresponding to the free particle, the free rotator, and

the free spherical top. Giving different values to this invariant amounts only in changing the zero-point energy of these systems and we get identical realizations.

5. CONCLUDING REMARKS

Summarizing the results obtained, we can give the following rules to build up the most general faithful canonical realization of a Lie group $\mathcal{G}$.

First, one takes the Poisson bracket relations among the generators y_τ and reduces the number of the nonzero $d_{\rho\sigma}$'s to a minimum. Then, one constructs the variables $\mathfrak{Q}_1, \cdots, \mathfrak{Q}_h, \mathfrak{P}_1, \cdots, \mathfrak{P}_h, \mathfrak{I}_1, \cdots, \mathfrak{I}_k$. Next, one prescribes definite values for some of the chosen invariants according to the criteria of Sec. 4.

When this is done, one introduces a set of new variables $Q_1, \cdots, Q_n, P_1, \cdots, P_n$ with abstractly defined Poisson brackets, achieving a scheme like the table (34) of Sec. 3. Finally, one performs an arbitrary constant canonical transformation.

To conclude, we would like to stress some aspects of the theory that could be interesting also for representation theory and therefore for quantum mechanics.

Precisely, we point out that if in Eqs. (6) the Poisson brackets $\{\ \}$ are replaced by the commutators and the y_τ's are interpreted as the infinitesimal operators of a projective representation of $\mathcal{G}$, the expressions $\mathfrak{I}_1(y_1, \cdots, y_r), \cdots, \mathfrak{I}_k(y_1, \cdots, y_r)$, when attention is paid to the order of the y_τ's, become the invariant operators. Therefore, our procedure can be clearly utilized for the determination of the number and for the actual construction of the invariant operators of any Lie group $\mathcal{G}$. This is of special interest in case $\mathcal{G}$ is not a semisimple group. The semisimple groups are indeed the only ones for which a general method was known up to now (see for instance Ref. 9).

In the same order of ideas, the variables $\mathfrak{P}_1, \cdots, \mathfrak{P}_h$ or $\mathfrak{Q}_1, \cdots, \mathfrak{Q}_h$ furnish directly a complete system of commuting observables acting within the irreducible representations. However, the analytic structure of such quantities is generally much more involved than that of the invariants and the questions of order may become essential ones. Actually, some of the corresponding operators might fail to exist.

ACKNOWLEDGMENTS

It is a great pleasure to thank Professors P. Caldirola and P. Gulmanelli for their kind interest, and Professors A. Loinger and G. Racah for very useful discussions.

[9] G. Racah, "Group Theory and Spectroscopy," Lectures, Princeton University (1951); (CERN Report 61-8).

S-Operator Theory. II. Quantum Electrodynamics*

ROBERT E. PUGH
Department of Physics, University of Toronto, Toronto, Canada
(Received 24 May 1965)

The formalism of S-operator theory is generalized to apply to all low-spin particles with particular emphasis on quantum electrodynamics. It is shown that only direct (nongradient) interactions are allowed, when spinor fields are involved. The Pauli interaction is explicitly shown to be forbidden. Problems of gauge invariance are discussed and commutation rules between the interpolating quantum fields and $\partial_\mu A_\mu^{\text{in}}(x)$ are derived. Finite S-matrix elements are easily calculated with complete convergence at each stage of the calculation.

I. INTRODUCTION

In a previous paper,[1] referred to hereinafter as I, the quantum theory of spin-zero chargeless particles was developed in terms of functional derivatives of the S-operator with respect to the free fields. It was shown that the equations

$$S^\dagger \delta^n S/\delta x_1 \cdots \delta x_n \equiv (-i)^n K_1 \cdots K_n \varphi(\varphi(x_1) \cdots \varphi(x_n)), \quad \text{(I.1)}$$

together with boundary conditions determined the theory. [Here $\delta/\delta x \equiv \delta/\delta\varphi_{\text{in}}(x)$.] The $n = 0$ part of this equation gave generalized unitarity[2] (unitarity off the mass shell); the $n = 1$ part defined the interpolating field,[3] $\varphi(x)$, while the $n = 2$ part defined the dynamics of the quantum theory. The parts of (I.1) with $n > 2$ were derivable from the $n = 0, 1, 2$ parts and thus said nothing new.

The purpose of this paper is twofold: first, to generalize the formalism of I to the interaction of arbitrary low spin particles (spin 0, $\frac{1}{2}$, and the photon), and second, to examine the special problems of gauge invariance in this framework.

In principle the formulation of quantum electrodynamics requires many more equations than does the formulation of scalar field theory because in quantum electrodynamics there are three fields interacting instead of one. In practice, however, it is possible, by a judicious choice of notation, to write most equations in a form that is independent of the type of field involved. The best notation for abbreviating quantum electrodynamics is discussed in Sec. II.

The calculus of functional derivatives with respect to free fields was discussed in I. This calculus is revised and generalized in Sec. III. In Sec. IV the dynamics of quantum electrodynamics is formulated and discussed. Secs. V and VI are devoted to problems of gauge invariance. A sample calculation is performed in Sec. VII just to show the simplicity of this formulation.

II. NOTATION

In quantum electrodynamics one deals with the electron field $\psi(x)$, its adjoint $\bar{\psi}(x)$, and the photon field $A_\mu(x)$. We let $\varphi(x)$ stand for any of these three fields and we write φ_i for $\varphi(x_i)$. If $\varphi_i = \psi(x_i)$, we say $i \sim F$, while if $\varphi_i = \bar{\psi}(x_i)$, we say $i \sim \bar{F}$, and if $\varphi_i = A_{\mu_i}(x_i)$ we say $i \sim B$. This can be written briefly as

$$\varphi_i = (A_{\mu_i}(x_i), \psi(x_i), \bar{\psi}(x_i)) \text{ as } i \sim (B, F, \bar{F}) \quad \text{(II.1)}$$

and the adjoint operator can be written

$$\bar{\varphi}_i = (A_{\mu_i}(x_i), \bar{\psi}(x_i), \psi(x_i)) \text{ as } i \sim (B, F, \bar{F}). \quad \text{(II.2)}$$

In the following, any triplet will refer to the order $(B, F, \bar{F})$, so we shall hereafter omit the phrase "as $i \sim (B, F, \bar{F})$" leaving it understood.

The differential operators K_i are defined as

$$K_i = (\Box_i, -\vec{\partial}_i - m, -\overleftarrow{\partial}_i + m). \quad \text{(II.3)}$$

Here $\vec{\partial}_i \equiv \gamma_\mu \partial/\partial x_i^\mu$ and $\overleftarrow{\partial}_i$ is the same as $\vec{\partial}_i$ except that it differentiates to the left instead of to the right. The free field equations can then be written in the concise form

$$K_i \varphi_i^{\text{in}} = 0, \quad \text{(II.4)}$$

where

$$K_i \varphi_i^{\text{in}} \equiv (\Box_i A_{\mu_i}^{\text{in}}(x_i), (-\vec{\partial}_i - m)\psi_{\text{in}}(x_i), \bar{\psi}_{\text{in}}(x_i)(-\overleftarrow{\partial}_i + m)). \quad \text{(II.5)}$$

It is convenient to introduce differential operators $\bar{K}_i$ that are complementary to K_i:

$$\bar{K}_i \equiv (1, -\vec{\partial}_i + m, -\overleftarrow{\partial}_i - m). \quad \text{(II.6)}$$

* This work was supported in part by the National Research Council of Canada.
[1] R. Pugh, J. Math. Phys. 6, 740 (1965).
[2] K. Nishijima, Phys. Rev. 119, 485 (1960).
[3] R. Pugh, Ann. Phys. (N. Y.) 23, 335 (1963).

These operators are defined so that

$$K_i\bar{K}_i = \Box_i - m_i^2 \tag{II.7}$$

with $m_i = (0, m, m)$.

The homogeneous Δ-functions, $\Delta(x_i; m_i)$, will be written simply as $\Delta(x_i)$:

$$\Delta(x_i) \equiv (D(x_i), \Delta(x_i; m), \Delta(x_i; m)) \tag{II.8}$$

so that

$$\bar{K}_i\Delta(x_i) = (D(x_i), -S(x_i), -S(-x_i)) \tag{II.9}$$

and

$$K_i\bar{K}_i\Delta(x_i) = 0. \tag{II.10}$$

The inhomogeneous Δ-functions are defined by

$$\begin{aligned}\Delta_R(x) &= \theta(x)\Delta(x),\\ \Delta_A(x) &= -\theta(-x)\Delta(x),\\ \Delta_C(x) &= \theta(x)\Delta_+(x) - \theta(-x)\Delta_-(x)\end{aligned} \tag{II.11}$$

so that

$$\bar{K}_i\Delta_R(x_i) = (D_R(x_i), -S_R(x_i), S_A(-x_i)), \tag{II.12}$$

$$\bar{K}_i\Delta_A(x_i) = (D_A(x_i), -S_A(x_i), S_R(-x_i)), \tag{II.13}$$

$$\bar{K}_i\Delta_C(x_i) = (D_C(x_i), -S_C(x_i), S_C(-x_i)). \tag{II.14}$$

These inhomogeneous Δ-functions satisfy

$$K_i\bar{K}_i\Delta_R(x_i) = K_i\bar{K}_i\Delta_A(x_i) = K_i\bar{K}_i\Delta_C(x_i) = -\delta(x_i). \tag{II.15}$$

Next we introduce the symbols

$$\delta_{iB} = (1, 0, 0), \quad \delta_{iF} = (0, 1, 0), \quad \delta_{i\bar{F}} = (0, 0, 1), \tag{II.16}$$

$$\epsilon_{iF} = \delta_{iF} - \delta_{i\bar{F}} = (0, 1, -1), \tag{II.17}$$

$$\tilde{\delta}_{ij} = \delta_{\mu_i\mu_j}\delta_{iB}\delta_{jB} + \delta_{iF}\upsilon_{j\bar{F}} + \delta_{i\bar{F}}\delta_{jF}, \tag{II.18}$$

$$\tilde{\delta}_j^i = \delta_{\mu_i\mu_j}\delta_{iB}\delta_{jB} + \delta_{iF}\delta_{j\bar{F}} + \delta_{i\bar{F}}\delta_{jF}. \tag{II.19}$$

With this notation the (anti) commutation rules for the free fields take the form

$$[\varphi_i^{\text{in}}, \varphi_j^{\text{in}}]_{\sigma_{ij}} = -i\tilde{\delta}_j^i\bar{K}_i\Delta(x_i - x_j). \tag{II.20}$$

Here σ_{ij} is a sign subscript: it is $-$ unless both i and j refer to F and/or $\bar{F}$ in which case it is $+$.

Finally, the functional derivative with respect to the free fields will be written as $\delta/\delta x_i$ meaning $\delta/\delta\varphi_i^{\text{in}}$:

$$\frac{\delta}{\delta x_i} \equiv \left(\frac{\delta}{\delta A_{\mu_i}^{\text{in}}(x_i)}, \frac{\delta}{\delta\psi_{\text{in}}(x_i)}, \frac{\delta}{\delta\bar{\psi}_{\text{in}}(x_i)}\right),$$

$$\frac{\delta}{\delta\bar{x}_i} \equiv \left(\frac{\delta}{\delta A_{\mu_i}^{\text{in}}(x_i)}, \frac{\delta}{\delta\bar{\psi}_{\text{in}}(x_i)}, \frac{\delta}{\delta\psi_{\text{in}}(x_i)}\right).$$

The functional differentiation with respect to spinor fields is always a differentiation with respect to a particular component of the spinor field. We shall not, however, put in the spinor indices; these indices can always be recovered from an equation by remembering that a spinor index maintains its association with a given space–time variable and by remembering that a functional derivative with respect to a spinor transforms like an adjoint spinor.

III. FUNCTIONAL DERIVATIVES

A. The Simple Case

The functional derivative of a normal product of field operators is easy to define; one need only take account of the anticommutation of fermion fields:

$$\begin{aligned}&(\delta/\delta x_i)\ {:}\varphi_{\text{in}}(y_1)\cdots\varphi_{\text{in}}(y_n){:}\\ &\equiv \sum_j \delta(x_i - y_j)\ \tilde{\delta}_{ij}(-1)^{P_{ij}}\ {:}\varphi_{\text{in}}(y_1)\cdots\Lambda_j\cdots\varphi_{\text{in}}(y_n){:}.\end{aligned} \tag{III.1}$$

(The symbol $\Lambda_{ij\cdots k}$ is used throughout to indicate that the ith, jth, $\cdots$, kth terms of a product or sequence are absent.) Here P_{ij} takes account of the anticommutation of spinor fields:

$$P_{ij} = |\epsilon_{iF}|\left\{1 - \sum_1^j |\epsilon_{kF}|\right\}. \tag{III.2}$$

This definition gives functional differentiation from the left.

B. The $\mathfrak{N}$-Operator

It was demonstrated in I that functional differentiation does not, in general, commute with space–time integration or differentiation. As a simple example we note that

$$(\delta/\delta x_i)K_j\varphi_j^{\text{in}} \neq K_j(\delta/\delta x_i)\varphi_j^{\text{in}}$$

unless both sides vanish. We should like to be able to interchange the order of functional and space–time differentiation somehow. To accomplish this we define an operator $\mathfrak{N}$ as follows: $\mathfrak{N}f(x_1\cdots x_n) = f(x_1\cdots x_n)$ minus all parts of f that are of the form $K_i^pK_j^q[\delta(x_i - x_j)g(x_1\cdots x_n)]$, with p and q non-negative integers and $p + q \geq 1$. We call $\mathfrak{N}f$ the *primitive part* of f.

It is now easy to see that functional differentiation will commute with space–time differentiation whenever the $\mathfrak{N}$-operator is present:

$$\begin{aligned}&\frac{\delta}{\delta y}\,\mathfrak{N}K_1\cdots K_nG(x_1\cdots x_n)\\ &\equiv \mathfrak{N}K_1\cdots K_n\frac{\delta}{\delta y}G(x_1\cdots x_n).\end{aligned} \tag{III.3}$$

In particular we have

$$\frac{\delta}{\delta y} K_x\varphi_{\text{in}}(x) = \frac{\delta}{\delta y} \mathfrak{N}K_x\varphi_{\text{in}}(x) = \mathfrak{N}K_x \frac{\delta\varphi_{\text{in}}(x)}{\delta y}$$
$$= \mathfrak{N}K_x\delta(x - y) = 0. \quad \text{(III.4)}$$

Finally we define the functional derivative of a functional of the form

$$F[f] \equiv \sum_{n=0}^{\infty} \int d^4x_1 \cdots d^4x_n \times f_n(x_1 \cdots x_n) :\varphi_{\text{in}}(x_1) \cdots \varphi_{\text{in}}(x_n):.$$

Since a functional is *equal* to its primitive part,

$$F[f] = F[\mathfrak{N}f], \quad \text{(III.5)}$$

we can avoid ambiguities by defining the functional derivative of a functional as follows:

$$\frac{\delta F[f]}{\delta y_i} \equiv \frac{\delta F[\mathfrak{N}f]}{\delta y_i}$$
$$\equiv \sum_{n=1}^{\infty} \sum_{j=1}^{n} \int d^4x_1 \cdots d^4x_n[\mathfrak{N}f_n(x_1 \cdots x_n)] \times \delta(y_i - x_j)\, \delta_{ij}(-1)^{P_{ij}} :\varphi_1^{\text{in}} \cdots \Lambda_j \cdots \varphi_n^{\text{in}}:.$$

The $\mathfrak{N}$-operator has some useful properties that should be noted. First it is a projection operator,

$$\mathfrak{N}(\mathfrak{N}f) = \mathfrak{N}f. \quad \text{(III.7)}$$

Next, if we write the $\mathfrak{N}$-operator that is to be applied to a generalized function of the variables, $x_1, \cdots, x_n$, as $\mathfrak{N}(x_1 \cdots x_n):$,

$$\mathfrak{N}f(x_1 \cdots x_n) \equiv \mathfrak{N}(x_1 \cdots x_n)f(x_1 \cdots x_n), \quad \text{(III.8)}$$

then the product of two primitive parts is a primitive part,

$$[\mathfrak{N}f][\mathfrak{N}g] = \mathfrak{N}(fg), \quad \text{(III.9)}$$

i.e.,

$$[\mathfrak{N}(x_1 \cdots x_n)f(x_1 \cdots x_n)][\mathfrak{N}(y_1 \cdots y_m)g(y_1 \cdots y_m)]$$
$$= \mathfrak{N}(x_1 \cdots x_n\, y_1 \cdots y_m)(f(x_1 \cdots x_n)g(y_1 \cdots y_m)).$$

Hence the multiplication table for the $\mathfrak{N}$-operators is

$$\mathfrak{N}(x_1 \cdots x_n)\mathfrak{N}(y_1 \cdots y_m) = \mathfrak{N}(x_1 \cdots x_n\, y_1 \cdots y_m). \quad \text{(III.10)}$$

From this it follows that if F and G are functionals,

$$[\mathfrak{N}F][\mathfrak{N}G] \equiv \mathfrak{N}(FG), \quad \text{(III.11)}$$

where we have written $\mathfrak{N}F$ for $F[\mathfrak{N}f]$. The validity of (III.11) rests upon the fact that the contractions of the free fields, $-i\Delta_+(x - y)$, do not contain any four-dimensional δ-function singularities. This would not be true in general if time ordering were introduced:

$$[(\mathfrak{N}\varphi(x))(\mathfrak{N}\varphi(y))]_+ \equiv [\mathfrak{N}(\varphi(x)\varphi(y))]_+ \not\equiv \mathfrak{N}[\varphi(x)\varphi(y)]_+.$$

C. Identity of Functionals

We remark that the fact that two functionals are equal does not imply that their functional derivatives are equal. If it happens that $\delta^nF/\delta x_1 \cdots \delta x_n = \delta^nG/\delta x_1 \cdots \delta x_n$, for all n, we say that F and G are identical and write $F \equiv G$. From our definition of functional differentiation it is clear that

$$F \equiv \mathfrak{N}F \quad \text{(III.12)}$$

so that

$$\langle 0|\, \delta^nF/\delta x_1 \cdots \delta x_n\, |0\rangle = \mathfrak{N}f_n(x_1 \cdots x_n), \quad n = 0, 1, 2, \cdots. \quad \text{(III.13)}$$

IV. DYNAMICS

A. Formulation

One assumes the existence of a unitary operator, S, that transforms the initial state vector into the final state vector. Our first axiom is that S is a functional of the free fields $\varphi_{\text{in}}(x)$. This axiom, plus Lorentz invariance, is exactly equivalent to the substitution law.

With the S-operator a functional of the free fields, one may calculate its functional derivatives. Thus one is able to *define* the current associated with a particular type of field by

$$S^\dagger\, \delta S/\delta\bar{x}_i \equiv ij(x_i). \quad \text{(IV.1)}$$

The corresponding interpolating quantum field is therefore *defined* up to a boundary condition by

$$S^\dagger\, \delta S/\delta\bar{x}_i \equiv -iK_i\varphi_i. \quad \text{(IV.2)}$$

We complete the definition of the interpolating field by taking the usual asymptotic condition[4] at $t = -\infty$. By integrating (IV.2) we have

$$\varphi(x_i) \equiv \varphi_{\text{in}}(x_i) - i \int d^4\xi_i \bar{K}_i\, \Delta_R(x_i - \xi_i) S^\dagger \frac{\delta S}{\delta\bar{\xi}_i}. \quad \text{(IV.3)}$$

Equation (IV.3) defines both the interpolating fields and their adjoints. If we demand as our second axiom that the adjoint of an interpolating field be related to the interpolating field in the same way that the adjoint of a free field is related to the free field, then this imposes a restriction on the S-operator:

[4] H. Lehmann, R. Symanzik, and W. Zimmermann, Nuovo Cimento 1, 205 (1955).

$$S^\dagger \, \delta S/\delta \bar{x}_i \equiv -(\delta S^\dagger/\delta \bar{x}_i)S. \tag{IV.4}$$

This equation expresses the form invariance of the interpolating field under charge conjugation. Equation (IV.4), together with unitarity, implies generalized unitarity:

$$S^\dagger S \equiv S S^\dagger \equiv 1. \tag{IV.5}$$

Next let us discuss the basic dynamical postulate of this formulation. It follows[1,4] from the asymptotic behavior of the interpolating fields that $S^\dagger \delta^n S/\delta x_1 \cdots \delta x_n$ is the same as $(-i)^n K_1 \cdots K_n(\varphi_1 \cdots \varphi_n)_+$ apart from possible terms that vanish on the mass shell. From (III.8), we are led to think that

$$S^\dagger \, \delta^n S/\delta \bar{x}_1 \cdots \delta \bar{x}_n \equiv (-i)^n \mathfrak{N} K_1 \cdots K_n(\varphi_1 \cdots \varphi_n)_+ \tag{IV.6}$$

is the correct form of the dynamical postulate, but this would be postulating too much since the $n = 0, 1, 2$, parts of (IV.6) will completely determine the higher functional derivatives of S. Thus we adopt as the basic dynamical postulate the equation of identity,

$$S^\dagger \, \delta^2 S/\delta \bar{x}_1 \, \delta \bar{x}_2 \equiv (-i)^2 \mathfrak{N} K_1 K_2(\varphi_1 \, \varphi_2)_+. \tag{IV.7}$$

To conclude this section we discuss two consequences of (IV.7). First it will be shown that (IV.6) is valid for all n, and therefore the dynamical axiom and generalized unitarity are consistent with the definition of the interpolating field. Secondly, we shall rewrite (IV.7) in terms of the currents to obtain a particularly simple functional differential equation.

B. Proof of Self-Consistency

Equation (IV.6) is proven by induction: apply $\delta/\delta \bar{x}_{n+1}$ to (IV.6) to get

$$\begin{aligned} &S^\dagger \, \delta^{n+1} S/o\bar{x}_{n+1} \cdots \delta \bar{x}_1 \\ &\equiv (-i)^{n+1} K_{n+1}\varphi_{n+1} \mathfrak{N} K_1 \cdots K_n(\varphi_n \cdots \varphi_1)_+ \\ &\quad + (-i)^n \mathfrak{N} K_1 \cdots K_n \sum_{i=1}^{n} \left(\varphi_n \cdots \frac{\delta \varphi_i}{\delta \bar{x}_{n+1}} \cdots \varphi_1\right)_+. \end{aligned} \tag{IV.8}$$

From (IV.3) and generalized unitarity we have

$$\begin{aligned} \frac{\delta \varphi_i}{\delta \bar{x}_{n+1}} &\equiv \delta(x_i - x_{n+1}) \, \tilde{\delta}^i_{n+1} - i \int d^4\xi_i \bar{K}_i \, \Delta_R(x_i - \xi_i) \\ &\times \left[S^\dagger \frac{\delta^2 S}{\delta \bar{x}_{n+1} \, \delta \bar{\xi}_i} - S^\dagger \frac{\delta S}{\delta \bar{x}_{n+1}} S^\dagger \frac{\delta S}{\delta \bar{\xi}_i} \right]. \end{aligned} \tag{IV.9}$$

From (III.11) and (IV.2), it is clear that $\mathfrak{N} K\varphi \equiv K\varphi$, so that we may write

$$\begin{aligned} &S^\dagger \frac{\delta^2 S}{\delta \bar{x}_{n+1} \, \delta \bar{\xi}_i} - S^\dagger \frac{\delta S}{\delta \bar{x}_{n+1}} S^\dagger \frac{\delta S}{\delta \bar{\xi}_i} \\ &\equiv \mathfrak{N} K_{n+1} K_i \{\theta(\xi_i - x_{n+1})[\varphi_{n+1}, \varphi(\xi_i)]_{\sigma_{n+1},i}\}. \end{aligned} \tag{IV.10}$$

When this is substituted into (IV.9) we may integrate by parts with respect to the differentiation in K_i and the $\theta(\xi_i - x_{n+1})$ and $\Delta_R(x_i - \xi_i)$ ensure the vanishing of the boundary terms. This gives us the result

$$\begin{aligned} \delta \varphi_i/\delta \bar{x}_{n+1} &\equiv \delta(x_i - x_{n+1}) \, \tilde{\delta}^i_{n+1} \\ &+ i \mathfrak{N} K_{n+1}\{\theta(x_i - x_{n+1})[\varphi_{n+1}, \varphi_i]_{\sigma_{n+1},i}\}. \end{aligned} \tag{IV.11}$$

Next we note that $K_{n+1}\varphi_{n+1} \mathfrak{N} K_1 \cdots K_n(\varphi_n \cdots \varphi_1)_+ \equiv \mathfrak{N} K_1 \cdots K_{n+1}\varphi_{n+1}(\varphi_n \cdots \varphi_1)_+$ because of (III.11). Thus when (IV.11) is substituted into (IV.8) and the terms are rearranged, we get

$$\begin{aligned} &S^\dagger \, \delta^{n+1} S/\delta \bar{x}_{n+1} \cdots \delta \bar{x}_1 \\ &\equiv (-i)^{n+1} \mathfrak{N} K_1 \cdots K_{n+1}(\varphi_{n+1} \cdots \varphi_1)_+. \end{aligned} \tag{IV.12}$$

This completes the proof that (IV.3), (IV.5), and (IV.7) are self-consistent.

C. The Projection Operators

In Sec. (IV.D) it is seen that the dynamical functional differential equation for the S-operator may be written in a very simple form in terms of a set of projection operators. Define

$$P_A \equiv K_1 K_2[\theta(x_2 - x_1)\bar{K}_1 \bar{K}_2 \Delta_R(x_1 - \xi_1)\Delta_A(x_2 - \xi_2)], \tag{IV.13}$$

$$P_R \equiv K_1 K_2[\theta(x_1 - x_2)\bar{K}_1 \bar{K}_2 \Delta_A(x_1 - \xi_1)\Delta_R(x_2 - \xi_2)], \tag{IV.14}$$

$$P_a \equiv K_1 K_2[\theta(x_2 - x_1)\bar{K}_1 \bar{K}_2 \Delta_R(x_1 - \xi_1)\Delta(x_2 - \xi_2)], \tag{IV.15}$$

$$P_r \equiv K_1 K_2[\theta(x_1 - x_2)\bar{K}_1 \bar{K}_2 \Delta(x_1 - \xi_1)\Delta_R(x_2 - \xi_2)]. \tag{IV.16}$$

The multiplication of these operators is defined by

$$P_i P_j \equiv \int d^4\xi_1 \, d^4\xi_2 P_i(x_1x_2; \xi_1\xi_2)P_j(\xi_1\xi_2; y_1y_2). \tag{IV.17}$$

It is then easy to prove that the P_i's form a complete set of orthogonal projection operators:

$$P_i P_j = \delta_{ij} P_j, \tag{IV.18}$$

$$\sum_i P_i = 1. \tag{IV.19}$$

Here the unit operator is $\delta(x_1 - \xi_1)\delta(x_2 - \xi_2)$. We shall use the abbreviation $P_i g(x_1x_2)$ to mean the expression

$$\int d^4\xi_1 \, d^4\xi_2 P_i(x_1x_2; \xi_1\xi_2)g(\xi_1\xi_2).$$

The complete set of eigenfunctions of these projection operators is not known, but the solutions of $P_A A = 0$ and $P_R R = 0$ can be listed. We examine the first of these two equations in some detail, the solutions of the second equation then being obvious. Using the formulas

$$\theta(x) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{dk}{k - i\epsilon} e^{ikx}, \tag{IV.20}$$

$$\Delta_R(x_i) = \frac{1}{(2\pi)^4}\int \frac{d^4p\, e^{ip\cdot x_i}}{\mathbf{p}^2 + m_i^2 - (p_0 + i\epsilon)^2}, \tag{IV.21}$$

$$\Delta_A(x_i) = \frac{1}{(2\pi)^4}\int \frac{d^4p\, e^{ip\cdot x_i}}{\mathbf{p}^2 + m_i^2 - (p_0 - i\epsilon)^2}, \tag{IV.22}$$

and writing

$$K^{\mathrm{T}}(p) = (1, -i\mathbf{p} + m, -i\mathbf{p} - m), \tag{IV.23}$$

$$A(x_1, x_2) = \int d^4p_1\, d^4p_2 \tilde{A}(p_1, p_2) e^{ip_1\cdot x_1 + ip_2\cdot x_2}, \tag{IV.24}$$

we have

$$P_A A = \frac{K_1 K_2}{2\pi i}\int d^4p_1\, d^4p_2\, e^{ip_1x_1 + ip_2x_2}\int_{-\infty}^{\infty} \frac{dk K^{\mathrm{T}}(p_1 - k)K^{\mathrm{T}}(p_2 + k)\tilde{A}(p_1 - k, p_2 + k)}{(k - i\epsilon)[\mathbf{p}_1^2 + m_1^2 - (p_1^0 - k + i\epsilon)^2][\mathbf{p}_2^2 + m_2^2 - (p_2^0 + k - i\epsilon)^2]}. \tag{IV.25}$$

Here we have written $(\mathbf{p}, p_0 - k)$ as $(p - k)$. Note that all the explicit poles in the integrand of (IV.25) are in the upper half k-plane. By closing the contour in the lower half k-plane, the integral will vanish if $\tilde{A}(p_1 - k, p_2 + k)$ has no poles in the lower half plane. (The possibility that $\tilde{A}$ might have poles in the lower half plane and zeros in the upper half plane at the positions of the explicit poles is ruled out by noting that $\tilde{A}(p_1, p_2)$ would then be zero.) A further condition on the integrand is that it is permissible to close the contour. This means that $K^{\mathrm{T}}(p_1 - k)K^{\mathrm{T}}(p_2 + k)\tilde{A}(p_1 - k, p_2 + k)$ must blow up less rapidly than k^4. Using (IV.23) we see that $\tilde{A}(p_1 - k, p_2 + k)$ must blow up less rapidly than k to the power $2 + \delta_{1B} + \delta_{2B} \equiv z(1, 2)$.

It is now obvious by inspection that R is a solution of $P_R R = 0$ provided that $\tilde{R}(p_1 - k, p_2 + k)$ has no poles in the upper half k-plane and blows up less rapidly than k to the power $z(1, 2)$.

Finally we examine the simultaneous eigenfunctions of P_A and P_R:

$$P_A b = P_R b = 0. \tag{IV.25}$$

From the above discussion it is clear that $\tilde{b}(p_1 - k, p_2 + k)$ can have no poles in the whole k-plane and must blow up less rapidly than k to the power $z(1, 2)$. Note that because of the completeness of the projection operators, b satisfies

$$(P_a + P_r)b = b. \tag{IV.26}$$

Then because of the homogeneous Δ-functions in P_a and P_r, b must depend on the times only through some derivative of $\delta(x_1^0 - x_2^0)$. Thus $\tilde{b}$ has the form

$$\tilde{b}(p_1 - k, p_2 + k) = \delta(p_1^0 + p_2^0)\tilde{c}(\mathbf{p}_1, \mathbf{p}_2, p_1^0 - k).$$

The analyticity of b then implies that $\tilde{c}$ is a polynomial in $(p_1^0 - \mathrm{k})$ of degree less than $z(1, 2)$. Therefore the general solution of (IV.25) is

$$b(x_1, x_2) = \sum_{n,m} \frac{\partial^n}{\partial x_1^{0n}} \frac{\partial^m}{\partial x_2^{0m}} [\delta(x_1^0 - x_2^0) g_{nm}(x_1, x_2)] \tag{IV.27}$$

with $g(x_1, x_2)$ an arbitrary distribution and with n and m restricted by

$$n + m < 2 + \delta_{1B} + \delta_{2B}. \tag{IV.28}$$

D. The Current Equation

If we eliminate the interpolating fields from the dynamical equation, (IV.7), by substituting (IV.3), the resulting equation for the S-operator is

$$\mathfrak{N}(P_A + P_R)S^\dagger \frac{\delta^2 S}{\delta x_1\, \delta x_2} \equiv \mathfrak{N}\left[P_R S^\dagger \frac{\delta S}{\delta x_1} S^\dagger \frac{\delta S}{\delta x_2} - \sigma_{12} P_A S^\dagger \frac{\delta S}{\delta x_2} S^\dagger \frac{\delta S}{\delta x_1}\right]. \tag{IV.29}$$

Using generalized unitarity this becomes

$$\mathfrak{N}(P_A + P_R)\frac{\delta}{\delta x_1}\left(S^\dagger \frac{\delta S}{\delta x_2}\right) \equiv -\mathfrak{N}P_A\left[S^\dagger \frac{\delta S}{\delta x_1}, S^\dagger \frac{\delta S}{\delta x_2}\right]_{\sigma_{12}}. \tag{IV.30}$$

According to (IV.25), the general solution of this integral equation is

$$\frac{\delta}{\delta x_1}\left(S^\dagger \frac{\delta S}{\delta x_2}\right) \equiv -i\mathfrak{N}b(x_1, x_2) - \mathfrak{N}P_A\left[S^\dagger \frac{\delta S}{\delta x_1}, S^\dagger \frac{\delta S}{\delta x_2}\right]. \tag{IV.31}$$

These equations may be rewritten in terms of

the currents defined by (IV.1). Equation (IV.30) becomes

$$\mathfrak{N}(P_A + P_B)\delta j_2/\delta\bar{x}_1 \equiv -i\mathfrak{N}P_A[j_1, j_2]_{\sigma_{12}}. \quad \text{(IV.32)}$$

while Equation (IV.31) becomes

$$\delta j_2/\delta\bar{x}_1 \equiv \mathfrak{N}b(x_1, x_2) - i\mathfrak{N}P_A[j_1, j_2]_{\sigma_{12}}. \quad \text{(IV.33)}$$

V. SUBSIDIARY CONDITION

Let the free photon wavefunctions be denoted by the positive energy function $f_\mu^\alpha(x)$:

$$\Box f_\mu^\alpha(x) = 0, \qquad \mu = 1, 2, 3, 4, \text{ and } \alpha = 1, 2, \cdots. \quad \text{(V.1)}$$

The subsidiary condition

$$\partial_\mu f_\mu^\alpha(x) = 0 \quad \text{(V.2)}$$

ensures that in free photon states the scalar and longitudinal photons occur with the same amplitude. Photon state vectors are constructed by use of the creation operators

$$A_{\text{in}}^\alpha = -i\int d^3x f_\mu^\alpha(x) \overleftrightarrow{\partial_0} A_\mu^{\text{in}}(x). \quad \text{(V.3)}$$

[Here $f\overleftrightarrow{\partial_0}g \equiv f\partial g/\partial t - (\partial f/\partial t)g$.] As a consequence of the subsidiary condition all state vectors satisfy

$$\partial_\mu A_\mu^{\text{in}(+)}(x)\Phi_{\text{in}} = 0, \quad \text{(V.4)}$$

where the superscript (+) denotes the destruction operator part of a free field.

VI. GAUGE INVARIANCE

The subsidiary condition ensures that scalar and longitudinal photons make no contributions to matrix elements of the form $(\Phi_{\text{in}}^{(\alpha)}, \Phi_{\text{in}}^{(\beta)})$. In order that they make no contribution to matrix elements of the form $(\Phi_{\text{in}}^{(\alpha)}, S\Phi_{\text{in}}^{(\beta)})$ it is necessary and sufficient that such matrix elements be invariant under a gauge transformation of *any* one free photon wavefunction:[3,5]

$$f_\mu^\alpha(x) \to f_\mu^\alpha(x) + \partial_\mu\Lambda_f(x), \quad \text{(VI.1)}$$

with $\Lambda_f(x)$ a positive energy solution of

$$\Box\Lambda_f(x) = 0. \quad \text{(VI.2)}$$

This gauge invariance means that the matrix element would have the same value if the photon wavefunction had no scalar and longitudinal parts. An alternate, but equivalent,[5] requirement is that the S-operator should be invariant under a gauge transformation of the free fields:

$$A_\mu^{\text{in}}(x) \to A_\mu^{\text{in}}(x) + \partial_\mu\Lambda(x), \quad \text{(VI.3)}$$

$$\psi_{\text{in}}(x) \to \psi_{\text{in}}(x). \quad \text{(VI.4)}$$

Here $\Lambda(x)$ is a *real* solution of (IV.2) whereas $\Lambda_f(x)$ was a positive energy solution.

With the notation $\delta/\delta y^\mu \equiv \delta/\delta A_\mu^{\text{in}}(y)$, we define the operator

$$\lambda \equiv \int d^4y\, \partial_\mu\Lambda(y)\frac{\delta}{\delta y^\mu}. \quad \text{(VI.5)}$$

Under the gauge transformation, (IV.3) and (IV.4), any functional F transforms according to

$$F \to F' \equiv e^\lambda F. \quad \text{(VI.6)}$$

The gauge invariance of the S-operator, $S' = S$, implies that

$$\lambda S = 0 \quad \text{(VI.7)}$$

and this, in turn, is equivalent to the statement

$$[\partial_\mu A_\mu^{\text{in}}(x), S] = 0 \quad \text{(VI.8)}$$

since $\Lambda(x)$ is an arbitrary real solution of the Klein–Gordon equation.

It is possible to impose a stronger condition on the theory than the simple gauge invariance of the S-operator. This stronger condition is that the interpolating fields shall transform according to

$$A_\mu(x) \to A'_\mu(x) \equiv e^\lambda A_\mu(x) = A_\mu(x) + \partial_\mu\Lambda(x), \quad \text{(VI.9)}$$

$$\psi(x) \to \psi'(x) \equiv e^\lambda\psi(x) = e^{-ie\Lambda(x)}\psi(x) \quad \text{(VI.10)}$$

under the gauge transformation, (IV.3), (IV.4), and (IV.6).

First consider the transformation of the photon interpolating field. According to (IV.9) we have

$$A_\mu(x) + \partial_\mu\Lambda(x) = \sum_{n=0}^{\infty}\frac{1}{n!}\int d^4\xi_1 \cdots d^4\xi_n \times \partial_{\mu_1}\Lambda(\xi_1)\cdots\partial_{\mu_n}\Lambda(\xi_n)\frac{\delta^n A_\mu(x)}{\delta\xi_1^{\mu_1}\cdots\delta\xi_n^{\mu_n}}. \quad \text{(VI.11)}$$

Since $\Lambda(x)$ is arbitrary, this equation must be satisfied term by term. The $n = 1$ part gives

$$\partial_\mu\Lambda(x) = \int d^4\xi\, \partial_\nu\Lambda(\xi)\frac{\delta A_\mu(x)}{\delta\xi^\nu} = i\int d^3\xi\, \partial_\nu\Lambda(\xi) \overleftrightarrow{\partial_0}[A_\nu^{\text{in}}(\xi), A_\mu(x)] \quad \text{(VI.12)}$$

$$= -i\int d^3\xi\Lambda(\xi) \overleftrightarrow{\partial_0}[\partial_\nu A_\nu^{\text{in}}(\xi), A_\mu(x)].$$

In the last step we made use of $\Box\Lambda = \Box A_\nu^{\text{in}} = 0$. By a similar argument the left side of (VI.12) may

[5] R. Pugh, Ann. Phys. 30, 422 (1964).

be written[6]

$$\partial_\mu \Lambda(x) = \partial_\mu \int d^3\xi \Lambda(\xi) \overleftarrow{\partial_0} D(\xi - x) = -\int d^3\xi \Lambda(\xi) \overleftarrow{\partial_0} \partial_\mu D(\xi - x). \tag{VI.13}$$

Thus we have

$$\int d^3\xi \Lambda(\xi) \overleftarrow{\partial_0} \{[\partial_\nu A_\nu^{\text{in}}(\xi), A_\mu(x)] + i \partial_\mu D(\xi - x)\} = 0,$$

and since Λ is an arbitrary solution of the Klein–Gordon equation, this implies

$$[\partial_\nu A_\nu^{\text{in}}(\xi), A_\mu(x)] = -i \partial_\mu D(\xi - x). \tag{VI.14}$$

This commutation relation is sufficient to ensure that all terms in (VI.11) with $n > 1$ vanish since the first functional derivative is transformable to a c-number, all higher derivatives vanish. The transformation $A_\mu \to A_\mu + \partial_\mu \Lambda$ is therefore equivalent to the commutation relation (VI.14).

Now consider the transformation of the ψ-field. Expand both $e^{-ie\Lambda(x)}\psi(x)$ and $e^{\lambda}\psi(x)$ and compare terms:

$$\sum_{n=0}^{\infty} \frac{(-ie\Lambda(x))^n}{n!} \psi(x) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \psi(x). \tag{VI.15}$$

The $n = 0$ part is satisfied automatically. We take the same steps as in (VI.12) and the $n = 1$ part gives

$$-ie\Lambda(x)\psi(x) = \int d^4\xi \, \partial_\mu \Lambda(\xi) \frac{\delta\psi(x)}{\delta\xi^\mu} = -i \int d^3\xi \Lambda(\xi) \overleftarrow{\partial_0} [\partial_\mu A_\mu^{\text{in}}(\xi), \psi(x)]. \tag{VI.16}$$

Using (VI.13) we see that

$$-i \int d^3\xi \Lambda(\xi) \overleftarrow{\partial_0} \{[\partial_\mu A_\mu^{\text{in}}(\xi), \psi(x)] - e D(\xi - x)\psi(x)\} = 0$$

so that

$$[\partial_\mu A_\mu^{\text{in}}(\xi), \psi(x)] = e D(\xi - x)\psi(x). \tag{VI.17}$$

One can now easily see that (VI.15) is satisfied for all n: using (VI.17) and (VI.13) we have

$$\lambda\psi(x) = -ie\Lambda(x)\psi(x), \tag{VI.18}$$

which makes (VI.15) obviously satisfied.

The commutator of $\bar{\psi}(x)$ with $\partial_\mu A_\mu^{\text{in}}(\xi)$ is obtained from (VI.17) by taking the Hermitian adjoint

$$[\partial_\mu A_\mu^{\text{in}}(\xi), \bar{\psi}(x)] = -e D(\xi - x)\bar{\psi}(x). \tag{VI.19}$$

[6] ∂_μ always differentiates the first possible variable on its right. Thus in the first line of (VI.13) $\partial_\mu = \partial/\partial x_\mu$, while in the second line, $\partial_\mu = \partial/\partial\xi_\mu$.

Hence in general we have

$$[\partial_\mu A_\mu^{\text{in}}(\xi), \varphi_i] = -i\delta_{iB} D(\xi - x_i) + e\epsilon_{iF} D(\xi - x_i)\varphi_i. \tag{VI.20}$$

The commutation rules between $\partial\mu A_\mu^{\text{in}}(\xi)$ and the interpolating fields impose restrictions on the interpolating fields and therefore on the S-operator. Using (VI.20), one easily arrives at the result

$$[\partial_\mu A_\mu^{\text{in}}(\xi), \delta^2 S/\delta\bar{x}_1\delta\bar{x}_2] = -e\mathfrak{N}SK_1K_2\{[\epsilon_{1F} D(\xi - x_1) + \epsilon_{2F} D(\xi - x_2)](\varphi_1\varphi_2)_+\}. \tag{VI.21}$$

This is the S-operator form of the Nishijima[2] generalization of the Ward–Takahashi relationship. It is equivalent to the commutation rules (VI.20) and to the gauge transformations (VI.9) and (VI.10).

So far we have merely *assumed* that (VI.9) and (VI.10) were possible; to complete the proof of gauge invariance, one must prove that the assumed gauge transformation is consistent with the dynamical parts of the theory. Such a proof was given in an earlier paper[5] and need not be repeated here. We remind the reader of the result: gauge invariance is automatically satisfied by all S-matrix elements except the vertex function. For the vertex function, gauge invariance is a boundary condition.

VII. SAMPLE CALCULATION IN PERTURBATION THEORY

If we expand the currents in powers of e, Eq. (IV.32) becomes

$$\frac{\delta j_2^{(n)}}{\delta\bar{x}_1} \equiv \mathfrak{N}b^{(n)}(x_1, x_2) - i\mathfrak{N}P_A \sum_{k=1}^{n-1} [j_1^{(k)}, j_2^{(n-k)}]_{\sigma_1}, \tag{VII.1}$$

where $b^{(n)}$ is that part of b that is of nth order in e. We have assumed that j has no zero-order part. The first-order current is then determined entirely by $b^{(1)}$. If 1 and 2 both refer to fermion variables, $b^{(1)}$ cannot contain any derivatives as was pointed out in Section (IV.D). Hence the only Lorentz-invariant solution is

$$\delta j_2^{(1)}/\delta\bar{x}_1 \equiv -ie\gamma \cdot A_{\text{in}}(x_1) \, \delta(x_1 - x_2). \tag{VII.2}$$

The factor i has been put in to satisfy generalized unitarity. (It is shown explicitly in the Appendix that the Pauli interaction is not allowed.) Integrating (VII.2) gives

$$j_F^{(1)}(x) \equiv -ie\gamma \cdot A_{\text{in}}(x)\psi_{\text{in}}(x) \tag{VII.3}$$

for the current associated with a fermion field. From this one obtains

$$S^{(1)} \equiv e \int d^4x \, :\bar{\psi}_{\text{in}}(x)\gamma \cdot A_{\text{in}}(x)\psi_{\text{in}}(x):, \tag{VII.4}$$

so that

$$j_F^{(1)}(x) \equiv ie\bar{\psi}_{\text{in}}(x)\gamma\cdot A_{\text{in}}(x), \qquad \text{(VII.5)}$$

$$j_{B_\mu}^{(1)}(x) \equiv -ie :\bar{\psi}_{\text{in}}(x)\gamma_\mu\psi_{\text{in}}(x):. \qquad \text{(VII.6)}$$

As a sample second-order calculation, we consider the current that contains the photon propagator as this case will demonstrate the use of the $\mathfrak{N}$-operator. To calculate the second-order photon current we start with

$$\begin{aligned}\delta j_{\mu_2}^{(2)}(x_2)/\delta x_1^{\mu_1} &\equiv \mathfrak{N}b^{(2)}(x_1, x_2) - i\mathfrak{N}P_A[j_{\mu_1}^{(1)}(x_1), j_{\mu_2}^{(1)}(x_2)] \\ &\equiv \mathfrak{N}b^{(2)}(x_1, x_2) \\ &\quad + ie^2\mathfrak{N}P_A\{i :\bar{\psi}_{\text{in}}(x_1)\gamma_{\mu_1}S(x_1 - x_2)\gamma_{\mu_2}\psi_{\text{in}}(x_2): \\ &\quad - \text{Tr}\,[\gamma_{\mu_1}S_+(x_1 - x_2)\gamma_{\mu_2}S_-(x_2 - x_1)] - (1/2)\}.\end{aligned} \qquad \text{(VII.7)}$$

[Here (1/2) means the previous terms with the subscripts 1 and 2 interchanged.] The first term in (VII.7) gives second-order Compton scattering while the Tr term gives the photon propagator. The boundary conditions that determine $b^{(2)}(x_1, x_2)$ are that the Fourier transform of any physical matrix element shall vanish for large energies and that the vertex functions shall be consistent with gauge invariance. No boundary condition is needed for the propagator.

If we examine the structure of P_A in Eq. (IV.13) we see that when the K_1K_2 differentiation is carried out any term in which $\theta(x_2 - x_1)$ is differentiated is of the same form as $b(x_1, x_2)$. Thus by appropriate choice of b, one can replace P_A by $\theta(x_2 - x_1)$ provided the boundary conditions are not violated. This replacement is consistent with the boundary conditions for the Compton terms in (VII.7) but would lead to an undefined product of distributions for the propagator terms. Hence we have

$$\begin{aligned}\delta j_{\mu_2}^{(2)}(x_2)/\delta x_1^{\mu_1} &\equiv e^2 :\bar{\psi}_{\text{in}}(x_1)\gamma_{\mu_1}S_A(x_1 - x_2)\gamma_{\mu_2}\psi_{\text{in}}(x_2): \\ &\quad + e^2 :\bar{\psi}_{\text{in}}(x_2)\gamma_{\mu_2}S_R(x_2 - x_1)\gamma_{\mu_1}\psi_{\text{in}}(x_1): \\ &\quad - ie^2\mathfrak{N}P_A\,\text{Tr}\,[\gamma_{\mu_1}S_+(x_1 - x_2)\gamma_{\mu_2}S_-(x_2 - x_1) \\ &\quad - \gamma_{\mu_2}S_+(x_2 - x_1)\gamma_{\mu_1}S_-(x_1 - x_2)].\end{aligned} \qquad \text{(VII.8)}$$

The Tr term may be explicitly evaluated; one obtains

$$\begin{aligned}&\frac{e^2}{3(2\pi)^6}\,\mathfrak{N}\int d^4p\, e^{ip\cdot(x_1 - x_2)}(p^2)^2(p_\mu p_\nu - \delta_{\mu\nu}p^2) \\ &\quad \times \int_{4m^2}^{\infty}\frac{d\kappa^2}{\kappa^4}\left(\frac{\kappa^2 - 4m^2}{\kappa^2}\right)^{\frac{1}{2}}\frac{\kappa^2 + 2m^2}{\kappa^2}\frac{1}{(\kappa^2 + p^2)_A} \\ &\quad + \frac{e^2}{3(2\pi)^6}\,\mathfrak{N}\int d^4p\, e^{ip\cdot(x_1 - x_2)}(p^2)^2(\delta_{\mu\nu} - \delta_{\mu 4}\delta_{\nu 4}) \\ &\quad \times \int_{4m^2}^{\infty}\frac{d\kappa^2}{\kappa^4}\left(\frac{\kappa^2 - 4m^2}{\kappa^2}\right)^{\frac{1}{2}}\frac{\kappa^2 + 2m^2}{\kappa^2}.\end{aligned}$$

Here the subscript A indicates that p_0 is replaced by $p_0 - i\epsilon$. Note that the second term is projected to zero by the $\mathfrak{N}$-operator. The first term also contains a part that is projected out by the $\mathfrak{N}$-operator; this part is obtained by writing

$$[\kappa^2(\kappa^2 + p^2)]^{-1} = [\kappa^2p^2]^{-1} - [p^2(\kappa^2 + p^2)]^{-1}.$$

The $(\kappa^2p^2)^{-1}$ term is projected to zero by $\mathfrak{N}$; the other term yields the "renormalized" photon propagator

$$\begin{aligned}\omega_A^{(2)}(x_1x_2)_{\mu\nu} &= \frac{ie^2}{3(2\pi)^6}\int d^4p\, e^{ip\cdot(x_1 - x_2)}p^2(p_\mu p_\nu - p^2\,\delta_{\mu\nu}) \\ &\quad \times \int_{4m^2}^{\infty}\frac{d\kappa^2}{\kappa^4}\left(\frac{\kappa^2 - 4m^2}{\kappa^2}\right)^{\frac{1}{2}}\frac{\kappa^2 + 2m^2}{\kappa^2(\kappa^2 + p^2)_A}.\end{aligned} \qquad \text{(VII.9)}$$

Thus we have

$$\begin{aligned}\delta j_{\mu_2}^{(2)}(x_2)/\delta x_1^{\mu_1} &\equiv e^2[:\bar{\psi}_{\text{in}}(x_1)\gamma_{\mu_1}S_A(x_1 - x_2)\gamma_{\mu_2}\psi_{\text{in}}(x_2): \\ &\quad + :\bar{\psi}_{\text{in}}(x_2)\gamma_{\mu_2}S_R(x_2 - x_1)\gamma_{\mu_1}\psi_{\text{in}}(x_1):] \\ &\quad + i\omega_A^{(2)}(x_1, x_2)_{\mu_1\mu_2}.\end{aligned} \qquad \text{(VII.10)}$$

Integration yields the result

$$\begin{aligned}j_\mu^{(2)}(x) &\equiv e^2\int d^4\xi[:\bar{\psi}_{\text{in}}(\xi)\gamma\cdot A_{\text{in}}(\xi)S_A(\xi - x)\gamma_\mu\psi_{\text{in}}(x): \\ &\quad + :\bar{\psi}_{\text{in}}(x)\gamma_\mu S_R(x - \xi)\gamma\cdot A_{\text{in}}(\xi)\psi_{\text{in}}(\xi):] \\ &\quad + i\int d^4\xi\omega_A^{(2)}(\xi, x)_{\mu\nu}A_\nu^{\text{in}}(\xi).\end{aligned} \qquad \text{(VII.11)}$$

To obtain the second-order S-operator, we functionally integrate the equation

$$\delta S^{(2)}/\delta x_\mu \equiv -S^{(1)\dagger}\,\delta S^{(1)}/\delta x_\mu + ij_\mu^{(2)}(x). \qquad \text{(VII.12)}$$

The term involving $S^{(1)\dagger}$ is equal to zero, but it should be retained if one wishes to maintain the identity: it has the effect of changing the advanced and retarded Green's functions into causal Green's functions. Hence we have

$$\begin{aligned}S^{(2)} &\equiv ie^2\int d^4\xi\, d^4\eta \\ &\quad \times :\bar{\psi}_{\text{in}}(\xi)\gamma\cdot A_{\text{in}}(\xi)S_c(\xi - \eta)\gamma\cdot A_{\text{in}}(\eta)\psi_{\text{in}}(\eta): \\ &\quad - \frac{1}{2}\int d^4\xi\, d^4\eta\omega_c^{(2)}(\xi, \eta)_{\mu\nu} :A_\mu^{\text{in}}(\xi)A_\nu^{\text{in}}(\eta): \\ &\quad + \text{terms independent of } A_\mu^{\text{in}}.\end{aligned} \qquad \text{(VII.13)}$$

Finally it should be emphasized that no divergent expressions entered the calculation at any stage.

VIII. DISCUSSION

In I, the theory of self-interacting scalar particles was developed. The basic equations [Eq. (I.1)] involved the φ-product of interpolating fields. In this paper we have used a projection operator $\mathfrak{N}$ rather than the φ-product for two reasons. First, this operator is useful: it makes the interaction unique and projects out non-gauge-invariant terms in the photon propagator. Second, it is conjectured that the $\mathfrak{N}$-operator may be fundamental in the theory of weak interactions where the high-energy behavior of matrix elements causes divergences: the $\mathfrak{N}$-operator may eliminate these unwanted parts of matrix elements and leave the theory manageable. It seems fairly certain that the $\mathfrak{N}$-operator and the φ-product are equivalent for renormalizable theories although no proof of this conjecture exists.

The simplicity of the basic equations in the functional differential form is a most attractive aspect of this formulation. In this paper we have primarily dealt with quantum electrodynamics, but the notation was so chosen that the equations are valid for any low-spin particles; the treatment of higher spins will require separate consideration.

The functional differential equation for the currents,

$$\mathfrak{N}(P_A + P_R)\,\delta j(x_2)/\delta \bar{x}_1 \equiv -i\mathfrak{N}P_A[j(x_1),\, j(x_2)]_{\sigma_{12}},$$

is, perhaps, the most intriguing: it is strongly reminiscent of basic quantum conditions such as

$$\partial j(x)/\partial x_\mu = -i[P_\mu,\, j(x)].$$

The currents and the free fields seem to be "conjugate" to one another.

In conclusion, from a formally simple and universal set of equations we have easily produced finite results in perturbation theory that are in agreement with the results of renormalization theory.

ACKNOWLEDGMENTS

The author would like to thank Professor F. J. Bellinfante of Purdue University and Professor A. Salam of the International Centre for Theoretical Physics, Trieste, for the their hospitality during his visits to their respective institutions, when some of this work was carried out.

APPENDIX

Here we shall show that the $\mathfrak{N}$-operator projects the Pauli interaction to zero. Let

$$P \equiv \int d^4x\, :\bar{\psi}_{\rm in}(x)(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)\partial_\mu A_\nu^{\rm in}(x)\psi_{\rm in}(x):$$

be the Pauli interaction. Before differentiating P, we must write it in the standard form

$$P \equiv -\int d^4x_1\, d^4x_2\, d^4x_3\left[\frac{\partial}{\partial x_3^\mu}\left(\delta(x_3 - x_2)\;\delta(x_3 - x_1)\right)\right] \times :\bar{\psi}_{\rm in}(x_1)(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)A_\nu^{\rm in}(x_3)\psi_{\rm in}(x_2):.$$

Then we have

$$\begin{aligned}\frac{\delta^2 P}{\delta\bar{x}_1\,\delta x_2} &\equiv \int d^4x_3\left[\frac{\partial}{\partial x_3^\mu}\left(\delta(x_3 - x_2)\;\delta(x_3 - x_1)\right)\right] \times (\gamma_\mu\gamma\cdot A_{\rm in}(x_3) - \gamma\cdot A_{\rm in}(x_3)\gamma_\mu) \\ &\equiv (-\vec{\partial}_2 + m)\;\delta(x_2 - x_1)\gamma\cdot A_{\rm in}(x_1) \\ &\quad - \gamma\cdot A_{\rm in}(x_1)(-\vec{\partial}_2 + m)\;\delta(x_2 - x_1) \\ &\quad + (-\vec{\partial}_1 - m)\;\delta(x_1 - x_2)\gamma\cdot A_{\rm in}(x_2) \\ &\quad - \gamma\cdot A_{\rm in}(x_2)(-\vec{\partial}_1 - m)\;\delta(x_1 - x_2).\end{aligned}$$

In this last form we see that each term involves a factor $K_1\delta(x_1 - x_2)$ or a factor $K_2\delta(x_2 - x_1)$. Hence we have the result

$$\mathfrak{N}\;\delta^2 P/\delta\bar{x}_1\;\delta x_2 \equiv 0$$

and therefore a Pauli interaction is forbidden.

Gauge Approximations in Mesodynamics

R. J. RIVERS*
International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy
(Received 3 May 1965)

In a nonperturbative theory for vector and scalar mesodynamics of the kind proposed by Salam, it is found that the "gauge approximations" introduced in this approach are the most sensible of a set of approximation schemes that preserve the scalar gauge properties of the amplitudes of the theory. The Green's functions used are those suggested by the functional derivative approach of the paper. The generalized identities of the Ward–Takahashi kind that these Green's functions satisfy are exhibited, as well as the decomposition of the general n-point function in terms of these Green's functions.

1. INTRODUCTION

A GAUGE technique has been proposed in recent papers[1,2] that gives a nonperturbative renormalizable theory of vector and scalar mesodynamics, which can be summarized as follows:

The scalar gauge character of the theory implies generalized Ward–Takahashi identities between $M_{(l)}^{(n+1)}$ and $M_{(l)}^{(n)}$, where $M_{(l)}^{(n)}$ is an l-meson (1 incident mesons and l emitted mesons), n-photon Green's function. Given $M_{(l)}^{(n)}$, $M_{(l)}^{(n+1)}$ can be decomposed into longitudinal and transverse parts in such a way that the longitudinal part is a functional of $M_{(l)}^{(n)}$ alone. Beginning with $M_{(l)}^{(0)}$ for any l, the following sequence can be constructed:

$$M_{(l)}^{(1)} = M_{(l)}^{(1)A}[M_{(l)}^{(0)}] + M_{(l)}^{(1)B},$$
$$M_{(l)}^{(2)} = M_{(l)}^{(2)AA}[M_{(l)}^{(0)}] + M_{(l)}^{(2)AB}[M_{(l)}^{(1)B}] + M_{(l)}^{(2)BB}, \tag{1.1}$$

etc. In conventional notation $M_{(1)}^{(0)} \equiv \Delta$, $M_{(1)}^{(1)} \equiv \Gamma$, $M_{(1)}^{(2)} \equiv C$.

If D is the photon propagator the set of Green's functions form a *basic* set, in that any Green's function can be expressed as a functional of these in a Dyson–Schwinger sense. Instead of this set, let the set of Green's functions D, $\{M_{(1)}^{(n)}\}$ be taken as the set from which all further Green's functions are to be constructed.

Having expressed all higher Green's functions in this way, the unitarity equations for $M_{(1)}^{(n)}$ can be written in the form

$$M_{(1)}^{(n)} = F[D, \Delta, \Gamma, C, \cdots, M_{(1)}^{(m)} \cdots]. \tag{1.2}$$

The *zeroth* gauge approximation is to take

$$\Gamma = \Gamma^A[\Delta], \qquad C = C^{AA}[\Delta], \cdots, \tag{1.3}$$

where Δ is obtained by explicitly solving the unitarity equations

$$\begin{aligned}\Delta &= G[D, \Delta, \Gamma, C, \cdots, M_{(1)}^{(m)} \cdots] \\ &\approx G[D, \Delta, \Gamma^A[\Delta], C^{AA}[\Delta], \cdots] \\ &= G[D, \Delta], \end{aligned} \tag{1.4a}$$

$$\begin{aligned}D &= H[D, \Delta, \Gamma, C, \cdots, M_{(1)}^{(m)} \cdots] \\ &\approx H[D, \Delta, \Gamma^A[\Delta], C^{AA}[\Delta], \cdots] \\ &= H[D, \Delta], \end{aligned} \tag{1.4b}$$

for Δ and D. [Solved by iteration beginning with one-meson many-photon unitarity in (1.4a), reducing (1.4a) to the form

$$\Delta = G[\Delta],$$

for example.]

Γ^B, C^{BB}, $\cdots$, etc. are then determined by their unitarity equations as

$$\begin{aligned}\Gamma^B &= -\Gamma^A + K[D, \Delta, \Gamma, C, \cdots] \\ &\approx -\Gamma^A + K[D, \Delta, \Gamma^A, C^{AA}, \cdots], \end{aligned} \tag{1.5a}$$

$$\begin{aligned}C^{AB} &= -C^{AA} + L[D, \Delta, \Gamma, C, \cdots] \\ &\approx -C^{AA} + L[D, \Delta, \Gamma^A, C^{AA}, \cdots], \end{aligned} \tag{1.5b}$$

etc.

In the *next* gauge approximation, the procedure is analogous, except that three equations have to be solved explicitly, those for Δ, D, and Γ.

Symbolically, these equations are

$$\Delta = G[D, \Delta, \Gamma, C^{AA} + C^{AB}, \cdots], \tag{1.6a}$$

$$D = H[D, \Delta, \Gamma, C^{AA} + C^{AB}, \cdots], \tag{1.6b}$$

$$\Gamma = K[D, \Delta, \Gamma, C^{AA} + C^{AB}, \cdots]. \tag{1.6c}$$

C^{BB} and corresponding terms for higher functions are then determined by their unitarity equations as

$$C^{BB} = -(C^{AA} + C^{AB}) + L[D, \Delta, \Gamma, C^{AA} + C^{AB}, \cdots], \tag{1.7}$$

etc.

* Permanent address: Imperial College of Science and Technology, London, England.

[1] A. Salam, Phys. Rev. 130, 1287 (1963).

[2] A. Salam and R. Delbourgo, Phys. Rev. 135, 1398 (1964).

Higher-order gauge approximations are analogously defined.

It is not the purpose of this paper to discuss the severe approximations that have to be made in (1.1) before solutions to Eqs. (1.4) and (1.5) are feasible but to consider the gauge properties of the theory and the approximations to the Green's functions of the theory that maintain these gauge properties. In Sec. 2 the identities satisfied by the n-point functions and their connected parts are exhibited. In Secs. 3–5 these connected parts are expanded in terms of the Green's functions previously mentioned and the generalized identities of the Ward–Takahashi kind that these Green's functions satisfy are determined. In Sec. 6 the approximation schemes for the Green's functions that maintain the gauge properties of the theory are discussed, and in the event of there not being spectral representations for these Green's functions, the gauge approximations outlined above are shown to be the most natural of such schemes.

Since the motivation of this paper is a nonperturbative solution of mesodynamics, the approach will be as far removed from perturbation theory as possible, although a perturbative approach like that of the ξ-limiting process[3] would yield the identities in a much shorter space. For convenience the terminology of scalar mesodynamics will be used, since the generalization of notation to the vector meson case is readily apparent.

2. THE n-POINT FUNCTIONS

The identities satisfied by the connected parts of the n-point functions are first established. Introduce the operator-valued generating functional of time-ordered operator products

$$\hat{T}\{J, \bar{\eta}, \eta\} = T \exp\left(i \int [J_\mu(x)A_\mu(x) + \bar{\eta}(x)\phi(x) + \eta(x)\phi^\dagger(x)]\, dx\right), \tag{2.1}$$

where $J_\mu(x)$, $\bar{\eta}(x)$, $\eta(x)$ are c-number functions of x. Define $\hat{T}_{\mu_1\cdots\mu_n}(x_1, \cdots, x_n; y_1 \cdots y_m; z_1 \cdots, z_r)$ by

$$\hat{T}_{\mu_1\cdots\mu_n}(x_1, \cdots, x_n; y_1 \cdots y_m; z_1 \cdots z_r) = \frac{\delta^{(n+m+r)}\hat{T}}{\prod_1^n \delta J_{\mu_i}(x_i) \prod_1^m \delta\bar{\eta}(y_i) \prod_1^r \delta\eta(z_i)} \tag{2.2}$$

and define $\tau_{\mu_1\cdots\mu_n}(x_1 \cdots, x_n; y_1 \cdots y_m; z_1 \cdots z_m)$ by

$$\tau_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_m; z_1 \cdots z_m) = (-i)^{n+2m} \times \langle \hat{T}_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_m; z_1 \cdots z_m)\rangle \Big|_{\substack{J=0\\ \eta=\bar{\eta}=0}}. \tag{2.3}$$

The generator $\hat{H}$ of the ρ-functions, the connected parts of the τ-functions, is defined by

$$\exp \hat{H}\{J, \bar{\eta}, \eta\} = \langle \hat{T}\{J, \bar{\eta}, \eta\}\rangle. \tag{2.4}$$

Starting from a renormalized gauge-invariant Lagrangian, a differential recursion relation can be obtained between the τ-functions.[4] This is that

$$(\Box_x \partial/\partial x_\mu)[\tau_{\mu\mu_1\cdots\mu_n}(x, x_1 \cdots x_n; y_1 \cdots y_m; z_1 \cdots z_m) - \sum_i \rho_{\mu\mu_i}(x, x_i)\tau_{\mu_1\cdots\mu_{i-1}\mu_{i+1}\cdots\mu_n}(x_1 \cdots x_{i-1}, x_{i+1} \cdots x_n; y_1 \cdots y_m; z_1 \cdots z_m)] = e \sum_i [\delta^4(x - z_i) - \delta^4(x - y_i)]\tau_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_m; z_1 \cdots z_m), \tag{2.5}$$

where e is the renormalized charge.

On the left-hand side of Eq. (2.5) one uses the recursion formula (arising from $\hat{T}_x = \hat{H}_x\hat{T}$)

$$\tau(x, x_1, \cdots x_n; y_1 \cdots y_m; z_1, \cdots z_m) = \rho(x, x_1 \cdots, x_n; y_1 \cdots y_m; z_1, \cdots z_m) + \sum_{\text{comb}} \rho(xx_1' \cdots x_k'; y_1' \cdots y_s'; z_1' \cdots z_s') \cdot \tau(x_{k+1}' \cdots x_n'; y_{s+1}' \cdots y_m'; z_{s+1}' \cdots z_m'), \tag{2.6}$$

where $x_1' \cdots x_n', y_1' \cdots y_m', z_1 \cdots z_m'$ is the permutation of $x_1 \cdots x_n, y_1 \cdots y_m, z_1 \cdots z_m$ and the summation is taken over all partitions like the above. (Indices have been dropped when no confusion can arise.)

Using the corresponding recursion formula arising from $\hat{T}_{y_1} = \hat{H}_{y_1}\hat{T}$ on the right-hand side of (2.5), it can easily be shown by induction that

$$(\Box_x \partial/\partial x_\mu)\rho_{\mu\mu_1\cdots\mu_n}(x, x_1 \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l) = e \sum_i [\delta^4(x - z_i) - \delta^4(x - y_i)] \cdot \rho_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l). \tag{2.7}$$

A corollary of (2.7) is that

$$(\Box_x \partial/\partial x_\mu)\rho_{\mu\mu_1\cdots\mu_n}(x, x_1, \cdots x_n) = 0, \qquad \text{all } n > 1. \tag{2.7a}$$

Taking the *full* meson and photon propagators as $i\Delta$ and $-iD_{\mu\nu}$, respectively, ρ can be expressed as

[3] T. D. Lee, Phys. Rev. 128, 899 (1962).

[4] K. Nishijima, Phys. Rev. 119, 485 (1960).

$$\rho_{\mu_1\cdots\mu_n}(x_1, \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l)$$

$$= i^{n+1}(-1)^l e^n \int \prod dx_i' \prod dy_i' \prod dz_i'$$

$$\times \left[\prod_1^l \Delta(y_i - y_i') \prod_1^l (z_i - z_i') D_{\mu_i\nu_i}(x_i - x_i') \right.$$

$$\left. \times \mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(x_1' \cdots x_n'; y_1' \cdots y_l'; z_1' \cdots z_l') \right]. \quad (2.8)$$

[That is, removing the propagators from ρ gives

$$i(-e)^n \mathfrak{M}^{(n)}_{(l)}].$$

$\mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(x_1, \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l)$ has the Fourier transform

$$\mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1' \cdots k_n')$$

$$\text{with} \quad \sum p' = \sum p + \sum k,$$

where the incoming and outgoing mesons have charge $+e$, momenta p, and p', respectively, and the photons have incoming momenta k and polarizations ν.

C and CPT invariance give

$$\mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(p_1, \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_n) = (-1)^n \mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(-p_1' \cdots -p_l'; -p_1 \cdots, -p_l; k_1, \cdots k_n)$$

$$= \mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(p_1' \cdots p_l'; p_1 \cdots, p_l; -k_1, \cdots -k_n). \quad (2.9a)$$

In addition,

$$\mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_a\cdots\nu_b\cdots\nu_n}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_a \cdots k_b \cdots k_n)$$

$$= \mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_b\cdots\nu_a\cdots\nu_n}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_b \cdots k_a \cdots k_n) \quad (2.9b)$$

and

$$\mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(p_1 \cdots p_i \cdots p_k \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_n)$$

$$= \mathfrak{M}^{(n)}_{(l)\nu_1\cdots\nu_n}(p_1 \cdots p_k \cdots p_i \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_n). \quad (2.9c)$$

A similar expression exists for interchange of p_i', p_k'.

Taking Fourier transforms of (2.7), it is seen that the following relation holds for *all* n and l except $l = 0$, and $n = 0$, $l = 1$, for which the right-hand side of (2.10) is not defined:

$$k_\mu \mathfrak{M}^{(n+1)}_{(l)\mu\mu_1\cdots\mu_n}(p_1 \cdots p_l; p_1' \cdots p_l'; k, k_1, \cdots k_n)$$

$$= \sum_i \Delta^{-1}(p_i)\, \Delta(p_i + k) \mathfrak{M}^{(n)}_{(l)\mu_1\cdots\mu_n}(p_1, \cdots, p_i + k, \cdots p_l; p_1' \cdots p_l', k_1 \cdots k_n)$$

$$- \sum_i \Delta^{-1}(p_i')\, \Delta(p_i' - k) \mathfrak{M}^{(n)}_{(l)\mu_1\cdots\mu_n}(p_1 \cdots p_l; p_1', \cdots p_i' - k, \cdots p_l'; k_1 \cdots k_n). \quad (2.10)$$

The case $n = 0$, $l = 1$, is the Ward–Takahashi identity

$$k_\mu \Gamma_\mu(p, p') = \Delta^{-1}(p') - \Delta^{-1}(p). \quad (2.10a)$$

A corollary of (2.10) is that for all mesons on the mass shell

$$k_\mu \mathfrak{M}^{(n+1)}_{(l)\mu\mu_1\cdots\nu_n}(p_1, \cdots, p_l; p_1', \cdots, p_l'; k, k_1, \cdots k_n) \equiv 0 \quad (2.11)$$

as required by gauge invariance.

It is necessary to consider the case $l = 1$ in detail.

3. THE GREEN'S FUNCTIONS OF $\mathfrak{M}^{(1)}_{(n)}$

Let the *one-photon* irreducible part of $\mathfrak{M}^{(n)}_{(1)}$ be $\mathfrak{M}^{[n]}_{(1)}$. [That is, all diagrams of $\mathfrak{M}^{(n)}_{(1)}$ containing photon poles $D(k)$, where k is a sum of external photon momenta, are omitted.] It follows from graphical considerations that $\mathfrak{M}^{[n]}_{(1)}$ satisfies (2.10) for $l = 1$. Denoting the one-*photon*, one-*meson* irreducible vertex with m-photons and one incident and emitted meson by $i(-e)^m M^{(m)}_{(1)}$, $\mathfrak{M}^{[n]}_{(1)}$ can be expressed in terms of $M^{(m)}_{(1)}$, $m = 1, 2, \cdots, n$ as

$$\mathfrak{M}^{[n]}_{(1)}(p, p'; k_1 \cdots k_n) = \sum_{\text{comb}} M^{(n_1)}_{(1)}(p, p_1; k_1' \cdots k_{n_1}')$$

$$\times [-\Delta(p_1)] M^{(n_2)}_{(1)}(p_1, p_2; k_{n_1+1}' \cdots k_{n_1+n_2}') \times \cdots \times [-\Delta(p_{s-1})] M^{(n_s)}_{(1)}(p_{s-1}, p'; k_{n-n_s+1}' \cdots k_n'), \quad (3.1)$$

where

$$\sum_i n_i = n, \qquad p_r = p_{r-1} + \sum_{n_{r-1}+1}^{n_r} k_i,$$

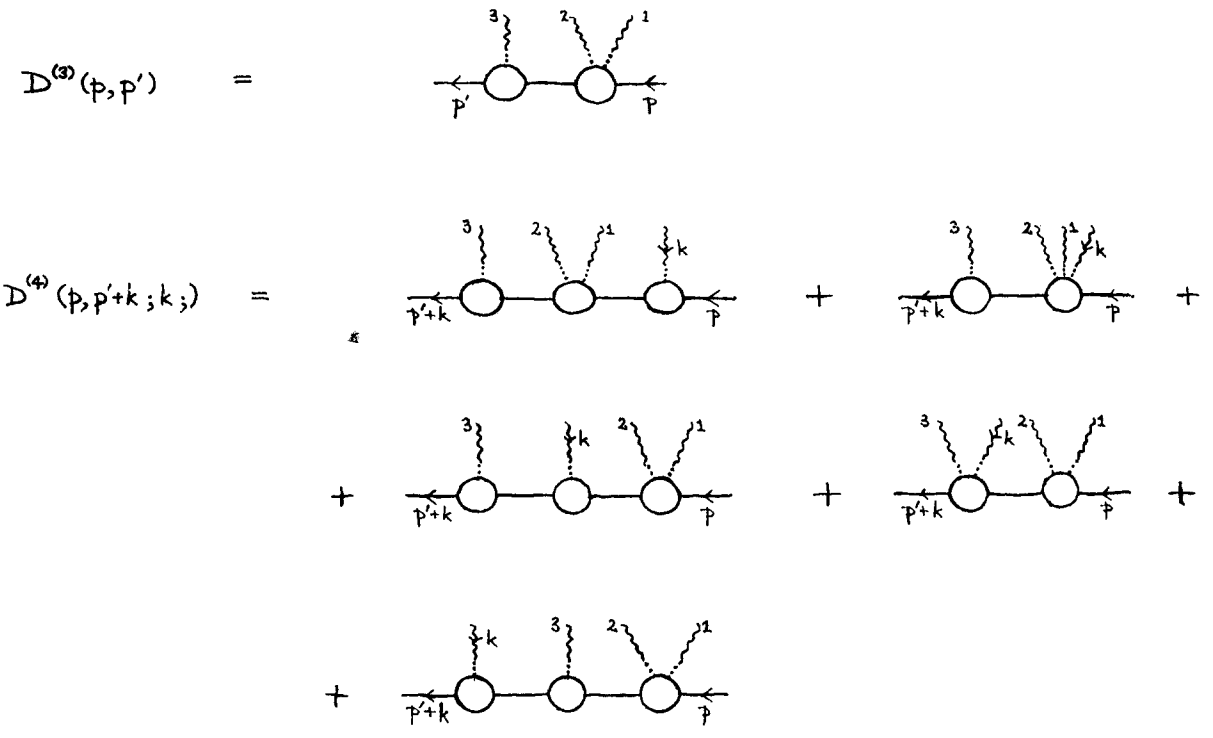

FIG. 1.

and where $k_1' \cdots k_n'$ is the permutation of $k_1 \cdots k_n$, the summation being taken over all possible partitions. (The photon indices have again been omitted for convenience.)

The symmetry properties of $M_{(1)}^{(n)}$ are the same as for $\mathfrak{M}_{(1)}^{(n)}$ and

$$M_{(1)\mu}^{(1)}(p, p') \equiv \Gamma_\mu(p, p'), \tag{3.2a}$$

$$M_{(1)\mu\nu}^{(2)}(p, p'; k, k') \equiv C_{\mu\nu}(p, p'; k, -k'), \tag{3.2b}$$

in the notation of Refs. 1 and 2. In a perturbation sense, $M_{(1)}^{(n)}$ is that function obtained by the insertion of n photon vertices in all possible ways in a meson proper self-energy blob.

The identities satisfied by the $M_{(1)}^{(n)}$ can now be obtained. Firstly, the following definitions are needed: Let $D_{\mu_1}^{(n)}{}_{\cdots\mu_n}(p, p'; k_1 \cdots k_n)$ [when there can be no confusion, abbreviated to $D^{(n)}(p, p')$] denote any diagram which connects a single meson line of incoming momentum p and outgoing momentum p', with photon vertices of incoming momenta $k_1, \cdots k_n$ and polarisations $\mu_1 \cdots \mu_n$ in which all general vertices are one-meson one-photon irreducible [i.e., any term on the right-hand side of (3.1)]. Then $D_{\mu_1}^{(n)}{}_{\cdots\mu_n}(p, p'; k_1 \cdots k_n)$ can be said to be symbolically of the form

$$D^{(n)}(p, p') = [(n_s)(n_{s-1}) \cdots (n_2)(n_1)], \tag{3.3}$$

where

$$\sum_{i=1}^{s} n_i = n.$$

For every such

$$D_{\mu_1}^{(n)}{}_{\cdots\mu_n}(p, p'; k_1 \cdots k_n), D_{\mu;\mu_1}^{(n+1)}{}_{\cdots\mu_n}(p, p'+k; k; k_1 \cdots k_n)$$

[abbreviated as $D_\mu^{(n+1)}(p, p' + k; k;)$] is defined as the sum of the set of diagrams obtained from $D_{\mu_1}^{(n)}{}_{\cdots\mu_n}(p, p'; k_1 \cdots k_n)$ by the insertion of a photon vertex of incoming momentum k and polarisation μ in all possible ways [i.e., the removal of a photon vertex μ from $D_\mu^{(n+1)}(p, p' + k; k;)$ gives a set of diagrams topologically equivalent to $D^{(n)}(p, p')$]. Figure 1 is an example of this procedure. (Straight lines denote mesons and wavy lines denote photons.)

The following Lemma then holds.

Lemma:

If, for $n = 1, 2, \cdots, m$,

$$k_\mu M_{(1)\mu\mu_1\cdots\mu_n}^{(n+1)}(p, p' + k; k, k_1, \cdots k_n) = M_{(1)\mu_1\cdots\mu_n}^{(n)}(p + k, p' + k; k_1 \cdots k_n) - M_{(1)\mu_1\cdots\mu_n}^{(n)}(p, p'; k_1, \cdots k_n), \tag{3.4}$$

then

$$k_\mu D_{\mu;\mu_1\cdots\mu_n}^{(n+1)}(p, p' + k; k; k_1 \cdots k_n) = e\Delta^{-1}(p' + k)\Delta(p')D_{\mu_1\cdots\mu_n}^{(n)}(p, p'; k_1 \cdots k_n) - eD_{\mu_1\cdots\mu_n}^{(n)}(p+k, p'+k; k_1, \cdots k_n)\Delta(p+k)\Delta^{-1}(p), \tag{3.5}$$

for all $D^{(n)}$ for which $n = 1, 2, \cdots, m$.

Proof: (a) If $D^{(n)}(p, p')$ is of the form $[(n)]$, then

$$k_\mu D_\mu^{(n+1)}(p, p' + k; k;)$$
$$= -i(-e)^{n+1}[k_\mu M_{(1)\mu}^{(n+1)}\ldots(p, p' + k;)$$
$$- k_\mu \Gamma_\mu(p, p' + k)\Delta(p + k)M_{(1)}^{(n)}(p + k, p' + k;)$$
$$- k_\mu \Gamma_\mu(p', p' + k)\Delta(p')M_{(1)}^{(n)}(p, p';)]$$
$$= e\Delta^{-1}(p + k)\Delta(p)D^{(n)}(p, p')$$
$$- eD^{(n)}(p + k, p' + k)\Delta(p + k)\Delta^{-1}(p), \quad (3.6)$$

for $n = 1, 2, \cdots, m$.

(*b*) If $D^{(n)}(p, p')$ is of the form $[(s) \cdots]$ where $1 \leq s < n$, assume that (3.5) holds for $n = 1, 2, \cdots r$ where $r < m$. Then $D^{(r+1)}(p, p')$ is of the form

$$D^{(r+1)}(p, p')$$
$$= -(-e)^s M_{(1)}^{(s)}(p'', p';)\Delta(p'')D^{(r-s+1)}(p, p''), \quad (3.7)$$

where $s \neq r + 1$ and

$$p'' = p' - \sum_{r-s+2}^{r+1} k_i.$$

This gives

$$D_\mu^{(r+2)}(p, p' + k; k;)$$
$$= -(-e)^s M_{(1)}^{(s)}(p'' + k, p' + k;)$$
$$\times \Delta(p'' + k)D_\mu^{(r-s+2)}(p, p'' + k; k)$$
$$+ e(-e)^s M_{(1)\mu}^{(s+1)}\ldots(p'', p' + k; k\cdots)$$
$$\times \Delta(p'')D^{(r-s+1)}(p, p'')$$
$$- e(-e)^s \Gamma_\mu(p', p' + k)$$
$$\times \Delta(p')M_{(1)}^{(s)}(p'', p';)D^{(r-s+1)}(p, p'') \quad (3.8)$$

which, under the hypothesis made, gives

$$k_\mu D_\mu^{(r+2)}(p, p' + k; k;)$$
$$= eD^{(r+1)}(p, p')\Delta^{-1}(p' + k)\Delta(p')$$
$$- eD^{(r+1)}(p + k, p' + k)\Delta(p + k)\Delta^{-1}(p). \quad (3.9)$$

Equation (3.9) is easily seen to be true for $D = [(1), (1)]$, and hence, the Lemma is true for D's of type b). Thus the Lemma is true. The following theorem can now be shown to be true.

Theorem I:

If $M_{(1)}^{(n)}$ is the one-photon one-meson irreducible part of $\mathfrak{M}_{(1)}^{(n)}$, where $\mathfrak{M}_{(1)}^{(n)}$ is defined as in (2.8), then

$$k_\mu M_{(1)\mu\mu_1\cdots\mu_n}^{(n+1)}(p, p' + k; k, k_1 \cdots k_n)$$
$$= M_{(1)\mu_1\cdots\mu_n}^{(n)}(p + k, p' + k; k_1 \cdots k_n)$$
$$- M_{(1)\mu_1\cdots\mu_n}^{(n)}(p, p'; k_1 \cdots k_n) \quad (3.10)$$

for all $n > 0$.

Proof:

The previous results are that

$$\Delta(p)k_\mu \mathfrak{M}_{(1)\mu\mu_1\cdots\mu_n}^{[n+1]}(p, p' + k; k, k_1 \cdots k_n)\Delta(p' + k)$$
$$= \Delta(p + k)\mathfrak{M}_{(1)\mu_1\cdots\mu_n}^{[n]}(p + k, p' + k; k_1 \cdots k_n)$$
$$\times \Delta(p' + k)$$
$$- \Delta(p)\mathfrak{M}_{(1)\mu_1\cdots\mu_n}^{[n]}(p, p'; k_1 \cdots k_n)\Delta(p') \quad (3.11)$$

for *all* n, and that if (3.10) holds for $n = 1, 2, \cdots m$ then (3.5) holds for $n = 1, 2, \cdots m$.

Assume that (3.10) is true for $n = 1, 2, \cdots m$. Take $n = m + 1$ in (3.11). Using the decomposition (3.1) for $\mathfrak{M}_{(1)}^{[n+2]}$ and $\mathfrak{M}_{(1)}^{[m+1]}$ and the Lemma (3.11) reduces, for this value of n, to the form

$$k_\mu M_{(1)\mu\mu_1\cdots\mu_{m+1}}^{(m+2)}(p, p' + k; k, k_1, \cdots k_{m+1})$$
$$- k_\mu \Gamma_\mu(p, p + k)\Delta(p + k)$$
$$\times M_{(1)\mu_1\cdots\mu_{m+1}}^{(m+1)}(p + k, p' + k; k_1 \cdots k_{m+1})$$
$$- k_\mu \Gamma_\mu(p', p' + k)\Delta(p')$$
$$\times M_{(1)\mu_1\cdots\mu_{m+1}}^{(m+1)}(p, p'; k_1, \cdots k_{m+1})$$
$$= \Delta^{-1}(p' + k)\Delta(p')$$
$$\times M_{(1)\mu_1\cdots\mu_{m+1}}^{(m+1)}(p, p'; k_1 \cdots k_{m+1})$$
$$- \Delta(p + k)\Delta^{-1}(p)$$
$$\times M_{(1)\mu_1\cdots\mu_{m+1}}^{(m+1)}(p + k; p' + k; k_1 \cdots k_{m+1}), \quad (3.12)$$

equivalent diagrammatically to Fig. 2.

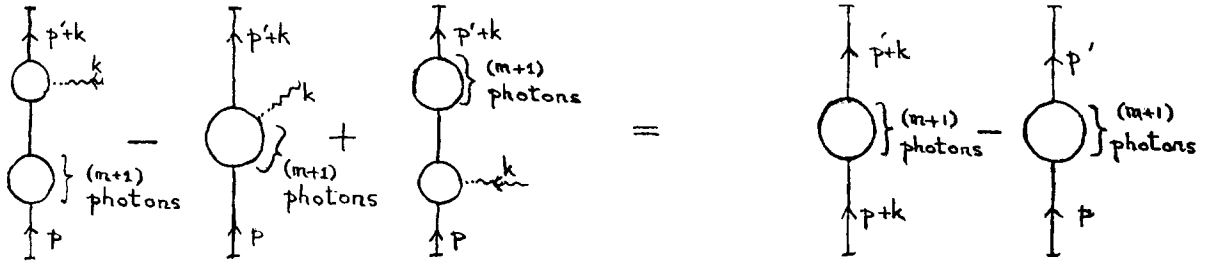

FIG. 2.

Ward's identity immediately gives

$$k_\mu M^{(m+2)}_{(1)\mu\mu_1\cdots\mu_{m+1}}(p, p'+k; k, k_1, \cdots k_{m+1}) = M^{(m+1)}_{(1)\mu_1\cdots\mu_{m+1}}(p+k, p'+k; k_1, \cdots k_{m+1}) - M^{(m+1)}_{(1)\mu_1\cdots\mu_{m+1}}(p, p'; k_1, \cdots k_{m+1}). \quad (3.13)$$

Thus if (3.10) holds for $n = 1, 2, \cdots m$, it holds for $n = m + 1$. Equation (3.10) can easily be shown to be true for $n = 1$. Hence, the theorem is proved.

A corollary to the theorem is that (3.5) is true for all $n > 0$. If the meson is on the mass shell initially and finally, then

$$k_\mu D^{(n+1)}_\mu(p, p'+k; k;) \equiv 0 \qquad \text{for all} \quad n > 0.$$

4. THE DECOMPOSITION OF $\mathfrak{M}^{(n)}_{(l)}$

The case $l > 1$ is now considered in detail, using further properties of functional derivatives.

For the sake of convenience the notation will be adopted that repetition of the same space–time variables implies integration over them, and that, where not stated explicitly, the space–time coordinates x, y, z correspond to functional differentiation with respect to J_μ, $\bar\eta$, and η, respectively.

In addition the functions $\hat\rho$ will be defined as the functional derivatives of $\hat H\{J, \bar\eta, \eta\}$ with respect to external sources so that

$$\hat\rho\,\Big|_{\substack{J=0\\ \eta=\bar\eta=0}} = \rho. \quad (4.1)$$

With this notation, it is seen that

$$\hat\rho_{\mu\nu}(x_1, x_2) = \frac{\delta\hat\rho_\mu(x_1)}{i\delta J_\nu(x_2)} = \frac{\delta\hat\rho_\nu(x_2)}{i\delta J_\mu(x_1)}. \quad (4.2)$$

Differentiating the identity

$$\delta\hat\rho(y, z) = -\hat\rho(y, z')\delta\hat\rho^{-1}(z', y')\hat\rho(y', z) \quad (4.3)$$

n times with respect to J_μ and using the relation (4.2), a decomposition of $\hat\rho_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y, z)$ is obtained in terms of

$$\hat M^{(r)}_{(1)\nu_1\cdots\nu_r}(x_1 \cdots x_r; y, z) \quad \text{and} \quad \hat\sigma^{(s)}_{\nu_1\cdots\nu_s}(x_1 \cdots x_s),$$

where

$$\hat\sigma^{(s)}_{\nu_1\cdots\nu_s}(x_1 \cdots x_s) = \frac{\delta}{\delta\hat\rho_{\nu_s}(x_3)} \cdots \frac{\delta}{\delta\hat\rho_{\nu_s}(x_s)}\,\hat\rho^{-1}(x_1, x_2) \quad (4.4)$$

and

$$\hat M^{(r)}_{(1)\nu_1\cdots\nu_r}(x_1 \cdots x_r; y, z) = i(-e)^{-r}\,\frac{\delta}{\delta\hat\rho_{\nu_1}(x_1)} \cdots \frac{\delta}{\delta\hat\rho_{\nu_r}(x_r)}\,\hat\rho^{-1}(y, z). \quad (4.5)$$

If

$$\sigma^{(s)}_{\nu_1\cdots\nu_s}(x_1, \cdots, x_s) = \hat\sigma^{(s)}_{\nu_1\cdots\nu_s}(x_1, \cdots, x_s)\,\Big|_{\substack{J=0\\ \eta=\bar\eta=0}} \quad (4.6)$$

and

$$M^{(r)}_{(1)\nu_1\cdots\nu_r}(x_1 \cdots x_r; y, z) = \hat M^{(r)}_{(1)\nu_1\cdots\nu_r}(x_1 \cdots x_r; y, z)\,\Big|_{\substack{J=0\\ \eta=\bar\eta=0}}, \quad (4.7)$$

then $\sigma^{(s)}$ is the one-photon irreducible s-photon purely-photon vertex and $M^{(r)}_{(1)}$ is the one-meson one-photon irreducible vertex whose Fourier transform is defined in Sec. 3.

Because the form of the identities is one of *photon–vertex insertion*, there is no need to begin from the basic identity (4.3) in the construction of $\mathfrak{M}^{(n)}_{(l)}$, and it is sufficient to start with $\hat\rho(y_1 \cdots, y_l; z_1, \cdots z_l)$ which can be expressed as

$$\rho(y_1, \cdots, y_l; z_1, \cdots, z_l) = i\rho(y_1, z_1') \cdots \rho(y_l, z_l')\rho(z_1, y_1') \cdots \rho(z_l, y_l') \cdot \hat{\mathfrak{M}}^{(0)}_{(l)}(y_1', \cdots, y_l'; z_1' \cdots z_l'), \quad (4.8)$$

where

$$\mathfrak{M}^{(0)}_{(l)}(y_1, \cdots y_l; z_1 \cdots z_l) = \hat{\mathfrak{M}}^{(0)}_{(l)}(y_1, \cdots, y_l; z_1, \cdots z_l)\,\Big|_{\substack{J=0\\ \eta=\bar\eta=0}}. \quad (4.9)$$

Differentiating each side of (4.8) n times with respect to J_μ and using (4.2), (4.4), and (4.5) an expansion of $\hat\rho_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l)$ (and hence an expansion of $\hat{\mathfrak{M}}^{(n)}_{(l)}$ by stripping off the external propagators) is obtained as a sum of simply connected diagrams, each one expressible in terms of

$$\hat M^{(r)}_{(1)}(x_1 \cdots x_r; y, z), \ \hat\sigma^{(s)}(x_1 \cdots x_s)$$

and

$$\hat M^{(m)}_{(l)\mu_1\cdots\mu_m}(x_1 \cdots x_m; y_1 \cdots y_l; z_1 \cdots z_l),$$

where

$$\hat M^{(m)}_{(l)\mu_1\cdots\mu_m}(x_1 \cdots x_m; y_1 \cdots y_l; z_1 \cdots z_l) = (-e)^{-m}\,\frac{\delta}{\delta\hat\rho_{\mu_1}(x_1)} \cdots \frac{\delta}{\delta\hat\rho_{\mu_m}(x_m)} \times \hat{\mathfrak{M}}^{(0)}_{(l)}(y_1 \cdots y_l; z_1 \cdots z_l). \quad (4.10)$$

Define $M^{(m)}_{(l)\mu_1\cdots\mu_m}(x_1 \cdots x_m; y_1 \cdots y_l; z_1 \cdots z_l)$ by

$$M^{(m)}_{(l)\mu_1\cdots\mu_m}(x_1 \cdots x_m; y_1 \cdots y_l; z_1 \cdots z_l) = \hat M^{(m)}_{(l)\mu_1\cdots\mu_m}(x_1 \cdots x_m; y_1 \cdots y_l; z_1 \cdots z_l)\,\Big|_{\substack{J=0\\ \eta=\bar\eta=0}}. \quad (4.11)$$

Diagrammatically it is easily seen that differentiation with respect to J_μ corresponds to the insertion of an external photon propagator in all possible ways in a

given diagram and that differentiation with respect to $\hat{\rho}_\mu(x)$ corresponds to the insertion of a single photon vertex. Thus $M^{(n)}_{(l)}$ corresponds to the insertion of n photon vertices in $\mathfrak{M}^{(0)}_{(l)}$. (Note that $M^{(0)}_{(l)} \equiv \mathfrak{M}^{(0)}_{(l)}$.)

Consider the above decomposition in more detail for a general ρ-function. In this decomposition the ρ-function is expressed as a sum of topologically distinct sets of simply connected diagrams, the number of diagrams in each set being just sufficient to yield the necessary symmetries in the external indices. For a given value of n and l, let these distinct sets be labeled by the index $i = 0, 1, 2, \cdots N(n, l)$, and let the sum of the diagrams in the ith set be $\mathcal{G}_{(i)\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1, \cdots, y_l; z_1, \cdots z_l)$. Then the $\mathcal{G}_{(i)}$ have the requisite symmetries in the external indices and

$$\rho_{\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l) = \sum_{i=0}^{N} \mathcal{G}_{(i)\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1, \cdots, y_l; z_1 \cdots z_l). \tag{4.12}$$

The value $i = 0$ is given to the set consisting of the single diagram in which all photon insertions have been made into $\mathfrak{M}^{(0)}_{(l)}$.

That is,

$$\mathcal{G}_{(0)\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l) = i(-e)^n \prod_1^n \rho_{\mu_i\nu_i}(x_i, x_i') \prod_1^l \rho(y_i, z_i') \prod_1^l \rho(z_i, y_i') \cdot M^{(n)}_{(l)\nu_1\cdots\nu_n}(x_1' \cdots x_n'; y_1' \cdots y_l'; z_1' \cdots z_l'). \tag{4.13}$$

Thus for $l > 1$, all $\mathcal{G}_{(i)}$ for which $i \neq 0$ contain only $M^{(m)}_{(l)}$ for which $m < n$.

5. THE GENERALIZED IDENTITIES

Reverting to momentum space, the Fourier transforms of the $M^{(n)}_{(l)}$ defined in Sec. 4 satisfy the following theorem.

Theorem II:

If $M^{(n)}_{(l)\mu_1\cdots\mu_n}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_n)$ is the Fourier transform of $M^{(n)}_{(l)\mu_1\cdots\mu_n}(x_1 \cdots x_n; y_1 \cdots y_l; z_1 \cdots z_l)$ as defined in Sec. 4 where p, p', k denote incoming momenta of mesons with charge e, outgoing momenta of mesons with charge e, and incoming momenta of photons with polarization μ, respectively, then

$$k_\mu M^{(n+1)}_{(l)\mu\mu_1\cdots\mu_n}(p_1 \cdots p_l; p_1' \cdots p_l'; k, k_1 \cdots k_n) = \sum_j M^{(n)}_{(l)\mu_1\cdots\mu_n}(p_1 \cdots, p_j + k, \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_n) - \sum_j M^{(n)}_{(l)\mu_1\cdots\mu_n}(p_1 \cdots p_l; p_1' \cdots, p_j' - k, \cdots p_l'; k_1 \cdots k_n), \tag{5.1}$$

for all n, l except $n = 0$, $l = 1$ for which the right-hand side of (5.1) is not defined. [In which case we have the Ward–Takahashi identity (2.10a).]

Proof:

Since Theorem I is the statement of Theorem II for the case $l = 1$, we need only consider $l > 1$. It is immediately seen from (2.10) and (2.10a) that (5.1) is true for $n = 1$, all l. The proof for general n is by induction in n.

Suppose (5.1) is true for all $l > 1$ and $n = 1, 2, \cdots m - 1$ for some $m > 2$. Consider the decomposition of $\rho_{\mu_1\cdots\mu_m}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_m)$ for arbitrary $l > 1$. From Eq. (4.10)

$$\rho_{\mu_1\cdots\mu_m}(p_1 \cdots, p_l; p_1' \cdots p_l'; k_1 \cdots k_m) = \sum_i \mathcal{G}_{(i)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_m), \tag{5.2}$$

where

$$\mathcal{G}_{(0)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_m) \; i^{m+1}(-1)^l e^m \Big[\prod_1^m D_{\mu_i\nu_i}(k)\Big]\Big[\prod_1^l \Delta(p_i)\Big]\Big[\prod_1^l \Delta(p_i')\Big] \cdot M^{(m)}_{(l)\nu_1\cdots\nu_m}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_m). \tag{5.3}$$

If, for $i \neq 0$, $g_{(i)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_m)$ is one of the diagrams in the ith set, then it has the form

$$g_{(i)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p_1' \cdots p_l'; k_1 \cdots k_m) \times \Big[\prod_1^m D^{-1}_{\mu_i\nu_i}(k_i)\Big] = i(-i)^m(-e)^r \prod_1^l [\Delta(p_i)\, D'^{(n_i)}(p_i, q_i)\, \Delta(q_i)] \times M'^{(r)}_{(l)\gamma_1\cdots\gamma_r}(q_1 \cdots q_l; q_1' \cdots q_l'; u_1 \cdots u_r) \times \prod_1^l [\Delta(q_i')\, D'^{(n'i)}(q_i', p_i')\, \Delta(p_i')], \tag{5.4}$$

where $\sum n_i + \sum n'_i + r = m$, and $u_1 \cdots u_r$ are some subset of $k_1 \cdots k_m$ with $\gamma_1 \cdots \gamma_r$ the corresponding polarizations.

The functions $D'^{(n)}(p, q)$ appearing on the right-hand side of Eq. (5.4) correspond to simply connected diagrams obtained from $D^{(n_0)}(p, q)$ as defined in Sec. 3 for some $n_0 \leq n$ by grafting on strings of purely-photon vertices in such a way that the total number of external photon vertices is n. The relationship between $M'^{(r)}_{(l)}$ and $M^{(r_0)}_{(l)}$ for some $r_0 \leq r < m$ is similarly defined.

With a little labor, using the lemma to Theorem I, Eqs. (2.7a) and (2.10a) and the induction assumption, it is seen that $g_{(i)}$ satisfies the identity

$$ik_\nu D^{-1}_{\nu\mu}(k) g_{(i)\mu;\mu_1\cdots\mu_m}(p_1 \cdots p_l; p'_1 \cdots p'_l : k; k_1, \cdots k_m)$$
$$= \sum_j g_{(i)\mu_1\cdots\mu_m}(p_1, \cdots, p_j + k, \cdots, p_l; p'_1 \cdots p'_l; k_1 \cdots k_m)$$
$$- \sum_j g_{(i)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p'_1, \cdots, p'_j - k, \cdots p'_l; k_1 \cdots k_m), \qquad (5.5)$$

where $g_{(i)\mu;\mu_1\cdots\mu_m}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_m)$ is the sum of the set of diagrams obtained from $g_{(i)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_m)$ by the insertion of a single photon propagator of momentum k and external polarization μ in all possible ways.

Thus, $\mathcal{G}_{(i)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_m)$ satisfies an identical identity for all $i \leq 0$, where

$$\mathcal{G}_{(i)\mu;\mu_1\cdots\mu_m} = \sum_{g\in\mathcal{G}} g_{(i)\mu;\mu_1\cdots\mu_m}. \qquad (5.6)$$

From Eqs. (2.8) and (2.10) it is seen that $\rho_{\mu_1\cdots\mu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_n)$ satisfies the identity

$$ik_\nu D^{-1}_{\nu\mu}(k) \rho_{\mu\mu_1\cdots\mu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l; k, k_1, \cdots k_n)$$
$$= \sum_j \rho_{\mu_1\cdots\mu_n}(p_1, \cdots, p_j + k, \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_n)$$
$$- \sum_j \rho_{\mu_1\cdots\mu_n}(p_1 \cdots p_l; p'_1 \cdots, p'_j - k, \cdots p'_l; k_1 \cdots k_n) \qquad (5.7)$$

for all n.

Hence, from (5.2) $\mathcal{G}_{(0)\mu_1\cdots\mu_m}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_m)$ satisfies (5.5). From the form of $\mathcal{G}_{(0)}$ as given in (5.3), it is seen that this implies that (5.1) holds for all l and $n = m$. Thus the theorem is proved.

A corollary to Theorem II is that Eq. (5.5) holds for all values of l and m. Removing the external propagators from $g_{(i)\mu_1\cdots\mu_n}$ and $g_{(i)\mu;\mu_1\cdots\mu_n}$ to give functions $h_{(i)\nu_1\cdots\nu_n}$ and $h_{(i)\nu;\nu_1\cdots\nu_n}$ defined by

$$g_{(i)\mu_1\cdots\mu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_n)\left[\prod_1^n D^{-1}_{\mu_i\nu_i}(k_i)\right]$$
$$= (-1)^l(-i)^n \prod_1^l \Delta(p_i) \prod_1^l \Delta(p'_i) h_{(i)\nu_1\cdots\nu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l; k_1 \cdots k_n), \qquad (5.8)$$

and

$$iD^{-1}_{\nu\mu}(k) g_{(i)\mu;\mu_1\cdots\mu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l : k; k_1 \cdots k_n)\left[\prod_1^n D^{-1}_{\mu_i\nu_i}(k_i)\right]$$
$$\equiv (-1)^l(-i)^n \prod_1^l \Delta(p_i) \prod_1^l \Delta(p'_i) h_{(i)\nu,\nu_1\cdots\nu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l : k; k_1 \cdots k_n), \qquad (5.9)$$

it is seen that for all the mesons on the mass shell

$$k_\nu h_{(i)\nu;\nu_1\cdots\nu_n}(p_1 \cdots p_l; p'_1 \cdots p'_l; k; k_1 \cdots k_n) \equiv 0: \qquad (5.10)$$

and can be taken as a definition of the gauge invariance of the theory.

6. THE APPROXIMATION SCHEMES THAT MAINTAIN GAUGE INVARIANCE

The function $M^{(n)}_{(l)}$ can be decomposed (nonuniquely) as

$$M^{(n)}_{(l)} = M^{(n)A}_{(l)} + M^{(n)B}_{(l)}, \qquad (6.1)$$

so that (symbolically)

$$k_i.M_{(l)}^{(n)A} = \sum_{\text{in}} M_{(l)}^{(n-1)} - \sum_{\text{out}} M_{(l)}^{(n-1)}, \qquad \text{all } k_i. \quad (6.2)$$

$$k_i.M_{(l)}^{(n)B} = 0,$$

For a given n, l define $M_{(l)}^{(m,n-m)}$ by the decomposition

$$M_{(l)}^{(n)A} = \sum_{m=1}^{n} M_{(l)}^{(m,n-m)}, \qquad M_{(l)}^{(n)B} = M_{(l)}^{(0,n)}, \quad (6.3)$$

where

$$k_i.M_{(l)}^{(m,n-m)} = \sum_{\text{in}} M_{(l)}^{(m-1,n-m)} - \sum_{\text{out}} M_{(l)}^{(m-1,n-m)}. \quad (6.4)$$

If ${}^{(r)}M_{(l)}^{(n)}$, the rth approximation for $M_{(l)}^{(n)}$, is defined as

$${}^{(r)}M_{(l)}^{(n)} = \sum_{s=0}^{r} M_{(l)}^{(n-s,s)} \quad \text{for } r < n, \qquad = M_{(l)}^{(n)} \quad \text{for } r \geq n, \quad (6.5)$$

it follows that

$$k_i.{}^{(r)}M_{(l)}^{(n)} = \sum_{\text{in}} {}^{(r)}M_{(l)}^{(n-1)} - \sum_{\text{out}} {}^{(r)}M_{(l)}^{(n-1)} \quad (6.6)$$

for all i.

For given arbitrary values of s and m, it is apparent that there is almost unlimited scope in setting up approximation schemes that preserve the symmetries and the gauge properties of $\mathfrak{M}_{(m)}^{(s)}$. The most general scheme that is at all meaningful is as follows.

(a) For $m = 1$ take

$$M_{(1)}^{(n)} = {}^{(r_1)}M_{(1)}^{(n)} \quad (6.7a)$$

everywhere that $M_{(1)}^{(n)}$ appears in the expansion of $\mathfrak{M}_{(1)}^{(s)}$ for arbitrary fixed r_1 and all n.

(b) For $m > 1$ take

$$M_{(1)}^{(n)} = {}^{(r_1)}M_{(1)}^{(n)}, \qquad M_{(m)}^{(n)} = {}^{(r_2)}M_{(m)}^{(n)} \quad (6.7b)$$

everywhere that $M_{(1)}^{(n)}$ and $M_{(m)}^{(n)}$ occur in the expansion for arbitrary fixed r_1 and r_2 and all n.

If the approximated $\mathfrak{M}_{(m)}^{(s)}$ in cases (a) and (b) are denoted by ${}^{(r_1)}\mathfrak{M}_{(1)}^{(s)}$, ${}^{(r_1,r_2)}\mathfrak{M}_{(m)}^{(s)}$, respectively, then (symbolically) in correspondence to Eq. (2.10),

$$k_i.{}^{(r_1)}\mathfrak{M}_{(1)}^{(s)} = \Delta^{-1}\,\Delta^{(r_1)}\mathfrak{M}_{(1)}^{(s-1)} - \Delta^{-1}\,\Delta^{(r_1)}\mathfrak{M}_{(1)}^{(s-1)}, \quad (6.8a)$$

$$k_i.{}^{(r_1,r_2)}\mathfrak{M}_{(m)}^{(s)} = \sum_{\text{in}} \Delta^{-1}\,\Delta^{(r_1,r_2)}\mathfrak{M}_{(m)}^{(s-1)} - \sum_{\text{out}} \Delta^{-1}\,\Delta^{(r_1,r_2)}\mathfrak{M}_{(m)}^{(s-1)} \quad (6.8b)$$

for all i.

Using these approximated $\mathfrak{M}_{(m)}^{(s)}$ in the full amplitude for a specific process will give an approximated amplitude having the correct gauge properties.

An approximation scheme of the type (6.7b) could only be of use if there were considerable knowledge of a spectral representation of $M_{(l)}^{(0)}$ for $l > 1$. In the absence of such representations $M_{(l)}^{(0)}$ must be expanded as the sum of a set of symmetrized[5] diagrams, each of which is expressible in terms of the $M_{(l)}^{(m)}$ vertices. The Dyson[6] expansion of $M_{(l)}^{(0)}$ is the most fundamental of these, using only Γ and C vertices. Further expansions can be made by suitable recombination of subsets of diagrams in this Dyson expansion. Although all such expansions are obviously equivalent in the Dyson sense, in general they become inequivalent when the vertices appearing in the diagrams of the expansion are approximated in gauge-approximation schemes. By maximum recombination of the Dyson expansion, a unique well-defined expansion can presumably be obtained. However, for the purpose of discussing the gauge properties of the diagrams in the expansion of $M_{(l)}^{(0)}$ any expansion can be considered.

Suppose $G_{(l)}^{(0)}$ is a symmetrized diagram in such an expansion. The contribution to $M_{(l)}^{(n)}$, arising from $G_{(l)}^{(0)}$, denoted by $G_{(l)}^{(n)}$, is obtained by grafting external propagators to $G_{(l)}^{(0)}$, making n photon-propagator insertions and then amputating all external propagators from the resulting diagrams in the manner previously discussed. Under the operation of photon-propagator insertion of a longitudinal $(n+1)$st photon propagator in all possible ways into this set of diagrams, it is easily seen that the parts of the diagrams that would be disconnected were all internal photon propagators served behave independently. This implies that the $G_{(l)}^{(n)}$ satisfy the same identities and symmetries as the $M_{(l)}^{(n)}$ to which they contribute, since it is only the meson lines passing right through the diagrams that contribute to the identities.

This has two consequences.

(1) If $\tilde{M}_{(l)}^{(0)}$ for $l > 1$ is an approximation of $M_{(l)}^{(0)}$ obtained by taking only a subset of symmetrized diagrams in some expansion of $M_{(l)}^{(0)}$ and $\tilde{M}_{(l)}^{(n)}$ is the approximation of $M_{(l)}^{(n)}$ obtained by n photon propagator insertions in $\tilde{M}_{(l)}^{(0)}$, then $\tilde{M}_{(l)}^{(n)}$ has the symmetries of $M_{(l)}^{(n)}$, and

$$k_i.\tilde{M}_{(l)}^{(n)} = \sum_{\text{in}} \tilde{M}_{(l)}^{(n-1)} - \sum_{\text{out}} \tilde{M}_{(l)}^{(n-1)} \quad (6.9)$$

for all i.

[5] Given a diagram G, the symmetrized diagram G is the sum of the set of diagrams obtained from G by permuting the external indices in all possible distinct ways.

[6] F. J. Dyson, Phys. Rev. 75, 1736 (1949).

(2) If ${}^{[r]}M^{(n)}_{(l)}$ is the approximation of $M^{(n)}_{(l)}$ obtained by taking

$$M^{(s)}_{(1)} = {}^{(r)}M^{(s)}_{(1)}$$

with arbitrary fixed r everywhere in some specified expansion of $M^{(n)}_{(l)}$, then

$$k_i.{}^{[r]}M^{(n)}_{(l)} = \sum_{\text{in}} {}^{[r]}M^{(n)}_{(l)} - \sum_{\text{out}} {}^{[r]}M^{(n)}_{(l)}. \qquad (6.10)$$

Note that the form of ${}^{(r)}M^{(n)}_{(l)}$ depends on the actual expansion of $M^{(n)}_{(l)}$ that has been used, and there is no correspondence between ${}^{[r]}M^{(n)}_{(l)}$ and ${}^{(r)}M^{(n)}_{(l)}$ defined in (6.5).

Obviously (1) and (2) can be combined.

There is again considerable scope in setting up approximation schemes that maintain the gauge properties of $\mathfrak{M}^{(s)}_{(m)}$ for arbitrary fixed s and m. In the absence of any further criteria the only meaningful scheme is to take

$$\begin{aligned} M^{(n)}_{(1)} &= {}^{(r)}M^{(n)}_{(1)}, \\ M^{(n)}_{(m)} &= {}^{[r]}M^{(n)}_{(m)} \end{aligned} \qquad (6.11)$$

in the expansion of $\mathfrak{M}^{(s)}_{(m)}$ for fixed $r(r = 0, 1, 2, \cdots)$. If the approximated $\mathfrak{M}^{(s)}_{(m)}$ in this approximation is denoted by ${}^{[r]}\mathfrak{M}^{(s)}_{(m)}$ then it satisfies the identity

$$k_i.{}^{[r]}\mathfrak{M}^{(s)}_{(m)} = \sum_{\text{in}} \Delta^{-1}\,\Delta^{[r]}\mathfrak{M}^{(s-1)}_{(m)} - \sum_{\text{out}} \Delta^{-1}\,\Delta^{[r]}\mathfrak{M}^{(s-1)}_{(m)} \qquad (6.12)$$

for all i. These are the gauge approximations discussed in Sec. 1, which thus preserve the gauge properties of the amplitudes of the theory.

CONCLUSION

The Green's functions used in this paper were chosen primarily because they arose naturally from the functional derivative methods used, and because scalar gauge invariance is more easily understandable in terms of photon propagator insertion into "basic" Green's functions than by working directly with the connected parts of the n-point functions. Moreover, the identities satisfied by these Green's functions have been shown to be of the form

$$k_i.M^{(n+1)}_{(l)} = \sum_{\text{in}} M^{(n)}_{(l)} - \sum_{\text{out}} M^{(n)}_{(l)}$$

without the presence of external propagators, thus simplifying the construction of the part of $M^{(n+1)}_{(l)}$ that is a functional of $M^{(n)}_{(l)}$, were $M^{(n)}_{(l)}$ known. In the present context, it is only the $M^{(n)}_{(1)}$ that concern us (although it is beyond the scope of this paper to decide how the decomposition of $M^{(n)}_{(l)}$ into the $M^{(n)}_{(1)}$ is to be made and to what extent an approximation for such a decomposition may be a "good" approximation) and so this advantage cannot be fully used.

Although it has been shown that there is wider scope for approximations that maintain the gauge properties of amplitudes with these Green's functions than with the n-point functions, this greater choice is essentially spurious, since there are no sensible reasons for invoking it. The only sensible approximation schemes within this framework are the gauge approximations discussed in Sec. 1.

The application of the same approach in general Lie gauge theories will lead to approximation schemes of similar form, but this is also beyond the scope of this work and will be the subject of a subsequent paper.

ACKNOWLEDGMENTS

The author would like to thank Professor A. Salam and the International Atomic Energy Agency for hospitality at the International Centre for Theoretical Physics. In addition, the author would like to thank Professor A. Salam for suggesting the work and for his encouragement.

Internal Symmetry and Mass Splitting*

RALPH ROSKIES†
Oak Ridge National Laboratory, Oak Ridge, Tennessee
(Received 1 June 1965)

We prove two theorems concerning the possibility of embedding the Lie algebra S of an internal symmetry group $\mathcal{S}$ and the Lie algebra P of the Poincare group $\mathcal{P}$ into a larger algebra G to obtain intrinsic mass splitting among the multiplets belonging to the irreducible representations of $\mathcal{S}$. We assume that G is a semi-direct sum of the invariant subalgebra S and of P. First, if $\mathcal{S}$ is a compact Lie group, then G is in fact a direct sum of S and P and there is no mass splitting. Second, if $\mathcal{S}$ is a semi-simple Lie group, with an appropriate definition of the mass operator, there is still no mass splitting. The relation of these results to previous papers is discussed.

INTRODUCTION

SEVERAL papers[1-5] have considered the problem of embedding the Poincare group $\mathcal{P}$ and an internal symmetry group $\mathcal{S}$ into a large group $\mathcal{G}$ to obtain intrinsic mass splitting among the multiplets which belong to the irreducible representations of $\mathcal{S}$.

One often adds the requirement that $\mathcal{G}$ be a product of $\mathcal{S}$ and $\mathcal{P}$; that is, every element in $\mathcal{G}$ can be written as a product of an element of $\mathcal{S}$ and an element of $\mathcal{P}$, and that $\mathcal{S}$ and $\mathcal{P}$ have no elements in common. Michel[4] has then shown that under certain weak assumptions on the commutator of elements of $\mathcal{S}$ and $\mathcal{P}$, $\mathcal{G}$ reduces to a semi-direct product (i.e., one of the groups $\mathcal{S}$ and $\mathcal{P}$ is an invariant subgroup of $\mathcal{G}$). That $\mathcal{S}$ be an invariant subgroup seems a most plausible physical condition. For let π denote an element of $\mathcal{P}$, σ an element of $\mathcal{S}$; then $\pi^{-1}\sigma\pi$ applied to a state signifies making a Lorentz transformation π, followed by an internal transformation σ, followed by the inverse Lorentz transformation π^{-1}. If the resultant state differed from the original one by more than a pure internal transformation, one would have to associate some space–time effects with σ, which is *not* what is meant by an internal transformation. Thus $\pi^{-1}\sigma\pi$ is an element of $\mathcal{S}$ and so $\mathcal{S}$ is an invariant subgroup. We are concerned with the Lie algebras (which we denote by the corresponding ordinary capital letters). Thus we assume that the algebra G is a semi-direct sum of the invariant subalgebra S and the subalgebra P.

It is sometimes suggested[3] that the internal symmetry group $\mathcal{S}$ be a compact Lie group. In this context, we shall prove:

Theorem 1: Let G be the semi-direct sum of the invariant subalgebra S and the Poincare algebra P. If $\mathcal{S}$ is a compact Lie group, then G is in fact a direct sum of S and P, and there is no mass splitting.

To obtain mass splitting, one might try to take $\mathcal{S}$ semi-simple, but not compact. In this regard one can show[3,4] that if G is the semi-direct sum of an invariant semi-simple subalgebra S and the Poincare subaglebra P, then

$$G = S \oplus P' \tag{1}$$

where P' is an algebra isomorphic to, but not identical to P. ($\oplus$ denotes direct sum.) The group element corresponding to a given Lorentz transformation in $\mathcal{P}'$ is not the same as that corresponding to the same transformation in $\mathcal{P}$, but is composed also of an element from $\mathcal{S}$ (see Eq. 5). An irreducible representation of G is the direct sum of an irreducible representation of S and an irreducible representation of P'. We concentrate on those irreducible representations of G in which the irreducible representation of S is *finite-dimensional*. The states of such a representation may be labeled by $\phi_m(p', s')$ where m denotes the row in the irreducible representation of S, and p', s' are four-momentum and spin labels belonging to the "rows" of the representation of P'.

The mass operator is $M^2 = P_\mu P^\mu$, where P_μ are the translation generators of P. We show that M^2 maps a state $\phi_m(p', s')$ into a linear combination of states *with the same* p', s'. Thus M^2 is essentially an operator (not necessarily self-adjoint) in the finite-dimensional vector space of the irreducible

* Research sponsored by the U. S. Atomic Energy Commission under contract with the Union Carbide Corporation.
† Present address: Yale University, New Haven, Connecticut.
[1] See references cited in L. O'Raifeartaigh, Phys. Rev. Letters 14, 332 (1965).
[2] L. O'Raifeartaigh, Phys. Rev. Letters 14, 575 (1965).
[3] U. Ottoson, A. Kihlberg, and J. Nilsson, Phys. Rev. 137, B658 (1965), hereinafter referred to as OKN.
[4] L. Michel, Phys. Rev. 137, B405 (1965).
[5] L. O'Raifeartaigh, Phys. Rev. 139, B1052 (1965).

representation of S. As such, M^2 has eigenvalues. We will prove:

Theorem 2: Let G be the semi-direct sum of the invariant subalgebra S and the Poincare algebra P. If S is semi-simple, then $G = S \oplus P'$ where P' is isomorphic to P. In an irreducible representation of G which is the direct sum of a finite-dimensional representation of S and an irreducible representation of P', the operator $M^2 = P_\mu P^\mu$ is essentially a finite-dimensional matrix whose eigenvalues are all equal. Their common value is the invariant $M'^2 = P'_\mu P'^\mu$ characterizing the irreducible representation of P'.

Thus if the eigenvalues of M^2 correspond to the physical masses (or to the location of the poles of the S matrix if the eigenvalues are complex), there is no mass splitting for the finite-dimensional representations of S, if S is semi-simple.

BACKGROUND

To put the theorems in context, we briefly review some previous papers. O'Raifeartaigh[2] has shown that if G is a Lie algebra with a finite number of parameters and if the mass operator is to be self-adjoint, then there is either no mass splitting or there is a continuum of masses. Hence to obtain a nontrivial discrete mass spectrum, either G must have an infinity of parameters or the mass operator is not self-adjoint. The first alternative is largely unexplored, and we concentrate on the second.

This possibility was considered in Ref. 3 even before O'Raifeartaigh's paper appeared. There the authors assume that G is a sum of S and P, and under conditions specified in Ref. 3, they deduce

$$G = S \oplus P',$$

where P' is isomorphic to, but not identical to P. The elements of P' are composed of elements from P and from S. This "interaction" of elements of P and S to produce the invariant P' gives rise to the mass splitting. In Ref. 5, this "redefinition" of P into P' is not considered to have any physical significance; i.e., the elements of P' and not of P are identified as the generators of physical Lorentz transformations. In what follows, we assume the interpretation of OKN[3] that P and not P' corresponds to the physical Lorentz transformations.

As an example, OKN claim that if one considers the $(1 + 1)$-dimensional Poincare group as $\mathcal{P}$, and $SU(4)$ as the group $\mathcal{S}$, then one can obtain a finite-dimensional representation of $\mathcal{S}$ which gives nontrivial mass splitting. In accord with Theorem 1, we wish to show that the group that should appear is not the compact group $SU(4)$, but its non-compact complex extension $SL(4)$. This is seen by computing their $[H_1, p_1]$ and recognizing that the result is not in the Lie algebra of $SU(4)$, but only in its complex extension. Furthermore, according to OKN the mass of a state is the *expectation value* of M^2 for that state. Perhaps a more reasonable definition is that the masses corrrespond to the *eigenvalues* of M^2, for eigenvalues are invariant under a change of basis, but expectation values are not. (M^2 is a matrix in a finite-dimensional vector space. It is generally not diagonalizable, but still has eigenvalues.) With eigenvalues as masses, we show in accord with Theorem 2 that there is no mass splitting.

It may appear that our Theorem 1 has already been proved in Ref. 4. However, the proof there shows that if S is compact, it is possible to write

$$G = S \oplus P'$$

with a redefinition of P into P'. We show that no redefinition is necessary.

After this paper was written, we learned that Sudarshan[6] proved a result similar to our Theorem 1. He showed that in any unitary representation of G, one has the direct sum structure of Theorem 1. The restriction to unitarity is necessary because he considers Lie algebras over the complex numbers, not over the reals as in this paper. Unitarity then imposes the appropriate reality condition. We wish to thank the referee for bringing Sudarshan's paper to our attention.

Note added in Proof. Roman and Koh [Nuovo Cimento **39**, 1015 (1965)] have proved a theorem similar to our Theorem 2, without the restriction to finite-dimensional representations. However, they must assume that every vector of a definite weight in the representation also has a definite mass (is an eigenvector of P^2).

PROOF OF THEOREM 1

We begin with

$$G = S \dashv P,$$

where $\dashv$ denotes semi-direct sum with S as invariant subalgebra. We suppose that S is the Lie algebra of a compact group $\mathcal{S}$, which we can assume

[6] E. C. G. Sudarshan, J. Math. Phys. **6**, 1329 (1965). In the Corollary it is assumed, but not explicitly stated, that S is compact.

to be connected. Since S is invariant, the transformation

$$\exp s \rightarrow (\exp x)\ \exp s(\exp x)^{-1}, \qquad x \in P, s \in S$$

is an automorphism of $\mathfrak{S}$, which is connected to the identity automorphism ($x = 0$) by

$$\exp s \rightarrow (\exp \alpha x)\ \exp s(\exp \alpha x)^{-1},$$

where α runs between 0 and 1. But the set of automorphisms of a connected compact group $\mathfrak{S}$ which are connected to the identity automorphism is the group of *inner* automorphisms induced by elements of $\mathfrak{S}$.[7] Thus, given x in P, there is an element $t(x)$ in S such that for all s in S,

$$(\exp x)\ \exp s(\exp x)^{-1} = [\exp t(x)]\ \exp s[\exp t(x)]^{-1}. \tag{2}$$

The infinitesimal form of (2) is that for all s in S

$$[x, s] = [t(x), s]. \tag{3}$$

[If S has a nontrivial center, then $t(x)$ is not defined by (3). But since S is the Lie algebra of a compact group, S can be written as the direct sum of its center and of a semi-simple algebra.[8] We take $t(x)$ to be within the semi-simple algebra; this defines it uniquely.]

The $t(x)$ form a representation of P since

$$\begin{aligned}[t([x, y]), s] = [[x, y], s] &= [x, [y, s]] + [y, [s, x]] \\ &= [t(x), [y, s]] + [t(y), [s, x]] \\ &= [t(x), [t(y), s]] - [t(y), [t(x), s]] \\ &= [[t(x), t(y)], s].\end{aligned}$$

Thus

$$t([x, y]) = [t(x), t(y)]. \tag{4}$$

Also from (3), $x - t(x)$ commutes with S. Thus, if we define the Lie algebra P' with generators

$$x' = x - t(x), \tag{5}$$

then

$$G = S \oplus P'; \tag{6}$$

P' is isomorphic to P, since using (3), (4), and (5) we find

$$\begin{aligned}[x', y'] &= [x - t(x), y - t(y)] \\ &= [x, y] - [x, t(y)] - [t(x), y] + [t(x), t(y)] \\ &= [x, y] - [t(x), t(y)] - [t(x), t(y)] \\ &\qquad + [t(x), t(y)] \\ &= [x, y] - t([x, y]) \\ &= [x, y]'.\end{aligned}$$

If in (5), $t(x)$ vanishes for all x, then $P \equiv P'$ and from (6) there is no mass splitting. Suppose there exists an x_0 such that $t(x_0)$ does not vanish. Then there is a finite-dimensional representation of the group $\mathfrak{S}$ where the generator $t(x_0)$ is not represented by the zero matrix.[9] The representation may be assumed unitary since $\mathfrak{S}$ is compact.[10] Then we have a finite-dimensional unitary representation of P where not every element is represented by the unit matrix. But this is impossible, since P has no nontrivial finite-dimensional unitary representations.[11] Thus if $\mathfrak{S}$ is compact, $P \equiv P'$ and there is no mass splitting.

Note that the proof used only the compactness of the Lie group $\mathfrak{S}$ and that P had no nontrivial finite-dimensional unitary representations.

PROOF OF THEOREM 2

For a semisimple group, the derivation leading to (6) is valid. Indeed, the mapping

$$s \rightarrow x(s) \equiv [x, s]$$

of S into itself satisfies

$$x([a, b]) = [x(a), b] + [a, x(b)].$$

Such a map is called a derivation, and Cartan has shown that all derivations of a semi-simple group are inner.[12] Thus, given x in P, there exists a $t(x)$ in S such that (3) is satisfied for all s in S. Also note that S has no center, and so in (3), $t(x)$ is uniquely defined. (5) becomes

$$\begin{aligned}P'_\mu &= P_\mu - t(P_\mu), \\ M'_{\mu\nu} &= M_{\mu\nu} - t(M_{\mu\nu}),\end{aligned} \tag{7}$$

where $M_{\mu\nu}$ are the generators of the homogeneous Lorentz transformations, and $t(P_\mu)$, $t(M_{\mu\nu})$ are elements of S, which form a representation of P. We denote $t(P_\mu)$ by Q_μ, $t(M_{\mu\nu})$ by $N_{\mu\nu}$.

[7] K. Iwasawa, Ann. Math. **50**, 507 (1949); see especially p. 514.
[8] L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, New Jersey, 1939), Theorem 86.
[9] Ref. 8, Theorem 28.
[10] Ref. 8, Theorem 23.
[11] E. Wigner, Ann. Math. **40**, 149 (1939).
[12] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962), p. 74.

Consider a finite-dimensional irreducible representation of S. We denote the matrices representing Q_μ, $N_{\mu\nu}$ by $D(Q_\mu)$, $D(N_{\mu\nu})$. These form a finite-dimensional representation of P and so $D(Q_\mu)$ is a nilpotent matrix.[13] In triangular form, $D(Q_\mu)$ looks like

$$D(Q_\mu) = \begin{bmatrix} 0 & x & x & x & \cdots & x \\ & 0 & 0 & x & x & \cdot \\ & & 0 & 0 & 0 & x & \cdot \\ & & & \ddots & & & \cdot \\ & & & & & & x \\ & & & & & & 0 \end{bmatrix}. \tag{8}$$

Since for different μ, the $D(Q_\mu)$ commute, they can all be triangularized simultaneously, and all have zeros along the main diagonal.

Suppose the basis in which $D(Q_\mu)$ assumes the form (8) is denoted by ϕ_m, $m = 1, 2, \cdots n$, where n is the dimension of the representation. Then $\phi_m(p', s')$ form a basis for the representation of G. The p' satisfy

$$p'_\mu p'^\mu = \text{const} = M'^2. \tag{9}$$

[13] See, for example, Ref. 2, Eqs. (9)–(12).

The effect of P'_μ on $\phi_m(p', s')$ is given by

$$P'_\mu \phi_m(p', s') = p'_\mu \phi_m(p', s'). \tag{10}$$

Thus from (7), P_μ applied to $\phi_m(p', s')$ is a linear combination of states with the same p', s'. Suppressing the p', s' variables, we have that P_μ is represented by the matrix

$$D(P_\mu) = p'_\mu \mathbf{1} + D(Q_\mu),$$

where $\mathbf{1}$ is the unit matrix in the representation space of S. Using (8), $D(P_\mu)D(P^\mu)$ becomes

$$D(M^2) = \begin{bmatrix} M'^2 & x & x & \cdots & x \\ 0 & M'^2 & x & & x \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & & M'^2 \end{bmatrix},$$

and so all the eigenvalues of $D(M^2)$ are M'^2, which establishes Theorem 2.

ACKNOWLEDGMENTS

I wish to thank Professor M. A. Melvin and Professor E. P. Wigner for helpful advice.

A Method of Solution for Resonant Nonlinear Coupled Oscillator Systems*

JOHN WATERS
University of Illinois, Urbana, Illinois† and
School of Physics, Georgia Institute of Technology, Atlanta, Georgia
AND
JOSEPH FORD
School of Physics, Georgia Institute of Technology, Atlanta, Georgia
(Received 22 June 1965)

A method of solution for coupled oscillator systems having commensurable frequencies $\omega_k = k\omega$ and suitable nonlinear couplings is presented. The method is that of perturbation expansions about periodic solutions. For simplicity, a three-oscillator system is discussed, but larger systems may be treated in the same way. The method provides results which are considerably better than those obtainable by any other currently available method, such as the perturbation method of Wigner–Brillouin. As presented, the method is rather unrefined, but it demonstrates that nonlinear oscillator systems of this resonant type are susceptible to analysis.

I. INTRODUCTION

SYSTEMS of nonlinear coupled oscillators find extensive application as mathematical models for a variety of physical systems, in fields ranging from molecular chemistry[1] and statistical mechanics[2] to astronomy.[3] In particular, the class of cubic-coupled oscillator systems in which the uncoupled frequencies are linearly commensurable, or nearly so, is of special interest because of widespread energy sharing in such systems.[4] But it is just this resonant class of systems for which the usual perturbation methods encounter serious difficulties[5] characterized by the appearance of terms having small denominators in all orders of the expansions.

This paper presents a new method of solution by perturbation expansions about periodic solutions, which directly applies to the extreme case of a nonlinear coupled oscillator system in which the uncoupled frequencies are exactly commensurable. The results obtained in first order by this method are considerably better than those obtainable by any other presently available method. Although in the present formulation, difficulties are encountered in higher orders due to the appearance of terms having small denominators, it is likely that this situation can be remedied.

For the sake of clarity, the method of solution is presented in terms of its application to a specific simple example, that of a system of three oscillators. Detailed calculations have also been carried out for a system of five oscillators; the essential results are the same. In principle, there is no limit to the size of the system which can be treated by this method. However, the practical limit is probably 10 or 15 oscillators, because of the computational labor involved.

Section II presents the system to be analyzed and provides a brief discussion of the difficulties encountered in the application of a familiar perturbation method, that of Wigner–Brillouin. In Sec. III, the existence of three periodic solutions for the three-oscillator system is established, and a first-order perturbation expansion about a periodic solution is introduced. In Sec. IV, the results of this first-order approximation are compared with the results of numerical integration of the equations of motion, for several appropriate sets of initial conditions. Finally, the difficulties encountered in making a second-order expansion are described in Sec. V, and some conclusions are offered in Sec. VI.

II. SYSTEM TO BE ANALYZED

Consider the three-oscillator system governed by the Hamiltonian

$$H = \tfrac{1}{2}(p_1^2 + p_2^2 + p_3^2) + \tfrac{1}{2}(\omega^2 q_1^2 + 4\omega^2 q_2^2 + 9\omega^2 q_3^2) + \alpha\omega^{-2}(\tfrac{1}{8}\sqrt{2})[(\sqrt{2} - 1)(\omega^2 q_1 q_1 q_2 - p_1 p_1 q_2 + q_1 p_1 p_2) + \tfrac{1}{8}(6\omega^2 q_1 q_2 q_3 + q_1 p_2 p_3 + 2p_1 q_2 p_3 - 3p_1 p_2 q_3)] \quad (1)$$

in which the uncoupled oscillator frequencies are exactly commensurable:

$$\omega_k = k\omega, \qquad k = 1, 2, 3.$$

* Work partly supported by the National Science Foundation. Portions of this work were presented in partial fulfillment of the requirements for the Ph.D. degree of J.W. at the Georgia Institute of Technology.
† Present address: Department of Computer Science.
[1] D. Bunker, J. Chem. Phys. 37, 393 (1962).
[2] R. Northcote and R. Potts, J. Math. Phys. 5, 383 (1964).
[3] M. Hénon and C. Heiles, Astron. J. 69, 73 (1964); G. Contopoulos, *ibid.* 68, 1, 763 (1963).
[4] J. Ford and J. Waters, J. Math. Phys. 4, 1293 (1963).
[5] E. A. Jackson, J. Math. Phys. 4, 551, 686 (1963).

This Hamiltonian is sufficient to illustrate the advantages of our methods over previous techniques. Hamiltonian (1) is an approximation to a more physically interesting[6] Hamiltonian

$$H = \tfrac{1}{2}(p_1^2 + p_2^2 + p_3^2) + \tfrac{1}{2}(\omega^2 q_1^2 + 4\omega^2 q_2^2 + 9\omega^2 q_3^2) + \alpha(\tfrac{1}{2}\sqrt{2})[(\sqrt{2} - 1)q_1 q_1 q_2 - (\sqrt{2} + 1)q_1 q_3 q_3 + 2q_1 q_2 q_3]. \quad (2)$$

Hamiltonians (1) and (2) are related by a theorem due to Birkhoff and Moser.[7] This theorem establishes the existence of a nonlinear canonical transformation which identifies the crucial resonant terms in Hamiltonian (2). Hamiltonian (1) was obtained from Hamiltonian (2) by retaining only resonant terms through third degree.

The application of a familiar perturbation scheme such as the Wigner–Brillouin method[4,5] to the system governed by Hamiltonian (1) results in the appearance of terms in the solution having small denominators in all orders, including first order. If truncated at any low order, such a result provides a very poor approximation to the general solution of the system, because terms of all orders need to be summed.

For nonresonant oscillator systems in which the uncoupled frequencies are not commensurable, it is possible in the Wigner–Brillouin method to eliminate those "self-resonant" terms having small denominators, which occur in even orders. But in the present case, it is not clear how the method can be adapted in general to provide elimination of the added "coupled-resonant" terms, which have small denominators because of commensurability of the uncoupled frequencies. The Wigner–Brillouin method is apparently useful for systems of the type (1) only in special situations in which the perturbed frequencies are also commensurable, leading to periodic solutions which are discussed in the following section.

III. METHOD OF SOLUTION

The equations of motion for the system governed by Hamiltonian (1) are

$$\dot{q}_1 = p_1 + C[(\sqrt{2} - 1)(-2p_1 q_2 + q_1 p_2) + \tfrac{1}{3}(2q_2 p_3 - 3p_2 q_3)]; \quad (3a)$$

$$\dot{p}_1 = -\omega^2 q_1 - C[(\sqrt{2} - 1)(2\omega^2 q_1 q_2 + p_1 p_2) + \tfrac{1}{3}(6\omega^2 q_2 q_3 + p_2 p_3)]; \quad (3b)$$

$$\dot{q}_2 = p_2 + C[(\sqrt{2} - 1)(q_1 p_1) + \tfrac{1}{3}(q_1 p_3 - 3p_1 q_3)]; \quad (3c)$$

$$\dot{p}_2 = -4\omega^2 q_2 - C[(\sqrt{2} - 1)(\omega^2 q_1 q_1 - p_1 p_1) + \tfrac{1}{3}(6\omega^2 q_1 q_3 + 2p_1 p_3)]; \quad (3d)$$

$$\dot{q}_3 = p_3 + \tfrac{1}{3}C(q_1 p_2 + 2p_1 q_2); \quad (3e)$$

$$\dot{p}_3 = -9\omega^2 q_3 - C(2\omega^2 q_1 q_2 - p_1 p_2), \quad (3f)$$

where $C = \frac{1}{8}\alpha\omega^{-2}\sqrt{2}$.

It has been discovered[4] that a set of three exact periodic solutions to Eqs. (3) can be found by assuming a trial solution of the form

$$q_k = A_k \cos k\tau;$$

$$p_k = -kA_k \sin k\tau,$$

where $\tau = (\omega + \alpha\omega^{-1}b)t$ and $k = 1, 2, 3$. The sets of approximate numerical parameters which characterize the three periodic solutions are given in Table I, where A_0 is an amplitude which is determined by the initial conditions. It has been verified that these periodic solutions are stable over a large number of oscillations, by numerical integration of Eqs. (3) starting from the three sets of initial conditions given in Table I.

Given the existence of a number of periodic solutions of the system, it is natural to attempt to obtain approximations to the general solution of the system in terms of expansions about these solutions. That is, assume a trial solution of the form

$$q_k = A_0\{(A_{k0} + \beta_1 B_{k1} + \beta_2 B_{k2}) \cos k\tau + \beta_1[A_{k1} \cos (k\tau + \tau_1) + A_{k-1} \cos (k\tau - \tau_1)] + \beta_2[A_{k2} \cos (k\tau + \tau_2) + A_{k-2} \cos (k\tau - \tau_2)]\}; \quad (4a)$$

$$p_k = -k\omega A_0\{(A_{k0} + \beta_1 B_{k1} + \beta_2 B_{k2}) \sin k\tau + \beta_1[A_{k1} \sin (k\tau + \tau_1) + A_{k-1} \sin (k\tau - \tau_1)] + \beta_2[A_{k2} \sin (k\tau + \tau_2) + A_{k-2} \sin (k\tau - \tau_2)]\}, \quad (4b)$$

where $\tau = [\omega + \alpha\omega^{-1}(b_0 + \beta_1 b_1 + \beta_2 b_2)]t + \theta$ and $\tau_i = \alpha\omega^{-1}c_i t + \theta_i$, and where $k = 1, 2, 3$. This is a trigonometric series expansion in powers of dimensionless expansion parameters β_1 and β_2, in which

TABLE I. Coefficients for three periodic solutions.

Coefficient	First solution	Second solution	Third solution
b	$0.07834A_0$	0	$-0.07834A_0$
A_1	$0.91364A_0$	$0.97922A_0$	$0.91364A_0$
A_2	$0.36959A_0$	0	$-0.36959A_0$
A_3	$0.16932A_0$	$-0.20280A_0$	$0.16932A_0$

[6] See Ref. 4, Eq. (13).

[7] G. D. Birkhoff, *Dynamical Systems* (American Mathematical Society Colloquium Publications, New York, 1927), p. 82; J. Moser, Commun. Pure Appl. Math. 11, 81 (1958), Lemma 4.

only zeroth- and first-order terms have been retained. Amplitude parameter A_0, expansion parameters β_1 and β_2, and phase angles θ, θ_1, and θ_2 are determined by the six initial conditions in a given situation. The substitution of Eqs. (4) into Eqs. (3) reduces the differential equations of motion to a set of algebraic equations which determine the A's, B's, b's, and c's of Eqs. (4) for all sets of initial conditions.

For the general case of a system of N oscillators, a similar trial solution having $N-1$ expansion parameters β_j would be employed.

In zeroth order, the result of substituting the assumed solution, Eqs. (4), into the equations of motion, Eqs. (3), is the same as when the periodic solutions were found. The first-order coefficients depend upon which periodic solution is chosen for a zeroth-order solution. Substitution of a zeroth-order solution into the algebraic equations and setting $B_{k1} = 0$ and $B_{k2} = 0$ produces a linear eigenvalue problem, from which the values of the frequency components c_j may be determined, along with the values of the coefficients A_{kj} for $j \neq 0$. The above choice of the B_{kj} causes the first-order frequency corrections b_1 and b_2 to vanish.

For the sets of initial conditions which result in periodic solutions, the expansion parameters β_1 and β_2 are zero. For a set of initial conditions in a small neighborhood of those of a periodic solution, the expansion parameters are small, and the first-order expressions of Eqs. (4) should provide a useful approximation to the general solution in this region. Some indication of the size of these neighborhoods relative to the amount of error involved in the first-order expansions is given in the next section.

IV. COMPARISON WITH NUMERICAL SOLUTIONS

In order to roughly determine the region of validity of the first-order approximate solutions for the particular three-oscillator system being considered, an approximate solution may be compared with the corresponding numerical solution of the equations of motion for several selected sets of initial conditions.

Numerical integration of the equations of motion has been carried out using a standard fourth-order Runge–Kutta method with a step size of 0.05. Certain checks, including a running calculation of the Hamiltonian and the reversal of several runs back to the initial conditions, indicate that the accuracy of the numerical solutions presented is better than $\pm 0.1\%$.

In the comparisons which follow, rather than plotting positions or momenta versus time, graphs of

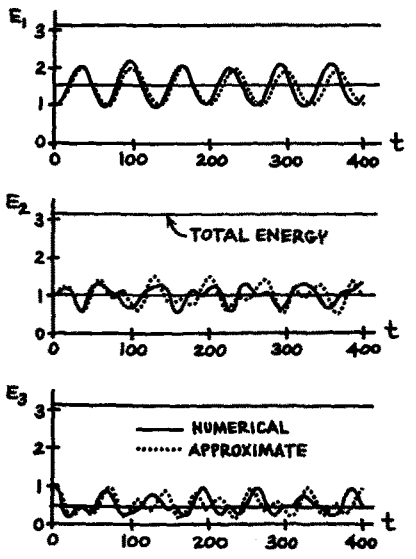

FIG. 1. Comparison of numerical and approximate solutions for initial conditions close to those of the first periodic solution.

energies,

$$E_k = \tfrac{1}{2}p_k^2 + \tfrac{1}{2}k^2\omega^2 q_k^2,$$

versus time are presented. This best displays the differences in long-term behavior of the various solutions. The long-term characteristics, dependent upon the nonlinear coupling and the uncoupled frequencies, are the most difficult to approximate for systems of the present type.

In Fig. 1, the first-order approximate solution is compared with the numerical solution for a set of initial conditions which are rather close to those of the first periodic solution, the first column of Table I. The constant-energy lines in Fig. 1 are those of this periodic solution. The expansion parameters are $\beta_1 = -0.15$ and $\beta_2 = -0.06$, so that a first-order expansion about this periodic solution is expected to provide a rather good approximation to the actual solution. Figure 1 confirms that it does.

Graphical presentations such as Fig. 1 are adequate for comparing multiperiodic functions which are dominated by one component, as is E_1. But when several components have comparable amplitudes, and there are errors in the determination of the frequencies and amplitudes of each, the over-all error becomes difficult to estimate, as in the case of the plots of E_2 and E_3 versus time. In these cases, the numerical solution may be harmonic-analyzed and its amplitude and frequency components compared with those of the first-order energy approximation. A sample of such a comparison is given in Table II for E_2 of Fig. 1.

TABLE II. Harmonic content of energy E_2 in Fig. 1.

Numerical solution		Approximate solution	
Frequency	Amplitude	Frequency	Amplitude
0.000	0.996	0.000	1.001
0.097	0.273	0.095	0.273
0.194	−0.127	0.189	−0.019
0.165	−0.101	0.172	−0.223
0.068	−0.058	0.077	−0.042

Since the first-order approximate solution for a situation close to a periodic solution is rather good, it is of interest to investigate how good the first-order approximation is for a more extreme situation. In the case of a two-oscillator system with a similar nonlinear coupling, the solution of which can be obtained exactly in terms of elliptic functions,[8] the system configuration which is most extreme from that of the periodic solutions is the one in which one of the oscillators initially has all of the energy. The same class of initial conditions is expected to be most extreme in this sense for larger systems.

Figure 2 compares the first-order approximation with the numerical solution for the situation in which all of the energy is initially given to the first oscillator; Figure 3 provides the same comparison for the case in which the third oscillator starts with most of the energy. The approximate solution presented in Fig. 2 is in error by about 10% in both the frequencies and the amplitudes. That shown in Fig. 3 is not as good; it is in error by about 20% in the frequencies and by perhaps 50% in the amplitudes.

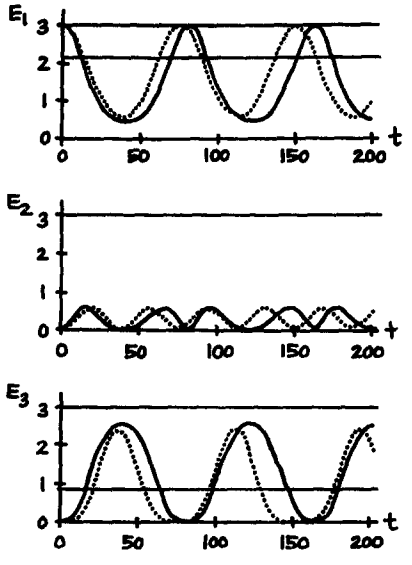

FIG. 2. Comparison of numerical (solid curves) and approximate (dotted curves) for extreme initial conditions. All energy was initially given to the first oscillator; the perturbation expansion is about the second periodic solution.

[8] B. Baker and E. Ross, Proc. Edinburgh Math. Soc. 39, 34 (1921).

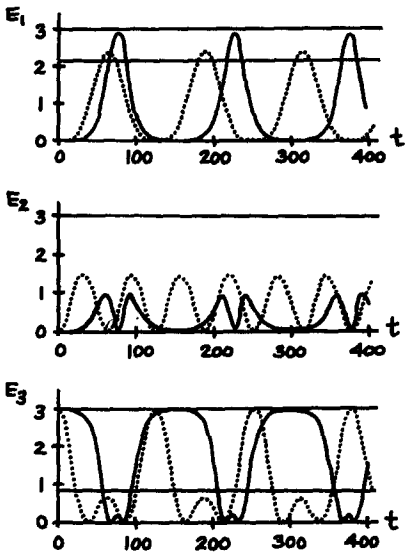

FIG. 3. Similar to Fig. 2, but most of energy was initially given to the third oscillator.

This brief survey suggests the likelihood that the first-order approximation to the general solution of this three-oscillator system is reasonably good for most sets of initial conditions, providing that the expansions are made about the appropriate periodic solutions in each case. The same conclusion may be made for a nonlinear five-oscillator system which has been similarly investigated.

V. HIGHER-ORDER APPROXIMATIONS

The difficulty of the appearance of terms having small denominators occurs in the present scheme in second and higher orders, just as it occurred in the Wigner–Brillouin perturbation method in first and higher orders when applied to a system having commensurable uncoupled frequencies.

If second-order terms of all possible kinds are added in a natural manner to the first-order expression for the assumed solution, Eqs. (4), it is found upon numerical solution that the amplitude coefficients of terms such as $\cos(k\tau + \tau_i - \tau_j)$ in which τ_i and τ_j are equal or very nearly equal, are quite large.

In the interest of convergence, it is necessary to prevent the appearance of such terms in the solution. The manner in which this should be done is not immediately clear. There are numerous arbitrary coefficients, such as the B_{kj} in Eqs. (4) and similar second-order coefficients, which have been

set zero for computational simplicity and for lack of a better choice. It is likely that these coefficients can be chosen in such a manner as to prevent the appearance of troublesome terms in higher orders, as is done in other perturbation schemes; however, our investigation of this point is incomplete at present.

VI. CONCLUSIONS

The method of solution by perturbation expansions about periodic solutions, which has been described here in terms of its application to a three-oscillator system, needs improvement with regard to the calculation of higher-order terms. Its strength lies in the fact that it can provide first-order results which are considerably better than those obtainable by other methods, principally because the problem of the appearance of terms having small denominators does not occur until second order in this method.

This method has been presented in order to demonstrate that nonlinear coupled oscillator systems of the type considered, having exactly commensurable uncoupled frequencies, are susceptible to analysis. It is hoped that this preliminary investigation will lead to the development of more practical and rigorous methods of analysis for these systems.

Model of Interacting Radiation and Matter. III. Multimode Gas Lasers*

CHARLES R. WILLIS
Boston University, Boston, Massachusetts
(Received 13 May 1965; final manuscript received 1 August 1965)

We extend the investigation of the long-time behavior of a model consisting of N two-level atoms interacting with a single electromagnetic cavity mode to include interaction with many cavity modes. We show that, as a consequence of the coupling between radiation modes produced by spatial density variations of the population inversion, there is no strictly stationary state possible for multimode behavior. However we obtain a stationary state by neglecting the rapidly oscillating terms. The steady-state population inversion is then a solution of an eigenvalue problem. The eigenvalue determinate is a function of the number of modes and the coupling between modes in addition to the usual dependence on frequencies and relaxation times. We explicitly solve for the unique eigenvalue in special cases. The corresponding eigenvector gives the steady-state mode intensity ratios. The absolute values of the steady-state intensities are determined by the energy conservation equation generalized to include pumping and dissipation. We also calculate the steady-state frequency shifts for each mode. The mode frequency shifts are practically independent of each other and have the same functional form as the single-mode frequency shifts.

I. INTRODUCTION

IN this paper we extend our model of interacting radiation and matter for single-mode gaseous lasers[1,2] to include multimode gaseous lasers. In I we showed that the lowest-order solution to our model of N stationary two-level systems in the presence of radiation is the self-consistent field approximation which we refer to as SCFA. In the absence of center-of-mass motion we found there is only one dimensionless dynamical constant in the theory, namely

$$\tilde{\gamma}N\tilde{\gamma} = \mathfrak{N}e^2 \, |\langle a| \, \boldsymbol{\varepsilon}\cdot\mathbf{r} \, |b\rangle|^2 \, (2\pi/\hbar\Omega)$$
$$= (4\pi)^{-1}\mathfrak{N}r_0\lambda^2 \equiv \alpha^2,$$

where $\mathfrak{N}$ is the number of two-level systems per unit volume, $\langle a| \, \boldsymbol{\varepsilon}\cdot\mathbf{r} \, |b\rangle$ is the dipole matrix element between the two levels of the atom, Ω is the cavity frequency, and λ is the wavelength of the radiation.

When we generalized our model to include dissipation, pumping, and the center-of-mass motion, we obtained only one new dimensionless dynamical constant, namely $\beta \equiv \alpha(\omega_0/\omega_D)$ where ω_0 is the frequency separation of the two atomic levels and ω_D is the Doppler width. The Doppler width is $[k^2(k_BT)m^{-1}]^{\frac{1}{2}}$ where k is the wavenumber of the radiation, T is the temperature, k_B is Boltzmann's constant, and m is the mass of the two-level system. In II we showed that when β is small compared with one, the electromagnetic field amplitudes vary adiabatically slowly compared with the center-of-mass motion. Since β is much less than one in gas lasers, the electromagnetic field amplitudes see mainly certain integrals of the velocity distribution, not the microscopic details of the center-of-mass motion. Consequently, for small α and β we solved the nonlinear equations for a single mode by a generalization of the Bogoliuboff–Kryloff[3] quasi-linear solution for nonlinear equations. We showed there is a unique stationary state and it is approached very rapidly. The theory is nonperturbative and, as long as α and β are small compared with one, there are no limits placed on the amplitude of the electromagnetic field.

When we try to extend the method for single-mode behavior in II to the problem of multimode behavior, a fundamental difficulty arises. As a consequence of the coupling between the modes produced by density variations, there is no strictly stationary state possible for multimode behavior. The reason for this is that the exact differential equations have time-dependent coefficients that depend on beat frequencies between the modes.

We find a unique steady state by dropping the small contribution from the rapidly oscillating terms. The solution for the steady state becomes an eigenvalue problem where the steady-state population inversion is the eigenvalue, and the mode intensities are eigenvectors.

* The research reported in this paper was sponsored in part by the U. S. Air Force Cambridge Research Laboratories, Office of Aerospace Research.

[1] C. R. Willis, J. Math. Phys. 5, 1241 (1964); hereafter referred to as I.

[2] C. R. Willis, J. Math. Phys. 6, 1984 (1965), hereafter referred to as II.

[3] N. Kryloff and N. Bogoliuboff, *Introduction to Nonlinear Mechanics*, (Princeton University Press, Princeton, New Jersey, 1947).

Many of our results are qualitatively similar to Lamb's.[4] The differences arise because Lamb includes center-of-mass–internal variable correlations and restricts his equations to near threshold. We on the other hand have no restriction on the magnitude of the field but neglect center-of-mass–internal correlations.

In Sec. II we introduce the Hamiltonian of our system and derive the equations of motion in the presence of dissipation and a pump. We then formally eliminate the matter dipole operators in the equations of motion for the electromagnetic field variables.

In Sec. III we introduce the SCFA and the adiabatic approximation in the equations of motion for the electromagnetic field variables. Although our treatment of the Hamiltonian, SCFA, and adiabatic approximation is self-contained, we refer the reader to I and II for a more thorough treatment.

We find the stationary state in Sec. IV and we solve the three-mode problem exactly.

We calculate the steady-state frequency shift in Sec. V and show that in the absence of noise the frequency shifts of the individual modes are only weakly dependent on each other. In Sec. VI we compare our results with Lamb's theory.[4]

II. HAMILTONIAN AND EQUATIONS OF MOTION

Our Hamiltonian for N two-level systems interacting with the electromagnetic field is

$$H_N = h(N) + H_{cm} + H_f + H_i, \tag{2.1}$$

where

$$h(N) = \frac{\hbar\omega_0}{2}\sum_{\alpha}^{N}\hat{\sigma}_\alpha, \qquad H_f = \sum_k \hbar\Omega_k(a_k^\dagger a_k + \tfrac{1}{2})$$

$$H_i = \hbar\omega_0 \sum_\alpha \sum_k \tilde{\gamma}_k\Gamma_k(X_\alpha)\{a_k^\dagger\sigma_\alpha + a_k\sigma_\alpha^\dagger\}$$

$$H_{cm} = \sum^{N}\frac{P_\alpha^2}{2M} + \frac{1}{2}\sum_\alpha\sum_\beta V(X_\alpha - X_\beta) + \sum_\alpha^N\sum_i^M U(X_\alpha - \eta_i).$$

The $a_k^\dagger$ and a_k are the usual creation and annihilation operators for the electromagnetic field in the kth cavity mode. The operators for the internal degrees of freedom of the αth atom are

$$\sigma_\alpha^\dagger = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \quad \sigma_\alpha = \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \quad \hat{\sigma}_\alpha = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

We drop the nonresonant terms from H_i as we did in II. The definitions of $\tilde{\gamma}_k$ and $\Gamma_k(X_\alpha)$ are

$$\tilde{\gamma}_k \equiv (\hbar\omega_0)^{-1}(\hbar\Omega_k)^{\frac{1}{2}}e\langle a|\,\boldsymbol{\varepsilon}_k\cdot\mathbf{x}\,|b\rangle(4\pi/V) \approx \tilde{\gamma}$$

$$\Gamma_k(X_\alpha) = E_k(X_\alpha)V^{\frac{1}{2}},$$

where the $E_k(X_\alpha)$ are the normalized eigenfunctions of the cavity, V is the volume and $\langle a|\,\boldsymbol{\varepsilon}_k\cdot\mathbf{x}\,|b\rangle$ is the matrix element of the dipole moment between the two relevant atomic wavefunctions, $\psi_a(\mathbf{x})$ and $\psi_b(\mathbf{x})$. Since the k dependence of $\tilde{\gamma}_k$ is negligible, we replace $\tilde{\gamma}_k$ by $\tilde{\gamma}$. We treat the center of mass classically as N atoms interacting with each other and with M pump atoms through two-body forces, $V(X_\alpha - X_\beta)$ and $U(X_\alpha - \eta_i)$ respectively. The X_α is the coordinate of the center of mass of the αth two-level system and η_i is the coordinate of the center of mass of the ith pump atom.

When we use Eq. (2.1) and the commutation relations, we obtain the following equations of motion for the radiation and matter operators

$$i\dot{a}_k - \Omega_k a_k + i(a_k/T_k) = \tilde{\gamma}\omega_0\sum_\alpha\sigma_\alpha\Gamma_k(X_\alpha), \tag{2.2a}$$

$$i\dot{\sigma}_\alpha - \omega_0\sigma_\alpha + i(\sigma_\alpha/T_2) = -\tilde{\gamma}\omega_0\sum_k a_k\hat{\sigma}_\alpha\Gamma_k(X_\alpha), \tag{2.2b}$$

$$i\dot{a}_k^\dagger + \Omega_k a_k + i(a_k^\dagger/T_k) = -\tilde{\gamma}\omega_0\sum_\alpha\sigma_\alpha^\dagger\Gamma_k(X_\alpha), \tag{2.2c}$$

$$i\dot{\sigma}_\alpha^\dagger + \omega_0\sigma_\alpha^\dagger + i(\sigma_\alpha^\dagger/T_2) = \tilde{\gamma}\omega_0\sum_k a_k^\dagger\hat{\sigma}_\alpha\Gamma_k(X_\alpha). \tag{2.2d}$$

We have introduced the phenomenological decay constants, T_2 and T_k, where T_2 represents the relaxation of the off-diagonal matrix elements of the matter density matrix and T_k is the radiation relaxation time for the matter-free cavity. The term T_k is usually written in the form $(\nu_k/2Q_k)^{-1}$ where Q_k is the Q of the kth mode.

When we rigorously eliminate the operators σ_α and $\sigma_\alpha^\dagger$ from Eqs. (2.2a), (2.2b), (2.2c), and (2.2d) by the same method we used in II, Sec. II, we obtain

$$\dot{b}_k + \frac{b_k}{T_k} = \tilde{\gamma}^2\omega_0^2\sum_\alpha\sum_{k'}e^{id_{kk'}t}\int_0^t\Gamma_k(\alpha,t)\Gamma_{k'}(\alpha,t')\times\{\hat{\sigma}_\alpha(t')b_{k'}(t')e^{-\nu_2(t-t')}e^{-i\Delta(k')(t-t')}\}\,dt', \tag{2.3a}$$

$$\dot{b}_k^\dagger + \frac{b_k^\dagger}{T_k} = \tilde{\gamma}^2\omega_0^2\sum_\alpha\sum_{k'}e^{-id_{kk'}t}\int_0^t\Gamma_k(\alpha,t)\Gamma_k(\alpha,t')\times\{\hat{\sigma}_\alpha(t')b_{k'}^\dagger(t')e^{-\nu_2(t-t')}e^{i\Delta(k')(t-t')}\}\,dt', \tag{2.3b}$$

where

$$\Delta(k) \equiv \omega_0 - \Omega_k, \qquad \nu_2 \equiv T_2^{-1}, \qquad d_{kk'} \equiv \Omega_k - \Omega_{k'}$$

$$b_k \equiv a_k e^{i\Omega_k t}, \qquad \Gamma_k(\alpha,t) \equiv \Gamma_k[X_\alpha(t)].$$

[4] W. E. Lamb, Phys. Rev. 134, A1429 (1964).

We note that Eq. (2.3b) is just the complex conjugate of Eq. (2.3a).

The equation of motion for $\hat{\sigma}_\alpha$ is

$$i\,\partial\hat{\sigma}_\alpha/\partial t + i(\hat{\sigma}_\alpha - \hat{\sigma}_0)/T_1 = 2\tilde{\gamma}\omega_0 \sum_k \{\Gamma_k(X_\alpha)\sigma_\alpha^\dagger a_k - \Gamma_k(X_\alpha)\sigma_\alpha a_k^\dagger\}, \quad (2.4)$$

where we add the combined dissipation and pumping terms. The constant T_1 is the relaxation time for the diagonal matrix elements of the density matrix and $N\hat{\sigma}_0$ is the population difference produced by the pump in the absence of radiation. If we have N permanent two-level systems, $N\hat{\sigma}_0$ can take any value from $-N$ to N, depending on the pump. However, when both levels of the two-level system are excited states of an atom, $N\hat{\sigma}_0$ really defines the number of two-level systems. Consequently, when we do not have permanent two-level systems, we take $\hat{\sigma}_0 = 1$. When we eliminate $\sigma_\alpha^\dagger$ and σ_α from Eq. (2.4) with the help of Eqs. (2.2a) and (2.2c), we obtain

$$\frac{\partial}{\partial t}\{\sum_\alpha \hat{\sigma}_\alpha + 2\sum_k b_k^\dagger b_k\} = -4\sum_k \frac{b_k^\dagger b_k}{T_k} - \frac{1}{T_1}\sum_\alpha (\hat{\sigma}_\alpha - \hat{\sigma}_0). \quad (2.5)$$

In the absence of dissipation, Eq. (2.5) is essentially the statement that the energy of the matter plus radiation is conserved. Equation (2.5) is exact, involves no approximations, and is rigorously independent of the center-of-mass coordinates. Equations (2.3a), (2.3b), and (2.5) are as far as we can go with an exact solution. In the next section we introduce an approximation procedure that allows us to solve the equations.

III. SELF-CONSISTENT FIELD APPROXIMATION

We obtain the SCFA solution by taking the trace of Eq. (2.3a) with a density matrix which consists of a product of N one-particle density matrices, multiplied by a product of density matrices one for each radiation oscillator, which in turn is multiplied by N one-particle center of mass distribution functions. The result is

$$\langle\dot{b}_k\rangle + \langle b_k\rangle/T_k = \omega_L^2 \sum_{k'} e^{id_{kk'}t} \times \int^t \langle\Gamma_k(t)\Gamma_{k'}(t')\rangle e^{-\nu_2(t-t')} \times e^{-i\Delta(k')(t-t')}\langle\hat{\sigma}(t')\rangle\langle b_{k'}(t')\rangle\, dt', \quad (3.1)$$

where $\omega_L^2 \equiv \tilde{\gamma}N\tilde{\gamma}\omega_0^2$ and $\langle 0\rangle$ represents the trace of 0 with the density matrix.

Since the Doppler frequency is of the order of 10^2–10^3 times larger than the collision frequency in the He–Ne laser, the dominate contribution to $\langle\Gamma_k(t)\Gamma_{k'}(t')\rangle$ comes from the free-particle motion. We showed in the Appendix of II how to calculate $\langle\Gamma_k(t)\Gamma_{k'}(t')\rangle$ including collision effects. For free-particle motion we obtain

$$\langle\Gamma_k(t)\Gamma_{k'}(t')\rangle = \int \sin k[X - V(t-t')]\sin(k'X) f_1'(X, V)\, dX\, dV \approx \tfrac{1}{2}\exp[-\tfrac{1}{2}\omega_D^2(k)(t-t')^2]\{C(k-k') - C(k+k')\} \approx \tfrac{1}{2}C(k-k')\exp[-\tfrac{1}{2}\omega_D^2(k)(t-t')^2], \quad (3.2)$$

where the steady-state distribution function is

$$f_1'(X, V) = (2\pi k_B T/m)^{-\frac{1}{2}}\rho(X)\exp[-(mV^2/2k_BT)]$$

and where the cosine transform of the density of atoms is

$$C(k) \equiv \int_0^L \cos kX\rho(X)\, dX.$$

In deriving Eq. (3.2) we assume that the steady-state velocity distribution is Maxwellian. The Doppler frequency, $\omega_D^2(k) \equiv k^2(k_BT)m^{-1}$, is k dependent but, since the variation over all excited modes is less than one part in 10^{-5}, we replace $\omega_D(k)$ by ω_D, the value of the Doppler frequency at the line center. We neglect $C(k+k')$ compared with $C(k-k')$ because $C(k+k')$ represents the Fourier transform of the density variation at distances of the order of one-half the wavelength of light; on the other hand, $C(k-k')$ represents density variations over distances of the order of the size of the laser. The value of $(k-k')$ for adjacent cavity modes is π/L where L is the length of the cavity. The density variations on this scale are produced by surface effects and spatial variations in the pumping mechanism.

When Eq. (3.2) is substituted in Eq. (3.1) and the resultant integral is made dimensionless with ω_D, we obtain

$$\langle\dot{b}_k\rangle + \langle b_k\rangle/T_k = \tfrac{1}{2}(\omega_L\beta)\sum_{k'} C(k-k') \times \exp(id_{kk'}t)\int^{\omega_D t}\exp(-\tau^2/2) \times \exp(-\bar{\nu}_2\tau)\exp[-i\bar{\Delta}(k')\tau] \times \langle\hat{\sigma}(t-\tau)\rangle\langle b_{k'}(t-\tau)\rangle\, d\tau, \quad (3.3)$$

where $\beta \equiv \omega_L/\omega_D$, $\bar{\nu}_2 \equiv \nu_2/\omega_D$, and $\bar{\Delta}(k') = \Delta(k')/\omega_D$. In general, a frequency with a bar over it represents the dimensionless ratio of the frequency to the Doppler frequency.

In a gas laser β is much less than one, thus $\langle b_k \rangle$ and $\langle \hat{\sigma} \rangle$ are slowly varying. Consequently, to lowest order in β and for times $t > \omega_D^{-1}$, we can simplify Eq. (3.3) in the following manner:

$$\langle \dot{b}_k \rangle + \langle b_k \rangle / T_k = \omega_L \beta \sum_{k'} C(k - k') \times \exp(i d_{kk'} t) \langle b_{k'}(t) \rangle \langle \hat{\sigma}(t) \rangle F(k'), \quad (3.4)$$

where

$$F(k') \equiv \frac{1}{2} \int_0^\infty \exp(-\tfrac{1}{2}\tau^2) \times \exp(-\bar{\nu}_2 \tau) \exp[-i\bar{\Delta}(k')\tau] \, d\tau$$

$$\equiv F_c(k') + iF_s(k') \equiv \frac{1}{2} \int_0^\infty \exp(-\tau^2/2) \times \cos[\bar{\Delta}(k')\tau] \exp(-\bar{\nu}_2 \tau) \, d\tau + i \frac{1}{2} \int_0^\infty \exp(-\tau^2/2) \times \sin[\bar{\Delta}(k')\tau] \exp(-\bar{\nu}_2 \tau) \, d\tau.$$

If we have only a single mode, $d_{kk} = 0$, $C(0) = 1$, and Eq. (3.4) reduces to Eq. (4.1a) of II. In II we showed there is a unique stationary state for a single mode and the effect of the nonlinearity is to cause the system to approach this stationary state rapidly. However, when we have more than one mode present, $d_{kk'}$ is not zero and a strictly stationary state is impossible. From Eq. (3.4) we see that the adiabaticity parameter that measures the slow variation of the amplitudes in the multimode case is not β but $\beta |C|$ where $|C|$ is the magnitude of $C(k - k')$. We expect $|C|$ to be usually less than one; however, this is not necessarily true. If $|C|$ is so large that $\beta |C|$ is greater than one, then the amplitudes would not be slowly varying and we would expect complicated multimode relaxation oscillations. Consequently, for Eq. (3.4) to hold we require $|C| \ll \beta^{-1}$. This condition is almost certainly satisfied.

When we replace the complex c-numbers, $\langle b_k \rangle$ and $\langle b_k^\dagger \rangle$, by $f_k e^{-i\phi_k}$ and $f_k e^{i\phi_k}$ in Eq. (3.4), we obtain

$$\dot{f}_k + (f_k/T_k) = (\omega_L \beta/2) \sum_{k'} \{\exp(iD_{kk'}t) F(k') + \exp(-iD_{kk'}t) F^*(k')\} C(k - k') f_{k'}(t) \langle \hat{\sigma}(t) \rangle, \quad (3.5a)$$

$$i f_k \dot{\phi}_k = \tfrac{1}{2} \omega_L \beta \sum_{k'} \{\exp(iD_{kk'}t) F(k') - \exp(-iD_{kk'}t) F^*(k')\} C(k - k') f_{k'}(t) \langle \hat{\sigma}(t) \rangle, \quad (3.5b)$$

where

$$D_{kk'} = d_{kk'} + [\phi_k(t) - \phi_{k'}(t)]/t.$$

In Sec. V, we show for $t > \omega_D^{-1}$ that $\phi_k(t) = \omega_s(k)t$, where $\omega_s(k)$ is the frequency shift in the steady state. Consequently, for $t > \omega_D^{-1}$ we have

$$D_{kk'} = d_{kk'} + \omega_s(k) - \omega_s(k') = [\Omega_k + \omega_s(k)] - [\Omega_{k'} + \omega_s(k')].$$

Thus $D_{kk'}$ is just the actual frequency difference between mode k and k'. In gas lasers the fractional frequency shift is of the order of 10^{-3} or less so the difference between $D_{kk'}$ and $d_{kk'}$ is actually very small.

We conclude this section with a discussion of the multimode problem when there is no coupling between modes; that is, $C(k - k')$ vanishes for $k \neq k'$. In the absence of coupling we obtain for each k

$$\dot{f}_k + (f_k/T_k) = \omega_L \beta \langle \hat{\sigma}(t) \rangle f_k(t) F_c(k'), \quad (3.6a)$$

$$\dot{\phi}_k = \omega_L \beta \langle \hat{\sigma}(t) \rangle F_s(k'), \quad (3.6b)$$

$$\frac{\partial}{\partial t} \left\{ \langle \hat{\sigma} \rangle + \frac{2}{N} \sum_{k'} f_{k'}^2 \right\} = \frac{\hat{\sigma}_0 - \langle \hat{\sigma} \rangle}{T_1} - \frac{4}{N} \sum_{k'} \frac{1}{T_{k'}} f_{k'}^2. \quad (3.6c)$$

We solved the above equations rigorously in II for the case of a single mode. The multimode steady state raises some difficulties. To see this, let us assume that we have m modes and we try to find the steady state in Eq. (3.6a) by setting $\dot{f}_k = 0$. We then obtain the following m conditions (one for each k) on the steady-state fractional population difference, $\langle \hat{\sigma} \rangle_s$,

$$\langle \hat{\sigma} \rangle_s = 1/\omega_L \beta F_c(k) T_k. \quad (3.7)$$

Since it is impossible to simultaneously satisfy these m conditions, there is no stationary state with more than one f_k which is nonzero. We can have m different stationary states with only one mode at a time excited. We obtain these solutions in the following manner. For each k we take $\langle \hat{\sigma} \rangle_s$ determined by Eq. (3.7), determine the corresponding f_k and ϕ_k in Eqs. (3.6a) and (3.6b), and set all $f_{k'}$ for $k' \neq k$ equal to zero. In other words, due to the homogeneity of Eq. (3.6a) the only steady-state solutions that exist for uncoupled modes are single-mode solutions with all other modes unexcited.

As the mode is chosen further and further away from the atomic frequency, ω_0, the corresponding $\langle \hat{\sigma} \rangle_s$ will require a larger and larger pump power due to the exponential decrease of $F_c(k)$ until finally $|\langle \hat{\sigma} \rangle_s|$ becomes equal to one and it becomes impossible to excite the mode. Consequently, only the modes near the line center have low enough thresholds to satisfy the steady-state conditions, Eq. (3.7).

The above situation is reasonable because taking $C(k - k') = 0$ for $k \neq k'$ corresponds, in effect, to a uniform spatial population inversion. Thus the conditions for stationarity put competing demands on $\langle \hat{\sigma} \rangle$ such that no single constant value $\langle \hat{\sigma} \rangle_s$ can satisfy them all. The only strictly time-independent answer is for one mode alone to oscillate. When $C(k - k') \neq 0$, we can have a time-independent stationary state $\langle \hat{\sigma} \rangle_s$ because the population inversion is spatially inhomogeneous and, in effect, different modes can sample different spatial Fourier-transform components of the population inversion. We show this in the next section.

Since Eqs. (3.6) are nonlinear, we cannot take linear combinations of the single-mode stationary states and get solutions. However, there may be time-dependent solutions of Eqs. (3.6) which look like time-dependent linear combinations of stationary states. This would mean that energy would occasionally flow from one mode to another mode through $\langle \hat{\sigma}(t) \rangle$ which couples the modes. This time-dependent flow of energy is what prevents $\langle \hat{\sigma}(t) \rangle$ from approaching $\langle \hat{\sigma} \rangle_s$ when $C(k - k') = 0$ for $k \neq k'$. We observe in the absence of spatial inhomogeneities that the kth modes sees the k'th mode not directly, but through $\hat{\sigma}(t)$ in the exact equations of motion, Eqs. (2.3a) and (2.3b). Thus, the family of solutions for $C(k - k') = 0$ for $k \neq k'$ represented by Eq. (3.7) is not some accidental property of an approximation scheme but is a property of the exact equations of motion.

IV. STEADY-STATE SOLUTIONS

We can find the steady-state amplitudes $(f_k)_s$, by setting $\dot{f}_k$ equal to zero in Eq. (3.5a) and solving the resulting equations for the $(f_k)_s$. In order to obtain a time-independent solution we drop the rapidly oscillating terms, exp $[2iD_{k'k''}t]$. The $D_{k'k''}$ are multiples of the cavity spacing and are of the order of 10^9 or greater and they make negligible contribution to the steady state.

The procedure outlined above leads to equations for $(f_k)_s$ which are linear and homogeneous in $(f_k)_s$ and the condition for the existence of a solution takes the form of an equation for $\langle \hat{\sigma} \rangle_s$. Thus the procedure is solved for $\langle \hat{\sigma} \rangle_s$, obtain the ratio of the amplitudes, and finally use the steady-state solution of Eq. (3.6c)

$$\langle \hat{\sigma}(t) \rangle + \frac{4T_1}{N} \sum_k \frac{f_k^2(t)}{T_k} = 1$$

to obtain the absolute values of the $(f_k)_s$.

We obtain exactly the same equations by a more direct method which has the advantage that the qualitative features of the equations are discernible without explicit solution. We set $\dot{f}_k$ equal to zero in Eq. (3.5a), square the resulting equations, neglect the rapidly oscillating terms and obtain

$$(f_k^2)_s/T_k^2 - (f_k^2)_s \langle \hat{\sigma} \rangle_s^2 \omega_L^2 \beta^2 F_c^2 = \frac{\omega_L^2 \beta^2}{2} \sum_{k' \neq k} (f_{k'}^2)_s \langle \hat{\sigma} \rangle_s^2 \, |F(k')|^2 \, C^2(k - k'), \qquad (4.1)$$

where a subscript s means steady state.

With a simple change of variables, Eq. (4.1) takes the following suggestive form:

$$[\lambda - A_k] x_k = \sum_{k' \neq k} A_{kk'} x_{k'}$$

where

$$x_k \equiv (f_k^2)_s, \qquad A_k = \omega_L^2 \beta^2 T_k^2 F_c^2(k), \qquad \lambda = \langle \hat{\sigma} \rangle_s^{-2}$$

$$A_{kk'} = \tfrac{1}{2} \omega_L^2 \beta^2 T_k^2 C^2(k - k') \, |F(k')|^2.$$

We can now state the problem of the stationary state very simply as an eigenvalue problem. The eigenvalues λ give the allowed values of $\langle \hat{\sigma} \rangle_s$ and the components of the eigenvectors give the ratio of the mode amplitudes in the steady state. The magnitude of the eigenvectors is obtained from the steady-state solution of Eq. (3.6c). The steady-state conditions do not give the frequency shifts. In Sec. V we find the steady-state frequency shifts.

Since A_k and $A_{kk'}$ decrease exponentially as the mode frequency Ω_k gets further away from the atomic frequency ω_0, the number of modes k in the system is equal to the number of modes contained in one to two Doppler widths. This number is usually under ten. For algebraic convenience we take $T_k = T_r$ independent of k and we assume one mode to have the same frequency as the atomic frequency. Even with these simplifications the algebra is formidable but the general features of the solution are qualitatively discernible.

As a specific example, let us consider the case where only five modes are important. The eigenvalue determinate is

$$\begin{vmatrix} (A_{-2}-\lambda) & a(2) & 0 & 0 & 0 \\ a(2) & (A_{-1}-\lambda) & a(1) & 0 & 0 \\ 0 & a(1) & (A_0-\lambda) & a(1) & 0 \\ 0 & 0 & a(1) & (A_1-\lambda) & a(2) \\ 0 & 0 & 0 & a(2) & (A_2-\lambda) \end{vmatrix}, \qquad (4.3)$$

where

$$A_i = A_{-i}; \quad A_0 \equiv \omega_L^2\beta^2T_r^2F_c^2(0) \equiv A_k(\Omega_k = \omega_0),$$

$$A_i \equiv \omega_L^2\beta^2T_r^2F_c^2(j\bar{D}); \quad \bar{D} \equiv (\Omega_{k-1} - \Omega_k)/\omega_D,$$

$$a(j) \equiv \tfrac{1}{2}\omega_L^2\beta^2T_r^2C^2(\pi/L)\ |F(j\bar{D})|^2.$$

The length of the cavity is L. In writing the determinate we assume that only the first cosine transform of the density is important. It is possible that the density variations at $k - k' = 2\pi/L$ are also important. In that case the determinate would contain terms proportional to $C^2(2\pi/L)$ which would appear on subdiagonals parallel to the subdiagonals containing the $a(j)$'s.

We can now see some of the qualitative features of the solution. Starting from A_0 and proceeding along the diagonal in either direction, we find the A_i's decrease rapidly. The same property holds along the parallels to the diagonal. Thus, the modes near the line center are tightly coupled to each other while the modes in the wings of the line are weakly coupled to each other.

Fortunately, we do not have to find all the eigenfunctions and all the eigenvalues of a nine- or ten-dimensional matrix to find the steady state. Since the magnitude of $\langle\hat{\sigma}\rangle$ must be less than or equal to one, then the only physically realizable eigenvalues are $\lambda \geq 1$. Second, the x_k's have to be positive so only eigenvectors with all positive coefficients are physically realizable. In practically all cases the combination of conditions will lead to only a single physical realizable eigenvalue (population inversion) and a single realizable eigenvector. As an example, the matrix in Eq. (4.3) is Hermitian so its eigenvectors are orthogonal. Thus, if there is an eigenvector which is a linear combination of all the x_k's with positive coefficients, it is unique. It is easy to show there is such an eigenvector.

We work out the problem of three modes because this is the only case that can be solved without any approximations. However, the three-dimensional problem is rich enough to include weak and strong coupling limits which we need in order to understand the qualitative features of the general solution. The matrix for the three-mode problem is

$$\begin{vmatrix} (A_1-\lambda) & a(1) & 0 \\ a(1) & (A_0-\lambda) & a(1) \\ 0 & a(1) & (A_1-\lambda) \end{vmatrix}.$$

The eigenvalues are

$$\lambda = A_1, \lambda_+, \lambda_-,$$

where

$$\lambda_\pm = \tfrac{1}{2}(A_1 + A_0) \pm \tfrac{1}{2}[(A_0 - A_1)^2 + 8a^2(1)]^{\frac{1}{2}}.$$

Only the eigenvalue λ_+ is greater than one and has all components of its eigenvector positive. This is a specific example of the uniqueness of the stationary state we discussed earlier.

The stationary value of $\langle\hat{\sigma}\rangle_s$ is

$$\langle\hat{\sigma}\rangle_s = \{\tfrac{1}{2}(A_1 + A_0) + \tfrac{1}{2}[(A_0 - A_1)^2 + 8a^2(1)]\}^{-\frac{1}{2}}. \tag{4.4}$$

The steady-state intensity ratios are

$$x_1 = \frac{x_0}{4}\left\{\frac{(A_1 - A_0) + [(A_1 - A_0)^2 + 8a^2(1)]^{\frac{1}{2}}}{a(1)}\right\}, \tag{4.5}$$

$$x_1 = x_{-1}.$$

The steady-state solution of Eq. (3.6c) is

$$\langle\hat{\sigma}\rangle_s + (4/N)(T_1/T_r)\ \textstyle\sum_{k'} f_{k'}^2 = 1. \tag{4.6}$$

We obtain the steady-state value of x_0 when we substitute Eqs. (4.4) and (4.5) in Eq. (4.6):

$$x_0 = \frac{1 - \{\tfrac{1}{2}(A_1 + A_0) + \tfrac{1}{2}[(A_0 - A_1)^2 + 8a^2(1)]^{\frac{1}{2}}\}^{-\frac{1}{2}}}{(4T_1/Na(1)T_r)[a(1) + \tfrac{1}{2}(A_1 - A_0) + \{[(A_1 - A_0)^2 + 8a^2(1)]\}^{\frac{1}{2}}]}.$$

The general solution, Eqs. (4.4), (4.5), and (4.7), for the steady-state amplitudes is complicated and not very illuminating. There are, however, two limits which indicate the range of behavior of the solution.

The first limit is weak coupling which we define by the inequality $2\sqrt{2}\ |a(1)| \ll |A_0 - A_1|$. In this limit the steady-state population inversion per molecule is

$$\langle\hat{\sigma}\rangle_s \approx \frac{1}{\omega_L\beta T_r F_c(0)}\left|1 - \frac{1}{4}\frac{C^2}{(\bar{D})^2} + \cdots\right|.$$

The first term in brackets is the steady-state population inversion we obtained in II for a single mode. The weak coupling limit implies $|C| \ll \bar{D}$; thus, the steady-state population inversion is only slightly different from the single-mode result. In the weak coupling limit the ratio of the intensity

in the off-resonant models to the intensity in the resonant mode is $(x_1/x_0) = (C^2/8\bar{D}^2)$. Thus, almost all the energy is in the tuned mode. Consequently, the weak coupling limit corresponds to the physical situation, where the tuned mode behaves almost independently of the weakly excited adjacent modes. However, the coupling, although weak, makes possible the existence of a unique stationary state.

In the opposite limit of strong coupling where $2\sqrt{2}\,|a(1)| \gg |A_0 - A_1|$, which implies $|C|^2 \gg \bar{D}^2$, we have

$$x_{-1} = x_1; \qquad x_1 = \tfrac{1}{2}\sqrt{2}x_0;$$
$$\langle\hat{\sigma}\rangle_s \approx [\omega_L\beta T_r F_c(0)(1 + \tfrac{1}{2}\sqrt{2}C^2)^{\frac{1}{2}}]^{-1}.$$

In this limit the modes adjacent to the center mode are excited to a large fraction, $\sqrt{2}/2$, of the center mode's intensity.

With the insight gained from the three-mode case let us look again at the general problem. First, we see that, for modes far from the center of the Doppler line, the modes are weakly coupled to adjacent modes and are weakly excited. Near the center of the line we need to know the value of $C(\pi/L)$, the cosine transform of the density of atoms for the wavenumber π/L. As yet we do not have sufficient information to evaluate $C(\pi/L)$ which depends on surface effects and detailed spatial variations of the pump excitation method. This spatial variation is probably a function of the intensity of the pump in gas lasers. The modes near the atomic frequency ω_0 are more than weakly coupled but probably not strongly coupled to each other. The values of $C(\pi/L)$ are probably in an intermediate range with properties closer to strong coupling than weak coupling; i.e., the adjacent modes are excited to an appreciable fraction of the intensity of the center of the line.

V. CALCULATION OF THE FREQUENCY SHIFT

The condition that we have a stationary state requires that $\dot{\phi}_k$ approaches a constant as t approaches infinity. We now solve Eq. (3.5) for ϕ_k and show that for long times $\phi_k(t)$ approaches $\omega_s(k)t$ where $\omega_s(k)$ is the steady-state frequency shift of the kth mode. We rewrite Eq. (3.5b) for $t > \omega_D^{-1}$ in the following form:

$$\dot{\phi}_k(t) = \omega_L\beta\langle\hat{\sigma}(t)\rangle F_s(k) + \omega_L\beta \sum_{k'\neq k} C(k - k')\frac{f_{k'}(t)}{f_k(t)}\langle\hat{\sigma}(t)\rangle \sin[D_{kk'}t + \chi(k')], \tag{5.1}$$

where

$$F(k') \equiv |F(k')| \exp[-i\chi(k')].$$

When we integrate t over a time long compared with the time required for the amplitudes to reach their stationary values, we obtain

$$\phi_k(t) - \omega_s(k)t = \omega_L\beta \sum_{k'\neq k}\frac{f_{k'}}{f_k}\langle\hat{\sigma}\rangle_s C(k - k')\,|F(k')| \times \left\{\frac{\sin\chi(k')t\sin D_{kk'}t}{D_{kk'}} + \frac{\cos\chi(k')t[1 - \cos D_{kk'}t]}{D_{kk'}}\right\}, \tag{5.2}$$

where

$$\omega_s(k) \equiv \omega_L\beta\langle\hat{\sigma}\rangle_s F_s(k).$$

The frequency shift $\omega_s(k)$ has the same functional form as the frequency shift we calculated for a single mode in II. However, the steady-state population difference per molecule, $\langle\hat{\sigma}\rangle_s$, is now different from the single-mode solution because it is the eigenvalue of Eq. (4.2) which is a functional of all the modes through the mode coupling. In the weak coupling limit, $\langle\hat{\sigma}\rangle_s$ approaches the single-mode result, $[\omega_L\beta T_k F_c(k)]^{-1}$, and $\omega_s(k)$ then approaches the single-mode frequency shift, $F_s(k)[T_k F_c(k)]^{-1}$.

The right-hand side of Eq. (5.2) consists of rapidly oscillating terms that make negligible contributions to the phase in times $t > D_{kk'}^{-1} \sim 10^{-9}$ seconds. For all coupling strengths we can write $\phi_k(t)$ in the following form:

$$\phi_k(t) = t\left\{\omega_s(k) + \sum_{k'\neq k}\left[\frac{M_{kk'}\sin D_{kk'}t}{D_{kk'}t} + \frac{N_{kk'}(1 - \cos D_{kk'}t)}{D_{kk'}t}\right]\right\}, \tag{5.3}$$

where $|M_{kk'}|,\ |N_{kk'}| \leq |\omega_s(k)|$. For $t \gg D_{kk'}$ the contribution from the summation over k' vanishes and we are left with the result that for all coupling strengths $\phi_k(t) \approx \omega_s(k)t$.

In the presence of noise there is a small noise induced frequency shift and $\{\phi_k^2(t) - \phi_k^2(0)\}_{av}$ is proportional to t where the average is over the noise ensemble. It is necessary to consider times $t \gg D_{kk'}^{-1}$ to obtain a result proportional to t for the mean-square fluctuation of the phase. Consequently, as we will show in a future publication the sum over k' of the rapidly oscillating terms in Eq. (5.3) makes no contribution to the stochastic wandering of the phase in the presence of noise.

The frequency difference between two modes is

$$D_{kk'} = (\Omega_k + \omega_s(k)) - (\Omega_{k'} + \omega_s(k')) = \Omega_k - \Omega_{k'} + \omega_L\beta\langle\hat{\sigma}\rangle_s\{F_s(k) - F_s(k')\}.$$

The only indication in the frequency shift that we have coupled modes is the fact that $\langle\hat{\sigma}\rangle_s$ depends in detail on the number of modes and the magnitude of their coupling. Consequently, the derivation of the frequency shift to second order in β derived in II applies to the multimode problem as well, and we have

$$D_{kk'} = \Omega_k - \Omega_{k'} + \omega_L\beta\langle\hat{\sigma}\rangle_s\{F_s(k)[1 - \beta\langle\hat{\sigma}\rangle_s F_c^1(k)] - F_s(k')[1 - \beta\langle\hat{\sigma}\rangle_s F_c^1(k')]\}, \quad (5.4)$$

where

$$F_c^1(k) \equiv \frac{1}{2}\int_0^\infty \tau \exp(-\tfrac{1}{2}\tau^2) \times \exp(-\bar{\nu}_2\tau)\cos[\bar{\Delta}(k)\tau]\, d\tau.$$

If we had N permanent two-level systems the frequency difference $D_{kk'}$ between two modes k and k' would be strictly independent of the magnitude of the pump; and depend only on the spatial density of the two-level systems. In a gas laser we do not have permanent two-level systems but they are created by the pump. Consequently, $C(k - k')$ probably depends on the pump power and thus, $D_{kk'}$ in Eq. (5.4) can have a dependence on pump power because $\langle\hat{\sigma}\rangle_s$ depends on all the $C(k - k')$'s.

VI. COMPARISON OF SCFA WITH LAMB'S THEORY

In this paper we have used the SCFA which means that we have neglected all correlations including those between internal variables $\hat{\sigma}_\alpha$ and the center-of-mass variables X_α, V_α, which are induced by the pump. By ignoring these correlations we are able to allow large exchanges of energy between matter and radiation and as a result we found the electromagnetic field in the steady state is proportional to the population inversion N and thus the field increases proportional to pump power. Lamb on the other hand uses perturbation theory and thus he is restricted to near threshold, $[1 - (N_t/N)] \ll 1$ where N_t is the threshold population inversion. However he includes correlations between the internal variables $\hat{\sigma}_\alpha$ and the center-of-mass variables induced by the pump. Thus the fundamental difference between Lamb and the SCFA consists in different treatments of the pump. Many properties of both solutions are similar, the threshold population inversion density are identical, the first-order frequency shifts are identical and both have Lamb dips. For a more complete analysis see Ref. 5.

In the multimode problem there is a fundamental difference. In Lamb's theory the dominate interaction between the modes comes from Doppler motion which introduces the frequencies $\omega_\alpha = k\cdot V_\alpha$. The frequencies ω_α in effect couple the modes so that even if the population inversion is perfectly uniform there is mode coupling. Lamb has some purely spatial effects which are proportional to $C(2\pi/L)$ and $C(4\pi/L)$. However these terms are small and are not the same as the spatial dependent coupling of the SCFA which is proportional to $C^2(\pi/L)$. In the SCFA there is no mode coupling for the purely spatial homogeneous case. In fact the total coupling is produced not by the Doppler motion, $\omega_\alpha = k\cdot V_\alpha$, but by spatial density variations with wavenumbers of the order of magnitude $(n\pi/L)$ where n is 1 or 2.

Both the Doppler motion coupling and the spatial density variation coupling are probably present at the same time. Thus a very useful experiment would be one in which the density of excited atoms is varied over spatial distances of the order of L, the size of the container. Also since the effect of pump-induced correlations should be less important at high pump power the mode coupling will probably vary as a function of N, i.e., the pump power.

VII. DISCUSSION

We have shown that a multimode gas laser does not have a strictly stationary state because of coupling between the modes. However, we found a unique time averaged stationary state which is a solution of an eigenvalue problem. The eigenvalues are the steady-state population inversion per molecule; i.e., the population inversion threshold. The components of the eigenvectors give the steady-state mode intensity ratios. As we increase the pump above threshold, $\langle\hat{\sigma}\rangle_s$ stays at its steady-state value and the additional energy goes into electromagnetic energy in the modes. The components of the steady-state eigenstate determine how the energy is distributed among the modes.

The steady-state frequency shifts of the modes are practically independent of each other. They are coupled indirectly to each other in that they all see the same population inversion, $\langle\hat{\sigma}\rangle_s$, which is a function of the coupling between modes, $C(k - k')$.

[5] C. R. Willis, "Models of Gas Lasers" (unpublished).

In II we were able to show the unique stationary state for a single mode is approached rapidly by using an extension of the Bogoliuboff–Kryloff quasi-linear theory for nonlinear equations. In the multi-mode case the approach to the steady state is complicated because there is no strictly stationary state and we now have systems of nonlinear equations. We cannot make a careful analysis of the approach to our time averaged steady state without a simultaneous study of the noise problem. We hope to do this in a future publication. Fortunately, the steady-state amplitudes and frequency shifts do not depend on the approach to the stationary state.

ACKNOWLEDGMENT

I wish to thank Samuel T. Scott for many profitable discussions.